

Федеральное государственное бюджетное учреждение науки
Центральный экономико-математический институт
Российской академии наук

На правах рукописи

Милкова Мария Александровна

РАЗРАБОТКА И ЭКСПЕРИМЕНТАЛЬНАЯ АПРОБАЦИЯ МЕТОДИКИ
ПОСТРОЕНИЯ ТЕМАТИЧЕСКОЙ МОДЕЛИ ПРИ РАБОТЕ С НАУЧНОЙ
ИНФОРМАЦИЕЙ

Специальность 08.00.13 – Математические и инструментальные методы
экономики

Диссертация

на соискание ученой степени кандидата экономических наук

Научный руководитель:
д.э.н., к.ф.-м.н.,
Козырев Анатолий Николаевич

Москва 2021

Содержание

ВВЕДЕНИЕ.....	4
ГЛАВА 1. ЭКОНОМИЧЕСКИЕ ПОДХОДЫ К РАБОТЕ С ИНФОРМАЦИЕЙ	16
1.1 Информация в рамках экономики информации и экономики знаний.....	16
1.2. Поведенческие особенности обработки информации и принятия решений	19
1.2.1. Ограниченная рациональность.....	19
1.2.2. Теория двойного процесса познания	22
1.2.3. Влияние поведенческих факторов на функционирование систем и рынков	24
1.2.4. Управление ограниченной рациональностью.....	25
1.3. Нейроэкономика: взгляд на принятие экономических решений со стороны нейронаук	28
1.3.1. Диффузная модель	28
1.3.2. Последствия нейробиологических особенностей принятия решений для экономики	30
1.4. Внимание к информации в рамках экономических наук.....	32
1.4.1. Ограниченное внимание.....	32
1.4.2. Экономика внимания.....	34
1.4.3. Экономические проблемы дефицита внимания	39
1.5. Инструменты для исследования поведенческих особенностей принятия решений.....	41
1.5.1. Экспериментальные методы.....	41
1.5.2. Полевые методы.....	42
Выводы Главы 1	44
ГЛАВА 2. МЕТОДИКА ИСПОЛЬЗОВАНИЯ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРИ РАБОТЕ С НАУЧНОЙ ИНФОРМАЦИЕЙ.....	48
2.1 Общее описание инструментария	49
2.2 Подход аддитивной регуляризации тематических моделей.....	52
2.2.1. Постановка задачи	52
2.2.2. Принцип регуляризации.....	54
2.2.3. Критерии качества	58
2.2.4. Мультимодальная АРТМ	60
2.2.5. Особенности применения тематического моделирования	61

2.2.6. Некоторые важные вопросы	65
2.2.7. Формализация алгоритма для использования в экономике знаний	67
Выводы Главы 2	69
ГЛАВА 3. АПРОБИРОВАНИЕ МЕТОДИКИ ПРИ РАБОТЕ С НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИЕЙ	71
3.1. Тематическое представление научных статей направления поведенческая и экспериментальная экономика	71
3.1.1. Формирование и описание выборки	72
3.1.2. Предобработка данных и формирование словаря	77
3.1.3. Построение тематической модели	78
1.5.3. Анализ чувствительности и другие тестирования	102
3.2. Разведочный поиск на примере патентного анализа для плана импортозамещения.....	104
3.2.1. Постановка задачи исследования	104
3.2.2. Построение модели	107
3.2.3. Результаты	108
3.3. Другие приложения.....	114
Выводы Главы 3	116
ЗАКЛЮЧЕНИЕ	118
СПИСОК ЛИТЕРАТУРЫ.....	123
Приложение 1. Методика построения тематической модели для коллекции научных публикаций.....	145
Приложение 2. Отбор данных для построения тематической модели	146
Приложение 3. Результаты тестирования тематической модели.....	147
Приложение 4. Скрипт предобработки данных	167
Приложение 5. Скрипт для вычисления когерентности тем	168
Приложение 6. Скрипт для построения тематической модели	169
Приложение 7. Скрипт для выгрузки результатов на веб-сайт.....	170

ВВЕДЕНИЕ

Актуальность темы исследования

Как отмечал Герберт Саймон еще 50 лет назад, «информационное перенасыщение с одной стороны рождает дефицит внимания с другой». Сегодня это высказывание особенно актуально, поскольку с развитием цифровых технологий и информация, и внимание все чаще воспринимаются как товар. В эпоху цифровой трансформации темпы накопления информации настолько стремительны, что это порождает сложности ее восприятия. Задача извлечения необходимой информации из больших массивов с целью принятия на ее основе решений все чаще отдается на откуп информационно-коммуникационным технологиям. Здесь речь идет о развитии различных поисковых, рекомендательных, вопросно-ответных и других систем, снижающих затраты экономического агента на поиск информации. Однако важно различать информацию коммерческую и информацию научную, а также информацию, на основе которой предполагается формирование новых знаний. Алгоритмы, работающие с коммерческой информацией (Amazon, Alibaba и др.) нацелены на сбор информации о пользователе и на вывод тех индивидуальных рекомендаций, которые будут соответствовать его профилю и истории совершенных покупок. Алгоритмы, работающие с информацией научной, а также той, которая так или иначе призвана обогащать тезаурус человека, устроены в настоящее время схожим образом. Различные рекомендательные сервисы, в том числе и в электронных библиотеках (Cyberleninka и др.) выдают в рекомендациях схожие по смыслу публикации к уже прочитанным. Поисковые системы (Google, Яндекс) внедряют и совершенствуют систему выдачи готовых ответов.

Стоит отметить также, что человек, как лицо принимающее решение на основе информации, обладает ограниченной рациональностью. Несмотря на развитие сервисов, оказывающих информационные услуги, склонность к когнитивным искажениям, социальное ускорение, клиповое мышление и другие факторы приводят к существенному снижению порога, на котором принимается решение о прекращении поиска информации. Цифровые технологии в настоящее время не только преобразуют данные в информацию, но и ставят перед собой цель преподнесения

информации в виде готовых знаний, стремятся исключить стимулы для самостоятельного анализа информации экономическим агентом.

Стремительный рост информации также остро ставит вопрос о распределении наиболее дефицитного в настоящее время ресурса – внимания. Конкуренция за него привела к росту новых технологий для привлечения и управления вниманием. В данной борьбе активно используются свойства человеческого мозга (например, автоматическая конформность, доминирование эмоциональных областей), а также склонность к принятию быстрых решений на основе эвристик.

Таким образом, становится актуальной разработка инструментов, позволяющих оказывать помощь в выборе информационных продуктов не на основе их простоты потребления или схожести с уже потребленными, а на основе их ценности с точки зрения смысла. Спрос на информационные продукты, представленные в виде научно-технической информации устроен иначе, чем спрос на коммерческие информационные продукты. Основным отличием является стремление к полноте и ценности извлекаемой информации, вместо обращения внимания исключительно на быструю, популярную или продвигаемую информацию.

Актуальными в этом смысле являются как разработка самих инструментов, включая модели, алгоритмы и программный код, так и методики их настройки и применения к конкретным областям человеческой деятельности. Разработка соответствующего инструментария имеет прямое отношение к экономической науке, так как, во-первых, он может быть рассмотрен в рамках функционирования экономики информационных продуктов, а именно, представляет собой способ организации потребления информационных продуктов, помощи выбора из множества. Во-вторых, инструментарий может быть применен к анализу экономической литературы (в данной работе это показано на примере такого информационного продукта как научные публикации по поведенческой и экспериментальной экономике). В-третьих, позволяет вычленять смысл из больших текстовых коллекций для принятия экономических решений (в данной работе это показано на примере организации смыслового поиска тех патентных документов, которые соответствуют реализации программы импортозамещения).

Степень научной разработанности проблемы

Степень научной разработанности темы рассматривается в рамках математических и инструментальных методов в экономике, принимая во внимание исследования, касающиеся обработки и поиска информации человеком, на стыке таких дисциплин как экономика информации, экономика знаний, поведенческая экономика, нейроэкономика. Собственно инструментальные методы лежат в области анализа текстов, анализа естественного языка, а именно, в области методов семантической компрессии информации (сжатого представления смысла), к которым относится тематическое моделирование. Научная разработанность темы описана в трёх ключевых направлениях: 1) ценность представления информации в надлежащем виде, 2) личностные препятствия к представлению и восприятию информации и 3) инструменты для извлечения ценной информации.

Вопросы экономической ценности информации, поиска информации как снижения неопределенности, затрат на сбор информации с учетом продолжительности поиска разрабатывались Дж. Стиглером, П. Нельсоном, М. Ротшильдом. Наделение информации определенными свойствами, неполнота информации, асимметрия, недостатки сбора информации и другие особенности отмечались в работах К. Эрроу, Дж. Акерлофа, Дж. Стиглица и др. Модели принятия решений на основе информации разрабатывались А. Бенерджи (модель стадного поведения), С. Бикчендани (модель информационных каскадов), модели, определяющие оптимальное взаимодействие раскрытия информации и конкуренции (М. Генцков и Е. Каменика), неприятия неоднозначности (Ф. Микарони), рационального невнимания (К. Саймс, Р. Рейс). Проблемы принятия решений на основе информации нашли обширное применение в теории игр (Р. Зельтен, А. Рубинштейн, Дж. Хасаньи и многие другие). В современных исследованиях учитывается различная ценность информации для разных покупателей, изучаются оптимальные механизмы раскрытия информации, продажи информации (К. Шапиро, Х. Вэриан, Д. Бергеманн, М. Бабая, А. Смолин). Среди отечественных ученых, занимающихся экономикой информации, экономикой информационных продуктов следует выделить работы В.Л. Макарова, А.Н. Козырева, М.И. Лугачева, К.Г. Скрипкина, Ю.Е. Хохлова.

Экономические аспекты производства, распространения и управления знаниями изучаются в рамках экономики знаний (Ф. Махлуп), информационной экономики (М. Порат). Среди отечественных ученых – существенный вклад в развитие экономики знаний оказали работы В.Л. Макарова, Г.Б. Клейнера, А.Е. Варшавского.

Ограниченная рациональность человека при обработке информации, психологические аспекты принятия решений, а также особенности принятия решений в условиях неопределенности и риска, в том числе на основе эвристик, изучались Г. Саймоном, У. Эдвардсоном, М. Алле, Д. Элсбергом, Д. Канеманом и А. Тверски, В. Смитом, Ф. Хайеком, Р. Хайнером, Ю. Эльстером, Г. Гигеренцером, Р. Талером, К. Санстейном и др. Поведенческие аспекты принятия решений применительно к анализу экономических организаций, рынков изучались О. Уильямсоном, Ж. Тиролем, Дж. Акерлофом, Р. Шиллером. Р. Талером, А. Шлейфером и др. Вопросы управления ограниченной рациональностью нашли отражения в теории подталкивания (Р. Талер, К. Санстейн, С. Бенартзи), изучению влияния медиа и социальных сетей (К. Санстейн, С. ДеллаВигна, Х. Элкот, М. Генцков, Д. Рэнд и др.)

В области нейроэкономики существенный вклад в понимание особенностей принятия решений на основе информации внесли работы П. Глимчера, К. Камерера, Е. Фехра, А. Рустичини и др. Среди российских ученых – В.А. Ключарева.

Изучение внимания к информации как важного звена в принятии экономических решений нашло отражение в направлении, получившем название «экономика внимания». Будучи предложенной Г. Франком, М. Голдхабером, экономика внимания развивалась в работах Т. Давенпорта, Дж. Бека, Т. Ву, Р. Лэнхема, А. Фестре и П. Гарусте, К. Камерера, а также отечественными учеными - А.Н. Козыревым, Г.Г. Почепцовым.

Отечественные ученые, отмечающие важность расширения охвата (или, что то же, обращения внимания) накапливающихся знаний в контексте функционирования экономики знаний – В.Л. Макаров, Г.Б. Клейнер, А.Е. Варшавский; среди отечественных ученых, занимающихся вопросами повышения эффективности

работы научного сообщества, стимулирования научно-исследовательской деятельности – Е.В. Балацкий, О.М. Писарева.

Работы, подчеркивающие и развивающие междисциплинарное сотрудничество экономики, психологии, социологии – Г.Б. Клейнер, В.М. Полтерович, Д.А. Жданов; среди поведенческих экономистов – А.В. Белянин, К. Паниди.

Вопросами автоматического понимания текстов занимается одна из наиболее динамично развивающихся областей искусственного интеллекта – анализ естественного языка (NLP). В данной работе представляют интерес методы семантической компрессии информации, реализованные в рамках тематического моделирования. В целом, развитию тематического моделирования способствовали работы Т. Хоффмана, Д. Блея. Отечественным ученым К.В. Воронцовым предложен подход на основе аддитивной регуляризации тематических моделей (АРТМ), который активно развивается в настоящее время им и его учениками: А. Потапенко, М. Апишевым, А. Яниной, В.Г. Булатовым и др.

Среди отечественных ученых, применяющих методы семантического структурирования контента научных электронных библиотек, в том числе с использованием тематического моделирования, стоит выделить работы Паринаова С.И., Кога-ловского М.В.

Среди ученых, включающих анализ естественного языка в экономический дискурс, стоит отметить работы С. Бейкера, С.И. Паринаова, И.О. Голощаповой, Д.О. Афанасьева, Е.А. Федоровой.

Вклад данной работы состоит как в расширении набора инструментов для анализа текстовой информации, так и в применении их к новым областям знаний. В том числе, расширяется область исследований, которые вводят в экономический дискурс анализ патентной информации, выходящей далеко за рамки патентной классификаций. В качестве ключевой темы выбрано «импортозамещение», поскольку именно эта тема весьма актуальна в сегодняшней политической ситуации. Другая интересная с точки зрения применения разработанных инструментов сфера – поведенческая экономика, которая непосредственно связана с экономикой знаний, информационной экономикой, нейроэкономикой и психологией. Кроме того,

именно для поведенческой экономики существуют исследования о том, кто и каким образом внес вклад в развитие данного направления (Аннотированная библиография П. Уаккера¹). Тем самым появляется возможность оценить результат работы программы, обучившейся без «учителя» с результатом работы квалифицированного специалиста. В том, что касается разработки собственно инструментов, вклад работы – это набор программ и методик для проведения экспериментов, представленных в открытом доступе в репозитории Github. Результаты применения методики выложены на специально созданном сайте.

Целью диссертации является адаптация тематического моделирования для выявления структуры научно-технической информации, способствующей эффективному получению знаний. В рамках работы необходимо разработать методику построения тематической модели (в концепции АРТМ) для научно-технической информации, позволяющую сконфигурировать среду для восприятия научно-технической литературы, так чтобы: 1) снизить затраты на поиск ключевой информации по исследуемой научной теме; 2) обеспечить полноту и ценность (с точки зрения смысла) извлекаемой информации 3) минимизировать сдвиги в восприятии (сместить фокус внимания с первой попавшейся, эмоциональной информации).

Задачи исследования:

1. Подготовить методическое обеспечение для непредвзятого анализа корпуса документов (в смысле эффектов искажённого восприятия) через формализацию процесса построения тематической модели в виде создания комплекса программ и рекомендаций, позволяющих выявлять тематическую структуру корпуса научных публикаций.

2. Экспериментально проверить авторскую методику на примере научных статей по поведенческой и экспериментальной экономике. При проверке выявить ключевые темы данного направления, их ключевые термины, авторов, ключевые публикации.

3. Проанализировать чувствительность полученных результатов.

¹ Peter P. Wakker. Annotated Bibliography – URL: <https://personal.eur.nl/wakker/refs/webfrncs.pdf> (дата обращения: 25.09.2021)

4. На основе результатов моделирования создать онлайн ресурс, позволяющий осуществлять навигацию по структуре коллекции научных публикаций по поведенческой и экспериментальной экономике.

5. Привести пример решения задачи другого типа, когда необходимо выявить не ключевую информацию внутри всей текстовой коллекции, а определить структуру только интересующего сложносоставного фрагмента. Более конкретно, в коллекции патентных документов выявить и структурировать только те, что соответствуют (по смыслу) пунктам планов по импортозамещению по 22 отраслям промышленности.

Объект и предмет исследования

Объект исследования – объединения отдельных исследователей, научных коллективов и потребителей научного знания, связанных общей тематикой исследований. *Предмет исследования* – информационное взаимодействие производителей и потребителей научного знания, объединенных единой тематикой исследования.

Научная новизна диссертационной работы заключается в разработке авторской интерпретации алгоритма тематического моделирования и комплекса вычислительных программ на его основе, позволяющих адаптировать применение тематического моделирования (в концепции ARTM) к экономическим исследованиям теоретического и прикладного характера, в том числе в области экономики знаний, поведенческой экономике, а также при выполнении прикладных экономических исследований.

В частности, научную новизну представляют следующие научные положения, выносимые на защиту:

1. Выявлены особенности потребления информации, на их основе разработаны и показаны в экспериментах способ организации информационных продуктов, представленных в виде научных публикаций и патентных документов, помогающий в выборе ценных (с точки зрения вычлененного смысла) продуктов. Научно обосновано, что подход к организации информационных продуктов на основе тематического моделирования, в отличие от подхода на основе алгоритмов

повсеместно распространенных поисковых систем, позволяет повысить эффективность работы научного сообщества (с точки зрения организации информационного взаимодействия производителей и потребителей научного знания).

2. Предложен формализованный подход к выявлению структуры больших коллекций научных публикаций: разработан комплекс программ, проведены эксперименты, представлен набор инструкций. Предложенная формализация алгоритма АРТМ, в отличие от общего алгоритма АРТМ, позволяет существенно снизить трудоемкость подбора параметров при построении тематической модели на базе научных публикаций.

3. На основе формализованного подхода АРТМ впервые выявлены машинным способом основные темы направления поведенческая и экспериментальная экономика (на базе научных публикаций), в том числе ключевые фразы, основные авторы и опорные публикации по каждой из тем. В сравнении с составляемой вручную аннотированной библиографией по поведенческой экономике, автоматизированный подход также выделяет релевантные темы, может уступать в точности выявленных основных фраз и авторов по темам, однако позволяет быстро определять адекватную структуру научной области там, где экспертное аннотирование отсутствует.

4. Предложен и апробирован подход к оценке результатов программы импортозамещения на основе патентных данных с помощью применения методов тематического моделирования. Предложенный подход патентного поиска позволяет находить релевантные документы сразу по большому числу пунктов планов импортозамещения по различным отраслям промышленности, что невозможно осуществлять стандартными средствами поисковых и аналитических систем.

Теоретическая значимость исследования заключается в следующем:

1. В рамках проводимого исследования, работа систематизирует результаты не только экономических, физико-математических и технических наук, но и соответствует критериям междисциплинарности: опирается на достижения психологических, социологических, нейроэкономических наук.

2. Разработанный подход организации результатов научной деятельности вносит вклад в экономику информации, экономику информационных продуктов в части расширения формализованного и измеримого описания информационных продуктов, которые могут выступать объектом экономических отношений.

3. Формализация построения тематических моделей является методическим вкладом в развитие подходов к анализу больших объемов текстовой информации, актуальных для принятия решений в экономической науке.

Практическая значимость исследования заключается в следующем:

1. Предложенная формализованная процедура тематического моделирования на основе аддитивной регуляризации представляет собой значимый вклад в практику выделения ключевой информации из больших текстовых коллекций и может быть использована в различных областях экономической науки. Например, при принятии решений о выборе информационных продуктов, обладающих определенной ценностью, а также при проведении прикладных исследований, в том числе при заключении крупных международных контрактов.

2. Предложенная методика работы с научной информацией представляет практическую ценность при проведении исследований – получении представления о структуре направления, выявлении ключевых авторов и опорных публикаций, (внутри поведенческой и экспериментальной экономики). Подход позволяет повысить эффективность работы научного сообщества путем организации информационного взаимодействия производителей и потребителей научного знания.

3. Предложенная методика применения тематического моделирования к анализу программы импортозамещения позволяет получить представление о состоянии всех отраслей экономики в разрезе соответствия плану импортозамещения на основе патентных данных.

Методы исследования

В диссертационном исследовании применялись методы научного познания, как на теоретическом, так и на эмпирическом уровне.

На теоретическом уровне к ключевым методам исследования относились: анализ, моделирование, синтез, обобщение.

На эмпирическом уровне ключевым методами являлись:

1. Информационный поиск (с помощью составления регулярных выражений) для извлечения необходимых для анализа данных;
2. Метод лингвистической предобработки текстовых данных для подготовки их к анализу: лемматизация, фильтрация, удаление стоп-слов.
3. Метод машинного обучения – тематическое моделирование на базе аддитивной регуляризации тематических моделей – для проведения мягкой кластеризации больших текстовых коллекций, выявления скрытых тем, а также ключевых слов и словосочетаний с максимальной содержательной нагрузкой в выявленных темах.
4. Метод визуализации данных – для получения наглядного представления о получаемых результатах.

Информационная база исследования

1. Научные статьи (название и аннотация) по поведенческой и экспериментальной экономике, находящиеся в репозитории Semantic Scholar², относящиеся к областям экономики и бизнеса, до 2020 года включительно. Всего 37352 статьи.
2. Патентные документы за период январь 2016 - июнь 2019 гг. из базы Роспатента³. Всего 152718 документов; отраслевые планы импортозамещения Минпромторг для 22 отраслей промышленности⁴.

Достоверность и апробация результатов работы

Результаты представления научных публикаций в области поведенческой и экспериментальной экономики верифицированы представителями предметной области.

Основные научные положения и результаты исследования были представлены на следующих конференциях и научных семинарах:

² Semantic Scholar Open Research Corpus // Интернет платформа Semantic Scholar: <https://www.semanticscholar.org> – URL: <http://s2-public-api-prod.us-west-2.elasticbeanstalk.com/corpus/download/> (дата обращения 22.04.2020)

³ Открытые реестры ФИПС // Федеральный институт промышленной собственности: <https://www.fips.ru/> – URL: <https://www.fips.ru/registers-web/> (дата обращения 22.04.2020)

⁴ Отраслевые планы импортозамещения. – Текст: электронный // Государственная информационная система промышленности: <https://gisp.gov.ru/> – URL: <https://gisp.gov.ru/plan-import-change/> (дата обращения 22.04.2020)

1. Милкова М.А. Инструментальное обеспечение поведенческой экономики. Научный семинар Института системного анализа РАН под рук. Лившица В.Н. – Январь 2020 г.

2. Милкова М.А. Инструментальное обеспечение экономики внимания. Научный семинар ЦЭМИ РАН под рук. ак. Макарова В.Л. – Март 2020 г.

3. Milkova M.A. Patent-based import substitution analysis with Additively Regularized Topic Models. X International Scientific and Practical Conference named after A. I. Kitov "Information Technologies and Mathematical Methods in Economics and Management" – October 2020.

4. Милкова М.А. Экономика внимания и ее инструментальное обеспечение. IV Российский экономический конгресс, тематическая конференция Поведенческая и экспериментальная экономика. Декабрь 2020 г.

5. Милкова М.А. Восприятие информации в период пандемии COVID-19. Международная научная школа-семинар «Системное моделирование социально-экономических процессов» им. С.С. Шаталина – Октябрь 2020 г.

Область исследования. Настоящее исследование соответствует паспорту научной специальности 08.00.13 – «Математические и инструментальные методы экономики» и соответствует требованиям следующих разделов:

2. Инструментальные средства:

2.6. Развитие теоретических основ методологии и инструментария проектирования, разработки и сопровождения информационных систем субъектов экономической деятельности: методы формализованного представления предметной области, программные средства, базы данных, корпоративные хранилища данных, базы знаний, коммуникационные технологии.

2.8. Развитие методов и средств аккумуляции знаний о развитии экономической системы и использование искусственного интеллекта при выработке управленческих решений.

Публикации основных результатов исследования. Основные положения и результаты диссертационного исследования отражены в 13 научных публикациях общим объемом 19,2 п.л. (в т.ч. авт. – 18 п.л.), в том числе в 3 статьях в изданиях,

включенных в перечень ВАК РФ, в 1 статье в издании, индексируемом в базе данных Scopus, оставшиеся – в других научных изданиях.

Объем и структура работы

Диссертационная работа содержит введение, три главы, заключение и библиографию общим объемом 172 стр., включая 9 таблиц, 13 рисунков и 7 Приложений.

ГЛАВА 1. ЭКОНОМИЧЕСКИЕ ПОДХОДЫ К РАБОТЕ С ИНФОРМАЦИЕЙ

Различные взаимоотношения между понятиями *информации, знания, неопределенность и принятие решений* являются ключевыми в теории информации, экономике информации, экономике знаний и теории игр, а добавление к этим понятиям особенностей восприятия информации человеком – ключевым в поведенческой и экспериментальной экономике, нейроэкономике. Обсуждению этих дисциплин в разрезе обработки информации человеком посвящена Глава 1.

1.1 Информация в рамках экономики информации и экономики знаний

Родоначальники теории информации, Клод Шеннон, Норберт Винер подошли к понятию информации по-разному, но так или иначе определяли ее через энтропию (неопределенность). По Шеннону, количество информации равно энтропии; энтропия содержит больше информации, чем структура, иначе говоря, информации тем больше, чем больше количество возможных интерпретаций. Винер же, напротив, определял информацию через уменьшение энтропии. Информация, по Шеннону, не есть значение, тогда как Винер предполагает некоторое значение информации, в рамках системы. Соотношение между информацией и знанием может быть различным в зависимости от контекста (Sveiby, 1994; Макаров и Клейнер, 2007, с.22).

Основоположником экономики информации считается Джордж Стиглер, который отмечал экономическую ценность информации, подразумевая, что поиск информации помогает уменьшить неопределенность и избежать ошибок при совершении сделки (Stigler, 1961). Оптимальная продолжительность поиска определяется из равенства предельной экономии предельным альтернативным издержкам поиска. Реклама (в том числе в виде идентификации продавцов и их цен) и репутация рассматривались как инструмент снижения затрат на поиск (Nelson, 1974). Позднее идеи Стиглера были развиты и даже пересмотрены большим числом последователей. Выделялись: неизвестность распределения цен, важность определения не оптимальной продолжительности поиска, но пересмотра на каждом шаге

решений о том, стоит ли продолжать поиск (Rothschild, 1974); наличие неполной, асимметричной информации и неопределенности качества товаров (Akerlof, 1970; Stiglitz, 1977); наличие информации, отличной от информации о текущих ценах, недостатки сбора информации, наделение информации определенными свойствами (Arrow, 1979, 1984); социальная ценность или ценность общественной информации для благосостояния (Ohlson, 1988) и др.

Современная экономическая теория рассматривает неопределенность и информацию с нескольких позиций. На индивидуальном уровне центральным вопросом является разработка стратегии принятий решений о необходимости дополнительного сбора информации. На уровне рынка главным вопросом является поиск общего равновесия, которое возникает, когда участники предпринимают такие «информационные» действия; рассматриваются проблемы групповых решений с учетом наличия разных убеждений, конфликта интересов (Hirshleifer and Riley, 1992). Были построены модели информационных каскадов – явления, когда индивиду оптимально принимать решение на основе действий предшествующего индивида без учета его собственной информации (Bikhchandani et al., 1992); модели стадного поведения – принятие решения на основе решений предыдущих индивидов, даже если это противоречит имеющейся информации (Banerjee, 1992); модели, определяющие оптимальное взаимодействие раскрытия информации и конкуренции (Gentzkow and Kamenica, 2017); неприятия неоднозначности (Maccheroni et al, 2006); модели рационального невнимания – когда агенты не могут обработать всю доступную информацию, однако могут выбрать, на какие именно фрагменты информации обращать внимание (Sims, 2003), в том числе модели, подразумевающие отложенное обновление информации, при невнимании между датами обновления (Reis, 2006a, 2006b); теории аукционов (Milgrom and Weber, 1982).

Проблемы принятия решений на основе информации нашли обширное применение в теории игр (Harsanyi, 1967; Selten, 1978), обсуждающей различные сценарии, например, зависимость равновесия от схожести интересов агентов при передаче информации (Crawford and Sobel, 1982), забывчивость (Piccione and Rubinstein, 1997) и др.

К современным задачам в области экономики информации относятся исследования управления атрибутами и стоимостью информационных продуктов (Shapiro and Varian, 1999), с учетом различной ценности информации для разных покупателей (Bergemann et al., 2018); изучаются оптимальные механизмы продажи информации, в том числе применительно к продаже информации о потенциальных потребителях рекламодателям (Babaioff et al., 2012), деятельности рекрутинговых агентств (Smolin, 2019). Подробный обзор современных достижений в области торговли информацией представлен в работе Козырева (Козырев, 2020).

Несколько иными являются цели исследований в рамках экономики знаний. В отличие от исследований по экономике информации, большая часть которых носит аналитический и математический характер, экономика знаний является преимущественно эмпирической и часто описательной (Mattessich, 1993). Экономика знаний прежде всего исследует экономические аспекты производства, распространения и управления знаниями. Основоположником экономики знаний считается Фриц Махлуп, охватывающий в своих исследованиях различные типы знаний, производство знаний и отрасли знаний, образование, исследования и разработки, средства коммуникации и информационные услуги, связь производства знаний с валовым национальным продуктом и многие другие темы (Machlup, 1962, 1980, 1982, 1984). Исследования Махлупа получили значительное развитие, в том числе в рамках «информационной экономики» (Porat, 1977). Так, Порат выделял «информационную деятельность» как включающую те конкретные отрасли и профессии, основной функцией которых является производство, обработка или передача экономически ценной информации. Схожим является определение «экономики, базирующейся на знаниях» – определенного типа экономики, где сектор знаний играет решающую роль, а производство знаний становится источником роста экономики (Макаров, 2004; Макаров и Клейнер, 2007).

1.2. Поведенческие особенности обработки информации и принятия решений

1.2.1. Ограниченная рациональность

Параллельно с развитием экономики информации и экономики знаний развивалось и понимание, что человек принимает решения, отклоняясь от оптимальной стратегии максимизации полезности или прибыли, учитывает личные установки, общественные нормы и другие внешние факторы.

Одними из первых отсылок к психологическим аспектам принятия решений можно считать работы Уорда Эдвардса, представляющего теорию принятия решений как исследовательскую тему в области психологии (Edwards, 1954). Также стоит выделить эксперименты Мориса Алле, демонстрирующие стремление человека в условиях риска и неопределенности к получению гарантированного результата (Allais, 1953); эксперименты Дэниела Эллсберга, показывающие склонность людей выбирать решение, для которого вероятность благоприятного исхода известна, вместо решения с неопределенной вероятностью благоприятного исхода (Ellsberg, 1961).

Существенный вклад в понимание поведенческих аспектов принятия решений оказали работы Герберта Саймона. Так, введенная им теория ограниченной рациональности (*bounded rationality*) подразумевает, что решения, принимаемые человеком, ограничены трактовкой решаемой проблемы, когнитивными способностями, а также временем для принятия решения. По Саймону, человек ищет решение, которое будет его удовлетворять, а отнюдь не оптимальное. Так называемая концепция «удовлетворенности» (*satisficing*) гласит, что организм будет выбирать первую альтернативу, которая соответствует его уровню желаний, а не тратить время на ранжирование всех альтернатив путем вычисления ожидаемых полезностей (Simon, 1956). Ключевым стало понимание того, что рациональность индивида определяется и обосновывается с точки зрения доступных индивиду знаний и методов расчета. Именно субъективность восприятия реальности, неполнота информации (или, напротив, ее переизбыток и высокие издержки поиска информации) являются определяющими звеньями принятия решений.

Понимание важной роли неопределенности как неопределенности окружающей среды, представляемое Ф. Найтом и другими учеными, также трансформировалось. Так, развивающий идеи Саймона Рональд Хайнер понимал структуру неопределенности как функцию от компетенций (способностей) индивида и сложности, нестабильности окружающей среды. Чем больше разрыв между компетенциями и сложностью, тем скорее индивид вырабатывает определенные «правила поведения» как средство борьбы с неопределенностью. Таким образом, неопределенность порождает предсказуемость в поведении (Heiner, 1983).

Рейнхард Зельтен, известный в первую очередь своим вкладом в теорию игр, одним из первых (еще в 50-х годах) развивал теорию ограниченной рациональности с помощью экспериментальных методов, тем самым представляя «описательную ветвь теории принятия решений и теории игр, серьезно опирающуюся на ограниченную рациональность человеческого поведения»⁵. Зельтен различал три корня человеческого поведения: мотивацию (движущая сила), адаптация (рутинная корректировка без рассуждений) и познание (рассуждение), особенно подчеркивая важное влияние познания на мотивацию и адаптацию, а в конечном итоге и на поведение (Selten and Stoecker, 1986). Модели Зельтена успешно предсказывали ряд поведенческих шаблонов в различных экспериментальных контекстах.

Юн Эльстер выделял следование социальным нормам как способ экономии сил при принятии решений (Elster, 1989). По Эльстеру, социальные нормы – не что иное, как следование индивидуальной, коллективной и генетической оптимизации.

Исследования Дэниела Канемана совместно с Амосом Тверски, основывающиеся на теории ограниченной рациональности Саймона, стали одними из ключевых. Предложенная теория эвристик и сдвигов (heuristics and bias) подразумевает, что, принимая решения в условиях неопределенности, человек склонен полагаться на ряд эвристик (репрезентативности, доступности, привязки и корректировки), которые упрощают процесс оценки вероятностей и прогнозирования значений величин до интуитивных оценок, что приводит к систематическим ошибкам (Tversky

⁵ Reinhard Selten - Biographical // NobelPrize.org. URL: <https://www.nobelprize.org/prizes/economic-sciences/1994/selten/biographical/> (дата обращения: 25.09.2021)

and Kahneman, 1974). Анализ принятий решений в условиях не только неопределенности, но и рисков, привел к разработке теории перспектив (prospect theory), показывающей, что тот или иной выбор делается человеком на основе субъективного представления о его правильности или полезности, без учета реальных фактов (Kahneman, Tversky, 1979). Так, изменение предпочтений может происходить даже под влиянием формулировки или контекста, так называемого фрейма (framing): формулирование одной и той же проблемы с точки зрения выгод или с точки зрения потерь приводит к разным результатам. Человек склонен избегать риска (risk-averse), если речь идет о выигрыше, и, напротив, принимать риск (risk-seeking), если речь идет о проигрыше. С эффектом неприятия потерь связан эффект наделенности (endowment effect), подразумевающий, что человек склонен выше оценивать те товары, которыми уже обладает, чем те, которыми может обладать (Thaler, 1980). Другим типом фрейма является склонность к предпочтению альтернатив, характеризующихся более полной информацией (Shafir, 1993). В некоторых экспериментах показано, что к принятию решений на основе эвристик, а не на основе тщательного анализа склонны и люди, обладающие квалификацией и опытом в рассматриваемой области (Tversky, Kahneman, 1971; Chen et al., 2007). Согласно подходу “heuristics and bias”, использование эвристик является естественной мерой оценивания и подразумевает экономию усилий взамен точности.

Впоследствии различными исследователями был значительно расширен набор возможных эвристик и сдвигов, показана возможность использования оптимизационных моделей для интеграции ограниченной рациональности в экономическую науку (Rabin, 2013). Построены модели, позволяющие учитывать различные отклонения от рациональности: межвременные несогласованности в предпочтениях (см. Thaler, 1981) с помощью гиперболического дисконтирования (Loewenstein and Prelec, 1992); преувеличение степени сходства будущих вкусов с настоящими (projection bias) (Loewenstein et al., 2003); модель предпочтений, зависящая от точки отсчета и неприятия потерь (Kőszegi and Rabin, 2006) и др.

В работе Стефано ДеллаВигна (DellaVigna, 2009) основные отклонения обосуждаются на основе эмпирических данных, вне лабораторий, показана

экономическая значимость отклонений. Отклонения обобщены на три (иногда пересекающихся) класса: нестандартные предпочтения (временные, рисковые, социальные), нестандартные убеждения (систематическая самоуверенность, преувеличение степени сходства будущих вкусов с настоящими); принятие нестандартных решений (упрощение проблемы на основе простых эвристик, эффект фрейминга, социальное давление, ограниченное внимание, влияние эмоций).

1.2.2. Теория двойного процесса познания

Отклонения поведения от рационального лежат в рамках так называемой теории двойного процесса познания (*dual process theory*) (James, 1980), согласно которой человеку свойственны два типа мышления — интуитивное понимание и логическое рассуждение. При анализе информации первое — подсознательное, автоматическое, непроизвольное — значительно превосходит второе по скорости, но уступает ему во внимании к деталям и способности к обучению. Данная теория развивалась в основном в области психологических наук (Petty and Cacioppo, 1986; Evans, 1984 и др.), однако нашла отражение и в экономических работах. Стоит обратить внимание, что с ростом понимания психологических, а также нейробиологических особенностей принятия решений (см. подробнее раздел 1.3), психологические науки все больше оказывают влияние на экономику (см., например, Rabin, 1998; Braesemann, 2019).

Так, в рамках «дуализма» Канемана и Тверски, ассоциативно-ориентированная система («Система 1») отвечает за быстрые, эмоциональные, интуитивные решения. Оценки, сделанные Системой 1, затем дополняются (а иногда и переопределяются) результатами более продуманной, последовательной и основанной на детальном анализе «Системы 2» (Kahneman, 2003).

Вернон Смит, работавший совместно с Канеманом, отмечал, в том числе опираясь на работы Хайека (Hayek, 1952), взаимодействие восприятия, памяти или опыта (Smith, 2003). Вводимое Смитом понятие экологической рациональности также учитывало это взаимодействие, а также контекст принятия решения. Смит обсуждает взаимодействие двух типов рациональности: конструктивистской и экологической. Первая подразумевает поведение (причем не только индивида, но и

рынка, института, любого коллектива) на основе максимизации своей выгоды при помощи сознательных размышлений, вторая – определяется внешней средой и порожаемым ей опытом.

Стоит отметить, что такой же термин – экологическая рациональность – был независимо введен представителем психологических наук - Гердом Гигеренцером, отмечающим, что поведение экологически рационально в той степени, в которой оно адаптировано к структуре внешней среды (Gigerenzer and Todd, 1999). Гигеренцером совместно с рядом экономистов разработана концепция быстрых и экономных эвристик (fast and frugal heuristics), согласно которой, в отличие от подхода Канемана и Тверски, эвристики не зависят от компромисса между точностью и усилием – их простота является причиной того, что они работают хорошо, должным образом сопоставляя эвристический подход с окружающей средой (Gigerenzer, 2008; Gigerenzer and Selten, 2001). Быстрые и экономные эвристики имеют дело с простыми, специфичными для конкретной задачи стратегиями принятия решений, которые можно использовать для выработки точных и объективных суждений. Основной лейтмотив подхода на основе простых эвристик гласит: в неопределенном мире меньшее часто оказывается большим. Эффект «меньшее есть большее» описывается через обратную U-образную связь между уровнем точности и объемом информации (сложностью вычислений или затраченным временем). U-образное соотношение подразумевает, что в определенный момент увеличение объема полученной информации не улучшает качество решения, а напротив, вредит ему. Причина, по которой эффект «меньшее есть большее» может привести к положительным результатам, заключается в том, что эвристики не применяются случайным образом. Они адаптированы к конкретной среде принятия решений, в которой они работают (Mousavi and Gigerenzer, 2014; Gigerenzer and Gaissmaier, 2011).

Подход быстрых и экономных эвристик был развит в обилии различных исследований, демонстрирующих, что результаты применения простых эвристик сопоставимы или даже превосходят результаты статистического анализа и машинного обучения, включая логистическую регрессию, множественную регрессию, деревья решений, случайные леса, метод опорных векторов (Czerlinski et al., 1999;

Buckmann and Şimşek, 2016; Brighton, 2006). Обзор быстрых и экономных эвристик, а также сопоставление этого подхода с подходом эвристик и сдвигов представлен в работе Милковой и коллег (Milkova et al., 2019(a)).

1.2.3. Влияние поведенческих факторов на функционирование систем и рынков

Влияние поведенческих факторов также исследовалось применительно к функционированию систем разного вида. Так, Оливер Уильямсон, исследуя экономические организации, отмечал ограниченную рациональность и оппортунизм (следование своим интересам, в том числе преднамеренное представление неполной или искаженной информации) как источник трансакционных издержек (Williamson, 1985). Жан Тироле формализует аналитическую модель коллективной репутации, коррупции (Tirole, 1996). Джордж Акерлоф и Роберт Дж. Шиллер исследовали влияние поведенческих факторов на эффективность финансовой системы и функционирования экономики в целом (Shiller, 1998; Akerlof, 2002). Авторы выделяют социальные явления, так называемые «animal spirits», управляющие экономикой: стремление к уверенности (ожидание положительного развития событий в будущем), коррупция (стимулы в периоды ажиотажа получать прибыль от доверчивых людей, верящих в вечный рост), денежные иллюзии (когнитивная неспособность учесть инфляцию или дефляцию цен или заработной платы), справедливость и истории (явление социальной психологии, когда люди, пытаясь понять что-либо, начинают рассказывать истории, необязательно приближенные к реальности; так, у каждого экономического пузыря есть история успеха, квазибесконечного роста какого-либо сектора экономики) (Akerlof and Shiller, 2009).

Влияние поведенческих особенностей на функционирование фондового рынка исследовалось в рамках направления, получившего название поведенческих финансов: описаны явления чрезмерного реагирования (Bondt and Thaler, 1985), диспозиции (склонность к слишком быстрой продаже прибыльных акций и слишком долгому удержанию убыточных) (Shefrin and Statman, 1985), влияния ограниченного внимания инвесторов на доходность акций (DellaVigna and Pollet, 2009),

увеличения торговой активности из-за чрезмерной самоуверенности и многие другие эффекты. См., например, обзоры Barberis and Thaler (2003), Shleifer, 2000.

1.2.4. Управление ограниченной рациональностью

Приложения знаний об особенностях человеческого поведения при принятии решений нашли отражение в работах Ричарда Талера, внесшего большой вклад в становление поведенческой экономики как научной дисциплины на основе обширного числа экспериментальных и эмпирических исследований поведения (см. обзоры Белянин, 2018; Паниди, 2017). Талером было впервые предложено использовать отклонения от рациональных стандартов как инструмент экономической политики, позволяющей повысить общественное благосостояние. Так, теория подталкивания (или либертарианский патернализм) (совместно с Кассом Санстейном) (Thaler and Sunstein, 2008), основанная на склонности людей к выбору опции по умолчанию (см. Johnson and Goldstein, 2003), предполагала подталкивание людей к принятию оптимальных решений, не ограничивая при этом свободу выбора. Теория эффективно применялась для проведения различных политик: повышения нормы пенсионных сбережений с помощью автоматического включения работника в программу сбережений с заданной долей дохода (Benartzi and Thaler, 2013); увеличения числа учащихся в колледжах, благодаря упрощению подачи федеральных заявок на финансовую помощь; снижения потребления электроэнергии с помощью рассылки писем с информацией о сравнении уровня электропотребления с соседями; повышения числа вакцинируемых от гриппа среди взрослых путем планирования конкретной даты вакцинации и др. (Benartzi et al., 2017).

Как отмечается в работе Милковой (2019(a)), убеждающие исследования политики либертарианского патернализма привели к ее распространению. В ряде правительств по всему миру сформировались группы «подталкивания» (Nudge units), представляющие собой группы экспертов по поведенческой науке, разрабатывающие поведенческие вмешательства, способные стимулировать желаемое поведение без ограничения выбора. Так, в 2010 году в Великобритании было создано «подразделение по подталкиванию» (Behavioral Insights Team), аналогичные структуры были введены в Австралии, Германии, Голландии, США и др. По данным

Всемирного банка, ожидается дальнейшее распространение поведенческих идей в ближайшие годы⁶. Учет поведенческих особенностей необходим также и в правовом регулировании, например, важность отражения пренебрежения вероятностью (склонности индивидов завышать низкие вероятности негативных событий в случае, если риски вызывают сильные эмоции, включая страх) в законе подчеркивается в работе Санстейна (Sunstein, 2001).

В последнее время растет обсуждение так называемого цифрового подталкивания (digital nudging), использование которого также, по мнению некоторых исследователей, способствует улучшению благосостояния. Речь идет о разработках специальных приложений, способствующих росту накоплений (Hershfield et al., 2019), мониторингу расходов на мобильном телефоне (Levi and Benartzi, 2020), разработке широкого спектра рекомендательных систем, цифровому подталкиванию к решениям, касающимся здоровья, устойчивого развития и образования. Исследователи в данном направлении придерживаются взглядов, согласно которым будущее цифрового подталкивания должно включать в себя широкий набор инструментов, включая Big Data и персонализацию⁷.

Противники данного подхода отмечают различные негативные аспекты и недобросовестное использование цифрового подталкивания. Так, в работах Хелбинга, Гигеренцера и коллег (Helbing, et al., 2019) введен термин Big Nudging, рассматривающий подталкивание в условиях стремительного накопления информации о каждом из индивидов. Увеличивающийся объем личной информации об индивидах, который часто собирается без их согласия, раскрывает привычки, намерения и другую личную информацию, позволяет манипулировать людьми. Такого рода Big Nudging приводит к тому, что индивиды делают выбор, который в ином случае бы не делали, например, покупают некоторые товары с завышенной ценой или ненужные товары, или, возможно, отдают свой голос определенной

⁶ Afif, Z., Islan, W. W., Calvo-Gonzalez, O., Dalton, A.G. Behavioral Science Around the World: Profiles of 10 Countries (English). 2019. eMBeD brief. Washington, D.C.: World Bank Group. URL: <http://documents1.worldbank.org/curated/en/710771543609067500/pdf/132610-REVISED-00-COUNTRY-PROFILES-dig.pdf> (дата обращения: 25.09.2021)

⁷ Benartzi, S. How Digital Tools and Behavioral Economics Will Save Retirement. Harvard Business Review, December 07, 2017. URL: <https://hbr.org/2017/12/how-digital-tools-and-behavioral-economics-will-save-retirement> (дата обращения 25.09.2021)

политической партии (Милкова, 2019(б)). Иными словами, «подталкивание» переходит в «толкание» (Sætra, 2019), или «тенденция заключается в переходе от программирования компьютеров к программированию людей» (Helbing, et. al., 2019).

Значительное внимание в экономических исследованиях уделяется влиянию медиа, социальных сетей на формирование экономических, политических предпочтений (Sunstein, 2017). Так, на основе охвата вещания канала Fox News было показано влияние СМИ на результаты голосования (DellaVigna and Kaplan, 2007); отмечалось доминирующее влияние друзей и сообществ в пропаганде (на примере исследования пропаганды геноцида против тутси в Руанде) (Yanagizawa-Drott, 2014); выделялась ключевая роль автора сообщения, а не текста передаваемого сообщения (Martin and Marks, 2019); показывалось влияние отзывов в социальных сетях на представления об услугах (в области гостиничного бизнеса – Kim and Park, 2017).

Восприятие информации через социальные сети играет значительную роль в получении новостных знаний (Allcott et al., 2019a). Ряд работ указывает на отрицательную корреляцию между интенсивностью использования социальных сетей и субъективным благополучием (Allcot et al., 2019a), отмечает возникновение зависимостей от приложений, усиление политической поляризации (Sunstein, 2017). Свойство транссортативности социальных сетей обуславливает усиление эффекта «иллюзии большинства», когда непопулярная идея может восприниматься как популярная большей частью людей (Ngo et al., 2020). Социальные сети являются основным каналом распространения фейковых новостей (Allcott and Gentzkow, 2017).

Тема влияния фейковых новостей на политику и экономику, распространение новостей получила отдельное внимание в недавних исследованиях (Allcott et al., 2019b). На основе публикаций в сети Твиттер показано, что фейковые новости способны распространяться гораздо быстрее правдивых за счет формулировки новизны и эмоциональной подачи (Vosoughi et al., 2018). К причине, почему фейковые новости имеют воздействие, относится наличие «ленивого» мышления – нежелания мыслить. Так, в исследовании Pennycook and Rand (2018) о восприятии фейковых новостей показано, что аналитическое мышление используется для оценки

правдоподобия заголовков, независимо от того, соответствует ли сообщение имеющейся у воспринимающего идеологии. Схожие результаты о важности осведомленности показывает и экспериментальное исследование Jost et al. (2020).

Более подробное обсуждение особенностей мыслительного процесса и его места в экономических исследованиях приведено в следующем разделе.

1.3. Нейроэкономика: взгляд на принятие экономических решений со стороны нейронаук

1.3.1. Диффузная модель

Изучением вопросов принятия решений как следствия текущего уровня активности нейронов в различных областях головного мозга занимается новая междисциплинарная область – нейроэкономика (Glimcher and Rustichini, 2004; Glimcher et al, 2009). В основе данного подхода лежит теория Бенджамина Либета, который на основе ряда экспериментов предположил, что каждому свободному действию человека (двигательной активности, например, желанию поднять руку) предшествует акт бессознательной нейронной активности мозга (Libet et al, 1983). Позже схожие эксперименты были проведены с помощью функциональной магнитно-резонансной томографии (фМРТ), позволившей предсказывать решение испытуемого за несколько секунд до осознания выбора (Soon et al., 2008).

Базовой нейроэкономической нейронной моделью теории принятия решения (Ключарев и др., 2011) является так называемая диффузная модель, согласно которой решение принимается при достижении пороговой величины разницы в активности нейронов, отвечающих за альтернативные варианты решений. В процессе выбора задействовано большое число нейронных сетей, расположенных в различных областях мозга, отвечающих как за сознательное, так и за различные виды бессознательного поведения. Оказывая влияние на одну из областей мозга, возможно в значительной мере влиять на результат выбора.

Так, взаимодействию рациональных и эмоциональных систем посвящено множество исследований, например, эксперименты на основе игр, предложенных

поведенческими экономистами - Ультиматум (Güth et al., 1982), Диктатор (Kahneman, Knetsch and Thaler, 1986). Если поведенческие теории объясняют иррациональное поведение игроков (отсутствие максимизации собственного выигрыша) важностью соблюдения социальных норм и этических стандартов, то нейроэкономические исследования находят причины таких предпочтений. Было показано, что эмоциональные зоны мозга имеют больший вклад по сравнению с когнитивными областями и являются причиной иррационального поведения (Knoch et al., 2006), то есть эмоциональная реакция на несправедливость будет преобладать по сравнению с рациональной реакцией (Sanfey et al., 2003). Отмечено также, что даже совсем небольшие изменения эмоционального состояния, которые не имеют непосредственного отношения к той задаче, которая решается, влияют на экономические решения человека (Sanfey and Chang, 2008).

С другой стороны, автоматическая непроизвольная Система 1 обеспечивает быструю реакцию, а затем возникает детальная обработка на основе Системы 2. Однако, когда необходимо принять быстрое решение, решение на основе Системы 1 опережает выводы, сделанные посредством работы Системы 2. (Ключарев и др., 2011).

Нейронаука указывает, что в основе принятия экономических решений лежит взаимодействие между автоматическими и управляемыми процессами, между когнитивными и эмоциональными системами. Взаимодействие Системы 1 и Системы 2 можно также проиллюстрировать на примере того, что опытные специалисты, работа которых может быть связана с принятием молниеносных решений в условиях стресса и риска, в критической ситуации редко выбирают между двумя опциями, потому что в большинстве случаев в голову приходит только один вариант. Варианты, которые отклонены, просто не представляются к выбору. Сомнения – феномен Системы 2. Однако результаты работы Системы 1 также базируются на опыте, полученном ранее в ходе работы Системы 2. Верно и обратное: результаты работы Системы 2 базируются на восприятии и оценке информации Системой 1. Современные исследования показывают, что роль «рационального» суждения серьезно переоценивается людьми. Различие между двумя системами познания не

может быть объяснено как простое несоответствие между сознательными и бессознательными процессами (Sloman, 2014). Действительно, посредством самоанализа возможно осознавать любую форму мышления. Однако мы можем осознавать различные этапы работы Системы 2 – вершину айсберга, которая является результатом работы Системы 1, подверженной искаженному восприятию информации (Gronchi and Giovannelly, 2018).

1.3.2. Последствия нейробиологических особенностей принятия решений для экономики

В работе Колина Камерера и коллег (Camerer et al., 2004) отмечается, что отсутствие у людей интроспективного доступа к источникам своих собственных суждений о поведении и склонность приписывать их управляемым процессам имеет важные последствия для экономики. Такая неосознанная дискриминация может объяснить, например, почему идентичные резюме для приема на работу для кандидатов со статистически «белыми», а не «афроамериканскими» именами имеют на 50% больше шансов получить обратный звонок (Marianne and Mullainathan, 2004). Второй класс последствий связан с явлениями очевидного самообмана и самонакрутки, когда, например, экономические агенты (инвесторы, потребители, предприниматели) слишком оптимистично оценивают свои шансы на успех. Например, внимание в значительной степени контролируется автоматическими процессами, однако внимание, в свою очередь, определяет, какую информацию мы потребляем. Если внимание постоянно обращено к информации, которая для нас предпочтительна, то будет отмечаться чрезмерно оптимистичное ощущение личных способностей и шансов на успех.

Отметим, что нейроэкономические исследования нашли приложения к экономической теории, в рамках которых становится возможным моделировать мозговую деятельность с дуальной системой организации (dual-system organization) (например, в Brocas and Carrillo (2008) – с учетом доступной информации в различных областях мозга; важности, придаваемой близким по времени событиям по сравнению с отложенными; относительной полезности, придаваемой привлекательным благам по сравнению с непривлекательными).

В работе Камерера и коллег также показывается, что рассматриваемые экономистами характеристики людей, такие как временные предпочтения (предпочтение скорого вознаграждения отложенному), склонность к риску, альтруизм не являются стабильными во времени и постоянными для разных видов деятельности, зависят от результатов взаимодействия когнитивных и аффективных систем, автоматических и контролируемых процессов. Это несоответствие частично является результатом того факта, что предпочтения зависят от состояния. Схожий вывод делается и о противопоставлении интеллекта и ограниченной рациональности, задействие которых, согласно нейронаукам, сильно зависит от предметной области. Отмечается также наличие связи между системами мотивации и удовольствия; активация при получении денег тех же областей мозга, что и при получении «основных подкреплений» (например, еды), что означает, что деньги приносят прямое удовольствие, а не просто оцениваются за возможность того, что на них можно что-то купить.

Нейроэкономические исследования также доказывают, что сходной с Системой 1 является автоматическая реакция, представляющая конформность – корректировку собственного мнения/поведения в пользу большинства (Klucharev et al., 2009). Эффект массового сознания является естественной функцией, так как с эволюционной точки зрения отклонение от «оптимальной» стратегии большинства наказывается в ходе естественного отбора. Однако также очевидно, что в современных реалиях следование автоматической конформности может привести к неблагоприятным последствиям.

В свете обсуждения нейробиологических аспектов принятия решений и воздействия на результат оценивания важно рассмотреть такое явление, как когнитивный диссонанс – психологически дискомфортное состояние, возникающее при столкновении ярко противоречивых когнитивных элементов (информации, опыта и т.д.) об одном и том же объекте, явлении, событии. Теория когнитивного диссонанса, разработанная американским психологом Леоном Фестингером (Festinger, 1957) определяет, что человек стремится уменьшить неудобное состояние путем изменений ощущений или социальных установок (Festinger and Carlsmith, 1959).

Важным с экономической точки зрения аспектом теории когнитивного диссонанса является вывод о том, что не только выбор объясняется результатом процесса оценивания, но и наоборот – оценки меняются в зависимости от осуществленного выбора. Более того, корректировка оценок тем больше, чем сильнее возникающий когнитивный диссонанс (Colosio et al., 2017).

Смежной к теории когнитивного диссонанса является так называемая парадигма свободного выбора (*free-choice paradigm*) (Brehm, 1956), иллюстрирующая переоценку альтернатив после осуществления выбора. Так, эксперименты показывают, что акт выбора между двумя одинаково привлекательными альтернативами вызывает изменение предпочтений: предпочтение выбранной альтернативы увеличивается, тогда как предпочтение отклоненной альтернативы уменьшается (Izuma, et al., 2010).

Важным открытием в области нейробиологии является существование зеркальных нейронов (Gallese and Goldman, 1998), феномен которых связан с тем, что в тот момент, когда мы наблюдаем за действием других, у нас происходит активация тех же самых нейронов. Наличие таких нейронов упрощает обучение, а также частично объясняет социальное влияние.

Таким образом, вышеприведенные выводы дают понимание природы неустойчивости человеческих предпочтений, возможности легкой манипуляции выбором путем воздействия, например, на эмоциональное состояние, скорость поступления информации. Важными аспектами являются также системы внимания, обсуждению которых посвящен следующий раздел.

1.4. Внимание к информации в рамках экономических наук

1.4.1. Ограниченное внимание

Феномен внимания, к которому обращались изначально философы (James, 1980), изучался в области психологических наук. Внимание рассматривалось как фильтр, который отсеивает ненужную информацию (Broadbent, 1958); фильтр, не отсеивающий, а только ослабляющий нерелевантные сигналы (Treisman, 1960); как

усилие по распределению ограниченных ресурсов (Kahneman, 1973); как цикл восприятия, предполагающий активное предвосхищение событий на основе существующих схем и последующую модификацию схем в процессе сбора информации (Neisser, 1976) и многие другие концепции. В соответствии с современной теорией, внимание рассматривается в контексте двух встречных информационных потоков: потока восходящего - управляемого текущими данными, и потока нисходящего – определяемого обработкой информации, связанной с опытом, конкретной задачей, ее контекстом (Lavie, 2005; Фаликман, 2018). На стыке этих потоков возникает правильно или неправильно решенная задача. Таким образом, большое число разнообразных источников информации, отвлекающих стимулов (distractors), особенно связанных с эмоциями, способны истощать внимание, а, значит, и результирующее познание, определяемое первоначальной мотивацией (Милкова, 2020(a)).

В экономический дискурс понятие внимания вошло после работ Саймона, который в научном труде по административному поведению (Simon, 1947) подчеркивал, что ограниченная рациональность зависит от ограниченной способности людей к вниманию и служит оправданием существования организаций и институтов как средства ориентации внимания, которые направляют поведение членов группы, предоставляют промежуточные цели, стимулирующие действия. Позднее Саймон подчеркивал роль зрительной памяти, указывая на ее пределы и организацию (Simon, 1969). Фокусировка внимания только на нескольких характеристиках облегчает решение проблемы. В теории обработки информации, по Саймону, именно внимание определяет ту узкую избирательную область, которая выделяется в общем проблемном пространстве, внутри которой осуществляется поиск решений до достижения порогового уровня удовлетворенности (Simon, 1978). Саймон рассматривал взаимодействие между информацией и когнитивными способностями человека: переизбыток информации рождает нехватку внимания и необходимость более эффективного перераспределения внимания между многочисленными источниками информации (Simon, 1971).

Описанием вклада внимания в процесс познания занимался и Фридрих Хайек. Внимание направлено или ограничено определенным классом событий,

которые мы ожидаем и которые, как следствие, воспринимаем более отчетливо (Науек, 1952, р.139). Хайек описывает внимание через его упреждающий характер, как «состояние готовности к возбуждению», которое не ограничивается сознательным уровнем. Тем самым знание, которым сознательно манипулирует любой индивидуальный разум, — это лишь небольшая часть знания, которое в каждый момент времени способствовало бы успеху действий (Науек, 1960, р.24). Таким образом, Хайек подчеркивает причинно-следственную связь между особыми обстоятельствами и конкретным результатом, который вызывают обстоятельства в контексте «готовности к возбуждению». На основе создания условий для коммуникаций Хайек предполагал использовать внимание как способ расширения пределов знания, имеющих фрагментарную природу.

Колин Камерер (Camerer, 2003) выделял ограниченное внимание как важное понятие поведенческой экономики, отмечал крайнюю дефицитность внимания как когнитивного ресурса. Ограниченность внимания, по Камереру, объясняет такие экономические явления, как организационная структура (разделение труда расширяет организационное внимание, но ограничивается необходимостью координации) и реклама (которая «захватывает» внимание). Так, «можно научиться обращать внимание на множество стимулов одновременно — как это делают занятые трейдеры на финансовых рынках и водители, пользующиеся мобильными телефонами, — но страдает долговременная память» (Camerer, 2003, р.16).

1.4.2. Экономика внимания

Изучение внимания как важного звена в принятии экономических решений нашло отражение в направлении, получившем название «экономика внимания» (см. обзоры Festré and Garrouste, 2015; Почепцов, 2017а; Козырев, 2019; Милкова, 2020). Сам термин «экономика внимания» был введен параллельно, независимо друг от друга, двумя учеными: Георгом Франком (Franck, 1993) и Майклом Голдхабером (Goldhaber, 1997а). Термин экономика внимания был выдвинут Голдхабером как альтернатива «информационной экономике». Отталкиваясь от определения Самуэльсона, что экономика — это изучение того, как общество использует ограниченные ресурсы, Голдхабер подчеркивает, что информация в настоящее

время предстает перед нами в избытке, а дефицитным ресурсом является внимание, поэтому именно экономика внимания является естественным названием экономики в киберпространстве (Goldhaber, 1997b). В теориях и Франка, и Голдхабера внимание больше рассматривается не как ограниченный ресурс, который необходимо распределять, а как основная человеческая потребность. С ростом общего благосостояния общества стремление к отличию создает спрос на характеристики, которые являются более избирательными, чем денежный доход, поэтому значительная часть сегодняшней деятельности связана именно с транзакциями внимания, а не с денежными транзакциями (Голдхабер, 1997а). К интерпретации внимания как базовой потребности (желание получить внимание есть движущая сила любой деятельности) Франк добавляет представление внимания как аналога денег (внимание в информационном обществе есть мера стоимости), как валюты (пользователи платят за услугу своим вниманием, которое затем перепродается рекламодателям), как капитала (внимание — долговременное богатство: чем больше внимания вы получаете сейчас, тем большим «капиталом» будете обладать при последующем привлечении внимания) (Franck, 1993; 2005). Схожее понимание внимания как новой валюты бизнеса отмечается и в книге Томаса Давенпорта и Джона Бека (Davenport and Beck, 2001).

Активное использование социальных сетей, рост числа блогов, youtube-каналов является четким свидетельством того факта, что внимание (а не информация или знания) стало критически важным экономическим ресурсом для принятия решений. Многие экономисты обращаются к проблеме ограниченного внимания и ее экономическим последствиям, однако представление экономики внимания как отдельной области стоит особняком и едва ли получило распространение. Можно сказать, что основное развитие теории внимания в экономических науках происходит в двух направлениях: первое относится к теории информации, второе – к поведенческой экономике (Festré and Garrouste, 2015).

Согласно первому подходу, экономика внимания представляет собой ответвление от экономики информации. Внимание рассматривается в соответствии с принципами теории коммуникации Шеннона (Shannon, 1948). Подход

предполагает полную рациональность и учитывает субоптимальности, вызванные асимметричной информацией и информационной перегрузкой. Ключевые эмпирические вопросы относятся к принятию решений в условиях информационного загрязнения. Этот подход, представленный в литературе, например, работой Шапиро и Вариана (Shapiro and Varian, 1999), рассматривает внимание как дефицитный экономический ресурс и обсуждает последствия этого для пользователей информации. Анализируются различные бизнес-стратегии (ценовая дискриминация или дискриминация продуктов при наличии сетевых внешних факторов, затраты на переключение, совместное использование продуктов и т.д.) для конкретизации информации, чтобы привлечь внимание пользователей и монетизировать его, предоставляя более релевантную информацию. Но обычно дефицит внимания асимметрично влияет на пользователей и поставщиков информации. Например, проблема рассылки спама – типичный случай внешнего воздействия или информационного загрязнения в контексте асимметричной информации между отправителем электронных писем, который знает содержание сообщения и навязывает его, и получателем, который не знает содержания сообщения.

Интересно отметить, что Гарольд Хотеллинг еще в 1938 году предостерегал от вредного воздействия избыточного спроса на внимание со стороны средств массовой информации и рекламного бизнеса и предлагал облагать рекламу налогом с целью уменьшения ее количества (Hotelling, 1938).

Применительно к макроэкономике ограниченность внимания формулируется как рациональное невнимание – экономическая теория, которая допускает, что люди обладают конечной способностью обрабатывать информацию (в Шенноновском смысле) (Sims, 2003, 2010). Ограниченность внимания ограничивается шумом, который не ухудшает экономической рациональности. Как подчеркивает Кристофер Альберт Саймс (Sims, 2003), предположение об ограниченном внимании (или рациональном невнимании) обеспечивает микрооснование для экономического поведения, которое достаточно похоже на знакомое предположение о рациональном ожидании, чтобы гарантировать его полезность и управляемость.

Концепция рационального невнимания была выделена как одна из многообещающих, идущих на смену теории рациональных ожиданий, особенно после кризиса 2008 года⁸. В области финансовой теории Хиршлейфер и Теох (Hirshleifer and Teoh, 2003) анализируют последствия ограниченного внимания инвесторов для информационной политики компаний и рыночного равновесия. Джозеф Фолкингер (Falkinger, 2008) разрабатывает модель общего равновесия, в которой агенты задуманы как пассивные приемники сигналов, характеризующиеся ограниченным вниманием, а поставщики информации стремятся привлечь их внимание, отправляя сообщения.

Согласно второму подходу, экономика внимания лежит в области поведенческой экономики и интерпретирует внимание в рамках концепции ограниченной рациональности Саймона. Стоит отметить, что здесь есть два разделения, о которых упоминалось в подразделе 1.2.1: ограниченная рациональность может быть рассмотрена как упрощение принятия решений взамен точности (Канеман) или же, напротив, как эффективная адаптация к окружающей среде (Гигеренцер).

Как отмечается в работе Милковой (2020а), принимая во внимание ограниченную рациональность человека, восприятием легко управлять в информационном обществе, что и происходит в настоящее время: броские заголовки, эмоциональная окраска новостей, распространение информации в социальных сетях, рост числа личных блогов и видеоканалов – все это является «эффективными» каналами распространения и привлечения внимания к информации. Возрастающая конкуренция за внимание привела к росту новых технологий для привлечения внимания, нацеленных на максимизацию рейтингов, наполняемость каналов и т.п. Внимание колеблется, переключается сначала на стиль, а потом через него – на содержание (Lanham, 2006).

Происходящие изменения характеризуются кардинальной сменой распределения внимания, регулирующего взаимодействие людей с миром как на

⁸ Speech by Jean-Claude Trichet, President of the European Central Bank, Opening address at the ECB Central Banking Conference Frankfurt, 18 November 2010. Reflections on the nature of monetary policy non-standard measures and finance theory. URL: <https://www.ecb.europa.eu/press/key/date/2010/html/sp101118.en.html> (дата обращения: 25.09.2021)

индивидуальном, так и на социальном уровне (Roda, 2019). То, как мы распределяем внимание, определяет, какие решения мы принимаем, что, в свою очередь, влияет на экономику, политику и этику.

На уровне коллективного внимания управление действует через средства массовой информации: медиасреда определяет будущее внимание. В работах Георгия Почепцова значительная роль уделяется анализу «информационных войн», которые, как он отмечает, стали привычным явлением в наше время: «информационный инструментарий перестал просто описывать события, как это было раньше, а стал создавать их» (Почепцов, 2017б). Таким образом, человек, попадая в поток информации, практически не способен ему противостоять. Кроме того, социальные медиа стали «машинами эмоций», которые регулярно производят более яркие, четкие и мощные эмоции, чем эмоции «настоящие». «Человек видит мир так, как ему диктуют определенные информационные решетки. Это — язык, выделяющий значимые аспекты мира» (Почепцов, 2017б).

В работах Тима Ву вводится понятие рынков внимания, а также брокеров внимания (*attention brokers*), занимающихся перепродажей человеческого внимания (Wu, 2015). Внимание, привлекаемое путем предоставления бесплатных новостей, различных сервисов, развлечений и т.п. затем перепродается рекламодателям за деньги. Примерами брокеров внимания «в чистом виде» являются социальные сети, поисковые системы, новостные сайты, телевизионные каналы. Деятельность брокеров играет ключевую роль в проведении операций на рынках внимания, и именно поэтому Тим Ву, будучи юристом, подчеркивает необходимость принимать рынки внимания всерьез, а также разработать правовые нормы, регламентирующие взаимоотношения на рынках такого рода. Ву отмечает, что антимонопольная политика не успела приспособиться и распознать сложности и угрозы, созданные на таких рынках (Wu, 2017). В «слепой зоне» антимонопольного регулирования находятся компании, которые предоставляют свои продукты "бесплатно", однако на самом деле конкурируют на рынках внимания.

Экономика внимания учитывает также роль брендов, определяющих коммуникационные процессы, структурирующие и формирующие осведомленность и

внимание. Важным аспектом является то, что бренды и реклама служат антиконкурентным целям в том плане, что они могут сдерживать переключение между продуктами (Wu, 2017).

1.4.3. Экономические проблемы дефицита внимания

В работе Милковой (2020а) обращается внимание на то, что возрастающая конкуренция за внимание также приводит к увеличению объема контента, который производится за меньшее время. Недавние исследования экспериментально подтверждают, что социальное ускорение истощает внимание, стремление к «новизне» заставляет коллективно переключаться между различными темами гораздо быстрее (Lorenz-Spreen et al., 2019).

В настоящее время ученые сходятся во мнении, что способ передачи информации определяет стиль мышления: господство аудиовизуальных средств определяет его клиповый характер. Из-за невозможности осмыслить бурный поток информации, человек фиксирует сиюминутные события, теряя возможность дифференцировать мусорную и полезную информацию, и потребляет более простые по форме сведения (Докука, 2013). Таким образом, носители клипового мышления могут легко поддаваться манипуляциям: «чем больше давление мозаичной культуры, тем меньшую роль играет логика, тем более восприимчиво сознание к манипуляции» (Кара-Мурза, 2004).

«При столь серьезном изменении инфосферы мы обречены на трансформирование собственного сознания, т.е. того, как мы осмысляем свои проблемы, как обобщаем информацию, каким образом предвидим последствия наших поступков и действий» (Тоффлер, 2004). В области нейронаук известно, что определенный род деятельности может приводить в перспективе к изменениям в некоторых отделах головного мозга. К примеру, в одном из исследований было показано, что способность лицензированных лондонских таксистов приобретать и использовать информацию о движении в большом и сложном для навигации городе привела к увеличению объема серого вещества в гиппокампе (Maguire, et al., 2006).

Одним из подходов к решению экономических проблем, вызванных дефицитом внимания, может служить использование теории подталкивания (Талер и

Санстейн). Однако если предложенные изначально принципы работы подталкивания призваны были служить для повышения благосостояния общества, то на рынках внимания склонность к выбору опции по умолчанию может иметь негативные последствия (см. обсуждение цифрового подталкивания, подраздел 1.2.2).

Применительно к задаче поиска информации важно отметить изменение принципов работы современных поисковых систем. Так, введенные структурные элементы на странице поисковой выдачи (поисковых систем Google, Яндекс) отображают прямые ответы на запросы пользователей и используют склонность человека выбирать решение по умолчанию (готовый ответ), а также переоценку выгод от получения быстрого ответа, при недооценивании негативных последствий в виде возможной деградации системного мышления (Милкова, 2019). Поисковые системы, отображая прямые ответы на запросы пользователя, параллельно показывают персонализированные рекламные объявления: чем дольше пользователь находится на странице SERP, тем выше вероятность, что он перейдет по рекламной ссылке.

Кроме того, познавательная система человека склонна искать подтверждения уже выдвинутым гипотезам, а не критически проверять их. «В целом человек сегодня не добывает знания, а получает их готовыми» (Почепцов, 2019). Можно сказать, что в настоящее время потребность в получении быстрого решения заменяет потребность свободного выбора, о чем свидетельствует и широкое распространение дискриминации цен (плата за быструю покупку – более высокая цена на товар).

Исследования в области поведенческой экономики показывают, что создание системы временных стимулов способно сначала изменить краткосрочное поведение, затем сформировать привычку и тем самым сохранить поведение в долгосрочной перспективе (например, применительно к выработке привычки посещения тренажерного зала – Charness and Gneezy, 2009; повышению качества работы отдела продаж – Friebel et al., 2017). То же касается и стимулов получения быстрого ответа в процессе поиска информации (например, в поисковой системе) и формирования соответствующей привычки.

Еще одним проблемным полем является игнорирование ключевого различия между информацией и знанием, которое становится заметным именно при принятии в расчет важной функции внимания. Кеннет Эрроу уже отмечал, что одна и та же информация может иметь очень разную ценность для разных людей (Arrow, 1974). Кеннет Боулдинг также подчеркивал, что «мы не можем рассматривать знание как просто накопление информации в хранилище, даже если все сообщения, получаемые мозгом, могут оставлять там какой-то вклад. Само знание следует рассматривать как структуру, очень сложную и часто довольно рыхлую структуру <...> с ее частями, связанными различными способами узлами разной степени силы. Сообщения непрерывно помещаются в эту структуру; некоторые из них проходят прямо через её пустоты <...> не производя в них заметных изменений. Иногда сообщения «прилипают» к структуре и становятся ее частью <...> Иногда, однако, сообщение, которое несовместимо с основным паттерном ментальной структуры, но которое имеет природу, в которую нельзя не поверить, попадает в структуру, которая затем вынуждена подвергнуться полной реорганизации» (Boulding, 1955, pp. 103-104). Таким образом, феномен внимания дает понимание того, что различие между информацией и знаниями не сводится к различию между потоком и запасом.

1.5. Инструменты для исследования поведенческих особенностей принятия решений

1.5.1. Экспериментальные методы

Изучение поведенческих особенностей принятия решений осуществляется с помощью ряда инструментов. Классическим является метод контролируемого эксперимента, преследующий цель лабораторной проверки предсказаний экономической теории и изучения поведения экономических агентов – направление, получившее название экспериментальной экономики.

Основополагающей работой в области экспериментальной экономики можно назвать работу Вернона Смита, представляющую серию экспериментальных игр для изучения гипотез неоклассической теории конкурентного рынка (Smith, 1962),

однако некоторые неформальные эксперименты, проводимые в области принятия решений индивидами, были и ранее (например, работы У. Эдвардса, М. Алле, Д. Эллсберга). В настоящее время эксперименты распространены почти во всех областях экономики, например, в рамках изучения принятия решений в условиях риска и неопределенности; теории игр; теории рынков; теории аукционов; экономики информации; теории общественных благ; теории внимания и др. (см. Kagel and Roth, 1995; Plott and Smith, 2008)

Последнее время развивается направление проведения онлайн-экспериментов на краудсорсинговых интернет-площадках, таких как Amazon MTurk, Prolific.ac, Gorilla.ac, опросов в Qualtrics. Аналогичная российская краудсорсинговая площадка – Яндекс.Толока. Результаты проведения различных экономических игр онлайн показывают результаты, схожие с лабораторными, но с меньшими затратами на вознаграждение участников (Amir and Rand, 2012).

К инструментам поведенческой экономики можно отнести также и различные методы из области нейробиологии, самыми популярными являются методы нейровизуализации (фМРТ и другие).

1.5.2. Полевые методы

В последнее время всё чаще проводятся исследования в реальных условиях, вне лабораторий (например, DellaVigna, 2009). К инструменту, направленному не на анализ принятия решений, а на управление этим процессом можно отнести метод подталкивания, позволяющему в реальных условиях оценить свою работоспособность.

С развитием цифровой среды появилась большая база для проведения эмпирических исследований, демонстрирующих особенности формирования суждений. В данном случае актуальными инструментами могут служить методы анализа естественного языка (natural language processing, NLP), позволяющие оценивать влияние текстового информационного потока на ожидания (в том числе экономические), принятие решений и ценообразование.

Так, показано влияние эмоциональной окраски новостей на рыночный курс рубля (Афанасьев и др., 2019), настроений комментариев в сети Твиттер – на

ценообразование предметов искусства (Федорова и др. 2020]). В работе Голощаповой и Андреева описано исследование оценки экономических ожиданий населения по комментариям к СМИ в социальных сетях (Голощапова и Андреев, 2017), в работе Бэйкера анализ уровня экономической неопределенности проводился на основе частоты упоминаний в новостях определенных слов (Baker et al., 2016).

Одним из подтверждений актуальности применения методов NLP для поведенческих исследований, нужд поведенческой политики может служить открытие в ведущих институтах соответствующих направлений. Например, Стэнфордский университет открывает направление, занимающееся изучением новой терминологии, используемой в СМИ, для поиска более убедительных терминов в плане воздействия на граждан с целью задействования терминов в экологической политике⁹. Проводятся мероприятия, посвященные применению текстовых методов анализа для поведенческих и социальных исследований¹⁰.

Использование методов NLP для выявления различных трендов реализовано, например, в ресурсе Google Books Ngram Viewer¹¹ — онлайн-сервисе, позволяющем строить графики частотности n-грамм (словосочетаний, языковых единиц) на основе источников в Google Books.

Отметим, что в целом область анализа естественного языка является одной из наиболее динамично развивающихся областей искусственного интеллекта¹². Одним из основных направлений развития (обзор всех сфер применения см., например, в Большакова, 2017) является совершенствование методов извлечения информации, понимаемых как — «автоматическое экстрагирование значимых для человека данных, как правило, из большого массива текстов, и преобразование их в структурированную форму, что облегчает их последующую обработку и анализ»

⁹ Stanford Earth. Using Psychology, Behavioral Economics, and Linguistics to Study Environmental Attitudes and Behaviors. URL: <https://earth.stanford.edu/using-psychology-behavioral-economics-and-linguistics-study-environmental-attitudes-and-behaviors> (дата обращения 25.09.2021)

¹⁰ Text Mining for Behavioral and Social Sciences Research, семинар Национального института здравоохранения (National Institutes of Health, NIH), Управление поведенческих и социальных исследований (Office of Behavioral and Social Sciences Research, OBSSR). URL: <https://videocast.nih.gov/watch=33304> (дата обращения 25.09.2021)

¹¹ URL: <https://books.google.com/ngrams> (дата обращения 25.09.2021)

¹² Конференции, посвященные искусственному интеллекту в 2020 году отмечают две доминирующие области: анализ естественного языка и компьютерное зрение. - Милкова, М.А. OpenTalks.AI: Конференция 20-21 февраля 2020 года // Цифровая экономика. 2020. 1(9), с. 76-79. DOI: 10.34706/DE-2020-01-08)

(Большакова и Ефремова, 2017). Однако достижения в данной области в основном используются для внедрения в сторонние системы (рекомендательные, вопросно-ответные, поисковые и др.), которые, в свою очередь, призваны снижать затраты пользователей на поиск информации, способствовать решению проблем, связанных с дефицитом внимания.

Стоит также отметить развитие практики применения методов текстового анализа к социальным наукам (Ignatow and Mihalcea, 2017), например, для анализа дискурса и связей в социальных сетях (Arishev et al., 2016; Митрофанова и др., 2014), измерения легитимности (Etter et al., 2018), анализа мнений и настроений (Krishna et al., 2018) и др.

На основе описанных в предыдущих разделах идей о склонности индивидов к выбору опции (решения, ответа) по умолчанию, получению фрагментарных знаний, озабоченности манипуляцией вниманием и возможными долгосрочными последствиями в виде формирования привычки получения быстрого решения и деградации системного мышления, представляется важным применение инструментов анализа текстов для нужд повышения эффективности распределения внимания, сбережения человеческого капитала.

Выводы Главы 1

1. Рассмотрены ключевые аспекты работы с информацией в рамках экономики информации и экономики знаний. Информация в экономике информации играет ключевую роль в принятии экономических решений. Выделяют поиск информации как снижение неопределенности; затраты на поиск; определение оптимальной продолжительности поиска; разработку стратегии принятия решений о необходимости дополнительного сбора информации; модели информационных каскадов, стадного поведения; неприятия неоднозначности; рационального невнимания. Информация может быть наделена различными свойствами (асимметричная, неполная, зашумленная). К современным задачам в области экономики информации относятся исследования управления атрибутами и стоимостью информационных продуктов, с учетом различной ценности информации для разных покупателей,

изучаются оптимальные механизмы раскрытия информации, продажи информации (Милкова, 2021).

Если экономика информации носит аналитический и математический характер, то экономика знаний является преимущественно эмпирической и часто описательной. Экономика знаний прежде всего исследует экономические аспекты производства, распространения и управления знаниями. Схожим понятием является экономика, основанная на знаниях.

2. Рассмотрены различные аспекты ограниченной рациональности человека при обработке информации. Показаны особенности принятия решений в условиях неопределенности и риска: подход эвристический и сдвигов, фрейминг, теория перспектив, подход быстрых и экономных эвристических и связанные с этими особенностями разнообразные эффекты. Поведенческие аспекты принятия решений показаны также применительно к анализу экономических организаций, рынков. Отклонения от рациональных стандартов могут использоваться как инструмент экономической политики, направленной на повышение благосостояния: используются меры подталкивания, в том числе цифрового подталкивания. В современных поведенческих исследованиях большое внимание уделено влиянию медиа, социальных сетей и сообществ.

Показаны ключевые характеристики принятия решений на стыке экономических, психологических и нейронаук. Так, нейронаука указывает, что в основе принятия экономических решений лежит взаимодействие между автоматическими и управляемыми процессами, между когнитивными и эмоциональными системами, причем роль «рационального» суждения серьезно переоценивается людьми.

Работы, подчеркивающие и развивающие междисциплинарное сотрудничество экономики, психологии и социологии встречаются и в современной отечественной экономической науке (Клейнер, 2003; Полтерович, 2011; Жданов, 2018)

3. Систематизированы исследования теории внимания в экономических науках, описано понятие экономики внимания. Выделено два направления развития: первое относится к теории информации, второе – к поведенческой экономике. В первом направлении внимание рассматривается как дефицитный экономический

ресурс и обсуждаются последствия этого для пользователей информации (конкретизация информации, чтобы привлечь внимание пользователей и повысить их полезность, дискриминация цен, дискриминация продуктов и др.). Отдельно развивается теория рационального невнимания.

С другой стороны, ограниченное внимание - важное понятие поведенческой экономики. Внимание интерпретируется в рамках концепции ограниченной рациональности. Ограниченность внимания, склонность к когнитивным искажениям при восприятии информации, клиповый характер мышления современного человека объясняет актуальность разработки инструментов для эффективного распределения внимания в условиях перенасыщения информацией.

4. Показаны различные инструменты поведенческой экономики, к которым относятся: метод контролируемого лабораторного эксперимента – для проверки предсказательной точности экономической теории и изучения поведения агентов; метод визуализации мозговой активности – для понимания нейробиологических принципов принятия решения; метод подталкивания – для управления принятием решений; анализ естественного языка (извлечение информации из текстов) – для проведения эмпирических исследований, демонстрирующих особенности формирования суждений, а также для автоматического извлечения из больших массивов текстов значимой для человека информации. Отмечено, что достижения в области анализа естественного языка в основном используются для внедрения в сторонние системы (рекомендательные, вопросно-ответные, поисковые и др.), которые, в свою очередь, призваны снижать затраты пользователей на поиск информации, способствовать решению проблем, связанных с дефицитом внимания.

5. В экономических исследованиях в настоящее время отведено место человеку, обладающему ограниченной рациональностью, учитывающему при принятии решений личные установки, общественные нормы и другие внешние факторы. Если моделированию ограниченной рациональности в экономических науках посвящен ряд работ, то развитию инструментов по управлению ограниченной рациональностью уделено значительно меньше внимания. Инструмент подталкивания показывает положительные результаты во множестве исследований, однако

применимость его значительно ограничена в том плане, что формирование привычки выбора опции по умолчанию имеет негативное долгосрочное воздействие на развитие человеческого капитала. Стоит отметить, что к подталкиванию, облегчающему выбор индивида, присоединяется и борьба за внимание как наиболее дефицитный ресурс.

Дизайн цифрового мира влияет на поведение и принимаемые решения, и использование достижений поведенческих наук в сочетании с цифровыми технологиями способно оказывать существенное воздействие. Таким образом, учитывая особенности поведения индивидов, инструменты поведенческой экономики должны подсказывать, каким образом необходимо сконфигурировать среду для восприятия информации, чтобы 1) снизить затраты на поиск ключевой информации; 2) обеспечить полноту и ценность (с точки зрения смысла) извлекаемой информации; 3) минимизировать сдвиги в восприятии (сместить фокус внимания с первой попавшейся, эмоциональной информации).

ГЛАВА 2. МЕТОДИКА ИСПОЛЬЗОВАНИЯ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ ПРИ РАБОТЕ С НАУЧНОЙ ИНФОРМАЦИЕЙ

Дефицит внимания в эпоху цифровых технологий остро ставит вопрос активного внедрения дополнительных подходов к поиску и представлению информации. Так, простого информационного поиска (на базе поисковых систем Google, Яндекс и т.п.) уже недостаточно для оперативного и адекватного получения информации (Милкова, 2019(б)). Например, при решении некоторых задач существует потребность при поиске информации формулировать длинный, разнородный запрос, при этом максимально учитывающий уже накопленные знания по анализируемой проблеме. Также стоит отметить, что при исследовании новых областей, формулирование поискового запроса в виде ключевых фраз может быть вовсе затруднительным.

Развитие подходов к поиску и представлению информации актуально как на уровне получения знаний и принятий решений отдельным индивидом, так и на макро уровне. Так, неупорядоченный характер работы с информацией, отсутствие необходимых навыков и инструментария является ключевым фактором, препятствующим распознаванию грядущих инноваций и предвидению их последствий (Миловидов, 2019).

Подход на основе тематического моделирования позволяет получить дорожную карту, структуру больших текстовых коллекций, объем которых превышает возможности человека для самостоятельного чтения и анализа. В данной главе приводится описание методологии тематического моделирования, а затем, с учетом обозначенных сложностей, представляется способ формализации алгоритма применительно к такому информационному продукту как научные публикации.

Тематическое моделирование является одним из подходов так называемого разведочного поиска — *exploratory search* (Marchionini, 2006; White and Roth, 2009). Если привычный нам итерационный поиск работает в формате «вопрос-ответ», отвечает на короткий четко сформулированный запрос, то разведочный поиск

характеризует отсутствие точной формулировки запроса и отсутствие единого ответа (Янина и Воронцов, 2016). Разведочный поиск в большей степени связан с полнотой (recall) (максимизация числа извлеченных релевантных объектов), чем с точностью (precision) (минимизация числа извлеченных нерелевантных объектов) и, таким образом, не очень хорошо поддерживается современными поисковыми системами, которые настроены на точность в первой странице результатов выдачи (Marchionini, 2006). Разведочный поиск требует активного участия человека в более продолжительном, аналитическом исследовательском процессе. Несмотря на высокий уровень современных поисковых систем, их концепция не эффективна для проведения разведочного поиска.

Разведочный поиск, как правило, охватывает определенные информационные ресурсы, на которых пользователи могут формулировать аналитические запросы, использовать комплексные службы просмотра результатов, позволяющие представить и осуществлять навигацию по структуре, основным темам, ключевым словам и другой информации.

2.1 Общее описание инструментария

Инструментарием, позволяющим проводить разведочный поиск, является тематическое моделирование (соответствующий поиск — тематический поиск) (Янина, Воронцов, 2016; Scherer et al., 2013; Grant et al., 2015). Тематическое моделирование – одно из активно развивающихся с конца 90-х годов направлений анализа больших объемов текстовой информации. Тематическая модель определяет структуру коллекции текстовых документов путем выявления скрытых тем в документах, а также термов (слов или словосочетаний), характеризующих каждую из тем.

Построение тематической модели рассматривается как задача одновременной кластеризации документов и слов по одному и тому же множеству кластеров – тем. Тема – результат би-кластеризации – одновременной кластеризации и слов, и документов по их семантической близости. Обычно выполняется мягкая

кластеризация – документ может принадлежать нескольким темам в различной степени. Таким образом, сжатое семантическое описание слова или документа представляет собой вероятностное распределение на множестве тем. Процесс нахождения этих распределений - тематическое моделирование (Daud et.al., 2010).

Исходными данными для построения тематической модели являются: 1) коллекция текстовых документов \mathcal{D} ; 2) множество слов (или, что то же, терминов) \mathcal{W} (представлено в виде словаря). Каждый документ $d \in \mathcal{D}$ есть последовательность слов $\mathcal{W} = \{w_1, \dots, w_{n_d}\}$, где n_d — длина документа. n_{dw} - число вхождений слова w в документ d . Необходимо описать каждую тему дискретным распределением на множестве слов, каждый документ — дискретным распределением на множестве тем. Считается, что коллекция документов — это последовательность слов, выбранных случайно и независимо из смеси таких распределений и необходимо восстановить компоненты смеси по выборке (Воронцов, Потапенко, 2012):

$$p(w|d) = \sum_{t \in T} p(w|t)p(t|d) = \sum_{t \in T} \varphi_{wt} \theta_{td} \quad (1)$$

Предположения тематического моделирования:

- гипотеза «мешка слов» (bag of words model) – предположение о том, что для выявления тематики текстов важна только частота слов в документах, но не их порядок;
- порядок следования документов в коллекции может быть любым;
- количество тем определяется заранее и не меняется
- в зависимости от подхода к построению тематической модели количество слов в теме может как задаваться заранее, так и выбираться экспертно уже после построения модели

Отметим, что в развитии тематического моделирования удастся отойти от перечисленных предположений, расширив границы применимости моделей. Однако в данной работе мы остаемся в рамках данных предположений.

В настоящее время существуют различные подходы к построению тематических моделей (см. обзоры Милкова, 2019; Boyd-Graber et al., 2017; Daud et al., 2010). Базовыми моделями являются Вероятностный латентно-семантический анализ

(Probabilistic Latent Semantic Analysis, PLSA) (Hofmann, 1999) и Модель скрытого размещения Дирихле (Latent Dirichlet Allocation, LDA) (Blei et al., 2003).

Тематическое моделирование применяется для решения широкого спектра задач: анализ научных статей, исторических документов, политических текстов, сентимент анализ, анализ социальных сетей и медиа публикаций и др. (см. общий обзор применений в Boyd-Graber et al., 2017).

Анализ литературы показывает, что LDA лидирует среди вероятностных тематических моделей благодаря многочисленным обобщениям, расширениям и приложениям к анализу коллекций текстовых документов (см. Милкова, 2019(б)). Однако в работах Воронцов и Потапенко (2012), Potapenko and Vorontsov (2013), в которых критически пересмотрен взгляд на PLSA и LDA, отмечено, что широкое распространение LDA объясняется скорее его чисто математическим удобством для байесовского обучения, подчеркивается, что априорные распределения Дирихле и их обобщения не имеют убедительных лингвистических обоснований. Более того, переход от порождающей модели к алгоритму настройки её параметров требует весьма громоздких выкладок, которые резко усложняются при введении более сложных априорных распределений или совместном моделировании нескольких языковых явлений.

По этим причинам мощный импульс получило развитие так называемого подхода Аддитивной регуляризации тематических моделей (Additive Regularization of Topic Models, ARTM) (Воронцов, Потапенко, 2014(а)), в основе которого лежит модель PLSA. ARTM – многокритериальный подход, в основе которого лежит представление задачи тематического моделирования как некорректно поставленной оптимизационной задачи, требующей введения регуляризатора – дополнительного критерия, учитывающего специфические особенности прикладной задачи или знания предметной области (Воронцов, Потапенко, 2014(а); Vorontsov and Potapenko, 2014(а)). ARTM представляет собой общий подход к тематическому моделированию как к задаче многокритериальной оптимизации (Воронцов, Потапенко, 2014(а)).

В работе Vorontsov and Potapenko (2014(a)) пересматриваются тематические модели, ранее разработанные в рамках байесовского подхода, для каждой из которых находится соответствующий регуляризатор, который приводит к тому же самому или очень похожему алгоритму обучения модели. По сравнению с байесовским подходом, ARTM радикально упрощает вывод алгоритма и позволяет комбинировать регуляризаторы в произвольных сочетаниях. Также, в недавних исследованиях показано превосходство ARTM над LDA по качеству выделенных тем (см., например, работу Apishev et.al., 2017, где ARTM и LDA сравниваются на примере мониторинга этнически обусловленного дискурса в социальных сетях).

Данная работа основывается именно на подходе аддитивной регуляризации как на наиболее гибком (обобщение всех подходов), эффективном (быстродействие) и удобном (наличие библиотеки Bigartm (Frei, Apishev, 2016) в программной среде Python) инструменте.

2.2 Подход аддитивной регуляризации тематических моделей

2.2.1. Постановка задачи

Подход аддитивной регуляризации разработан К.В. Воронцовым и коллегами и основан на представлении задачи тематического моделирования как задачи стохастического матричного разложения заданной матрицы частот слов (термов) в документах

$$F = (p_{wd})_{W \times D}, \quad p_{wd} = p(w|d) = \frac{n_{dw}}{n_d},$$

Задача сводится к поиску приближённого представления заданной матрицы частот F в виде произведения двух матриц меньшего размера:

матрицы термов в темах Φ

$$\Phi = (\varphi_{wt})_{W \times T}, \quad \varphi_{wt} = p(w|t) = \frac{n_{wt}}{n_t}$$

матрицы тем в документах Θ

$$\Theta = (\theta_{td})_{T \times D}, \quad \theta_{td} = p(t|d) = \frac{n_{td}}{n_d}$$

$$F \approx \Phi \Theta$$

Матрицы F, Φ, Θ - стохастические, т.е. имеют неотрицательные нормированные столбцы, представляющие дискретные распределения.

Для оценивания параметров Φ, Θ тематической модели (1) по коллекции документов \mathcal{D} максимизируется логарифм правдоподобие выборки при ограничениях неотрицательности и нормированности столбцов матриц Φ, Θ :

$$L(\Phi, \Theta) = \ln \prod_{d \in \mathcal{D}} \prod_{w \in \mathcal{W}} p(w|d)^{n_{dw}} = \sum_{d \in \mathcal{D}} \sum_{w \in \mathcal{W}} n_{dw} \ln \sum_{t \in \mathcal{T}} \varphi_{wt} \theta_{td} \rightarrow \max_{\Phi, \Theta}$$

$$\sum_{w \in \mathcal{W}} \varphi_{wt} = 1, \quad \varphi_{wt} \geq 0, \quad (2)$$

$$\sum_{t \in \mathcal{T}} \theta_{td} = 1, \quad \theta_{td} \geq 0$$

В работах Воронцова и Потапенко (Воронцов, Потапенко 2014(a); Vorontsov and Potapenko, 2014(b)) сформулирована и доказана теорема о том, что стационарная точка данной оптимизационной задачи удовлетворяет системе уравнений со вспомогательными переменными p_{tdw}, n_{wt}, n_{td} .

$$p_{tdw} = p(t|d, w) = \frac{n_{tdw}}{n_{dw}} - \text{вероятность, что терм } w \text{ из документа } d \text{ принадлежит теме } t$$

жит теме t

По формуле Байеса вероятность p_{tdw} можно записать как:

$$p(t|d, w) = \frac{p(w, t|d)}{p(w|d)} = \frac{p(w|t)p(t|d)}{p(w|d)} = \frac{\varphi_{wt}\theta_{td}}{\sum_s \varphi_{ws}\theta_{sd}} \quad (3)$$

$$\varphi_{wt} = \frac{n_{wt}}{n_t}, \quad n_{wt} = \sum_d n_{tdw} = \sum_d n_{dw} p_{tdw}, \quad n_t = \sum_{w \in \mathcal{W}} n_{wt} \quad (4)$$

$$\theta_{td} = \frac{n_{td}}{n_d}, \quad n_{td} = \sum_w n_{tdw} = \sum_w n_{dw} p_{tdw}, \quad n_d = \sum_{t \in \mathcal{T}} n_{td} \quad (5)$$

Для решения системы уравнений (3)-(5) используется EM-алгоритм. На первом этапе выбирается начальное приближение для $\varphi_{wt}, \theta_{td}$. Отметим, что существуют различные способы для формирования начальных приближений, однако самая распространенная рекомендация - задание приближений нормированными случайными векторами из равномерного распределения. На E-шаге (3) вычисляются вспомогательные переменные p_{tdw} . На M-шаге (4), (5) вычисляются частотные оценки максимального правдоподобия для искомым условных вероятностей $\varphi_{wt}, \theta_{td}$. Вычисления (3)-(5) продолжаются в цикле до сходимости.

2.2.2. Принцип регуляризации

Ключевым моментом является не единственность стохастического матричного разложения $\Phi\theta$, определяемого с точностью до невырожденного преобразования $\Phi\theta = (\Phi S)(S^{-1}\theta)$ (задача является некорректно поставленной). Выбор преобразования S никак не контролируется и зависит от случайного начального приближения. Согласно теории регуляризации¹³, решение такой задачи возможно определить и сделать устойчивым. Для этого к основному критерию добавляют дополнительный критерий — регуляризатор, учитывающий специфические особенности данной задачи и знания предметной области.

Таким образом, наряду с правдоподобием (2) требуется максимизировать r критериев- регуляризаторов $R_i(\Phi, \theta)$, $i=1, \dots, r$ с неотрицательными коэффициентами регуляризации τ_i . Таким образом, оптимизационная задача примет вид:

$$\begin{aligned} L(\Phi, \theta) &= \sum_{d \in D} \sum_{w \in d} n_{dw} \ln \sum_{t \in T} \varphi_{wt} \theta_{td} + R(\Phi, \theta) \rightarrow \max_{\Phi, \theta} \\ R(\Phi, \theta) &= \sum_{i=1}^r \tau_i R_i(\Phi, \theta) \\ \sum_{w \in W} \varphi_{wt} &= 1, \quad \varphi_{wt} \geq 0, \\ \sum_{t \in T} \theta_{td} &= 1, \quad \theta_{td} \geq 0 \end{aligned} \quad (6)$$

В работах Vorontsov and Potapenko (2014(b)), Воронцов и Потапенко (2014(a)) показано, что точка локального экстремума задачи (6) удовлетворяет системе уравнений (7-9) со вспомогательными переменными p_{tdw} , при исключении из решения нулевых столбцов матриц Φ, θ (функция $R(\Phi, \theta)$ непрерывно дифференцируема).

Перепишем равенства (3)-(5), введя оператор нормировки $norm$, который преобразует произвольный вектор в вектор вероятностей дискретного распределения путем обнуления отрицательных элементов и нормировки $p_i = norm_{i \in I}(x_i) = \frac{(x_i)_+}{\sum_{j \in I} (x_j)_+}$, $i \in I$:

$$p_{tdw} = norm_{t \in T}(\varphi_{wt} \theta_{td}) \quad (7)$$

$$\varphi_{wt} = norm_{w \in W} \left(n_{wt} + \varphi_{wt} \frac{\partial R}{\partial \varphi_{wt}} \right), \quad n_{wt} = \sum_d p_{tdw} n_{dw} \quad (8)$$

¹³ Тихонов, А. Н., Арсенин, В. Я. Методы решения некорректных задач. М.: Наука, 1986.

$$\theta_{td} = \text{norm}_{t \in T} \left(n_{td} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \right), \quad n_{td} = \sum_w n_{dw} p_{tdw} \quad (9)$$

Таким образом, для решения задачи (6) применяется так называемый регуляризованный EM-алгоритм, E-шаг (7) которого остается прежним, как в задаче (2), а M-шаг заменяется регуляризованными уравнениями (8-9).

В результате решения системы уравнения (7-9) некоторые столбцы матриц Φ, Θ могут оказаться нулевыми, что соответствует вырожденным темам и документам. В случае, если тема вырождена, считается, что она не релевантная и исключается из модели. Аналогично, если документ является вырожденным, считается, что определить его тему невозможно (например, потому что документ слишком короткий), он исключается из модели.

Вырожденность является следствием чрезмерного разреживающего действия регуляризатора R . Если вырожденность нежелательна, ее можно избежать путем уменьшения коэффициента регуляризации. В целом, коэффициенты τ_i необходимо подобрать экспериментально, чтобы найти компромисс между всеми критериями.

Задача адекватного описания текстов на естественном языке и выделения интерпретируемых тем, накладывает на вероятностную тематическую модель большое число требований. Рассмотрим регуляризаторы, направленные на учет лингвистических особенностей текста и повышение интерпретируемости тем.

В работах Воронцова и Потапенко предлагается конструировать регуляризаторы на основе дивергенции Кульбака-Лейблера (KL-дивергенция, относительная энтропия).

Дивергенции Кульбака-Лейблера – несимметричная функция расстояния между дискретными распределениями, мера вложенности двух распределений. Она неотрицательна и равна нулю тогда и только тогда, когда два распределения совпадают.

Дивергенции Кульбака-Лейблера распределения Q относительно P определяется как:

$$KL(P||Q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \quad (10)$$

Таким образом, минимизировать дивергенцию Кульбака-Лейблера есть то же, что и максимизировать правдоподобие модели распределения Q по эмпирическому распределению P (Воронцов, Потапенко, 2014).

Регуляризатор разреживания

Регуляризатор разреживания формализует предположение о том, что каждый документ относится к небольшому числу тем, равно как и терм характеризует небольшое число тем. Чем больше разрежено распределение, тем меньше его энтропия (мера неопределенности). Максимальной энтропией обладает равномерное распределение. Таким образом, необходимо максимизировать KL-дивергенцию между распределениями φ_t , θ_d и равномерными распределениями $\tilde{\beta}_w$, $\tilde{\alpha}_t$ соответственно (Воронцов, 2014):

$$\sum_{t \in T} KL_w(\tilde{\beta}_w || \varphi_{wt}) \rightarrow \max_{\Phi}, \quad \sum_{d \in D} KL_t(\tilde{\alpha}_t || \theta_{td}) \rightarrow \max_{\Theta}, \quad \text{где}$$

Индекс w, t в обозначениях KL_w , KL_t указывает на индекс суммирования при вычислении KL-дивергенции.

$$\tilde{\beta} = (\tilde{\beta}_w)_{w \in W} = \left(\frac{1}{|W|} \right),$$

$$\tilde{\alpha} = (\tilde{\alpha}_t)_{t \in T} = \left(\frac{1}{|T|} \right)$$

Складывая два функционала с коэффициентами β_0 , α_0 и удаляя из суммы константы, получаем регуляризатор

$$R(\Phi, \Theta) = -\beta_0 \sum_{t \in T} \sum_{w \in W} \tilde{\beta}_w \ln \varphi_{wt} - \alpha_0 \sum_{d \in D} \sum_{t \in T} \tilde{\alpha}_t \ln \theta_{td} \rightarrow \max \quad (11)$$

Тогда формулы M-шага (8), (9) будут иметь вид (переопределим $\beta_w = \beta_0 \tilde{\beta}_w$, $\alpha_t = \alpha_0 \tilde{\alpha}_t$):

$$\varphi_{wt} = \text{norm}_{w \in W} (n_{wt} - \beta_w) \quad (12)$$

$$\theta_{td} = \text{norm}_{t \in T} (n_{td} - \alpha_t), \quad (13)$$

В работе Воронцов и Потапенко (2012) отмечено, что применение разреживания на первых итерациях может приводить к снижению качества модели, поэтому регуляризатор разреживания рекомендуется подключать на последних этапах, когда алгоритм близок к сходимости.

Регуляризатор сглаживания

Если разреживающий регуляризатор формализует специфические для каждой темы термы, то сглаживающий регуляризатор наоборот, призван сгруппировать слова общей лексики. Объединим общеупотребительные термы в фоновые темы B , тогда как специфические термы будут находиться в предметных темах S . Предполагается, что фоновые темы присутствуют во всех документах и содержат с ненулевой вероятностью высокочастотные термы.

Таким образом, для фоновых тем регуляризатор сглаживания можно представить, минимизировав дивергенцию Кульбака-Лейблера между фоновыми компонентами распределений φ_t , θ_d и заданными распределениями $\tilde{\beta}_w$, $\tilde{\alpha}_t$ соответственно:

$$\sum_{t \in T} KL_w(\tilde{\beta}_w || \varphi_{wt}) \rightarrow \min_{\Phi} , \sum_{d \in D} KL_t(\tilde{\alpha}_t || \theta_{td}) \rightarrow \min_{\Theta}$$

Складывая два функционала с коэффициентами β_0 , α_0 , удаляя из суммы константы, и перейдя от \min к \max , получаем регуляризатор:

$$R(\Phi, \Theta) = \beta_0 \sum_{t \in T} \sum_{w \in W} \tilde{\beta}_w \ln \varphi_{wt} + \alpha_0 \sum_{d \in D} \sum_{t \in T} \tilde{\alpha}_t \ln \theta_{td} \rightarrow \max \quad (14)$$

Как видно, формула сглаживающего регуляризатора (14) отличается от разреживающего регуляризатора (11) только знаком.

Аналогично (12), (13), формулы М-шага будут иметь вид:

$$\varphi_{wt} = \text{norm}_{w \in W} (n_{wt} + \beta_w) \quad (15)$$

$$\theta_{td} = \text{norm}_{t \in T} (n_{td} + \alpha_t), \quad (16)$$

Таким образом видно, что сглаживание и разреживание описываются одинаково (отличаются знаком параметров β_w , α_t).

Регуляризатор декоррелирования тем

Для улучшения интерпретируемости тем, логично предположить, что темы должны быть максимально различными: каждая предметная тема характеризуется своим набором специфических термов. Для формализации требования различности тем вводится регуляризатор, который минимизирует ковариации между φ_t , φ_s :

$$R(\Phi) = -\gamma \sum_{t \in T} \sum_{s \in T \setminus t} \text{cov}(\varphi_t, \varphi_s) \rightarrow \max,$$

$$\text{cov}(\varphi_t, \varphi_s) = \sum_{w \in W} \varphi_{wt} \varphi_{ws}$$

Тогда формула М-шага для φ_{wt} будет иметь вид (критерий не зависит от Θ , поэтому для θ_{td} формула М-шага не меняется):

$$\varphi_{wt} = \text{norm}_{w \in W} (n_{wt} - \gamma \varphi_{wt} \sum_{s \in T \setminus t} \varphi_{ws}) \quad (17)$$

Стратегии регуляризации

Эксперименты на реальных данных показывают (Vorontsov and Potapenko, 2014), что комбинирование регуляризаторов сглаживания, разреживания и декоррелирования приводят к построению интерпретируемой модели. Однако регуляризаторы могут конфликтовать друг с другом, ухудшать качество модели или приводить к ее вырождению (Воронцов, Потапенко, 2014а). В работе Воронцов, Потапенко (2014б) выработаны рекомендации, согласно которым регуляризаторы декоррелирования и сглаживания целесообразно включать в начале, а регуляризатор разреживания, когда модель начинает сходиться.

Основная задача при построении аддитивно регуляризованной тематической модели заключается в подборе траектории регуляризации – функции коэффициента регуляризации от номера итерации и критериев качества модели. Траектории регуляризации подбираются, анализируя их влияние на критерии качества модели в ходе итераций.

Далее рассмотрим основные меры качества тематических моделей.

2.2.3. Критерии качества

Измерение качества тематической модели ведется по совокупности критериев – оценивается *точность* и *интерпретируемость* модели.

В качестве критерия *точности* описания коллекции документов используется перплексия. Перплексия показывает, насколько хорошо модель приближает наблюдаемые частоты появления слов в документах. Точность модели тем выше, чем меньше перплексия.

$$\text{perplexity}(\mathcal{D}; p) = \exp \left(-\frac{1}{n} L(\Phi, \Theta) \right) = \exp \left(-\frac{1}{n} \sum_{d \in \mathcal{D}} \sum_{w \in \mathcal{d}} n_{dw} \ln p(w|d) \right)$$

Интерпретируемость модели оценивается степенью разреженности матриц Φ и Θ (долей нулевых элементов в матрице), а также когерентностью.

Тема считается когерентной, если характеризующие ее наиболее вероятные слова неслучайно часто встречаются вместе. В ряде работ установлено, что именно когерентность коррелирует с человеческими оценками интерпретируемости тем (Mimno, et.al. (2011), Newman, et.al. (2010)).

Когерентность темы t (coherence) выражается через среднюю поточечную взаимную информацию (Pointwise Mutual Information, PMI) по всем парам k наиболее вероятных слов темы t :

$$coherence = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=1}^k PMI(w_i, w_j)$$

где w_i — i -й термин в порядке убывания φ_{wt} ; $PMI(w_i, w_j)$ оценивает, насколько термы w_i, w_j не случайно встречаются рядом (Потапенко, 2018).

$$PMI(w_i, w_j) = \ln \frac{p(w_i, w_j)}{p(w_i)p(w_j)} = \ln \frac{n_{w_i, w_j}^{|D|+\varepsilon}}{n_{w_i} n_{w_j}},$$

число наиболее вероятных слов в теме k обычно берется около 10-15;

ε вводится для избегания нуля под логарифмом (Stevens et al., 2012), в работе $\varepsilon = 10^{-6}$.

Дополнительно, в работе Воронцов и Потапенко (2014а) предлагается контролировать чистоту и контрастность темы. В работе предполагается, что интерпретируемая тема имеет ядро W_t — множество слов, имеющих высокую условную вероятность $p(t|w) = \varphi_{wt} \frac{n_t}{n_w}$ для данной темы (например, больше 0.25):

$$W_t = \{w \in W | p(t|w) > 0.25\}$$

Размер ядра - $kernel_t = |W_t|$

Чистота темы показывает, насколько определяющими являются термины внутри темы - вычисляется суммарная вероятность термов ядра темы:

$$purity_t = \sum_{w \in W_t} p(w|t)$$

Контрастность темы показывает, насколько хорошо ядро темы отличает её от остальных тем, т.е. вероятность встретить термины ядра именно в данной теме.

$$contr_t = \frac{1}{|W_t|} \sum_{w \in W_t} p(t|w)$$

Общие размер ядра, чистота и контрастность модели вычисляются как средние по всем темам t .

2.2.4. Мультимодальная АРТМ

Помимо построения модели на основе текста, АРТМ подход позволяет строить так называемую мультимодальную тематическую модель. Под модальностями понимаются метаданные, так или иначе характеризующие тематику текста. К модальностям могут относиться: биграммы (n-граммы), тэги, классы, авторы, цитируемые или цитирующие документы и т.п.

Пусть M – множество модальностей. Каждая модальность имеет свой словарь термов W^m , $m=1, \dots, M$. В мультимодальной АРТМ функция логарифма правдоподобия вводится для каждой модальности (Vorontsov et. al., 2015):

$$L_m(\Phi^m, \theta) = \sum_{d \in D} \sum_{w \in W^m} n_{dw} \ln p(w|d) \rightarrow \max_{\Phi^m, \theta}$$

Где n_{dw} – число вхождений термина $w \in W^m$ в документ d .

Таким образом, в мультимодальной модели матрица Φ определяется для каждой модальности, а матрица θ является общей. Оптимизационная задача представляет собой максимизацию взвешенной суммы лог-правдоподобий и r-регуляризаторов при условиях нормировки и неотрицательности столбцов матриц Φ^m, θ (Янина, Воронцов, 2016):

$$\sum_{m \in M} \frac{\tau_m}{n_m} \sum_{d \in D} \sum_{w \in W^m} n_{dw} \ln \sum_{t \in T} \varphi_{wt} \theta_{td} + \sum_{i=1}^r \tau_i R_i(\Phi, \theta) \rightarrow \max_{\Phi, \theta}$$

$$\sum_{w \in W^m} \varphi_{wt} = 1, \quad \varphi_{wt} \geq 0,$$

$$\sum_{t \in T} \theta_{td} = 1, \quad \theta_{td} \geq 0$$

Где τ_m – вес модальности m , $n_m = \sum_{d \in D} \sum_{w \in W^m} n_{dw}$ – нормировочный множитель.

2.2.5. Особенности применения тематического моделирования

Отметим, что основная сфера применения методов тематического моделирования – исследования в социальной сфере: анализ новостей, публикаций в социальных сетях и блогах с целью анализа дискурса, основных тем, настроений. Тематическое моделирование также используется для библиометрического анализа определенных областей научных знаний (см., например, Asmussen and Møller, 2019; Lim and Buntine, 2015). Помимо классификации работ по сфере применения, их также целесообразно делить по применяемому методу. В настоящее время лидирующим является подход на основе метода латентного размещения Дирихле (LDA), однако его недостатком является существенная сложность включения в модель иных модальностей, помимо текста (к которым может относиться различная мета-информация, например, авторы, теги, ссылки и т.п.). Поэтому работы на основе LDA обычно используют не более двух характеристик при построении модели (например, автор-тематическая модель - Rosen-Zvi, et.al., 2004; учет цитирования документов - Cohn and Hofmann, 2001; учет внешних и внутренних ссылок документа - Gruber, et.al., 2008; тематическая модель автор-конференция - Tang, et.al., 2008). Кроме того, существуют подходы к структурному исследованию корпуса научных знаний на основе создания связей между внутренними семантическими категориями (Паринов, Когаловский, 2011). Отдельным направлением является тематическое моделирование контекстов цитирований публикаций (Паринов, 2021).

Альтернативным подходом является аддитивная регуляризация тематических моделей. Сравнение методов LDA и АРТМ представлено в работах (Potapenko and Vorontsov, 2013). Обзор работ на основе LDA приведен в Asmussen and Møller (2019), в Таблице 1 приведены работы на основе метода АРТМ.

Таблица 1 – Работы о применении и тестировании подхода АРТМ

Статья	Данные	Объем данных	Описание. Стратегия регуляризации, коэффициенты
Янина, Воронцов (2016).	Определение тематик статей на habrahabr.ru	132157 статей	Использование модальностей слов (1.0), авторов (0.5), комментаторов (0.75), тегов (15.0), хабов (10.0). Использование регуляризаторов декоррелирования ($1e+8$), разреживания (-1.5), сглаживания (0.5)

Apishev et al. (2016).	Выявление этнически обусловленного контента на платформе LiveJournal	1.38 млн. документов	Использование только модальности слов. Выделение основных и фоновых тем. Использование экспертного словаря этнонимов. Сравнительный анализ 8 различных моделей.
Chirkova, Vorontsov (2016)	Построение иерархической модели (hARTM).	Тестирование на коллекции Wikipedia (3665223 статей); статей postnauka (1728 статей)	Разработка концепции иерархической аддитивной регуляризации.
Милкова (2020)	Разведочный поиск патентных документов (из базы Роспатента), соответствующих пунктам плана импортозамещения по 22 отраслям промышленности	152718 патентных документов	Использование модальности слов (1.0), наиболее частотных биграмм (5.0). Сглаживание по словарям с наименованиями товаров для импортозамещения (1e+8)
Ianina, Vorontsov (2020)	Разработка и тестирование системы для разведочного поиска схожих по тематике к задаваемым пользователем документам	Тестирование на основе данных habrahabr.ru (175143 статей); TechCrunch (759324 статей); триплетов (статья – похожие статьи – непохожие статьи) от Dai et al. (на основе 963564 статей arxiv.org)	Тестирование различных моделей: tf-idf, bm-25, GloVe, fasttext, CNN, MaLSTM, BERT, ARTM, hARTM
Gorshkov et al. (2021).	Определение тематик сообществ в сети V Kontakte	6967 сообществ.	Модель строится на основе одиночных слов. Приводится сравнение моделей LDA и ARTM. Применение регуляризаторов разреживания матриц Φ и θ . Значения коэффициентов регуляризации не приведены.

Подчеркнем, что цель настоящей диссертации – использовать методы тематического моделирования не просто для получения информации о структуре коллекции, а для конфигурирования среды так, чтобы она способствовала информационному взаимодействию производителей и потребителей научного знания, объединенных единой тематикой исследования.

Схожую цель преследуют авторы в работе (Ianina, Vorontsov, 2020) в рамках которой разработана система для разведочного поиска научных публикаций в базе arXiv. Система позволяет осуществлять поиск тематически похожих документов к загружаемым пользователем подборкам статей.

Ключевые сложности применения АРТМ для научных исследований

Зарубежными и отечественными учеными подчеркивается сложность построения тематических моделей не специалистами в области технических и компьютерных наук (Lee et al., 2017; Булатов, 2020). Ключевыми нерешенными проблемами являются формализация алгоритма, подбор параметров модели, обеспечение интерпретируемости результатов. В модели АРТМ исследователю необходимо не только экспериментально установить оптимальное число тем, но и выбрать оптимальную стратегию регуляризации, которая включает экспериментальный подбор коэффициентов для каждого из регуляризаторов каждой модальности. Несмотря на то, что общая стратегия регуляризации предложена в работах Воронцова и Потапенко (2014б), подбор коэффициентов «вслепую» значительно усложняет процесс построения модели. Имеющиеся публикации, представляющие тематическую модель для той или иной задачи, не объясняют, как именно выбирается диапазон для перебора значений коэффициентов. Выбор оптимального коэффициента делается на основе значений критериев качества модели, однако неясно насколько сильно меняется состав топовых слов тем (для каждой модальности). Высокий барьер для входа в область тематического моделирования представителей смежных специалистов (экономистов, социологов, психологов) отмечается в работах Boyd-Graber et al. (2017), Булатов (2020), Vulatov et al. (2020)

Ряд работ отмечают необходимость разработки программ для построения тематических моделей, предоставляющих пользователям возможности напрямую уточнять тематическую модель: разделять, объединять, удалять темы, явное указывать набор слов, которые должны встречаться в одной теме (Lee et al., 2017). Несмотря на то, что существуют задачи, которые исходя из своих формулировок требуют включение априорной информации о составе тем (см. работы Милкова, 2020б

Arpisev et al., 2016), участие пользователя в формировании тем не позволяет избежать влияния ограниченной рациональности на формирование новых знаний.

Стоит выделить созданную в недавнем времени систему TopicNet (Bulatov et al., 2020) – верхнеуровневую надстройку над общей библиотекой для работы с ARTM (BigARTM), облегчающую работу с тематическими моделями. Однако система ориентирована на пользователей, не имеющих потребности в контроле всех параметров модели, больше ориентирована на решение задач в коммерческих, а не научных целях.

Также стоит выделить систему для разведочного поиска научных публикаций в базе arXiv (Ianina and Vorontsov, 2020). Система позволяет осуществлять поиск тематически похожих документов к загружаемым пользователем подборкам статей.

В данной работе преследуется цель формализовать процедуру построения тематических моделей, дать четкие инструкции и набор вспомогательных инструментов, позволяющих включить тематическое моделирование в инструментарий исследователей из смежных областей знаний.

В недавних работах Воронцова и коллег предложен подход на основе введения относительных коэффициентов регуляризации (Булатов, 2020), позволяющих получить некоторую интерпретацию коэффициентов и тем самым облегчить их подбор.

Как было описано в подразделе 2.3.2 для регуляризаторов сглаживания и разреживания формула М-шага имеет вид:

$$\varphi_{wt} = \text{norm}_{w \in W}(n_{wt} + \tau \beta_w),$$

$$\beta_w = \left(\frac{1}{|W|}\right)\text{- равномерное распределение}$$

$\tau > 0$ для регуляризатора сглаживания, $\tau < 0$ для регуляризатора разреживания.

$$\varphi_{wt} = \frac{n_{wt} + \tau \beta_w}{\sum_{w \in W} n_{wt} + \tau \beta_w} = \frac{n_{wt} + \tau \beta_w}{n_t + \tau} \quad (18)$$

Влияние регуляризации можно описать как притягивание (в случае сглаживания) или отдаление (в случае разреживания) распределения n_{wt}/n_t , полученного

как оценка максимального правдоподобия к равномерному распределению β_w с некоторым весом λ . φ_{wt} может быть записана как выпуклая комбинация этих двух распределений:

$$\varphi_{wt} = (1 - \lambda) \frac{n_{wt}}{n_t} + \lambda \beta_w, \quad 0 \leq \lambda \leq 1 \quad (19)$$

Приравнявая (18) и (19), выражая τ :

$$\tau = \frac{n_t \lambda}{(1 - \lambda) |W|}$$

Таким образом, выражение М-шага имеет вид:

$$\varphi_{wt} = \text{norm}_{w \in W} \left(n_{wt} + n_t \frac{\lambda}{(1 - \lambda)} \beta_w \right),$$

Величина $\frac{\lambda}{(1 - \lambda)}$ определяет, во сколько раз регуляризатор влияет на оценку φ_{wt} больше, чем коллекция. Однако, чем больше значение n_t (число слов в теме), тем сильнее будет регуляризация. Возможно усреднение коэффициентов регуляризации по всем темам:

$$\varphi_{wt} = \text{norm}_{w \in W} \left(n_{wt} + \frac{n}{|T|} \frac{\lambda}{(1 - \lambda)} \beta_w \right)$$

Относительный коэффициент показывает, во сколько раз регуляризатор влияет на оценку сильнее, чем коллекция.

2.2.6. Некоторые важные вопросы

Предобработка данных

Решение любой поставленной задачи, связанной с обработкой текстовой информации, включает обязательный этап предобработки данных. Как правило, предобработка представлена следующими этапами (Милкова, 2018):

1. Токенизация текста – разбивка на отдельно значимые единицы (токены), например, слова.
2. Начальная фильтрация – удаление из списка полученных токенов знаков пунктуации, не участвующих в последующем анализе. Для некоторых задач актуально удаление всей пунктуации, для некоторых могут быть оставлены, например, точки (когда анализ чувствителен к нахождению слов в разных предложениях) или тире (при извлечении определений).

3. Стемминг/лемматизация текста. Лемматизация - приведение токенов к нормальной (словарной) форме. Так, например, словоформы «цифровые», «цифровых», «цифровыми» преобразовываются к одной лемме «цифровой». Стемминг подразумевает нормализацию словоформы к ее квази-основе. Так, например, вышеприведенные словоформы будут усечены до формы «цифров». Выбор того или иного способа нормализации определяется морфологией анализируемого языка. Так, в силу особенностей русского языка, для него, как правило, выбирается лемматизация, так как приведение итоговых слов в виде усеченных квази-основ значительно усложняет восприятие. Для английского языка часто приемлемо работает стемминг.

4. Фильтрация – исключение стоп-слов. Первичный список стоп-слов обычно состоит из предлогов, союзов, междометий и т.п. Для решения некоторых задач экспертно формируется дополнительный список стоп-слов, включающий слова, заведомо не несущие смысловой нагрузки.

5. Составление n-gram (как правило, берутся биграммы, $n=2$) – сочетания подряд идущих n токенов

Отдельным этапом является преобразование исходных пред обработанных данных в формат, пригодный для дальнейшего построения тематической модели. В данной работе используется преобразование к формату Vowpal Wabbit (описание формата см. в Разделе 3.1.2).

Масштабируемость

Актуальным вопросом является определение тематик новых документов по построенной модели. Для определения вектора тем нового документа используется найденная матрица Φ , для всех $w \in d_{new}$ находятся вероятности $p(t|d_{new}) = p(d_{new}|w)p(w|t)$

Когда в потоке текстовых документов встречается слово, неизвестное для тематической модели, то необходимо принимать решение как учитывать вклад этого слова в тему нового документа. Допустимо предположение, что темы «нового слова», впервые появившегося в ВТМ, связаны с темами документа, где это слово встретилось. Для того чтобы определить эту связь, находится вектор тем

документов без учета нового слова. Для формирования тематического вектора для нового слова используется произведение Адамара тематических векторов документов, где это слово встретилось (Карпович, 2016).

Программные реализации и визуализация

В настоящее время существует большое число программных реализаций методов тематического моделирования (Милкова, 2019б). Наиболее распространенными и активно поддерживаемыми реализациями являются библиотеки Python: Gensim, Scikit-learn, BigARTM (Vorontsov et.al., 2015). Также стоит отметить библиотеки на других языках: Vowpal Wabbit (C++), Mallet (Java), Matlab Topic Modeling Toolbox.

Важным вопросом является визуализация результатов тематического моделирования. Разработано большое число средств визуализации тематических моделей: Termite System (Chuang et.al. 2013), TIARA (Wei et.al.2010), HierarchicalTopics (Dou et.al.2013) и др. Подробный обзор средств визуализации мы представлен в работе Айсина (2015), где систематизированы основные актуальные инструменты.

2.2.7. Формализация алгоритма для использования в экономике знаний

Основным источником получения знаний являются научные публикации, число которых стремительно растет, что затрудняет как своевременное информирование о недавних достижениях, так и снижает качество самостоятельного получения знаний молодыми учеными. Актуальными вопросами являются: какие публикации являются основополагающими для изучаемой подобласти? Какие авторы публикуются по данной теме? Какова динамика развития темы?

Предлагается включить следующие модальности при построении тематической модели на основе научных публикаций: слова, биграммы, авторы, список литературы (ссылки).

Введение каждой дополнительной модальности улучшает качество модели. Анализ результатов при добавлении модальностей (слова, авторы, ссылки) приведен в Таблице 8 Приложения 3.

Весы модальностей: 0.5 для слов; 1.0 для биграмм, авторов и ссылок.

Анализ чувствительности результатов при изменении весов модальностей приведен в Таблице 9 Приложения 3.

Таблица 2 – Формализация алгоритма построения тематических моделей

Шаг	Описание шага	Контролируемые меры качества	Критерии отбора
1	<p>Отбор числа тем.</p> <p>Исходя из общего понимания задачи определяется диапазон возможного числа тем. Внутри диапазона с равным шагом выбирается 4-5 значений числа тем. Например: 5, 10, 15, 20 тем (для определения подтем направления); число тем значительно увеличивается в случае расширения анализируемой области знаний. Делается 10 проходов (1-10) по коллекции без регуляризации и с включенным регуляризатором декоррелирования. Тестируются несколько значений относительного коэффициента: 0.005, 0.01, 0.05, 0.1, 0.15.</p>	<p>Ключевые показатели: Средняя когерентность (по 15 наиболее частотным биграмм)</p> <p>Перплексия</p> <p>Дополнительно контролируемые показатели:</p> <p>Разреженность матриц Φ и Θ.</p>	<p>Выбираются два эксперимента, по которым были получены результаты с наибольшей когерентностью, из них выбирается эксперимент с наименьшей перплексией.</p> <p>Особенность: Когерентность обратно коррелирует с перплексией</p>
2	<p>Подбор коэффициента сглаживания для фоновой темы (или отказ от него).</p> <p>Перебор значений по сетке, 10 итераций. К выбранному числу тем и значению коэфф. декоррелирования добавить одну фоновую тему. Сетка для относительного коэффициента сглаживания [0.01, 0.05, 0.1, 0.15, 0.2]</p>		
3	<p>Подбор коэффициентов разреживания.</p> <p>Делается дополнительно 10 проходов (11-21 итерации), тестируется коэффициент разреживания для матриц Φ и Θ. Тестируемые значения относительного коэффициента для матрицы Φ в диапазоне от -0.1 до -0.8 с шагом 0.1. Для каждого значения рассчитывается по формуле $\tau = \frac{n}{ D T } \frac{\lambda}{(1-\lambda)}$ соответствующее значение абсолютного коэфф. разреживания для матрицы Θ.</p>		

4	Анализ когерентности для выбранной стратегии регуляризации	Значения когерентности по темам, интерпретируемость тем	Если существуют не интерпретируемые и/или плохо интерпретируемые темы, проводится тестирование для близкого числа тем (± 2 темы)
5	Выбранная стратегия регуляризации тестируется в окрестности выбранного числа тем (± 2 темы)	Аналогично шагам 1-3.	

Схема общей методики построения тематической модели научных публикаций приведена в Приложении 1.

Выводы Главы 2

В данной главе представлено описание подхода тематического моделирования как разведочного поиска информации. Существуют несколько подходов к построению тематических моделей, доминирующим из которых является модель скрытого размещения Дирихле. Однако в данной модели оказывается сложно совмещать построение кластеризации сразу по нескольким параметрам. Именно это требование – возможность учета при построении модели различной метаинформации (модальности) является ключевым при решении задачи о представлении больших объемов текстовых данных, которые сами по себе являются многомодальными (т.е. характеризуются сразу множеством параметров, например, помимо непосредственно слов, авторством, ссылками и т.п.). Подход, основанный на аддитивной регуляризации, позволяет учитывать это ключевое требование. Кроме того, регуляризация способствует получению устойчивых результатов.

Подход АРТМ представляет задачу тематического моделирования как задачу матричного разложения (матрицы термов в темах и матрицы тем в документах). Для устранения неоднозначности в разложении матриц в модели вводятся регуляризаторы, определение стратегии введения которых является актуальной исследовательской задачей. Для оценивания параметров тематической модели максимизируется логарифм правдоподобия с дополнительными критериями-регуляризаторами. Данная задача решается с помощью так называемого регуляризованного EM-алгоритма. Основные критерии качества модели – перплексия, средняя

когерентность тем, доля разреженных элементов матриц. Основные регуляризаторы – разреживания, сглаживания, декоррелирования.

Зарубежными и отечественными учеными подчеркивается сложность построения тематических моделей не специалистами в области технических и компьютерных наук. Ключевыми нерешенными проблемами являются формализация алгоритма, подбор параметров модели, обеспечение интерпретируемости результатов. В модели ARTM исследователю необходимо не только экспериментально установить оптимальное число тем, но и выбрать оптимальную стратегию регуляризации, которая включает экспериментальный подбор коэффициентов для каждого из регуляризаторов каждой модальности. Несмотря на то, что общая стратегия регуляризации предложена, подбор коэффициентов «вслепую» значительно усложняет процесс построения модели. Имеющиеся публикации, представляющие тематическую модель для той или иной задачи, не объясняют, как именно выбирается диапазон для перебора значений коэффициентов.

В данной работе представлен способ формализации процедуры построения тематических моделей, даны четкие инструкции и набор вспомогательных инструментов, позволяющих включить тематическое моделирование в инструментарий исследователей из смежных областей знаний. Применительно к решению задачи о представлении больших объемов научных публикаций предложено включать следующие модальности: слова, биграммы, авторы, ссылки (список статей, на которые ссылается публикация). Веса модальностей 0.5, 1.0, 1.0, 1.0 соответственно. На первом этапе осуществляется перебор по сетке оптимального числа тем вместе с относительным коэффициентом декоррелирования. На втором этапе тестируется значение коэффициента сглаживания фоновой темы, который добавляется к коэффициенту декоррелирования. Проводится 10 итераций. На 11-20 итерации добавляется регуляризатор разреживания для матриц. На каждом из этапов контролируется перплексия и средняя когерентность тем. Выбирается два значения с наибольшей когерентностью, из них выбирается значение, соответствующее наименьшей перплексии.

ГЛАВА 3. АПРОБИРОВАНИЕ МЕТОДИКИ ПРИ РАБОТЕ С НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИЕЙ

В данной главе представлена реализация формализованного во второй главе алгоритма построения аддитивно регуляризованных тематических моделей для решения задачи получения общего представления о большой коллекции текстовых документов (на примере научных статей по поведенческой и экспериментальной экономике). Также решена другая задача, имеющая отношение к разведочному поиску документов согласно априорно задаваемой информации (на примере патентного анализа для оценки программы импортозамещения). Несмотря на то, что эти задачи не связаны между собой, они демонстрируют два ключевых подхода к поиску и представлению информации в текстовых коллекциях, объем которых слишком велик для обработки человеком. Первая задача ставит перед собой цель дать представление о структуре текстовой коллекции. Вторая задача ставит цель осуществления поиска информации по смыслу (по темам, каждая из которых характеризуется своим набором слов).

И первая, и вторая задача актуальны в рамках функционирования экономики информационных продуктов, а именно, представляют собой способ организации потребления информационных продуктов, информационного взаимодействия производителей и потребителей научного знания.

Задачи решаются на основе описанного в Главе 2 подхода ARTM. Анализ проводился в среде Python, использовалась библиотека BigARTM, а также собственные процедуры, позволяющие адаптировать работу библиотеки к решению исследуемых задач.

3.1. Тематическое представление научных статей направления поведенческой и экспериментальная экономика

Актуальность выбранной задачи объясняется, с одной стороны, стремительным ростом общего числа публикаций, с другой стороны, снижением среднего качества подаваемых рукописей (так, в ведущих экономических журналах процент

приема снизился с 15% в 1990 году до 6% в 2012 году (Card and DellaVigna, 2013)), снижением продуктивности исследовательской работы молодых ученых (Conley and Önder, 2014). При этом работа в области науки, как показано в некоторых исследованиях, представляется важной и востребованной (Варшавский и др., 2020).

Вопросы о повышении эффективности работы научного сообщества, пересмотра методов оценки качества научных журналов, стимулирования научно-исследовательской деятельности ставились неоднократно (например, Balatsky et al., 2021; Писарева, 2012). В данной работе повышение эффективности деятельности научного сообщества представляется с совершенно другой стороны. Речь идет об обеспечении информационного взаимодействия производителей и потребителей научного знания. Важно отметить, что несмотря на то, что поиск и навигация по базам научных публикаций в последнее время значительно усовершенствованы (доступен семантический поиск публикаций, внедрены различные рекомендательные системы), получение общего представления о структуре изучаемого направления, ключевых и схожих публикациях, авторах внутри подтем является актуальной для научного сообщества задачей.

3.1.1. Формирование и описание выборки

Исходные данные для проведения исследования брались из базы Semantic Scholar, предоставляющей открытый исследовательский корпус публикаций¹⁴. База содержит более 180 млн. научных статей¹⁵.

Хранение архива, а также последующий отбор необходимых публикаций осуществлялись в базе MongoDB.

Перечень полей архива, загруженных в БД приведен в Таблице 3.

Таблица 3 – Поля архива научных публикаций

Поле	Расшифровка
_id	ID публикации

¹⁴ Semantic Scholar Open Research Corpus // Интернет платформа Semantic Scholar: <https://www.semanticscholar.org> – URL: <http://s2-public-api.prod.s2.allenai.org/corpus/download/> (дата обращения 31.05.2020)

¹⁵ Перечень издательств, входящих в базу Semantic Scholar // Интернет платформа Semantic Scholar: <https://www.semanticscholar.org> – URL: <https://pages.semanticscholar.org/publisher-partners> (дата обращения 22.04.2020)

authors	Авторы публикации (ФИО и id автора). Пример записи: [{"name": "Andrew F. Cooper", "ids": ["47729306"]}, {"name": "Richard Higgott", "ids": ["48746135"]}]
fieldsOfStudy	Область знаний (статья должна относиться к области Экономики или Бизнеса, однако в дополнение может включать и другие области). Пример записи: ["Business", "Sociology"]
inCitations	Список статей, цитирующих данную статью, представлен списком id статей. Пример записи: ["602dcf26a1d3003fb72ff8348d643aa71c66fce1", "43ab86e8c8d7cdadaf38830149771247aefe5472"]
journalName	Название журнала.
outCitations	Список литературы, представлен списком id статей. Пример записи (фрагмент): ["6337412de0fc1f45e6bb285065991910ce15d2c5", "c07f62da086f096dd9c2bda670aae7f61af09082", ...]
paperAbstract	Абстракт статьи
title	Название статьи
year	Год публикации

Отбор статей по поведенческой и экспериментальной экономике осуществлялся внутри разделов Экономика (Economics), Бизнес (Business) и проводился по ключевым словам, список которых составлялся экспертно (Таблица 4).

Таблица 4 – Список ключевых слов, по которым осуществлялся отбор статей для исследования

behavioral economics, prospect theory, nudge theory, nudging, bounded rationality, cognitive bias, heuristics, behavioral insights, experimental economics, neuroeconomics, behavioral models, behavioral finance, behavioral corporate finance, behavioral decision-making in finance, dictator game, ultimatum game, trust game, prisoner's dilemma, social behavior, pro-social behavior, psychology, behavioral models, behavioral theory, behavioral approach, behavioral perspective, behavioral game, behavioral experiment, experimental science, experimental game, experimental approach, dual-process theory, ecological rationality, cognitive dissonance, framing, hyperbolic discount, mental accounting.

Поиск по ключевым словам осуществлялся по Названию и Аннотации статей, производился стемминг поискового слова, формировались отдельные выборки для каждого поискового запроса с последующим объединением и удалением дубликатов. Пример запросов к базе MongoDB приведен в Приложении 1.

Итоговая выборка состояла из 37352 статей. Дальнейший анализ производился в среде Python 2.7.

Общие характеристики данных (число публикаций по годам – Рисунок 1; число публикаций авторов – Рисунок 2, Рисунок 3; число публикаций по областям – Таблица 5; число цитирований – Рисунок 4, Рисунок 5; публикации по журналам – Рисунок 6) приведены ниже.

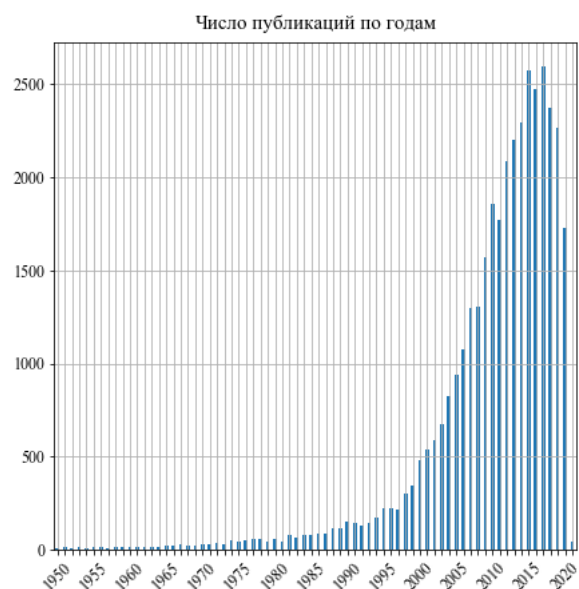


Рисунок 1 – Число публикаций по годам

Таблица 5 – Число статей по областям, 20 наиболее частотных

Область	Число публикаций
Economics	23048
Business	10258
Computer Science, Economics	1022
Business, Computer Science	509
Business, Medicine	463
Economics, Medicine	448
Economics, Psychology	383

Business, Economics	185
Economics, Sociology	136
Economics, Political Science	134
Business, Psychology	81
Economics, Mathematics	74
Business, Political Science	68
Computer Science, Economics, Mathematics	63
Business, Sociology	47



Рисунок 2 – Распределение числа публикаций авторов

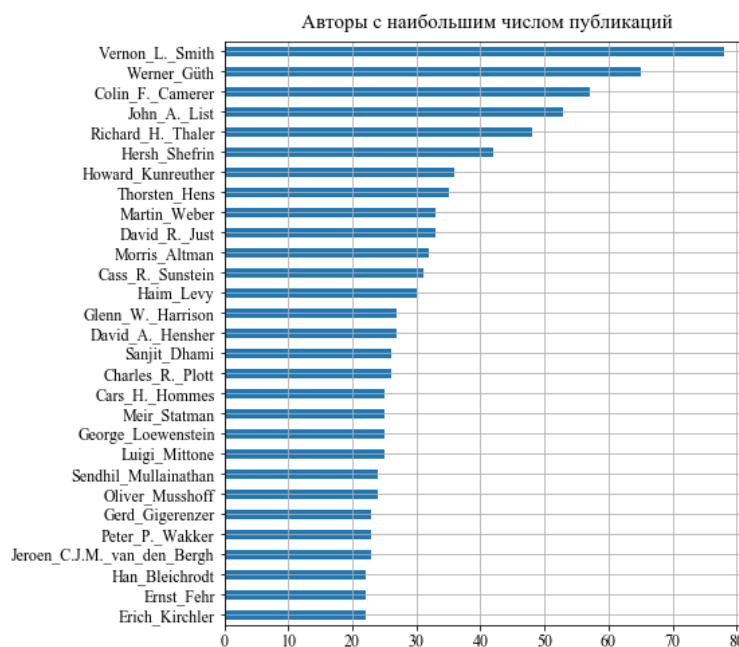


Рисунок 3 – Авторы с наибольшим числом публикаций

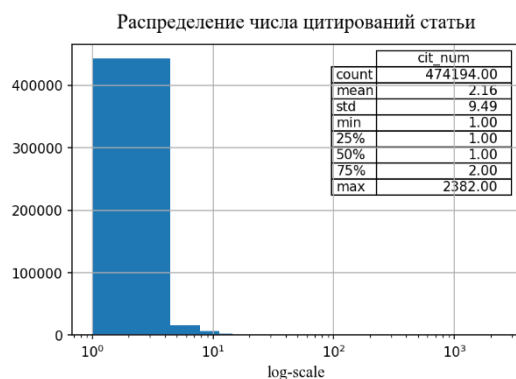


Рисунок 4 – Распределение числа цитирований статьи

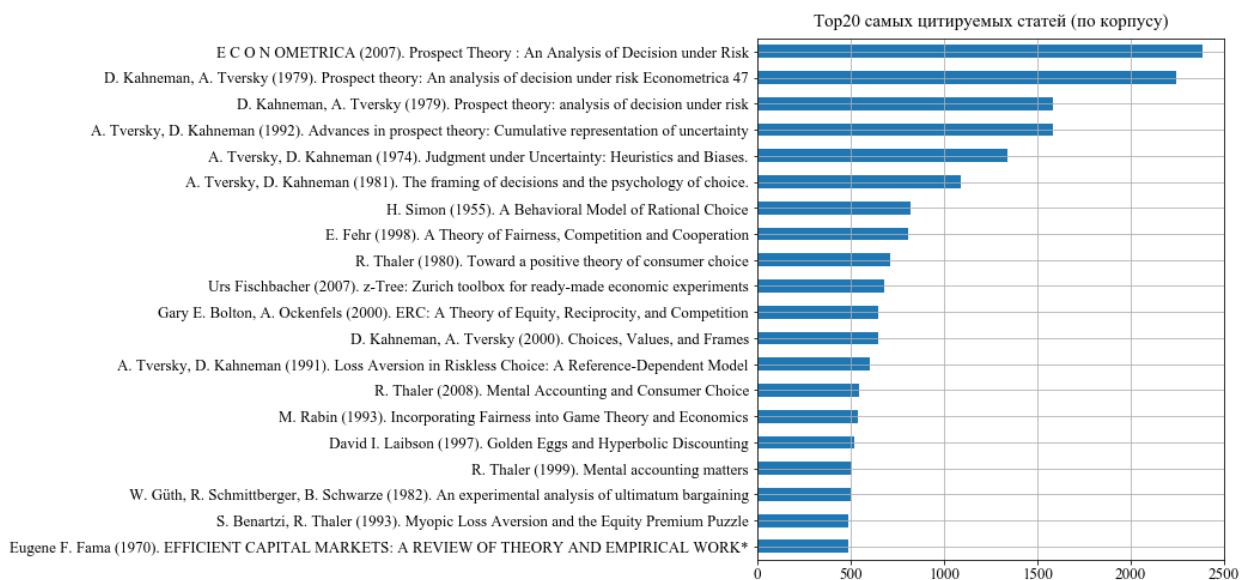


Рисунок 5 – Top20 самых цитируемых статей

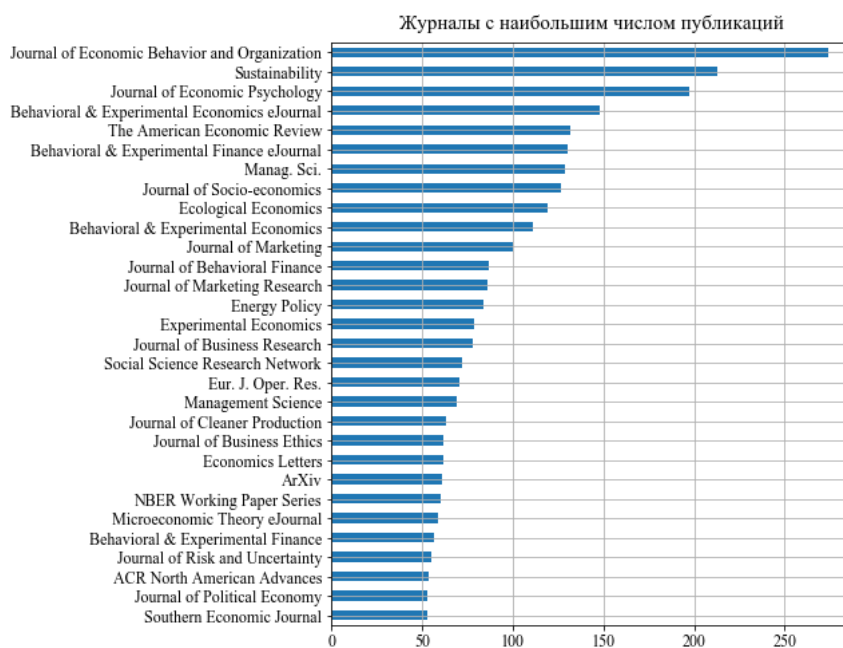


Рисунок 6 – Журналы с наибольшим числом публикаций

3.1.2. Предобработка данных и формирование словаря

Предобработка данных включала:

Для Названия и Аннотации: Токенизацию, лемматизацию, удаление стоп слов (список стоп слов приведен в Приложении 1), формирование биграмм (также к списку биграмм были добавлены униграммы 'nudge', 'nudging', 'framing', 'heuristic', 'psychology', 'neuroeconomics', так как они представляют собой устоявшиеся термины, используемые в поведенческой и экспериментальной экономике, вклад этих униграмм может быть приравнен к вкладу биграмм).

Использовались библиотеки Python: nltk, collections.

Для Авторов: Преобразование записи имени автора как Name_Surname (или Name1_Name2_Surname; в случае нескольких авторов их имена перечисляются через пробел). Например: Chris_Starmer Robin_P._Cubitt

Для Списка литературы: удаление нечисловых и небуквенных символов.

Предобработанные данные сохранялись в формате Vowpal Wabbit (.vpw), подразумевающим следующие правила записи данных:

Каждый документ представлен в одной строке

Строка имеет формат:

```
Id |Modality1 token11: count12 token12: count12 ... |ModalityN tokenN1:
countN1 tokenN2:countN2 ...
```

Где id – id документа; Modality – название модальности (например, биграмы, авторы, ссылки); token – токены, характеризующие данную модальность данного документа (для биграмм – перечень всех биграмм документа); count – частота токена в документе (по умолчанию берется равной 1).

В данном исследовании включались следующие модальности: bigrams (биграмы), authors (авторы), references (список литературы)

Фрагмент строки файла .vpw:

```
6002ec42db586f9a16ead048 |authors Vernon_L._Smith |bigrams economy_ex-
change:1 individual_motivation:1 serve_aggregate:1 different_experiment:2 ...|refer-
ences fa8614b50698f044b7788e24f8612f2d59880c36 ...
```

Дальнейшая работа строится с использованием библиотеки для построения аддитивно-регуляризованных тематических моделей – BigARTM¹⁶.

Данные в формате .vprw разбиваются на фрагменты – батчи (batches), также формируется единый словарь.

Словарь представляет собой перечень всех токенов всех модальностей с указанием tf токена – частоты встречаемости токена в коллекции, df – числа документов, в которых токен встретился, а также значения $token_value$ - нормированная частота (tf/n_m).

`token, class_id, token_value, token_tf, token_df`

Пример:

`management_make, bigrams, 2.74887224805e-06, 49.0, 48.0`

При сглаживании/разреживании введенный коэффициент регуляризации умножается на значение $token_value$.

Исходный объем словаря – 2486993.

Были убраны слова и биграммы, встречающиеся только однажды ($df=1$), а также слова с частотой больше $df>5000$; авторов – $df>630$; ссылок – $df>12000$. Объем словаря сокращен до 1050400.

Число слов - 36485, число биграмм – 497933, число авторов – 54287, число ссылок – 461695.

3.1.3. Построение тематической модели

Построение тематической модели проводилось согласно описанному в разделе 2.3.7 алгоритму. Средняя когерентность тем рассчитывалась на основе 15 наиболее вероятных биграмм темы.

Шаг 1. Отбор числа тем.

Результаты тестирования числа тем приведены ниже (Рисунок 7).

¹⁶ URL: <https://github.com/bigartm/bigartm> (дата обращения: 25.09.2021)

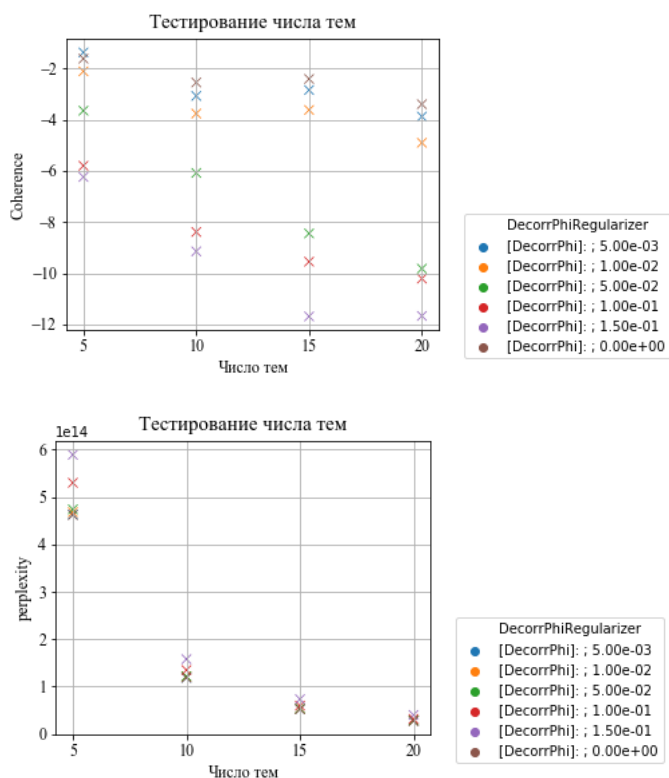


Рисунок 7 – Результаты работы модели (10 проходов) для числа тем 5, 10, 15, 20 со значениями относительного коэффициента регуляризатора декоррелирования [0.005, 0.01, 0.05, 0.1, 0.15], а также без регуляризатора (бордовый цвет): значения средней когерентности тем и перплексии

Таблица со всеми значениями контролируемых метрик качества при различном числе тем приведена в Таблице 1 Приложения 3.

Интересным является наблюдение, что качество модели (оцениваемое на основе значений средней когерентности и перплексии) тем выше, чем меньше значение относительного коэффициента регуляризатора декоррелирования, а наилучшей является модель вовсе без использования регуляризатора.

Объяснение этого факта, возможно, кроется в том, что сама общая тематика исходных данных (поведенческая и экспериментальная экономика) подразумевает пересечение в терминологии и библиографических источниках в подтемах, что не позволяет искусственно увеличивать расстояние между подтемами без потери качества.

Оптимальное (наибольшая средняя когерентность, наименьшая перплексия, невырожденность модели) значение числа тем – 15.

Шаг. 2. Результаты тестирования значений относительного коэффициента сглаживания для фоновой темы показали ухудшение качества модели (значение средней когерентности колебалось в диапазоне [-3.02, -3.25]). Таблица со всеми значениями контролируемых метрик качества при различных значениях относительного коэффициента сглаживания для 1 фоновой темы приведены в Таблице 2 Приложения 3. Итерации проводились для 15 основных тем (число которых определено на Шаге 1). Значения относительного коэффициента сглаживания фоновой темы перебирались по сетке [0.01, 0.05, 0.1, 0.15, 0.2].

Неэффективность введения фоновой темы объясняется отсеиванием высокочастотных и низкочастотных термов на стадии предобработки.

Увеличение числа фоновых тем, сглаживание матрицы Θ не улучшают контролируемые метрики качества модели.

Шаг. 3. Результаты работы модели на первых 10 итерациях сохранялись, проводилось тестирование регуляризатора разреживания для основных тем на 11-20 итерациях. Значения относительного коэффициента для матрицы Φ перебирались по сетке [-0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8].

Отметим, что в силу особенностей технической реализации библиотеки BigARTM, лежащей в основе данной работы, относительные коэффициенты регуляризации введены только для матриц Φ . Для регуляризации матрицы Θ работают абсолютные значения. Однако согласно формуле взаимосвязи относительного и абсолютного коэффициентов (см. подраздел 2.3.5) для матрицы Θ имеем:

$$\tau = \frac{n}{|D||T|} \frac{\lambda}{(1-\lambda)},$$

$$|D| = 37352; |T| = 15; n = 1050400$$

Таблица перевода для некоторых значений относительного коэффициента в абсолютный приведена ниже.

Таблица 6 – Перевод относительных значений коэффициента регуляризации (λ) и абсолютные (τ)

λ	τ
0.1	0.21

0.2	0.47
0.3	0.8
0.4	1.25
0.5	1.87
0.6	2.81
0.7	4.37
0.8	7.5

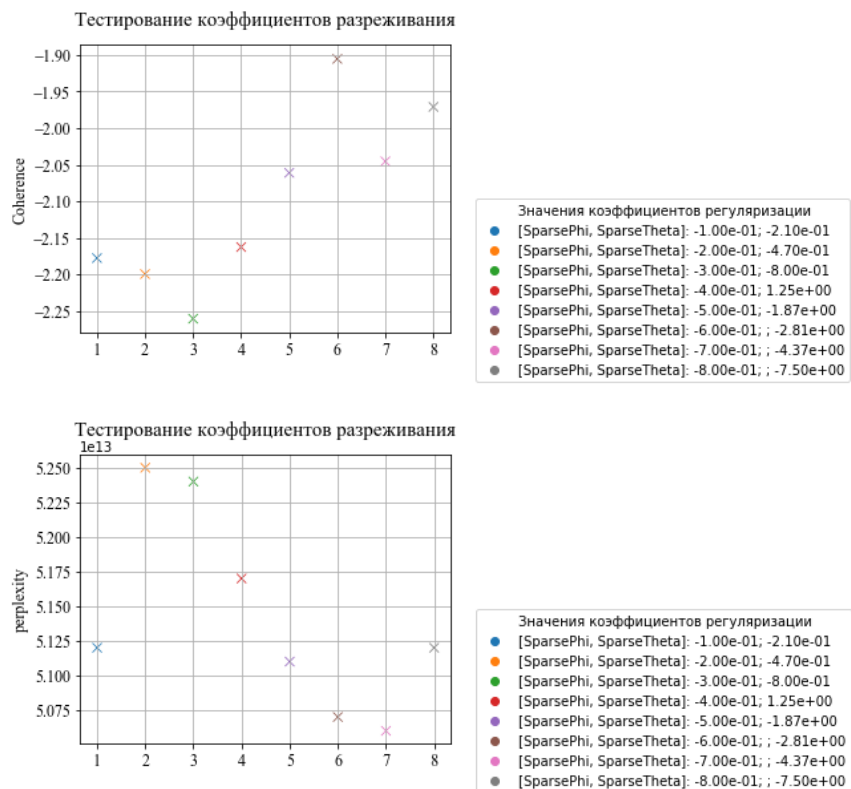


Рисунок 8 – Тестирование регуляризатора разреживания для основных тем. Значения относительного коэффициента разреживания матрицы Φ [-0.1, -0.2, -0.3, -0.4, -0.5, -0.6, -0.7, -0.8], соответствующие им значения абсолютного коэффициента разреживания матрицы Θ [-0.21, -0.47, -0.80, -1.25, -1.87, -2.81, -4.37, -7.5]

Таблица со всеми значениями контролируемых метрик качества при различных коэффициентах регуляризатора разреживания приведена в Приложении 3 (Таблица 3).

Оптимальное (наибольшая средняя когерентность, наименьшая перплексия) значение коэффициента установлено равным -0.6 для матрицы Φ , -2.81 для матрицы Θ .

Значения когерентностей для каждой из 15 тем при выбранной стратегии регуляризации приведены в Таблице 7.

Таблица 7 – Значения средней когерентности тем

Тема	Средняя когерентность
d1	1,2152
d2	-2,8643
d3	-1,8377
d4	-0,2487
d5	-0,7582
d6	0,3043
d7	-4,8092
d8	-1,0585
d9	-5,0754
d10	-6,1920
d11	-2,0757
d12	1,7037
d13	-4,5176
d14	-1,4540
d15	-0,9093
ср. знач.	-1,9052

Видно, что наименьшие значения когерентности для тем 10, 9, 7.

Тестирование выбранной стратегии для числа тем 13, 14, 16, 17 показывает, что качество модели для числа тем 14, 16, 17 не улучшается. Для числа тем равное 13, качество модели немного повышается (средняя когерентность равна -1.81, перплексия $6.69e+13$), состав тем 1-13 остается прежним, исключаются темы 14,15. В Таблице 4 Приложения 3 приведены значения измеряемых метрик. Итоговое решение о выборе оптимального числа тем (15, 13) выбирается экспертно.

Результатом работы модели являются выделенные ключевые биграммы, авторы и ссылки для каждой из тем. Важно, что процесс выделения тем является эвристическим, и интерпретация тем может быть не всегда однозначной. Интерпретация темы выражается в возможности ее ручного именованя. В результате построения данной модели все кроме одной темы (номер 10) являются интерпретируемыми.

Приведенный далее анализ чувствительности результатов показывает, что хорошо интерпретируемые темы устойчивы к различным начальным инициализациям модели. Неоднозначно интерпретируемые темы могут меняться в зависимости от начального приближения.

Хорошо интерпретируемые темы

Тема 1. Ограниченная рациональность, теория перспектив

Изучение отклонений от теории ожидаемой полезности, ограниченная рациональность, теория перспектив, отношения к риску, неприятие риска, неприятие потерь

Биграммы:

prospect_theory, decision_making, loss_aversion, expected_utility, reference_point, utility_theory, cumulative_prospect, risk_aversion, gain_loss, risk_preference, framing, utility_function, risk_averse, bounded_rationality, risk_attitude

Авторы:

Thorsten_Hens, Haim_Levy, David_A._Hensher, Enrico_G._De_Giorgi, Peter_P._Wakker, Marc_Oliver_Rieger, Han_Bleichrodt, Sanjit_Dhami, Glenn_W._Harrison, Mohammed_Abdellaoui, Ali_al-Nowaihi, David_Peel, Ulrich_Schmidt, Colin_F._Camerer

Ключевые ссылки:

A. Tversky, D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

A. Tversky, D. Kahneman (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model

S. Benartzi, R. Thaler (1993). Myopic Loss Aversion and the Equity Premium Puzzle

M. Allais (1953). Le comportement de l'homme rationnel devant le risque : critique des postulats et axiomes de l'ecole americaine

R. Thaler (1980). Toward a positive theory of consumer choice

J. Quiggin (1982). A theory of anticipated utility

D. Prelec (1998). The Probability Weighting Function

W. Deming, J. Neumann, Oscar Morgenstern (1944). Theory of Games and Economic Behavior.

Botond Kőszegi, M. Rabin (2004). A Model of Reference-Dependent Preferences

Тема 2. Про-экологическое поведение

Изучение влияния моделей поведения человека на окружающую среду. Вопросы об изменении климата, эффективности использования энергии, использования возобновляемой энергии, землепользования, проблемы пищевых отходов и др.

Биграммы:

framing, climate_change, energy_efficiency, bounded_rationality, renewable_energy, food_waste, decision_making, low_carbon, energy_consumption, greenhouse_gas, energy_policy, land_use, climate_policy, supply_chain, long_term

Авторы:

Jeroen_C.J.M._van_den_Bergh, Brian_Wansink, Nicolai_J._Foss, Haider_A._Khan, Libby_Weber, Timothy_J._Foxon, Andreas_Löschel, Jennie_C._Stephens, Massimo_Filippini, Ulrich_Witt, Kai_P._Purnhagen, Victor_P._Goldberg, Nicholas_Frank_Pidgeon, Kenneth_Gillingham

Ключевые ссылки:

- I. Ajzen (1991). The theory of planned behavior
- H. Allcott (2011). Social Norms and Energy Conservation
- W. Abrahamse, L. Steg, C. Vlek, T. Rothengatter (2005). A review of intervention studies aimed at household energy conservation
- L. Steg, C. Vlek (2009). Encouraging pro-environmental behaviour : An integrative review and research agenda
- F. Geels (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study
- A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.
- H. Simon (1955). A Behavioral Model of Rational Choice
- E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk
- P. C. Stern (2000). New Environmental Theories: Toward a Coherent Theory of Environmentally Significant Behavior
- P. C. Stern, T. Dietz, Troy D. Abel, G. Guagnano, L. Kalof (1999). A Value-Belief-Norm Theory of Support for Social Movements: The Case of Environmentalism
- P. Schultz, J. Nolan, R. Cialdini, Noah J. Goldstein, Vladas Griskevicius (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms
- A. Jaffe, R. Stavins (1994). The energy-efficiency gap What does it mean?
- A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

Тема 3. Ограниченная рациональность, методы подталкивания и поведение потребителя

Влияние ограниченной рациональности, когнитивных искажений, методов подталкивания, ментального учета, архитектуры выбора, психологических и социальных факторов на поведение потребителя.

Биграммы:

nudge, nudging, framing, decision_making, bounded_rationality, behavioral_economics, cognitive_bias, consumer_behavior, social_behavior, long_term, choice_architecture, rational_choice, mental_accounting, field_experiment, consumer_psychology

Авторы:

Cass_R._Sunstein, Olivia_S._Mitchell, Jeffrey_R._Brown, George_Loewenstein, Anran_Chen, David_Laibson, Shmuel_I._Becher, Steven_Haberman, Raphaële_Préget, Sophie_Thoyer, Tingliang_Huang, Gilles_Grolleau, Sendhil_Mullainathan, Brian_E._Roe

Ключевые ссылки:

R. Thaler, C. Sunstein (2008). Nudge: Improving Decisions About Health, Wealth, and Happiness

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

H. Allcott (2011). Social Norms and Energy Conservation

E. Johnson, D. Goldstein (2003). Do Defaults Save Lives?

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk Econometrica 47

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

P. Schultz, J. Nolan, R. Cialdini, Noah J. Goldstein, Vladas Griskevicius (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms

D. Kahneman, A. Tversky (2000). Choices, Values, and Frames

W. Samuelson, R. Zeckhauser (1988). Status quo bias in decision making

Chris Arney (2015). Nudge: Improving Decisions about Health, Wealth, and Happiness

R. Thaler, S. Benartzi (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving

Тема 4. Общие вопросы поведенческой и экспериментальной экономики

Переход от неоклассической экономики к поведенческой, теория рационального выбора, экспериментальная экономика и ее подходы

Биграммы:

behavioral_economics, decision_making, bounded_rationality, framing, economic_theory, credit_card, law_economics, experimental_economics, behavioural_economics, neoclassical_economics, rational_choice, human_behavior, experimental_approach, social_psychology, long_term

Авторы:

Richard_H._Thaler, Sendhil_Mullainathan, Morris_Altman, Colin_F._Camerer, Gerd_Gigerenzer, D._Wade_Hands, Todd_J._Zywicki, Peter_E._Earl, David_R._Just, Herbert_A._Simon, Alan_Lewis, Cass_R._Sunstein, Vernon_L._Smith, Mark_D._White

Ключевые ссылки:

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

H. Simon (1955). A Behavioral Model of Rational Choice

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

David I. Laibson (1997). Golden Eggs and Hyperbolic Discounting

R. Thaler (1980). Toward a positive theory of consumer choice

D. Kahneman (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics

Colin Camerer, S. Issacharoff, G. Loewenstein, Ted O'Donoghue, M. Rabin (2003). Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism'

R. Thaler, C. Sunstein (2008). Nudge: Improving Decisions About Health, Wealth, and Happiness

D. Kahneman, J. Knetsch, R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

R. Thaler, S. Benartzi (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving

Colin Camerer, G. Loewenstein (2003). Behavioral Economics: Past, Present, Future

Тема 5. Поведенческие факторы в корпоративном управлении

Корпоративное управление, человеческий капитал, управление человеческими ресурсами, управление рисками компании, социальная психология

Биграммы:

framing, bounded_rationality, corporate_governance, human_resource, transaction_cost, resource_management, decision_making, supply_chain, human_capital, risk_management, competitive_advantage, cost_economics, social_psychology, strategic_management, behavioral_economics

Корпоративное управление, управление человеческими ресурсами, управление рисками компании

Авторы:

Hrabrin_Bachev, Wing-Keung_Wong, Nicolai_J._Foss, Donald_C._Langevoort, Alain_Verbeke, Bart_Nooteboom, S._Niggol_Seo, Luis_Ballesteros, Dan_Lovallo, Daniel_A._Levinthal, Nicholas_Dew, Peter_W._Liesch, Kin_Lam, Yuri_Biondi

Ключевые ссылки:

J. Barney (1991). Firm Resources and Sustained Competitive Advantage

W. Dugger (1987). The Economic Institutions of Capitalism

S. Winter, R. Nelson (1982). An evolutionary theory of economic change

R. Coase (1937). The Nature of the Firm

M. C. Jensen, W. Meckling (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure

D. Teece, G. Pisano, A. Shuen (1997). DYNAMIC CAPABILITIES AND STRATEGIC MANAGEMENT

W. Ouchi, O. Williamson (1977). Markets and Hierarchies: Analysis and Antitrust Implications.

Wesley M. Cohen, Daniel A. Levinthal (1990). ABSORPTIVE CAPACITY: A NEW PERSPECTIVE ON LEARNING AND INNOVATION

Paul J. DiMaggio, W. Powell (None). The iron cage revisited institutional isomorphism and collective rationality in organizational fields

K. Eisenhardt (1989). Building theories from case study research

O. Williamson (1996). The mechanisms of governance

M. Shubik, R. Cyert, J. March (1965). A Behavioral Theory of the Firm.

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

B. Wernerfelt (1984). A Resource-Based View of the Firm

Тема 6. Экспериментальная экономика

Изучение социальных предпочтений, экспериментальный подход на основе игр «Ультиматум», «Общественное благо», «Диктатор», «Доверие»; равновесие в играх.

Биграммы:

bounded_rationality, ultimatum_game, dictator_game, public_good, game_theory, trust_game, experimental_economics, social_preference, nash_equilibrium, decision_making, laboratory_experiment, agent_based, experimental_study, behavioral_economics, experimental_evidence

Авторы:

Werner_Güth, Vernon_L._Smith, Cars_H._Hommes, Colin_F._Camerer, John_A._List, Gary_Charness, Ernst_Fehr, Axel_Ockenfels, Steffen_Huck, Gary_E._Bolton, Charles_A._Holt, Daniel_Houser, Reinhard_Selten, David_G._Rand

Ключевые ссылки:

E. Fehr (1998). A Theory of Fairness, Competition and Cooperation

Gary E. Bolton, Axel Ockenfels (2000). ERC: A Theory of Equity, Reciprocity, and Competition

U. Fischbacher (2007). z-Tree: Zurich toolbox for ready-made economic experiments

M. Rabin (1993). Incorporating Fairness into Game Theory and Economics

W. Güth, R. Schmittberger, Bernd Schwarze (1982). An experimental analysis of ultimatum bargaining

Joyce E. Berg, J. Dickhaut, K. McCabe (1995). Trust, Reciprocity, and Social History

Colin Camerer (2003). Behavioral Game Theory: Experiments in Strategic Interaction

Robert Forsythe, J. Horowitz, N. Savin, M. Sefton (1994). Fairness in Simple Bargaining Experiments

G. Charness, M. Rabin (2000). Understanding Social Preference with Simple Tests

E. Hoffman, K. McCabe, Keith Shachat, V. Smith (1994). Preferences, Property Rights, and Anonymity in Bargaining Games

J. Kagel, A. Roth (1995). Handbook of Experimental Economics

E. Fehr, S. Gächter (1999). Cooperation and Punishment in Public Goods Experiments

A. Falk, U. Fischbacher (2006). A Theory of Reciprocity

A. Roth, Vesna Prasnikar, M. Okuno-Fujiwara, S. Zamir (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study

Тема 8. Влияние маркетинговой стратегии на потребителя

Влияние маркетинговой стратегии, бренда, обслуживания на принятие решений о покупке и удовлетворенность потребителя.

Биграммы:

decision_making, framing, consumer_behavior, purchase_intention, investment_decision, consumer_psychology, marketing_strategy, brand_equity, message_framing, study_wa, influence_consumer, structural_equation, purchase_decision, consumer_decision, consumer_behaviour

Авторы

Gordon_R._Foxall, Shun-Yao_Tseng, Thomas_Rudolph, Arch_G._Woodside, Philip_M.J._Reckers, Chun-Tuan_Chang, Victoria_K._James, Chyan_Yang, Adrian_Mitroi, Jaakko_Aspara, Jiayi_Balasuriya, Jorge_M._Oliveira-Castro, Tim_Böttger, Michel_Laroche

Ключевые ссылки:

C. Fornell, D. Larcker (1981). Evaluating structural equation models with unobservable variables and measurement error.

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

M. Fishbein, I. Ajzen (1977). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

I. Ajzen (1991). The theory of planned behavior

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

James C. Anderson, D. Gerbing (1988). STRUCTURAL EQUATION MODELING IN PRACTICE: A REVIEW AND RECOMMENDED TWO-STEP APPROACH

V. Zeithaml (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence:

R. M. Baron, D. A. Kenny (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.

R. Thaler (2008). Mental Accounting and Consumer Choice

L. Festinger (1957). A Theory of Cognitive Dissonance

Fred D. Davis (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology

Тема 9. Поведенческая теория фирмы

Исследования поведения внутри фирм, семейный бизнес. Управление человеческими ресурсами, социально-психологические исследования, корпоративная социальная ответственность

Биграммы:

framing, supply_chain, behavioral_theory, corporate_social, social_responsibility, theory_firm, customer_satisfaction, health_care, bounded_rationality, public_sector, decision_making, family_firm, long_term, firm_performance, ecosystem_service

Авторы:

Robert_M._Wiseman, Henrich_R._Greve, Luis_R._Gomez-Mejia, Josip_Kotlar, Thomas_Zellweger, En_Xie, Nils_Stieglitz, Jonathan_P._O'Brien, Philip_Bromiley, Li_Xie, Ronald_J._Gilson, Todd_M._Alessandri, Songcui_Hu, Richard_A._Bettis

Ключевые ссылки:

P. Bromiley (1991). Testing a Causal Model of Corporate Risk Taking and Performance

M. Shubik, R. Cyert, J. March (1965). A Behavioral Theory of the Firm.

H. Greve (1998). Performance, Aspirations, and Risky Organizational Change

H. Greve (2003). A Behavioral Theory of R&D Expenditures and Innovations: Evidence from Shipbuilding

J. Singh (1986). Performance, Slack, and Risk Taking in Organizational Decision Making

M. C. Jensen, W. Meckling (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure

R. Cyert, J. March (1963). A Behavioral Theory of the Firm

E. Johnsen (1964). Richard M. Cyert & James G. March, A Behavioral Theory of The Firm, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963, 332 s.

Daniel A. Levinthal, J. March (1981). A model of adaptive organizational search

B. Staw, Lance E. Sandelands, J. Dutton (1981). Threat-rigidity effects in organizational behavior: A multilevel analysis.

L. Bourgeois (1981). On the Measurement of Organizational Slack

J. March, Z. Shapira (1987). Managerial perspectives on risk and risk taking

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

Dinesh N. Iyer, K. Miller (2008). Performance Feedback, Slack, and The Timing of Acquisitions

Тема 11. Поведенческая экономика и налогообложение

Применение экспериментальных подходов к анализу поведения в контексте уплаты налогов, склонности к уклонению от уплаты налогов

Биграммы:

framing, behavioral_economics, tax_compliance, experimental_economics, tax_evasion, decision_making, income_tax, economic_theory, tax_system, long_term, social_psychology, tax_rate, job_satisfaction, social_science, game_theory

Авторы:

John_B._Davis, Morris_Altman, Erich_Kirchler, Vernon_L._Smith, John_A._List, Edward_J._McCaffery, Francesco_Guala, Paul_Webley, Joel_Slemrod, Luigi_Mittone, Simon_James, Dale_J._Menkhaus, James_Alm, Ariel_Rubinstein

Ключевые ссылки:

M. Allingham, A. Sandmo (1972). Income tax evasion: a theoretical analysis

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk Econometrica 47

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

J. Alm, G. McClelland, W. Schulze (1992). Why do people pay taxes

E. Kirchler (2007). The Economic Psychology of Tax Behaviour

S. Yitzhaki (1974). Income tax evasion: A theoretical analysis

S. Watanabe (1986). Income Tax Evasion: A Theoretical Analysis

B. Torgler (2002). Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments

V. Smith (1962). An Experimental Study of Competitive Market Behavior

C. Clotfelter (1983). Tax Evasion and Tax Rates: An Analysis of Individual Returns

G. Yaniv (1999). Tax Compliance and Advance Tax Payments: A Prospect Theory Analysis

E. Kirchler, E. Hoelzl, Ingrid Wahl (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework

Тема 12. Поведенческие финансы

Влияние поведенческих факторов на работу финансовых рынков.

Биграммы:

behavioral_finance, stock_market, financial_market, investment_decision, stock_price, individual_investor, decision_making, disposition_effect, stock_return, efficient_market, stock_exchange, investor_sentiment, finance_theory, asset_pricing, capital_market

Авторы:

Hersh_Shefrin, Martin_Weber, Meir_Statman, Michael_Pompian, Nicholas_Barberis, Zamri_Ahmad, Avaniidhar_Subrahmanyam, Robert_J_Shiller, Ezzeddine_Ben_Mohamed, David_A_Hirshleifer, Victor_Ricciardi, John_R_Nofsinger, Henriette_Prast, Enrico_Maria_Cervellati

Ключевые ссылки:

E. Fama (1970). EFFICIENT CAPITAL MARKETS: A REVIEW OF THEORY AND EMPIRICAL WORK*

W. D. Bondt, R. Thaler (1985). Does the Stock Market Overreact

N. Barberis, A. Shleifer, R. Vishny (1997). A Model of Investor Sentiment

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

Kent D. Daniel, D. Hirshleifer, Avanidhar Subrahmanyam (1998). Presentation Slides for 'Investor Psychology and Security Market Under and Overreactions'

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

Narasimhan Jegadeesh, S. Titman (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency

Brad M. Barber, Terrance Odean (1998). Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment

N. Barberis, R. Thaler (2002). A Survey of Behavioral Finance

Terrance Odean (1998). Are Investors Reluctant to Realize Their Losses

E. Fama, K. French (1993). Common risk factors in the returns on stocks and bonds

Brad M. Barber, Terrance Odean (2000). Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors

J. D. de Long, A. Shleifer, L. Summers, R. Waldmann (1990). Noise Trader Risk in Financial Markets

Тема 13. Экспериментальные подходы к анализу готовности платить

Анализ намерений о покупке, готовности платить с помощью методов экспериментальной экономики.

Биграммы:

framing, information_security, willingness_pay, decision_making, experimental_economics, behavioral_economics, purchase_intention, business_model, social_behavior, price_discount, small_business, social_medium, field_experiment, real_estate, framing_effect

Авторы:

Kent_B._Monroe, Alessandro_Acquisti, Dhruv_Grewal, Eyal_Gamliel, Robert_Slonim, Vernon_L._Smith, Jayson_L._Lusk, Agnès_Festré, Wiktor_L._Adamowicz, Julia_M._Puaschunder, Anna_S._Mattila, James_C._Cox, Mario_Macis, Andrei_Simonov

Ключевые ссылки:

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

R. Thaler (2008). *Mental Accounting and Consumer Choice*

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

E. C. O. N. *Ometrica* (2007). Prospect Theory : An Analysis of Decision under Risk

S. Chen, K. B. Monroe, Y. Lou (1998). The effects of framing price promotion messages on consumers' perceptions and purchase intentions

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

A. Krishna, Richard A. Briesch, D. Lehmann, Hong Yuan (2002). A Meta-Analysis of the Impact of Price Presentation on Perceived Savings

V. Morwitz, E. Greenleaf, E. Johnson (1998). Divide and Prosper: Consumers' Reactions to Partitioned Prices

D. Kahneman, A. Tversky (2000). Choices, Values, and Frames

H. Simon (1955). A Behavioral Model of Rational Choice

U. Gneezy, A. Rustichini (2000). Pay Enough or Don't Pay at All

Albert J. Della Bitta, K. B. Monroe, J. M. McGinnis (1981). Consumer Perceptions of Comparative Price Advertisements

M. Yadav, K. B. Monroe (1993). How Buyers Perceive Savings in a Bundle Price: An Examination of a Bundle's Transaction Value:

Тема 14. Игровые подходы и теория перспектив в решении различных проблем

Экспериментальные подходы к анализу принятых решений в различных областях. Эволюционная теория игр.

Биграммы:

framing, real_estate, behavioral_economics, evolutionary_game, endowment_effect, decision_making, estate_market, prospect_theory, game_theory, bounded_rationality, social_capital, framing_effect, time_preference, loss_aversion, risk_preference

Авторы:

Oliver_Musshoff, Quang_Nguyen, Oliver_Mußhoff, Matthias_Sutter, Sytner_Christin_Maart, Daniel_Hermann, Narayan_B._Mandayam, H._Vincent_Poor, Joachim_Winter, Stein_T._Holden, Shreyas_Sundaram, Walid_Saad, Ashish_R._Hota, Nicholas_Barberis

КЛЮЧЕВЫЕ ССЫЛКИ:

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk Econometrica 47

A. Tversky, D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

C. Holt, Susan K. Laury (2002). Risk Aversion and Incentive Effects

David I. Laibson (1997). Golden Eggs and Hyperbolic Discounting

R. Thaler (1980). Toward a positive theory of consumer choice

D. Kahneman, J. Knetsch, R. Thaler (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias

D. Kahneman, J. Knetsch, R. Thaler (1990). Experimental Tests of the Endowment Effect and the Coase Theorem

Botond Köszegi, M. Rabin (2004). A Model of Reference-Dependent Preferences

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

Ted O'Donoghue, M. Rabin (1999). Doing It Now or Later

H. Simon (1955). A Behavioral Model of Rational Choice

Тема 15. Политика в отношении современных проблем

Государственная политика в отношении современных проблем. Использование больших данных, продовольственная безопасность, проблемы устойчивого развития, экологии, водных ресурсов

Биграммы:

framing, sustainable_development, food_security, decision_making, natural_resource, environmental_policy, public_policy, climate_change, policy_maker, european_union, policy_making, big_data, financial_crisis, public_good, water_resource

Авторы:

Jeffrey_Bennett, Clive_L._Spash, John_Rolfe, Les_Levidow, Joseph_E._Stiglitz, Theresa_Selfa, Achim_Spiller, Nassim_Nicholas_Taleb, Ronald_Labonté, Pierre_Desmet, Jari_Lyytimäki, Emmanuel_Flachaire, Carmen_Bain, Paul_Chaney

Ключевые ссылки:

R. Entman (1983). Framing: Toward Clarification of a Fractured Paradigm

Robert D. Benford, D. Snow (2000). Framing Processes and Social Movements: An Overview and Assessment

E. Goffman (1974). Frame analysis: An essay on the organization of experience

John W. Kingdon (1984). Agendas, alternatives, and public policies

P. Hall (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain

I. Ajzen (1991). The theory of planned behavior

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

W. Gamson, A. Modigliani (1989). Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach

S. Iyengar (1991). Is anyone responsible? How television frames political issues.

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

S. Sudman, R. Mitchell, R. Carson (1989). Using Surveys to Value Public Goods: The Contingent Valuation Method.

D. Scheufele (1999). Framing as a theory of media effects
M. Hajer (1995). The Politics of Environmental Discourse
D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

Неинтерпретируемые или плохо интерпретируемые темы

Отметим, что не интерпретируемой является одна тема (номер 10). Тема 7 является плохо интерпретируемой.

Тема 7. Влияние фрейминга и других факторов на покупательную способность; электронная коммерция

Международная электронная коммерция, фрейминг, межвременной выбор, дисконтирование. Экспериментальные подходы.

Биграммы:

framing, e_commerce, decision_making, intertemporal_choice, cross_border, bounded_rationality, discount_rate, border_e, framing_effect, experimental_economics, behavioral_economics, hyperbolic_discounting, local_government, economic_theory, long_term

Авторы:

Taiki_Takahashi, Julia_M._Puaschunder, David_Laibson, Rudolf_Richter, George_Ainslie, Jonathan_Baron, Rob_Ranyard, Robert_J._Bloomfield, Sanjit_Dhami, Ali_al-Nowaihi, Charles_R._Plott, Kevin_McCabe, Kirsten_I._M._Rohde, Shunsuke_Managi

Ключевые ссылки:

David I. Laibson (1997). Golden Eggs and Hyperbolic Discounting
E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk
G. Loewenstein, D. Prelec (1992). Anomalies in Intertemporal Choice: Evidence and an Interpretation
R. Thaler (1981). Some empirical evidence on dynamic inconsistency

S. Frederick, G. Loewenstein, Ted O'Donoghue (2002). Time Discounting and Time Preference: A Critical Review

P. Samuelson (1937). A Note on Measurement of Utility

A. Tversky, D. Kahneman (1981). The framing of decisions and the psychology of choice.

R. H. Strotz (1955). Myopia and Inconsistency in Dynamic Utility Maximization

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

R. Thaler (2008). Mental Accounting and Consumer Choice

Ted O'Donoghue, M. Rabin (1999). Doing It Now or Later

U. Ben-Zion, A. Rapport, J. Yagil (1989). Discount rates inferred from decisions: an experimental study

Тема 10.

Биграммы:

supply_chain, framing, decision_making, behavioral_economics, public_health, health_insurance, long_term, nudging, optimal_policy, lot_sizing, routing_problem, order_quantity, health_care, lead_time, hedge_fund

Авторы:

Howard_Kunreuther, Mark_V._Pauly, Cass_R._Sunstein, Stacey_McMorrow, Frank_J._Fabozzi, Xiaobo_Zhao, Svetlozar_T._Rachev, Anya_Samek, Jayson_L._Lusk, B._Douglas_Bernheim, Abootaleb_Shirvani, Stefan_Minner, Ding_Du, Luk_N._Van_Wassenhove

Ключевые ссылки:

E. C. O. N. Ometrica (2007). Prospect Theory : An Analysis of Decision under Risk

M. Schweitzer, Gérard P. Cachon (2000). Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence.: Experimental Evidence.

D. Kahneman, A. Tversky (1979). Prospect theory: An analysis of decision under risk *Econometrica* 47

A. Tversky, D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases.

A. Tversky, D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty

D. Kahneman, A. Tversky (1979). Prospect theory: analysis of decision under risk

Gary E. Bolton, Elena Katok (2008). Learning-by-Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback

A. Bostian, C. Holt, A. Smith (2008). Newsvendor "Pull-to-Center" Effect: Adaptive Learning in a Laboratory Experiment

Michael Becker-Peth, Elena Katok, U. W. Thonemann (2013). Designing Buyback Contracts for Irrational But Predictable Newsvendors

H. Simon (1955). A Behavioral Model of Rational Choice

David I. Laibson (1997). Golden Eggs and Hyperbolic Discounting

R. Thaler (2008). Mental Accounting and Consumer Choice

Teck-Hua Ho, Noah Lim, T. Cui (2010). Reference-Dependence in Multi-Location Newsvendor Models: A Structural Analysis

Xuanming Su (2007). Bounded Rationality in Newsvendor Models

Результаты работы модели выложены на специально созданном сайте <http://behavioral.site>.

Отметим, невозможность интерпретации небольшого числа тем не является причиной для признания неэффективности алгоритма.

В Таблице 5 Приложения 3 приведены примеры документов по каждой из представленных выше тем.

1.5.3. Анализ чувствительности и другие тестирования

Чувствительность модели проверялась на основе 10 запусков модели для 15 тем, с различными начальными инициализациями матрицы Φ нормированными случайными величинами.

Значения средней когерентности для тем приведены в Таблице 8.

Таблица 8 – Анализ чувствительности. Значения средней когерентности

Номер запуска	Средняя когерентность
1	-1.6582
2	-2.7488
3	-1.8428
4	-2.4951
5	-2.8328
6	-3.0041
7	-2.5849
8	-3.2852
9	-2.3434
10	-2.2601
Исходн. модель	-1.9052

Среднее значение отклонения средней когерентности равно 0.6. Минимальное 0.247, максимальное – 1.38.

Значения контролируемых метрик качества для 10 запусков приведены в Таблице 6 Приложения 3.

Также было отмечено, что оптимальные значения коэффициентов регуляризации могут меняться в зависимости от начального приближения.

Пример ключевых биграмм для одной из моделей, построенной в ходе анализа чувствительности приведен в Таблице 7 Приложения 3. Видно, что для 12 из 15 тем исходной модели находятся аналогичные по интерпретации темы для модели, построенной по новой инициализации.

Другие тестирования

Также проводилось тестирование на оптимальный состав модальностей. Показано, что именно включение слов, биграмм, авторов и ссылок дает наилучшие результаты (наибольшая средняя когерентность) (Таблица 8 Приложения 3).

В Таблице 9 Приложения 3 приведены также результаты тестирования весов модальностей. Для биграмм вес фиксировался равным 1.0. Для слов значения весов перебирались равными [0.5, 1]. Для авторов значения весов перебирались равными [1.0, 1.5, 2.0], для ссылок – [0.5, 1.0]. Наилучшие значения когерентности и перплексии достигаются при весах слов, биграмм, авторов, ссылок равными 0.5, 1.0, 1.0, 1.0 соответственно.

3.2. Разведочный поиск на примере патентного анализа для плана импортозамещения

Задача получения представления о ключевых аспектах больших текстовых коллекций может быть сформулирована не только как задача мягкой кластеризации всей коллекции документов, но и как задача кластеризации только определенной части документов по априорно задаваемым правилам. Остальные документы при этом не представляют интереса и отбрасываются. Если подход, описанный в разделе 3.1. относится к первому типу задач (кластеризация всей коллекции), то в данном разделе представлен пример решения задачи второго типа.

3.2.1. Постановка задачи исследования

В связи с реализацией государственной программы по достижению технологического суверенитета России, в 2015 году был законодательно утвержден ряд положений, регламентирующий планы мероприятий по импортозамещению в 22 отраслях (далее – Программа). В каждой из отраслей составлен перечень товаров и технологий, по которым установлен свой показатель доли импорта к 2020 году¹⁷.

Исследование проводилось в преддверии окончания установленного для реализации Программы срока, и было важно представить некоторые результаты импортозамещения, основанные на анализе патентных данных (Милкова, 2020(б)). Несмотря на исключительную важность проведения подробного анализа по каждому из направлений развития, полезно иметь и общую структуру результатов.

¹⁷ Минпромторг России. Отраслевые планы импортозамещения. URL: <https://gisp.gov.ru/plan-import-change/> (дата обращения: 25.09.2021)

Подход, охватывающий сразу все отрасли, позволит как продемонстрировать результаты выполнения программы в целом, так и даст общее представление о состоянии различных отраслей экономики (на основе патентных данных).

Ключевыми документами Программы являются планы импортозамещения (далее – Планы), содержащие перечень продуктов/ технологий для импортозамещения, фактические показатели доли импорта до начала реализации программы и планируемые показатели к 2020 году.

Данные показатели по каждой из 22 отраслей визуализированы на Рисунке 9 (диаграммы Box-Whiskers). Наиболее «амбициозными» отраслями, заявляющими снижение доли импорта на 80% и более (медианные значения) являются: Автомобильная промышленность; Транспортное машиностроение; Фармацевтическая промышленность; Гражданское авиастроение. К отраслям, обозначившим минимальные показатели снижения доли импорта (не более 30%, медианные значения) относятся: Детские товары; Станкоинструментальная промышленность; Энергетическое машиностроение, кабельная и электротехническая промышленность; Нефтегазовое машиностроение.

Общий план импортозамещения подразумевал снижение доли импорта с фактического медианного значения 90% до планового медианного значения 15% (Рисунок 10).

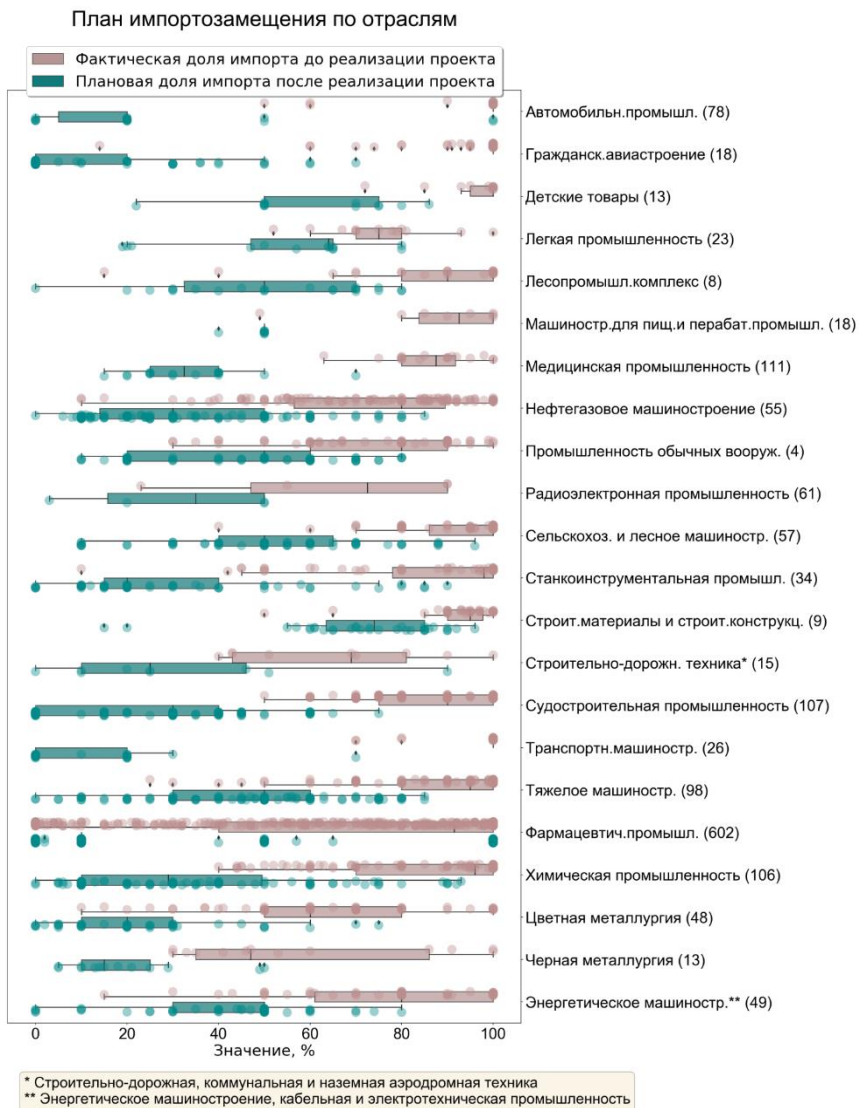


Рисунок 9 – Диаграмма Box-Whiskers по отраслям для фактической доли импорта до реализации плана по импортозамещению и плановой доли импорта после реализации программы. В скобках приведено число пунктов в каждом из отраслевых планов

План импортозамещения общий, доля импорта
(22 отрасли, 1553 продукта/технологии)

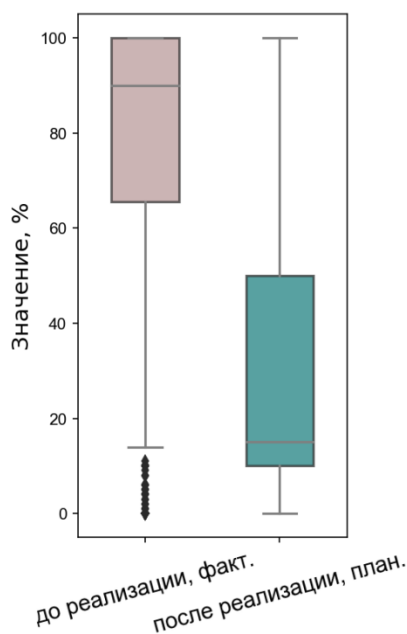


Рисунок 10 – Диаграмма Box-Whiskers для общей доли импорта до реализации Программы и плановой доли после реализации

В настоящее время, многие поисковые и аналитические системы (например, Яндекс.Патенты, Google Patents, Patseer) используют передовые достижения компьютерной лингвистики, в том числе методы семантического анализа текстов. Современные поисковые системы способны находить схожие патенты, смежные патенты (в которых упоминается тот или иной интересующий пользователя документ, или на которые он ссылается). Поиск похожих патентов осуществляется не только по ключевым словам, но и по смыслу. Стоит отметить, что современные системы патентного поиска и аналитики удобны, если необходимо получить информацию о небольшом числе объектов. В случае если интересующих объектов много, осуществление поиска требует значительных временных вложений.

Для анализа реализации плана по импортозамещению необходимо получить информацию по всем 1553 пунктам Плана, что, очевидно, требует принципиально иного подхода к осуществлению патентного поиска.

3.2.2. Построение модели

В качестве метода, позволяющего решить данную проблему, было выбрано тематическое моделирование. Результатом построенной тематической модели

является, с одной стороны, мягкая кластеризация патентных документов по заданному числу тем, а с другой стороны – выявленный набор слов или словосочетаний, характеризующий с определенными вероятностями каждую из тем. Кроме того, в модель можно добавлять некоторую априорную информацию об интересующих терминах, на основе которых необходимо сформировать темы (концепция ARTM). Таким образом, для нашей задачи ядром каждой из тем были слова и словосочетания из отраслевых планов импортозамещения.

Для построения тематической модели были собраны патенты на изобретения и полезные модели, выданные за 3,5-летний период (январь 2016- июнь 2019 гг.)- всего 152718 документов: 120768 изобретений и 31950 полезных моделей.

Тематическая модель строилась на основе Названий и Абстрактов патентов, представленных в виде униграмм (т.е.одиночных слов). В отдельную модальность были выделены наиболее частотные биграммы (двусловные словосочетания с частотой встречаемости в Названии и Абстракте более или равной 2), оптимальный вес модальности биграмм определялся экспериментально и был выбран равным $\tau_m = 5$.

Были введены 22 регуляризатора сглаживания, по одному для каждой из тем. Коэффициент сглаживающих регуляризаторов (абсолютные значения) был выбран равным $1e + 7$.

Итоговая модель обладала следующими метриками качества: доля разреженных элементов матриц униграмм $\Phi^1 = 0,994$, биграмм $\Phi^2 = 0,998$, $\Theta = 0,818$. Размер ядра равен 628, средняя чистота тем 0,992, средняя контрастность тем 0,976. Общее число итераций: 40.

3.2.3. Результаты

Итогом построенной модели стала кластеризация патентных документов: выбраны патентные документы по каждой из 22 отраслей в соответствии с темой, характеризующей набором слов и словосочетаний из соответствующего Плана.

Помимо стандартных автоматически вычисляемых метрик, качество модели также оценивалась с помощью ассессоров, определяющих, насколько релевантным является отобранный документ. Патентному документу ставилось в соответствие

значение quality $q=1$ в случае, если патент точно соответствовал одному из заявленных в Плате пункту импортозамещения; значение $q=0,5$ присваивалось в случае, если патент связан с одним из пунктов Плате; $q=0$ – если не соответствовал ни одному из пунктов Плате. Данная методика успешно использовалась в работах (Янина, Воронцов, 2016; Arishev, Koltsov et al., 2016) и позволяет, единожды сделав разметку результатов поиска, многократно вычислять оценки качества тематического поиска для различных реализаций модели.

Для документов со значениями $q=1$, $q=0,5$ была выделена ключевая фраза/слово, характеризующая как документ, так и его принадлежность к тому или иному пункту Плате. Какие именно позиции импортозамещались в каждой из отраслей наглядно показано на диаграммах Sankey (диаграммы приведены для двух отраслей: Рисунок 11 – Черная металлургия и Рисунок 12 – Лесопромышленный комплекс).

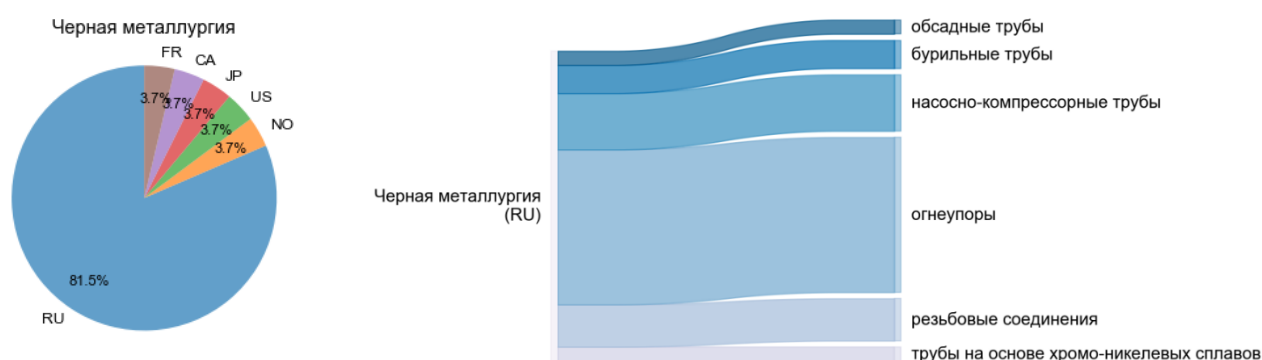


Рисунок 11 – Черная металлургия. Распределение по странам-патентообладателям и категориям патентования российскими патентообладателями

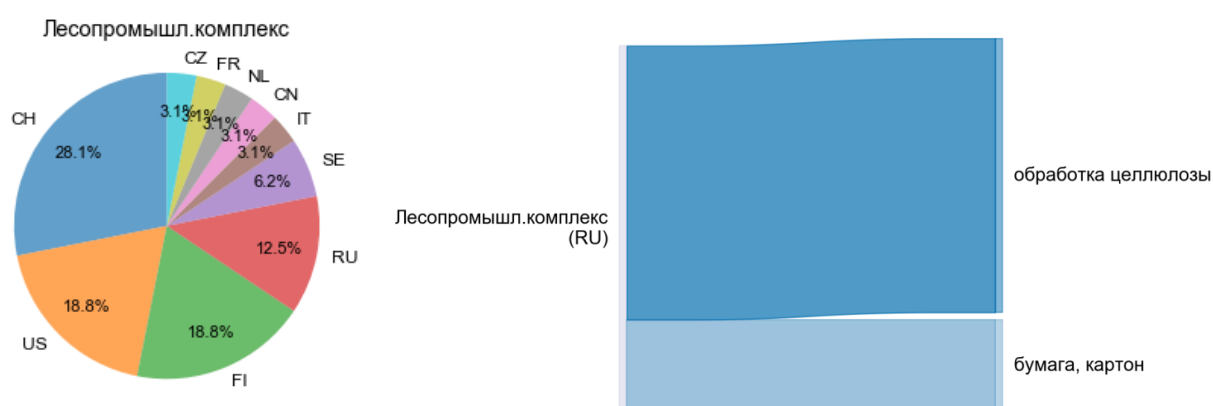


Рисунок 12 – Лесопромышленный комплекс. Распределение по странам-патентообладателям и категориям патентования российскими патентообладателями
 В Таблице 9 агрегированы результаты для всех отраслей.

Таблица 9 – Характеристики импортозамещения на основании патентных данных

Отрасль	Процент российских патентов	Категории импортозамещения (RU)	Число категорий, k	Средний балл, q	Суммарн. значение балла
Автомобильн. промышл.	74,1%	двигатель внутреннего сгорания	1	0,93	18,5
Гражданск. авиационное		-	-	-	-
Детские товары	95,2%	мебель для детей; игры и игрушки; спортивные комплексы; детская одежда; детское творчество	5	0,95	19,0
Легкая промышленность	52,2%	нетканые материалы; защитная одежда; переработка шерсти	3	0,96	11,0
Лесопромышл. комплекс	12,5%	обработка целлюлозы; бумага, картон	2	0,75	3,0
Машиностр. для пищ. и перабат. промышл.	100,0%	обработка зерновых	1	0,83	2,5
Медицинская промышленность	41,7%	стерилизация и дезинфекция; эндоскопические аппараты; иглы инъекционные; имплантируемые насосы	4	0,90	4,5

Нефтегазовое машиностроение	78,3%	катализаторы гидроочистки; бурение скважин; катализаторы гидрокрекинга; переработка углеводородного сырья; гидроразрыв пласта; катализаторы каталитического крекинга	6	1,00	18,0
Промышленность обычных вооруж.	82,4%	патроны; спортивное оружие	2	0,79	11,0
Радиоэлектронная промышленность		-	-	-	-
Сельскохозяй. и лесное машиностр.	83,3%	подшипники; зерноуборочный комбайн; пресс подборщик	3	0,70	17,5
Станкоинструментальная промышл.	78,9%	фрезерный станок; токарный станок; расточный станок; шпиндели; финишное шлифование; гидроабразивная резка; станки чпу	7	0,93	14,0
Строит. материалы и строит. конструкц.	96,6%	керамическая масса для плитки; теплоизоляционные материалы; щебеночно-мастичные асфальтобетоны	3	0,68	19,0
Строительно-дорожн. техника*	93,3%	дорожное покрытие; гидравлическое оборудование; фронтальные погрузчики; бульдозеры; фронтальный погрузчик; экскаватор; прицеп и полуприцеп; крановое шасси; коммунальная техника	9	0,79	11,0
Судостроительная промышленность	92,9%	двигатель; гребневой винт	2	0,50	6,5
Транспортн. машиностр.	62,7%	вагон-цистерна; тормозная система; тележки вагона; крытый вагон	4	0,81	26,0
Тяжелое машиностр.	50,0%	крепь горная; холодильные установки	2	0,75	1,5

Фармацевтич. промышл.	56,3%	инозин + никотинамид + рибофлавин + янтарная кислота; висмут калий аммоний цитрат; дротаверин; йогексол; лопинавир + ритонавир; этилметилгидроксипиридина сукцинат; рокурония бромид; дигоксин; 1 карбамоилметил 4 фенил 2 пирролидон; фенспирид; изониазид; лапшаконитина гидробромид; иммуноглобулин стандартный; бромдигидрохлорфенил-бензодиазепин; десмопрессин; финголимод; анастрозол	17	0,94	17,0
Химическая промышленность	87,5%	лакокрасочные материалы; уплотнительные материалы; эпоксидный композит; клеевые материалы; полиэтилентерефталат; сверхвысокомолекулярный полиэтилен; полимерные композиты	7	0,79	11,0
Цветная металлургия	88,6%	алюминиевый сплав; алюминий, электролиз; алюминиевая лигатура; гидроксид алюминия; алюминиевый порошок; алюминиевая фольга; алюминиевые прутки; анодная масса	8	0,81	25,0
Черная металлургия	81,5%	огнеупоры; насосно-компрессорные трубы; резьбовые соединения; бурильные трубы; трубы на основе хромоникелевых сплавов; обсадные трубы	6	0,89	19,5
Энергетическое машиностр.**	50,0%	трансформаторы тока	1	1,00	1,0

Подобный вариант кластеризации удобен, так как позволяет емко представить структуру коллекции патентных документов, сформированную уже на основе только тех слов и словосочетаний, которые представляют для нас интерес.

В зависимости от числа найденных релевантных патентных документов, степени их соответствия заявленному плану, а также общему количеству пунктов Плана, каждая из отраслей получила свой рейтинг: $Score = \frac{\sum(q) \cdot k}{N}$, где k – число категорий – различных позиций Плана, по которым были найдены релевантные патентные документы ($q=1, q=0,5$); N – общее число пунктов Плана. Результаты ранжирования отраслей представлены на Рисунке 13.

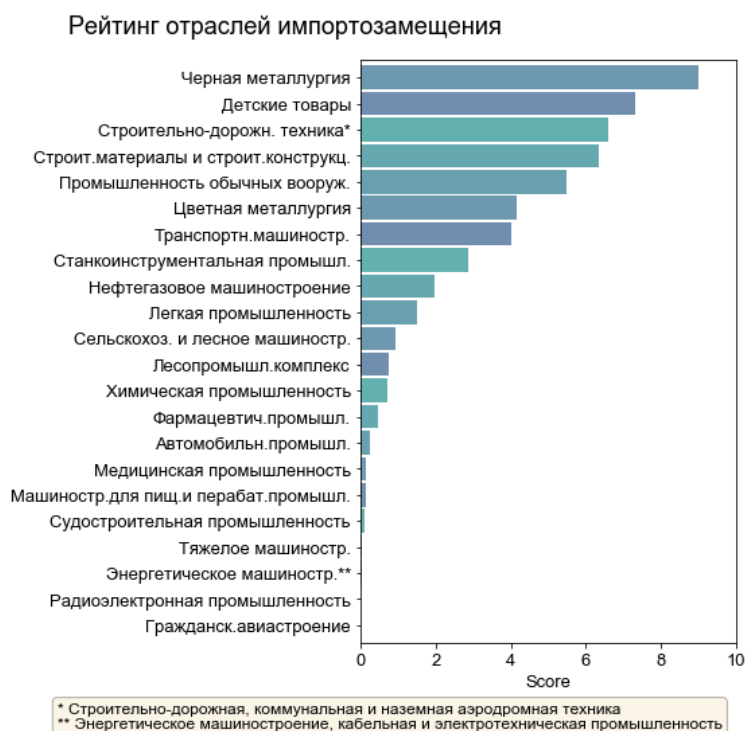


Рисунок 13 – Рейтинг отраслей импортозамещения на основании соответствия выданных патентов заявленному плану импортозамещения отрасли

Итак, в контексте выданных патентов на изобретения и полезные модели, к отраслям, продемонстрировавшим лучшие показатели импортозамещения можно отнести Черную металлургию; Детские товары; Строительно-дорожную, коммунальную и наземную аэродромную технику; Строительные материалы и строительные конструкции; Промышленность обычных вооружений; Цветную металлургию; Транспортное машиностроение.

Отрасли, которые на данный момент не способны соответствовать плану импортозамещения (на основании выданных патентных документов) – это Гражданское авиастроение; Радиоэлектронная промышленность; Энергетическое машиностроение, кабельная и электротехническая промышленность; Тяжелое

машиностроение; Судостроительная промышленность; Машиностроение для пищевой и перерабатывающей промышленности; Медицинская промышленность.

Важно, что полученная структура позволяет при необходимости детализировать результаты. Например, выявлять долю индивидуальных патентообладателей, которые не смогут стать основными агентами захвата ниш рынка и не смогут конкурировать с крупными иностранными компаниями; долю не действующих патентов и т.д.

3.3. Другие приложения

В рамках работы по изучению и апробации различных инструментов текстового анализа с целью описания информационных продуктов на основе семантической компрессии информации (сжатого представления смысла), автором также были проведены следующие исследования:

1. Извлечение ключевых терминов направления Цифровая экономика (на основе анализа правительственных документов, выпущенных в 2013-2018 гг.) (Милкова, 2018). Применение графового алгоритма TextRank к каждому из документов позволило построить графы семантических связей между ключевыми словами, на основании которых возможно выделение ключевых терминов. Задача извлечения ключевых терминов из текста является основополагающей при изучении отраслевой терминологии. Построенные графы семантических связей наглядно показывают, как меняются в близких по времени создания правительственных документах лексические связи одних и тех же слов. Употребляя одни и те же слова в разных словосочетаниях и контекстах, авторы текстов уходят от чётких границ терминов. Тем самым, размывается их содержание, исчезает терминологический статус.

Представленный анализ позволяет сделать некий вывод, что терминология цифровой экономики на данный момент имеет характер не сложившейся, незавершенной совокупности, что является следствием нерешенности проблем государственного и законодательного плана. С другой стороны, с точки зрения терминоведения очевидно, что термины цифровой экономики обладают признаками

«концептуализованной научной картины мира»¹⁸ – отражают фундаментальность и глобальность как происходящих на данный момент, так и грядущих изменений всех сфер человеческой деятельности.

2. Выделение ключевых слов и суммаризация текстов (выделение ключевых предложений) нормативных документов, регламентирующих вопросы питания и предупреждения железодефицитной анемии (Милкова и др., 2021). Предметом исследования в данной работе являлось извлечение значимой информации из российских нормативных документов, регламентирующих вопросы обеспечения достаточного количества микронутриентов, в частности, железа. Сформированная выборка (по ключевым словам, из справочно-правовой системы Консультант Плюс) обобщает разнородную информацию: нормы потребления для различных категорий граждан, список рекомендованных продуктов, особенности налогообложения сельскохозяйственных товаропроизводителей, распределение субсидий на стимулирование подотраслей агропромышленного комплекса и др. Рассматриваемые документы, как правило, достаточно объемные и содержат нормативы различного рода, поэтому важной задачей является емкое извлечение ключевой информации. Кроме того, некоторые из нормативов часто обновляются, и актуальной задачей является автоматическое фиксирование изменений ключевых параметров. Выделение ключевых слов и ключевых предложений реализовано на основе алгоритма TextRank. Важным аспектом работы алгоритма является способ определения весов ребер (или меры близости вершин – слов или предложений). В данной работе для определения меры близости предложений использовался алгоритм VM-25, позволяющий присваивать более высокий вес связям между предложениями, схожими не частотными словами.

3. Представленные в данной главе результаты имеют отношение к использованию текстовых методов анализа как инструментов поиска информации, минимизируя сдвиги восприятия. Другим, более ранним полем исследований автора явилась работа по минимизации сдвигов восприятия в разрезе принятия

¹⁸ Ордокова, Ф.М. Принципы формирования отраслевой терминологии (На материале терминов экономики сельского хозяйства) / Диссертация на соискание степени кандидата филологических наук: 10.02.19. Нальчик, 2004

многокритериальных решений. Предлагался инструмент, позволяющий избежать упрощения принятия решений на основе эвристик – использовался метод аналитических сетей (в том числе, его упрощенной форме – метода анализа иерархий). В основе инструмента – методология АНР/АНР¹⁹, которая также, как и подход Тверски и Канемана, опирается на принятие решений на основе эвристик, однако не упрощает исследуемую проблему, а представляет ее в виде сетевой или иерархической структуры, которая включает в себя все рассматриваемые альтернативы, оценочные критерии, их характеристики, а также взаимосвязи (Milkova et al., 2019(a), 2019(b)). Решение принимается на основе попарных сравнений, производимых по 9-балльной шкале и последующим синтезом результатов с учетом всех весов и связей (в том числе обратных) между критериями и альтернативами.

Разработана и зарегистрирована система Multichoice, позволяющая проводить полный цикл работы алгоритма АНР/АНР: построения сети (или иерархии), проведение попарных сравнений, синтез результатов, анализ чувствительности²⁰.

Выводы Главы 3

Данная глава демонстрирует адекватность применения подходов семантической компрессии информации, которые позволяют извлекать наиболее значимую информацию и преобразовать ее в структурированную форму с целью облегчения понимания и дальнейшего использования и анализа. Отметим, что в целом, методы смысловой компрессии не отменяют дальнейшее линейное чтение интересующих текстов, однако с учетом знаний ключевых положений всех документов.

Применение формализованного алгоритма АРТМ к задаче выявления структуры научных публикаций направления поведенческая и экспериментальная экономика позволило выявить ключевые подтемы направления. Кроме того, по каждой из тем определены ключевые фразы, опорные публикации (статьи, ссылки на

¹⁹ Analytic Hierarchy Process, Analytic Network Process. Методология разработана Томасом Саати. См., например: Saaty, T. L. How to make a decision: the Analytic Hierarchy Process // European Journal of Operational Research. 1990. Vol. 48, pp. 9–26.

²⁰ Свидетельство о регистрации программы для ЭВМ №2016616698, Multichoice / Милкова М.А., Андрейчикова О.Н. // Федеральная служба по интеллектуальной собственности, 17.06.2016

которые чаще всего присутствуют в публикациях по данной теме), авторы, публикующиеся по данной теме. Ключевые фразы предоставляют информацию о терминологии темы и могут быть в дальнейшем использованы для, например, осуществления итерационного поиска (часто эффективность итерационного поиска снижена, так как терминология области может быть заранее неизвестна исследователю). Ключевые ссылки показывают, с какими публикациями, возможно, стоит познакомиться в первую очередь при исследовании данной темы. Ключевые авторы – какие авторы в основном публиковались по данной теме (отбор авторов не учитывает цитирования).

Результаты работы модели представлены на специально созданном сайте <http://behavioral.site>

Построенный формализованный алгоритм представляет собой набор программ на языке Python и позволяет использовать его для проведения аналогичных исследований (построение тематической кластеризации научных статей любой области).

Возможно дальнейшее расширение практики использования формализованного алгоритма АРТМ (библиографические исследования, форсайт исследования), а также любых других исследований, ставящих перед собой задачу получения знаний из больших коллекций текстовых документов.

В рамках второго исследования была решена задача анализа программы импортозамещения по 22 отраслям промышленности на основе патентных данных. Подход АРТМ позволяет осуществлять поиск релевантных патентных документов по отраслям, каждая из которых характеризуется своим набором продуктов/технологий для импортозамещения. Выбранная подборка документов позволяет составить некий рейтинг отраслей на основе соответствия программе импортозамещения.

Анализ показал, что в контексте выданных патентов на изобретения и полезные модели, к отраслям, продемонстрировавшим лучшие показатели импортозамещения можно отнести Черную металлургию; Детские товары; Строительно-дорожную, коммунальную и наземную аэродромную технику; Строительные материалы

и строительные конструкции; Промышленность обычных вооружений; Цветную металлургию; Транспортное машиностроение. Отрасли, которые на данный момент не способны соответствовать плану импортозамещения (на основании выданных патентных документов) – это Гражданское авиастроение; Радиоэлектронная промышленность; Энергетическое машиностроение, кабельная и электротехническая промышленность; Тяжелое машиностроение; Судостроительная промышленность; Машиностроение для пищевой и перерабатывающей промышленности; Медицинская промышленность.

Кроме того, по каждой из отраслей составлен перечень продуктов/технологий, по которым отрасль демонстрирует возможности для импортозамещения в разрезе выданных патентов.

Автором также решены следующие задачи представления ключевой информации для описания информационных продуктов: 1) анализ терминологии направления Цифровая экономика на основе правительственных документов; 2) анализ политики России в области питания и продовольствия на основе нормативных документов. Несколько иным исследованием является разработка системы для снижения сдвигов восприятия при принятии решений на основе метода аналитических сетей/иерархий.

ЗАКЛЮЧЕНИЕ

В процессе проведения диссертационного исследования были получены следующие результаты.

1. В данной работе предлагается способ организации потребления информационных продуктов, помощи выбора из множества. В основе лежат идеи (Г. Саймон, Ф. Хайек), что 1) внимание ограничено определенным классом результатов, которые мы ожидаем видеть, 2) внимание определяет ту узкую избирательную область, которая выделяется в общем проблемном пространстве, внутри которой осуществляется поиск решений, 3) знания имеют фрагментарную природу.

Дизайн цифрового мира влияет на поведение и принимаемые решения, и использование достижений поведенческих наук в сочетании с цифровыми технологиями способно оказывать существенное воздействие. Таким образом, учитывая особенности потребления информации экономическими агентами разработан и показан в экспериментах способ организации информационных продуктов, позволяющий сконфигурировать среду для восприятия информации, чтобы 1) снизить затраты на поиск ключевой информации по исследуемой научной теме; 2) обеспечить полноту и ценность (с точки зрения смысла) извлекаемой информации; 3) минимизировать сдвиги в восприятии (сместить фокус внимания с первой попавшейся, эмоциональной информации). Применительно к текстовой информации, данным требованиям удовлетворяет подход аддитивной регуляризации тематических моделей. Таким образом, именно инструментарий тематического моделирования (для семантической компрессии текстовой информации) может оказать помощь как в управлении ограниченной рациональностью, так и в информационном взаимодействии производителей и потребителей научного знания.

2. Формализован подход к организации потребления информационных продуктов, помощи выбора из множества. В частности, на основе коллекции англоязычных научных статей по поведенческой и экспериментальной экономике из базы SemanticScholar проведены расчеты, показывающие, что наилучшей стратегией построения тематической модели (с точки зрения однозначной интерпретации тем; наибольшее значение когерентности) является:

- Построение модели на основе следующей метаинформации (включения следующих модальностей): 1) одиночных слов; 2) двухсловных словосочетаний (биграмм) Названий и Аннотаций статей; 3) авторов статей; 4) списков использованной литературы. Веса модальностей предлагается брать равными 0.5 для одиночных слов; 1.0 для биграмм и ссылок, 2.0 для авторов.

- Использование в качестве контролируемых метрик качества средней когерентности тем, перплексии. Также контролируется степень разреженности матриц Φ , Θ .

- Выбор оптимального числа тем совместно с относительным коэффициентом декоррелирования. Для этого осуществляется 10 итераций с включенным относительным регуляризатором декоррелирования, значения которого перебираются по сетке. Число тем также перебирается по сетке, определяемой экспертно. При этом эффективность включения регуляризатора декоррелирования зависит от природы исходных данных. В случае, если выделяемые темы могут быть сильно различными между собой, декоррелирование оправдано. В нашем же случае общая тематика исходных данных (поведенческая и экспериментальная экономика) подразумевает пересечение в терминологии и библиографических источниках в подтемах, что не позволяет искусственно увеличивать расстояние между подтемами без потери качества.

- На втором этапе для выбранного числа тем (и коэффициента декоррелирования, если он был подключен) аналогичным образом определяется значение коэффициента сглаживания фоновой темы, группирующей внутри себя не тематизируемые слова, слова общей лексики, мусорные слова. Если таковые были выделены на этапе предобработки документов, введение регуляризатора сглаживания не приводит к улучшению качества модели.

- На третьем этапе проводится дополнительно 10 итераций для разреживания матриц Φ и Θ (обнуления низкочастотных термов коллекции). Для этого аналогичным образом по сетке подбирается значение коэффициента разреживания для матрицы Φ и Θ .

3. На основе формализованного алгоритма осуществлена мягкая кластеризация статей по поведенческой и экспериментальной экономике, выделены для каждой из тем ключевые биграммы, авторы и ссылки. Все кроме одной темы являются интерпретируемыми, совпадают с экспертным представлением о подтемах направления, устойчивы к изменению начальных инициализаций матриц Φ и Θ , поддаются экспертному именованию. Без какого-либо экспертного вмешательства в состав тем, на основе алгоритма выделены следующие подтемы (именованы вручную): *Ограниченная рациональность, теория перспектив; Про-экологическое поведение; Ограниченная рациональность, методы подталкивания и поведение*

потребителя; Общие вопросы поведенческой и экспериментальной экономики; Поведенческие факторы в корпоративном управлении; Экспериментальная экономика; Влияние фрейминга и других факторов на покупательную способность, электронная коммерция; Влияние маркетинговой стратегии на потребителя; Поведенческая теория фирмы; Поведенческая экономика и налогообложение; Поведенческие финансы; Экспериментальные подходы к анализу готовности платить; Игровые подходы и теория перспектив в решении различных проблем; Политика в отношении современных проблем

Проведен анализ чувствительности результатов, показывающий устойчивость состава интерпретируемых тем в зависимости от начальных инициализаций модели.

Автоматизированное выделение подтем научной области позволяет 1) составить представление о структуре области; 2) дать представление о ключевых, опорных публикациях подтемы; 3) показать, какие авторы публикуются по каждой из подтем; 4) сделать тематическую подборку публикаций. Подход позволяет повысить эффективность работы научного сообщества путем организации информационного взаимодействия производителей и потребителей научного знания. Также, в свете обсуждений теории подталкивания, в особенности цифрового подталкивания, подход, основанный на использовании тематического моделирования текстовых коллекций документов, позволяет осуществлять своего рода «подталкивание к изучению», предоставляя структуру исследуемого направления и его ключевые положения. Это – первый шаг к получению знаний, не отменяющий дальнейшее углубление в рассматриваемую тему.

4. Создан набор программ на языке Python, позволяющий решать аналогичные задачи тематической кластеризации для других коллекций научных публикаций. Скрипты выложены в репозитории Github: <https://github.com/behavioral-econ-codes/Publications>

5. Создан сайт <http://behavioral.site> с результатами тематической кластеризации научных статей направления поведенческая и экспериментальная экономика.

6. Другой решенной задачей с помощью методов аддитивной регуляризации является анализ программы импортозамещения на основе патентных данных. Руководствуясь пунктами планов импортозамещения по 22 отраслям промышленности (всего 1553 продукта/технологий) возможно осуществить патентный поиск релевантных документов по каждой из отраслей. Данный подход является своего рода «крупным планом» патентного поиска, который может служить как конечной целью, так и являться отправной точкой для более детального анализа.

СПИСОК ЛИТЕРАТУРЫ

1. Афанасьев Д. О., Федорова Е. А., Рогов О. Ю. О влиянии тональности новостей в международных СМИ на рыночный курс российского рубля: текстовый анализ // Экономический журнал ВШЭ. 2019. Т. 23. № 2. С. 264–289.
2. Белянин, А.В. Ричард Галер и поведенческая экономика: от лабораторных экспериментов к практике подталкивания (Нобелевская премия по экономике 2017 года) // Вопросы экономики. 2018. №1. С. 5-25
3. Большакова, Е.И. Компьютерная лингвистика: задачи, подходы, ресурсы: учеб. пособие Автоматическая обработка текстов на естественном языке и анализ данных / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017 — 269 с.
4. Большакова, Е.И., Ефремова Н.Э. Извлечение информации из текстов: портрет направления: учеб. пособие Автоматическая обработка текстов на естественном языке и анализ данных / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017 — 269 с.
5. Булатов, В.Г. Методы оценивания качества и многокритериальной оптимизации тематических моделей в библиотеке TopicNet. Диссертация на соискание ученой степени канд. технич. наук: 05.13.18 / Булатов Виктор Геннадьевич — М., 2020. — 147 с.
6. Варшавский, А. Е., Винокурова, Н. А., Кочеткова, Е. В. Представления ученых о том, какие профессии являются наиболее подходящими для их детей и внуков. Народонаселение. 2020. № 23(1). С. 119-134. <https://doi.org/10.19181/population.2020.23.1.10>
7. Воронцов, К. В., Потапенко, А. А. Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. 2012. 4(4). С. 693–706.
8. Воронцов К. В., Потапенко, А.А. Аддитивная регуляризация тематических моделей // Электронный ресурс. 2014. С. 1-21. URL:

<http://www.machinelearning.ru/wiki/images/2/28/Voron14mlj-rus.pdf> (дата обращения: 29.09.2021)

9. Воронцов, К.В., Потапенко, А.А. Регуляризация вероятностных тематических моделей для повышения интерпретируемости и определения числа тем. Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4 — 8 июня 2014 г.). 2014(6). С. 676-687.

10. Воронцов, К.В. Аддитивная регуляризация тематических моделей коллекций текстовых документов // Доклады академии наук. 2014. 3(456). DOI: 10.7868/S0869565214090096

11. Голощапова, И.О., Андреев М.Л. Оценка инфляционных ожиданий российского населения методами машинного обучения // Вопросы экономики. 2017. № 6. С. 71-93.

12. Докука, С.В. Клиповое мышление как феномен информационного общества // Общественные науки и современность. 2013. № 2. С. 169-176.

13. Жданов, Д.А. Пространство слаборациональных предпочтений: индивидуальные и национальные особенности // Экономическая политика. 2018. Т.13, №4. С. 108-133. <https://doi.org/10.18288/1994-5124-2018-4-06>

14. Кара-Мурза, С.Г. Манипуляция сознанием. М.: Изд-во: Эксмо, 2004 — С. 832.

15. Карпович, С.Н. Тематическая модель с бесконечным словарем // Информационно-управляющие системы. 2016. № 6. С. 43-49. <https://doi.org/10.15217/issn1684-8853.2016.6.43>

16. Клейнер Г.Б. К Методологии Моделирования Принятия Решений Экономическими Агентами // Экономика и математические методы. 2003. Vol. 39(2).

17. Ключарев, В.А., Шмидс, А., и Шестакова, А.Н. Нейроэкономика: нейробиология принятия решений // Экспериментальная психология. 2011. 4(2). С. 14–35

18. Козырев, А.Н. Утопия и антиутопия экономики внимания // Цифровая экономика. 2019. 1(5). С.82-93. <https://doi.org/10.34706/DE-2019-01-08>

19. Козырев, А.Н. Современное состояние исследований в области торговли информацией // Цифровая экономика. 2020. 1(9). 63-75. <https://doi.org/10.34706/DE-2020-01-07>
20. Макаров, В.Л. Экономика знаний: уроки для России // Россия и современный мир. 2004. №1. С. 5-24.
21. Макаров, В.Л., Клейнер, Г.Б. Микроэкономика знаний // Отд. Обществ. Наук РАН, Центр. Эконом.-мат. Ин-т. – М.: ЗАО «Издательство «Экономика», 2007 – 204 с.
22. Милкова М.А. Извлечение ключевых терминов направления «Цифровая экономика»: графоориентированный подход // Цифровая экономика. 2018. № 4(4). С. 57-65. DOI:10.34706/DE-2018-04-06
23. Милкова М.А. Теория подталкивая и ее искажения в информационной среде // Цифровая экономика. 2019(а). № 4(8). С. 21-26. DOI: 10.34706/DE-2019-04-02
24. Милкова М.А. Тематические модели как инструмент «дальнего чтения» // Цифровая экономика. 2019(б). № 1(5). С. 57-70. DOI:10.34706/DE-2019-01-06
25. Милкова М.А. Феномен внимания в информационной среде: экономика внимания // Цифровая экономика. 2020(а). № 3(11). С. 73-87. DOI: 10.34706/DE-2020-03-08
26. Милкова М.А. Инновационный подход к поиску информации на примере патентного анализа плана импортозамещения // Экономическая наука современной России. 2020(б). № 1(88). С. 143-157. DOI: 10.33293/1609-1442-2020-1(88)-143-157
27. Милкова М.А. Информация и ограниченная рациональность выбора в цифровой экономике // Цифровая экономика. 2021а. № 13(1). С. 69-88. DOI: 10.34706/DE-2021-01-08
28. Милкова М.А., Неволин И.В., Пигорев Д.П. Извлечение ключевой информации из нормативных документов о политике продовольствия и питания в

России // Экономическая наука современной России. 2021б. № 2(93). С. 101-114.
DOI: 10.33293/1609-1442-2021-2(93)-101-114

29. Миловидов, В. Услышать шум волны: что мешает предвидеть инновации? // Форсайт. 2019. № 1(12). С. 88–97.

30. Митрофанова, О.А., Шиморина, А.С., Кольцов, С.Н. Кольцова, О.Ю. Моделирование семантических связей в текстах социальных сетей с помощью алгоритма LDA (на материале русскоязычного сегмента Живого Журнала) // Структурная и прикладная лингвистика Кн. 10. СПб.: Издательство Санкт-Петербургского государственного университета, 2014. С. 151-168.

31. Паниди, К.А. Нобелевская премия по экономике – 2017: вклад Ричарда Талера // Экономический журнал ВШЭ. 2017. 21 (4). С. 702–720.

32. Паринов, С.И., Когаловский, М.В. Технология семантического структурирования контента научных электронных библиотек // Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». 2011. Воронеж, Россия.

33. Паринов, С.И. Тематическое моделирование контекстов цитирований из научных публикаций: структура научного потребления автора // Программная инженерия. 2021. vol. 12, no. 3. С. 140—149.

34. Писарева О.М. Организационно-экономические механизмы стимулирования научно-исследовательской деятельности в вузах: методология и практика: Монография / под научной ред. В.Я. Афанасьева. – М., ГУУ, 2012. – 142 с.

35. Полтерович, В.М. Становление общего социального анализа // Общественные науки и современность. 2011. №2. С. 101-111.

36. Потапенко, А.А. Семантические векторные представления текста на основе вероятностного тематического моделирования. Диссертация на соискание ученой степени канд. ф.-м. наук: 05.13.17 / Потапенко Анна Александровна — М., 2018. — 147 с.

37. Почепцов, Г.Г. На первое место выходит не информационная экономика, а экономика внимания // Regla. 2017(a). 5 (323).

38. Почепцов, Г.Г. Человек за информационной решеткой // Regla. 2017(6). 13 (331).
39. Почепцов, Г. От человека разумного к человеку расслабленному // Regla. 2019. 14 (367).
40. Тоффлер, Э. Третья волна. М.: ООО "Фирма "Издательство АСТ", 2004— 261 с.
41. Фаликман, М. Современная психология внимания: от науки к практике // Лекция в рамках совместного цикла лекций программы «Когнитивные науки и технологии: от нейрона к познанию» департамента психологии НИУ ВШЭ и центра Архэ на базе ВШЭ, 19 апреля 2018 года. URL: <https://www.youtube.com/watch?v=gsOvflwMVRs> (дата посещения: 25.09.2021)
42. Федорова, Е.А., Зарипова, Д.В., Демин, И.С. Факторы формирования цены на предметы искусства с применением текстового анализа новостей в Твиттере // Экономическая наука современной России. 2020. №2. С. 114-131.
43. Янина, А.О., Воронцов, К.В. Мультимодальные тематические модели для разведочного поиска в коллективном блоге // Машинное обучение и анализ данных. 2016. 2(2). С. 173-186.
44. Akerlof, G. A. The Market for Lemons: Quality Uncertainty and the Market Mechanism // The Quarterly Journal of Economics. 1970. vol. 84, no. 3, pp. 488–500.
45. Akerlof, G.A. Behavioral Macroeconomics and Macroeconomic Behavior // The American Economic Review. 2002. vol. 92, no. 3, pp. 411–433.
46. Akerlof, G.A., Shiller, R.J. Animal spirits: How human psychology drives the economy, and why it matters for global capitalism. Princeton university press, 2009 – 264 P.
47. Allais, M. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de lécole américaine // Econometrica: Journal of the Econometric Society. 1953. Vol. 21, No. 4, pp. 503-546 <https://doi.org/10.2307/1907921>
48. Allcott, H., Gentzkow, M. Social Media and Fake News in the 2016 Election // Journal of Economic Perspectives. 2017. 31 (2), pp. 211-236.

49. Allcott, H., Braghieri, L., Eichmeyer, S., Gentzkow, M. The welfare effects of social media. Working Paper // National Bureau of economic research. Cambridge, MA. 2019(a). <https://doi.org/10.3386/w25514>
50. Allcott, H., Gentzkow, M., Yu, C. Trends in the diffusion of misinformation on social media // Research & Politics. 2019(b). 6(2), 205316801984855. <https://doi.org/10.1177/2053168019848554>
51. Amir, O., Rand, D.G. Economic games on the internet: The effect of \$1 stakes // PloS ONE. 2012. 7(2): e31461. <https://doi.org/10.1371/journal.pone.0031461>
52. Apishev, M., Koltcov, S., Koltsova, O., Nikolenko, S., Vorontsov, K. Mining Ethnic Content Online with Additively Regularized Topic Models // Computación y Sistemas. 2016. Vol. 20. No. 3, pp. 387–403.
53. Apishev, M., Koltsov S., Koltsova, O., Nikolenko, S., and Vorontsov, K. Additive Regularization for Topic Modeling in Sociological Studies of User-Generated Texts. In: Sidorov G., Herrera-Alcántara O. (eds) Advances in Computational Intelligence. MICAI 2016. Lecture Notes in Computer Science, vol 10061. Springer, Cham. 2017, pp 169-184
54. Arrow, K. J. The limits of Organization. William W. Norton & Company, 1974 – 86 P.
55. Arrow, K. J. The economics of information. In: M. L. Dertouzos & J. Moses (Eds.), The computer age: A twenty year view. Cambridge, MA MIT Press. 1979, pp. 306-317.
56. Arrow, K. J. The economics of information. Collected papers of Kenneth J. Arrow, vol. 4. Cambridge, MA: The Belknap Press of Harvard University Press, 1984 – 296 P.
57. Asmussen, C.B., Møller, C. Smart literature review: a practical topic modeling approach to exploratory literature review // Journal of Big Data. 2019. 6:93. <https://doi.org/10.1186/s40537-019-0255-7>
58. Babaioff. M., Kleinberg, R. and Leme, R. Paes. Optimal Mechanisms for Selling Information. In: Proceedings of the 13th ACM Conference on Electronic Commerce, EC 12, 2012. pp. 92-109.

59. Baker, S., R., Bloom, N., Davis, S.J. Measuring Economic Policy Uncertainty // *The Quarterly Journal of Economics*. 2016. 131 (4), pp. 1593–1636, <https://doi.org/10.1093/qje/qjw024>
60. Balatsky E.V., Ekimova, N.A., Tretyackova, O.V. Evaluation Methods of Scientific Economic Journals Quality // *Journal of Institutional Studies*. 2021. Vol. 13 (no. 2), pp. 27-52. DOI: 10.17835/2076-6297.2021.13.2.027-052
61. Banerjee, A. A simple model of herd behavior // *Quarterly Journal of Economics*. 1992. 107, pp. 797–818
62. Barberis, N. and Thaler, R.H. A Survey of Behavioral Finance. In: *Handbook of the Economics of Finance, Volume 1B, Financial Markets and Asset Pricing*, ed. George M. Constantinides, Milton Harris and René M. Stulz, pp. 1053–1123. Amsterdam; London and New York: Elsevier, 2003.
63. Benartzi, S., Thaler, R. Behavioral Economics and the Retirement Saving Crisis. // *Science*. 2013. 339(6124), pp. 1152–1153.
64. Benartzi, S., Beshears, J., Milkman, K.L., Sunstein, C.R., Thaler, R.H., Shankar, M., Tucker-Ray, W., Congdon W.J., Galing S. Should Governments Invest More in Nudging? // *Psychological Science*. 2017. 28(8), 1041–1055.
65. Bergemann, D., Bonatti, A., Smolin, A. The Design and Price of Information // *American Economic Review*. 2018. Vol. 108, pp. 1-48.
66. Bikhchandani, S., Hirshleifer, D. and Welch, I. A theory of fads, fashion, custom and cultural change as informational cascades // *Journal of Political Economy*, 1992. Vol. 100, pp. 992–1026.
67. Blei, D., Ng, A., and Jordan, M. Latent Dirichlet allocation // *Journal of Machine Learning Research*. 2003. № 3, pp. 993-1022.
68. Bondt de, W.F.M., Thaler, R. Does the Stock Market Overreact? // *The Journal of Finance*. 1985. Vol. 40, no. 3, pp. 793–805. <https://doi.org/10.2307/2327804>
69. Boulding, K. E. Notes on the Information Concept. *Explorations* [Toronto]. 1955. 6, pp. 103-112.

70. Boyd-Graber, J., Hu, Y., Mimno, D. Applications of Topic Models // Foundations and Trends in Information Retrieval. 2017. Vol. 11, No. 2-3, pp 143-296. <http://dx.doi.org/10.1561/15000000030>
71. Braesemann, F. How behavioural economics relates to psychology – some bibliographic evidence // Journal of Economic Methodology. 2019. Vol. 26, Issue 2, pp. 133-146
72. Brehm, J. W. Postdecision changes in the desirability of alternatives // The Journal of Abnormal and Social Psychology. 1956. 52(3), pp. 384-389.
73. Brighton, H. Robust inference with simple cognitive models. In: C. Lebiere, & B. Wray (Eds.), Between a rock and a hard place: Cognitive science principles meet AI-hard problems: Papers from the AAAI Spring Symposium (pp. 17–22). Menlo Park, CA: AAAI Press, 2006.
74. Broadbent, D.E. Perception and Communication. London: Pergamon, 1958 – 344 p.
75. Brocas, I. Carrillo, J.D. The Brain as a Hierarchical Organization // American Economic Review. 2008. 98 (4), pp. 1312-46. <https://doi.org/10.1257/aer.98.4.1312>
76. Buckmann, M., Şimşek, Ö. Decision Heuristics for Comparison: How Good Are They? // Proceedings of Machine Learning Research. 2016. 58, pp. 1-11.
77. Bulatov, M., Egorov, E., Veselova, E., Polyudova, D., Alekseev, V., Goncharov, A., Vorontsov, K. TopicNet: Making Additive Regularisation for Topic Modelling Accessible // Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pp 6745–6752, Marseille, 11–16 May 2020.
78. Camerer, C. The behavioral challenge to economics: understanding normal people // Conference Series; [Proceedings], Federal Reserve Bank of Boston. 2003, vol. 48(Jun).
79. Camerer, C. F, Loewenstein, G. F. and Prelec, D. Neuroeconomics: How Neuroscience Can Inform Economics. 2004. Available at SSRN: <https://ssrn.com/abstract=590965>

80. Card, D., DellaVigna, S. Nine Facts about Top Journals in Economics // *Journal of Economic Literature*. 2013. 51 (1), pp. 144-61, <https://doi.org/10.1257/jel.51.1.144>
81. Charness, G., Gneezy, U. Incentives to exercise // *Econometrica*. 2009. 77(3), pp. 909-931. <https://doi.org/10.3982/ECTA7416>
82. Chen, G., Kim, K. A., Nofsinger, J. R., & Rui, O. M. Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors // *Journal of Behavioral Decision Making*. 2007. Vol. 20, pp. 425-451.
83. Chirkova, N. A., Vorontsov, K. V. Additive regularization for hierarchical multimodal topic modeling // *Journal Machine Learning and Data Analysis*. 2016. Vol. 2(2), pp. 187-200.
84. Cohn, D., Hofmann, T. The missing link- a probabilistic model of document content and hypertext connectivity. In: *Advances in Neural Information Processing Systems (NIPS)*, Cambridge, MA, MIT Press, 2001.
85. Colosio, M., Shestakova, A., Nikulin, V.V., Blagovechtchenski, E., and Klucharev, V. Neural Mechanisms of Cognitive Dissonance (Revised): An EEG Study // *The Journal of Neuroscience*. 2017. 37(20), pp. 5074–5083.
86. Conley, J.P., Önder, A.S. The Research Productivity of New PhDs in Economics: The Surprisingly High Non-success of the Successful // *Journal of Economic Perspectives*. 2014. 28 (3), pp. 205-16.
87. Czerlinski, J., Gigerenzer, G., Goldstein, D. G. How good are simple heuristics? In: G. Gigerenzer, P. M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press, 1999.
88. Daud, A., Li, J., Zhou, L., and Muhammad, F. Knowledge discovery through directed probabilistic topic models: a survey. In *Proceedings of Frontiers of Computer Science in China*, pp. 280-301. — перевод на русский К. В. Воронцов, А. В. Темлянец и др. 2010.
89. Davenport, T.H., Beck, J.C. *The Attention Economy: Understanding the New Currency of Business*. Harvard Business Press, 2001 – 272 P.

90. DellaVigna, S., Kaplan, E. The Fox News Effect: Media Bias and Voting // The Quarterly Journal of Economics. 2007. Vol. 122, Issue 3, pp 1187–1234, <https://doi.org/10.1162/qjec.122.3.1187>
91. DellaVigna, S., Pollet, J.M. Investor inattention and Friday earnings announcements // The Journal of Finance. 2009. Vol. 64, Issue 2, pp. 709-749. <https://doi.org/10.1111/j.1540-6261.2009.01447.x>
92. DellaVigna, S. Psychology and Economics: Evidence from the Field // Journal of Economic Literature, 2009. 47(2), pp. 315–372. <https://doi.org/10.1257/jel.47.2.315>
93. Ellsberg, D. Risk, ambiguity, and the savage axioms // Quarterly Journal of Economics. 1961. 75(4), pp. 643–669. <https://doi.org/10.2307/1884324>
94. Elster, J. Social Norms and Economic Theory // The Journal of Economic Perspectives. 1989. Vol.3, no.4, pp. 89–117.
95. Etter, M., Colleoni, E., Illia, L., Meggiorin, K., dEugenio, A. Measuring Organizational Legitimacy in Social Media: Assessing Citizens Judgments With Sentiment Analysis // Business & Society. 2018. Vol. 57, pp. 60 - 97.
96. Evans, J. S. B. T. Heuristic and analytic processes in reasoning // British Journal of Psychology. 1984. Vol. 75(4), pp. 451-468.
97. Falkinger, J. Limited attention as a scarce resource in information-rich economies. The Economic Journal. 2008. 118(532), pp. 1596-1620. <https://doi.org/10.1111/j.1468-0297.2008.02182.x>
98. Festinger, L. A theory of cognitive dissonance. Stanford, CA: Stanford University Press, 1957 – 291 P.
99. Festinger, L, Carlsmith, J.M. Cognitive Consequences of Forced Compliance // Journal of Abnormal and Social Psychology. 1959. Vol. 58, pp. 203-210.
100. Festré, A., Garrouste, P. The Economics of Attention: A History of Economic Thought Perspective // Oeconomia. 2015. Vol. 5-1, pp. 3-36. <https://doi.org/10.4000/oeconomia.1139>
101. Franck, G. Ökonomie der Aufmerksamkeit // Merkur. 1993. 47(9/10), pp. 748-761.

102. Franck, G. Mental Capitalism. In: Shamiyeh M., DOM Research Laboratory (eds) What People Want, pp. 99-114. Birkhäuser Basel, 2005.
103. Friebel, G., Heinz, M., Krueger, M., Zubanov, N. Team incentives and performance: Evidence from a retail chain // *American Economic Review*. 2017. Vol. 107(8), pp. 2168-2203. <https://doi.org/10.1257/aer.20160788>
104. Gallese, V., Goldman, A. Mirror neurons and the simulation theory of mind-reading // *Trends in Cognitive Science*. 1998. Vol.2, Issue 12, pp. 493-501 [https://doi.org/10.1016/S1364-6613\(98\)01262-5](https://doi.org/10.1016/S1364-6613(98)01262-5)
105. Gentzkow, M., Kamenica, E. Competition in Persuasion // *Review of Economic Studies*. 2017. Vol. 84, pp. 300–322. <https://doi.org/10.1093/restud/rdw052>
106. Gigerenzer, G. Why Heuristics Work // *Perspectives of psychological science*. 2008. Vol. 3(1), pp. 20-29.
107. Gigerenzer, G. Todd, P.M. Ecological rationality: the normative study of heuristics. In Gigerenzer, Gerd; Todd, Peter M.; The ABC Research Group (eds.). *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press. 1999, pp. 487–497
108. Gigerenzer, R. *Selten Bounded Rationality the Adaptive Toolbox*. Cambridge, MIT Press, 2001 – 394 P.
109. Gigerenzer, G., Gaissmaier, W. Heuristic decision making // *Annual Review of Psychology*. 2011. Vol. 62, pp. 451–482.
110. Glimcher, P. W., Rustichini, A. Neuroeconomics: the consilience of brain and decision // *Science*. 2004. Vol. 306, pp. 447–452.
111. Glimcher, P.W. Camerer, C.F. Fehr, E. Poldrack, R.A. Neuroeconomics: Decision Making and the Brain. Elsevier, 2009 – 560 P. <https://doi.org/10.1016/C2011-0-05512-6>
112. Goldhaber, M.H. The Attention Economy and the Net. *First Monday*. 1997(a). Vol. 2 (4). <https://doi.org/10.5210/fm.v2i4.519>
113. Goldhaber, M.H. Attention Shoppers! 1997(b). URL: <https://www.wired.com/1997/12/es-attention/> <https://journals.uic.edu/ojs/index.php/fm/article/view/519>

114. Gorshkov S., Ilyushin E., Chernysheva A., Goiko V., Namiot D. Using topic modeling for communities clusterization in the vkontakte social network // *International Journal of Open Information Technologies*. 2021. Vol.9 №5, pp. 12-17.
115. Grant, C.E., George C.P., Kanjilal, V., Nirkhiwale, S., Wilson, J.N., Wang, D.Z. A topic-based search, visualization, and exploration system. FLAIRS Conference. AAAI Press, 2015, pp. 43–48.
116. Gronchi, G., Giovannelli, F. Dual Process Theory of Thought and Default Mode Network: A Possible Neural Foundation of Fast Thinking // *Frontiers in Psychology*. 2018. Vol. 9, 1237.
117. Gruber, A., Rosen-Zvi, M., and Weiss, Y. Latent topic models for hypertext. In: *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, Helsinki, Finland, 2008.
118. Güth W., Schmittberger R., Schwarze B. An experimental analysis of ultimatum bargaining // *Journal of Economic Behavior & Organization*. 1982. Vol. 3, No. 4, pp. 367—388.
119. Harsanyi, J.C. Games with incomplete information played by Bayesian players // *Management Science*. 1967. Vol. 4 No3, pp. 159-182.
120. Hayek, F. A. *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. Chicago, IL: The University of Chicago Press, 1952.
121. Hayek, F. A. *The Constitution of Liberty*. Chicago, IL: The University of Chicago Press, 1960. <https://doi.org/10.7208/chicago/9780226320519.001.0001>
122. Heiner, R.A. The Origin of Predictable Behavior // *The American Economic Review*. 1983. Vol. 73, no. 4, pp. 560–595.
123. Helbing, D., Frey, B.S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., Hoven, J., Zicari, R.V., Zwitter, A. Will Democracy Survive Big Data and Artificial Intelligence? In: Dirk Helbing eds. *Towards digital enlightenment: Essays on the dark and light sides of the digital revolution*. Springer. 2019.
124. Hershfield, H. Shu, S., Benartzi, S. Temporal Reframing and Participation in a Savings Program: A Field Experiment. 2019. Available at SSRN: <https://ssrn.com/abstract=3097468> or <http://dx.doi.org/10.2139/ssrn.3097468>

125. Hirshleifer, D. Teoh, S.H. Limited attention, information disclosure, and financial reporting // *Journal of Accounting and Economics*. 2003. Vol. 36(1-3), pp. 337-386.
126. Hirshleifer, J., Riley, J. *The Analytics of Uncertainty and Information* (Cambridge Surveys of Economic Literature). Cambridge: Cambridge University Press, 1992. <https://doi.org/10.1017/CBO9781139167635>
127. Hofmann, T. Probabilistic Latent Semantic Analysis // *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Stockholm, 1999, pp. 289–296.
128. Hotelling, H. The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica*. 1938. Vo. 6(3), pp. 242-269. DOI : 10.2307/1907054
129. James, W. *The Principles of Psychology*. New York, NY: Henry Holt and Company, 1890.
130. Johnson, E. J., Goldstein, D. G. Do defaults save lives? // *Science*. 2003. Vol. 302, pp. 1338–1339.
131. Jost, P.J., Pünder, J., Schulze-Lohoff, I. Fake news - Does perception matter more than the truth? // *Journal of Behavioral and Experimental Economics*. 2020. 85, 101513, <https://doi.org/10.1016/j.socec.2020.101513>
132. Ianina, A. Vorontsov, K. Hierarchical Interpretable Topical Embeddings for Exploratory Search and Real-Time Document Tracking // *International Journal of Embedded and Real-Time Communication Systems*. 2020. 11(4), pp. 134-152. <https://doi.org/0.4018/IJERTCS.2020100107>
133. Ignatow, G., Mihalcea, R. *Text Mining: A Guidebook for the Social Sciences*. SAGE Publications, Inc, 2017. <https://www.doi.org/10.4135/9781483399782>
134. Izuma, K., Matsumoto, M., Murayama, K., Samejim, K., Sadato, N., and Matsumoto, K. Neural correlates of cognitive dissonance and choice-induced preference change // *Proceedings of the National Academy of Sciences*. 2010. 107(51), pp. 22014-22019.

135. Kagel, J.H., Roth, A.E. The Handbook of Experimental Economics. Princeton University Press, 1995. <https://doi.org/10.2307/j.ctvzsmff5>
136. Kahneman, D. Attention and effort. Englewood Cliffs, NJ: Prentice-Hall, 1973 – 246 P.
137. Kahneman, D, Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979. Vol. 47(2), pp. 263-292.
138. Kahneman D., Knetsch J.L., Thaler R.H. Fairness and the assumptions of economics // *Journal of Business*. 1986. Vol. 59, No. 4, Part 2, pp. S285—S300.
139. Kahneman, D. A perspective on judgment and choice: Mapping bounded rationality // *American Psychologist*. 2003. Vol. 58(9), pp. 697–720.
140. Kim, W.G., Park, S.A. Social media review rating versus traditional customer satisfaction: Which one has more incremental predictive power in explaining hotel performance? // *International Journal of Contemporary Hospitality Management*. 2017. Vol. 29 No. 2, pp. 784-802. <https://doi.org/10.1108/IJCHM-11-2015-0627>
141. Klucharev, V., Hytonen, K., Rijpkema, M., Smidts, A. and Fernandez, G. Reinforcement learning signal predicts social conformity // *Neuron*. 2009. Vol. 61, pp. 140–151.
142. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E. Diminishing reciprocal fairness by disrupting the right prefrontal cortex // *Science*. 2006. Vol. 314, pp. 829–832.
143. Köszegi, B., Rabin, M. A Model of Reference-Dependent Preferences // *The Quarterly Journal of Economics*. 2006. Vol. 121, no. 4, pp. 1133–1165.
144. Krishna, A., Aich, A., Akhilesh, V., Hegde, C. Analysis of Customer Opinion Using Machine Learning and NLP Techniques // *International Journal of Advanced Studies of Scientific Research*. 2018. Vol. 3(9). Available at SSRN: <https://ssrn.com/abstract=3315430>
145. Lanham, R.A. The economics of attention. Style and substance in the age of information. — Chicago, 2006 — 362 p.
146. Lavie, N. Distracted and confused?: Selective attention under load // *Trends in Cognitive Sciences*. 2005. 9(2), pp. 75-82.

147. Lee, T.Y., Smith, A., Seppi, K. Elmqvist, N., Boyd-Graber, J., Findlater, L. The human touch: How non-expert users perceive, interpret, and fix topic models // *International Journal of Human-Computer Studies*. 2017. Vol. 105, pp. 28—42.
148. Levi, Y., Benartzi, S. Mind the App: Mobile Access to Financial Information and Consumer Behavior. 2020. Available at SSRN: <https://ssrn.com/abstract=3557689> or <http://dx.doi.org/10.2139/ssrn.3557689>
149. Lorenz-Spreen, P., Mønsted, B.M., Hövel, P., Lehmann, S. Accelerating dynamics of collective attention // *Nature Communications*. 2019. 10, 1759. <https://doi.org/10.1038/s41467-019-09311-w>
150. Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act // *Brain*. 1983. 3(106), pp. 623–642. DOI: 10.1093/brain/106.3.623
151. Lim, K.W., Buntine, W. Bibliographic Analysis with the Citation Network Topic Model // *Proceedings of the Sixth Asian Conference on Machine Learning*, PMLR 39. 2015, pp. 142-158
152. Loewenstein, G., Prelec, D. Anomalies in Intertemporal Choice: Evidence and an Interpretation // *The Quarterly Journal of Economics*. 1992. Vol. 107, no. 2, pp. 573-597
153. Loewenstein, G., O'Donoghue, Y., Rabin, M. Projection Bias in Predicting Future Utility // *The Quarterly Journal of Economics*. 2003. Volume 118, Issue 4, pp. 1209–1248. <https://doi.org/10.1162/003355303322552784>
154. Maccheroni, F., Marinacci, M., & Rustichini, A. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences // *Econometrica*. 2006. Vol. 74(6), pp. 1447-1498
155. Machlup, F. The production and distribution of knowledge in the United States. Princeton, N J: Princeton University Press, 1962.
156. Machlup, F. Knowledge: Its creation, distribution, and economic significance (vol. 1, Knowledge and knowledge production). Princeton, NJ: Princeton University Press, 1980.

157. Machlup, F. Knowledge, its creation, distribution, and economic significance (vol. 2, The branches of learning). Princeton, NJ: Princeton University Press, 1982.
158. Machlup, F. Knowledge, its creation, distribution, and economic significance (vol. 3, The economics of information and human capital). Princeton, NJ: Princeton University Press, 1984.
159. Maguire, E.A., Woollett, K., Spiers, H.J. London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis // *Hippocampus*. 2006. Vol. 16(12), pp. 1091-1101.
160. Marchionini, G. Exploratory search: From finding to understanding // *Communications of the ACM*. 2006. Vol 49(4), pp. 41–46. <https://doi.org/10.1145/1121949.1121979>
161. Martin, S., Marks, J. Messengers: Who We Listen to, Who We Dont and Why. In: A. Samson (Ed.) *The Behavioral Economics Guide*, 2019 (with an Introduction by Uri Gneezy), pp. 105-113. Retrieved from [https://www. behavioraleconomics.com](https://www.behavioraleconomics.com)
162. Marianne, B. Mullainathan, S. Whats in a Name for Black Job Seekers? // NBER Working Paper. 2004. 9873
163. Milgrom, P. R., R. J. Weber A theory of auctions and competitive bidding // *Econometrica*. 1982. Vol. 50(5), pp. 1089–1122.
164. Milkova, M., Andreichikova, O., Andreichikov, A. Decision making under uncertainty: a heuristics overview and the analytic network process. // *Psychology Journal of the Higher School of Economics*. 2019(a). Vol. 16. N 4, pp. 730–751. DOI: 10.17323/1813-8918-2019-4-730-751
165. Milkova, M., Andreichikova, O., Andreichikov, A. At the junction of mathematics and psychology: cognitive orientation of the AHP/ANP and new perspectives of structuring complexity // *International Journal of the Analytic Hierarchy Process*. 2019(b). Vol. 11(1). DOI: 10.13033/ijahp.v11i1.611
166. Mousavi, S., Gigerenzer, G. Risk, uncertainty, and heuristics // *Journal of Business Research*. 2014. Vol. 67, pp. 1671–1678.
167. Neisser, U. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. W.H. Freeman, 1976 – 230 p.

168. Nelson, P. Advertising as Information // *Journal of Political Economy*. 1974. Vol. 82, no. 4, pp. 729–754. www.jstor.org/stable/1837143
169. Ngo, S-C., Percus, A.G., Burghardt, K., Lerman, K. The transortative structure of networks. // *Proceedings of the Royal Society A*, 2020, 476(2237) <https://doi.org/10.1098/rspa.2019.0772>
170. Ohlson, J. A. The social value of public information in production economies. In G. A. Feltham, A. H. Amershi, & W. T Ziemba (Eds.), *Economic analysis of information and contracts: Essays in honor of John E. Butterworth* (pp. 95-119). Boston, MA: Kluwer Academic Publishers, 1988.
171. Petty, R.E., Cacioppo, J.T. The Elaboration Likelihood Model of Persuasion // *Advances in Experimental Social Psychology*. 1986. Vol. 19, pp. 123-205.
172. Piccione, M., Rubinstein, A. On The Interpretation of Decision Problems with Imperfect Recall // *Games and Economic Behavior*. 1997. Vol. 20, pp. 3-24. <https://doi.org/10.1006/game.1997.0536>
173. Plott, C., Smith, V. *Handbook of Experimental Economics Results, Volume 1*. Amsterdam: North-Holland Publications, 2008 – 1184 P.
174. Potapenko, A. A., Vorontsov, K. V. Robust PLSA Performs Better Than LDA. 35th European Conference on Information Retrieval, ECIR-2013, Moscow, Russia, 24-27 March 2013. — *Lecture Notes in Computer Science (LNCS)*, Springer Verlag-Germany, pp. 784–787.
175. Porat, M.U. *The information economy: definition and management*. — Office of Telecommunications (DOC), Washington, DC, 1977 — 319 P.
176. Rabin, M. Psychology and Economics // *Journal of Economic Literature*. 1998. Vol. 36, issue 1, pp. 11-46.
177. Rabin, M. Incorporating Limited Rationality into Economics // *Journal of Economic Literature*,. 2013. Vol. 51 (2), pp. 528-43.
178. Reis, R. Inattentive consumers // *Journal of Monetary Economics*. 2006(a). Vol. 53(8), pp. 1761-1800. <https://doi.org/10.3386/w10883>
179. Reis, R. Inattentive producers // *The Review of Economic Studies*. 2006(b). Vol. 73(3), pp. 793-821. <https://doi.org/10.1111/j.1467-937X.2006.00396.x>

180. Roda, C. A Roadmap of Studies in Attention and Digital Technology. In: W. Doyle, C. Roda (eds). *Communication in the Era of Attention Scarcity*. Palgrave Macmillan, 2019, pp. 7-20. <https://doi.org/10.1007/978-3-030-20918-6>
181. Rothschild, M. Searching for the lowest price when the distribution of prices is unknown // *Journal of Political Economy*. 1974. Vol. 82(4), pp. 689-711.
182. Rosen-Zvi M., Griffiths T., Steyvers M., Smyth P. The author-topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty in artificial intelligence*. UAI '04 - Arlington, Virginia, United States: AUAI Press, 2004, pp. 487-494.
183. Sætra, H.S. When nudge comes to shove: Liberty and nudging in the era of big data // *Technology in Society*, 2019. Vol. 59, 101130. <https://doi.org/10.1016/j.tech-soc.2019.04.006>
184. Sanfey, A. G., Chang, L. J. Multiple systems in decision making // *Annals of the New York Academy of Sciences*. 2008. Vol. 1128, pp. 53–62.
185. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., Cohen J. D. The neural basis of economic decision making in the Ultimatum Game // *Science*. 2003. Vol. 300, pp. 1755–1758.
186. Selten, R. The chain store paradox // *Theory and Decision*. 1978. Vol. 9, pp. 127–159 <https://doi.org/10.1007/BF00131770>
187. Selten, R., R. Stoecker End Behaviour in Sequences of Finite Prisoners Dilemma Supergames // *Journal of Economic Behaviour and Organization*. 1986. Vol. 7, pp. 47-70.
188. Simon, H. A. *Administrative Behavior: A Study of Decision-making Processes in Administrative Organizations*. Chicago, IL: Macmillan, 1947.
189. Simon, H. A. Rational choice and the structure of the environment // *Psychological Review*. 1956. Vol. 63(2), pp. 129–138.
190. Simon, H. A. *The Science of the Artificial (Karl Taylor Compton Lectures)*. Cambridge, MA: MIT Press, 1969.

191. Simon, H. A. Designing organizations for an information-rich world. In Martin Greenberger (ed.), *Computers, Communication, and the Public Interest*. Baltimore, MD: The Johns Hopkins Press, pp. 37-72. 1971.
192. Simon, H. A. Rational decision-making in business organizations. Nobel Memorial Lecture. *Economic Sciences*, 1978, pp. 343-371.
193. Sims, C.A. Implications of rational inattention // *Journal of Monetary Economics*. 2003. Volume 50, Issue 3, pp. 665-690 [https://doi.org/10.1016/S0304-3932\(03\)00029-1](https://doi.org/10.1016/S0304-3932(03)00029-1).
194. Sims, C.A. Rational Inattention and Monetary Economics. In: Benjamin M. Friedman, Michael Woodford (Eds). *Handbook of Monetary Economics*, Elsevier, Volume 3, 2010, pp. 155-181. <https://doi.org/10.1016/B978-0-444-53238-1.00004-1>.
195. Shafir, E. Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition*. 1993. Vol. 21, pp. 546–556.
196. Shannon, Claude E. A mathematical theory of communication // *Bell System Technical Journal*. 1948. Vol. 27, pp. 379-423 and 623-656.
197. Shefrin, H., Statman, M. The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence // *The Journal of Finance*. 1985. Vol. 40, no. 3, pp. 777–790 <https://doi.org/10.2307/2327802>
198. Shiller, R.J. Human Behavior and the Efficiency of the Financial System // NBER Working Paper. 1998. No 6375. <https://doi.org/10.3386/W6375>
199. Shapiro, C., Varian, H.R. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press, 1999.
200. Scherer, M., T. von Landesberger, Schreck, T. Topic modeling for search and exploration in multivariate research data repositories. *Research and Advanced Technology for Digital Libraries: Conference (International) on Theory and Practice of Digital Libraries Proceedings*. Eds. T. Aalberg, C. Papatheodorou, M. Dobрева, G. Tsakonas, and C. J. Farrugia. Berlin–Heidelberg: Springer, 2013, pp. 370–373.
201. Shleifer, A. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press, 2000 – 224 P. <https://doi.org/10.1093/0198292279.001.0001>

202. Smith, V. L. An Experimental Study of Competitive Market Behavior // *Journal of Political Economy*. 1962. Vol.70, Issue 2, pp. 111-137.
203. Smith, V. L. Constructivist and Ecological Rationality in Economics // *The American Economic Review*. 2003. Vol. 93(3), pp. 465-508.
204. Smolin, A. Disclosure and Pricing of Attributes // Available at SSRN. 2019: <https://ssrn.com/abstract=3318957> or <http://dx.doi.org/10.2139/ssrn.3318957>
205. Soon, C.S., Brass, M., Heinze, H-J., Haynes, J-D. Unconscious determinants of free decisions in the human brain // *Nature Neuroscience*. 2008. Vol. 11, pp. 543–545.
206. Stevens, K., Kegelmeyer, P., Andrzejewski, D., Buttler, D. Exploring topic coherence over many models and many topics // *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning / Association for Computational Linguistics*, 2012, pp. 952–961.
207. Stigler, G.J. The Economics of Information // *Journal of Political Economy*. 1961. Vol. 69, No. 3, pp. 213-225
208. Stiglitz, J. E. Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market // *Review of Economic Studies*. 1977. Vol. 44(3), pp. 407-430
209. Sveiby, K-E. (1994). What is information? Retrieved June 4 (1998): 2014.
210. Sunstein, C.R. Probability Neglect: Emotions, Worst Cases, and Law // *The Yale Law Journal*. 2001. Vol.112(1), pp. 61-107
211. Tang, J., Zhang, J., Yao, L., L.i, J., Zhang, L., and Su, Z.. ArnetMiner: extraction and mining of academic social networks. In: *Proceedings of ACM SIGKDD*, 2008.
212. Thaler, R. H. Toward a positive theory of consumer choice // *Journal of Economic Behavior and Organization*. 1980. Vol. 1, pp. 39–60.
213. Thaler, R. Some empirical evidence on dynamic inconsistency // *Economics Letters*. 1981. Volume 8, Issue 3, pp. 201-207. [https://doi.org/10.1016/0165-1765\(81\)90067-7](https://doi.org/10.1016/0165-1765(81)90067-7).
214. Thaler, R. H., Sunstein, C. R. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press, 2008.

215. Tirole J. A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality // *The Review of Economic Studies*. 1996. Vol. 63, No. 1 <https://doi.org/10.2307/2298112>
216. Treisman, Anne M. Contextual cues in selective listening // *Quarterly Journal of Experimental Psychology*. 1960. Vol. 12(4), pp. 242-248.
217. Trichet, Jean-Claude. Opening address at the ECB Central Banking Conference. 2010. Frankfurt, 18 November 2010.
218. Tversky, A., Kahneman, D. Belief in the law of small numbers. *Psychological Bulletin*. 1971. Vol 76, pp. 105–110.
219. Tversky, A., Kahneman, D. Judgment under uncertainty: Heuristics and biases // *Science*. 1974. Vol 185(4157), pp. 1124–1131.
220. Crawford, V.P., Sobel, J. Strategic Information Transmission // *Econometrica*, 1982. Vol. 50, No. 6, pp. 1431-1451 <https://doi.org/10.2307/1913390>
221. Vorontsov K. V., Potapenko A. A. Additive regularization of topic models // *Machine Learning*. 2014(a). Vol. 101, pp. 303-323.
222. Vorontsov, K.V., Potapenko A. A. Tutorial on Probabilistic Topic Modeling: Additive Regularization for Stochastic Matrix Factorization // *Analysis of Images, Social Networks and Texts. Third International Conference, AIST*. 2014(b). Vol. 436, pp. 29-46
223. Vorontsov K., Frei O., Apishev M., Romov P., Dudarenko M. BigARTM: Open Source Library for Regularized Multimodal Topic Modeling of Large Collections. In: Khachay M., Konstantinova N., Panchenko A., Ignatov D., Labunets V. (eds) *Analysis of Images, Social Networks and Texts. AIST 2015*, pp. 370-381. *Communications in Computer and Information Science*, vol 542. Springer, Cham. https://doi.org/10.1007/978-3-319-26123-2_36
224. Yanagizawa-Drott, D. Propaganda and Conflict: Evidence from the Rwandan Genocide // *The Quarterly Journal of Economics*. 2014. Vol. 129 (4), pp. 1947–1994, <https://doi.org/10.1093/qje/qju020>
225. White R.W., Roth R.A. Exploratory search: Beyond the query-response paradigm. In: G. Marchionini (ed) *Synthesis Lectures on Information Concepts Retrieval*

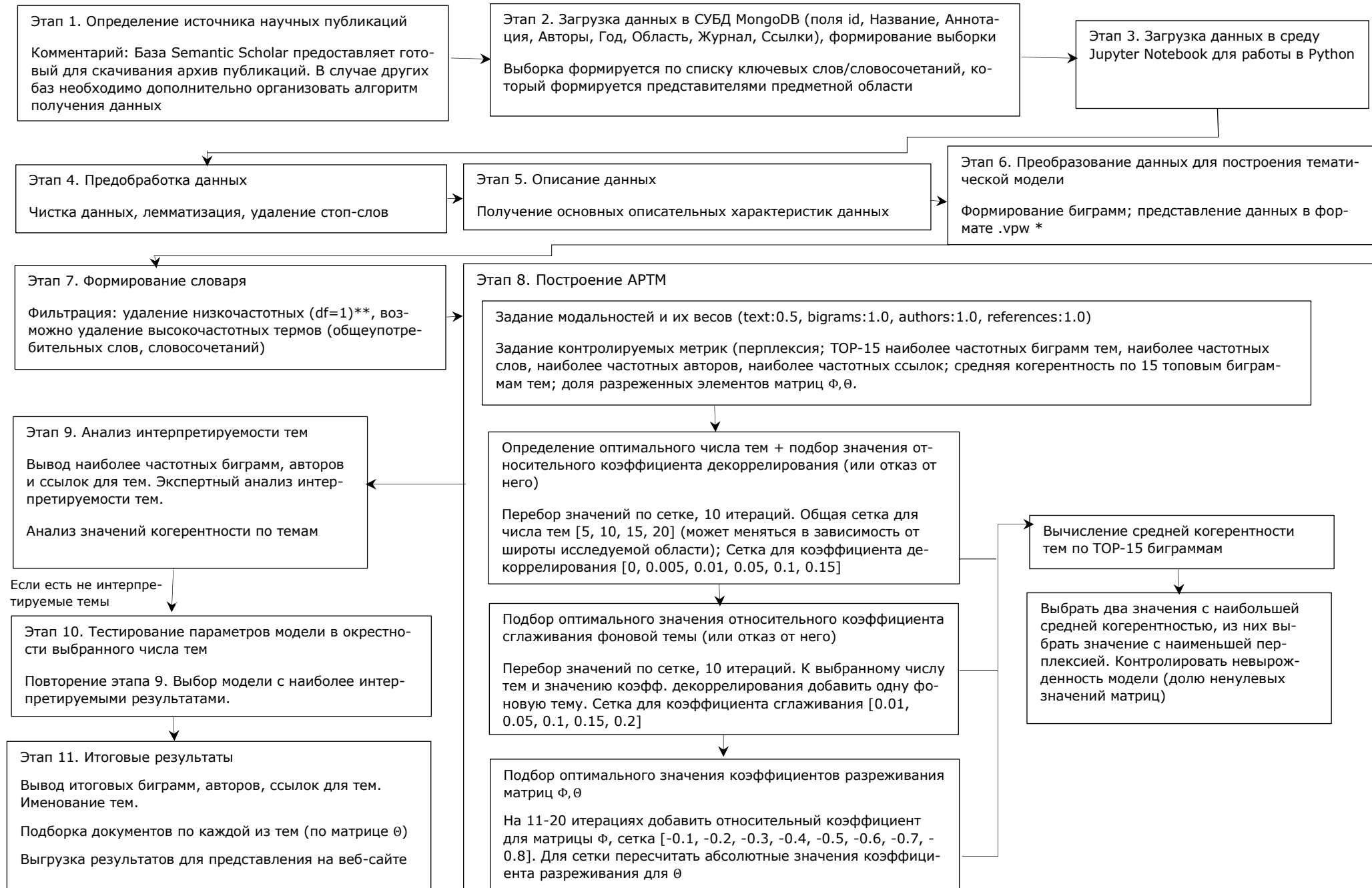
and Services 1(1). Morgan & Claypool Publishers, 2009 – 98 P.
<https://doi.org/10.2200/S00174ED1V01Y200901ICR003>

226. Williamson, O.E. Behavioral Assumptions. In: O.E. Williamson. The Economic Institutions of Capitalism. Firms, Markets, Relational Contracting. N.Y.: The Free Press, 1985, pp.44–52

227. Wu, T. Attention Brokers. 2015. Retrieved from NYU Law
http://www.law.nyu.edu/sites/default/files/upload_documents/Tim%20Wu%20-%20Attention%20Brokers.pdf (дата посещения: 25.09.2021)

228. Wu, T. Blind spot: the attention economy and the law. Antitrust Law Journal. 2017. Available at SSRN: <https://ssrn.com/abstract=2941094>

Приложение 1. Методика построения тематической модели для коллекции научных публикаций



Примечание. * Шаблон представления данных в формате .vprw: Id |Modality1 token11: count12 token12: count12 ... |ModalityN tokenN1: countN1 tokenN2:countN2 ...

** df – document frequency, частота встречаемости термина в документах

Приложение 2. Отбор данных для построения тематической модели

Пример запроса к базе MongoDB:

```
db.Ec_Bu.aggregate(  
  [  
    { $match: { $text: { $search: "\"behavioral economics\"" } } },  
    { $merge : { into: { coll: "Ec_Bu__be__t" }, on: "_id", whenMatched: "keepExisting", whenNotMatched: "insert" } }  
  ],  
  { allowDiskUse: true }  
)
```

Список стоп-слов, добавляемых дополнительно к стандартному списку стоп-слов библиотеки nltk:

[et_al, e_g, de_los, de_la, case_study, p_p, main_result, study_provides, research_question, better_understanding, per_cent, per_caput, de_l, en_el, en_la]

Приложение 3. Результаты тестирования тематической модели

Таблица 1. Результаты экспериментов по выбору оптимального числа тем. Итерации 1-10.

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность
128	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-03	4.63e+14	0.278	0.0002	153455.5938	0.5929	0.7999	-1.3712
129	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-02	4.68e+14	0.2774	0.0003	153209.4063	0.5939	0.8042	-2.1089
130	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-02	4.74e+14	0.2961	0.0014	151833.7969	0.6004	0.8301	-3.6454
131	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-01	5.3e+14	0.3341	0.0018	149058.2031	0.6139	0.867	-5.7978
132	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.50e-01	5.89e+14	0.4012	0.0028	144331.7969	0.637	0.9011	-6.2368

13 3	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-03	1.19e+14	0.514	0.0003	81446.8984	0.526	0.6527	-3.0701
13 4	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-02	1.21e+14	0.5135	0.0003	81352.5	0.5267	0.6591	-3.7592
13 5	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-02	1.22e+14	0.5342	0.0015	80979.5	0.5308	0.6973	-6.0784
13 6	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-01	1.35e+14	0.5696	0.0018	80511.7969	0.5372	0.7518	-8.3848
13 7	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.50e-01	1.58e+14	0.6127	0.0025	79610.6016	0.5476	0.7971	- 9.14299999999999 9
13 8	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-03	5.27e+13	0.6435	0.0004	55016.332	0.4998	0.5729	-2.8398

13 9	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-02	5.34e+13	0.6436	0.0005	54901.2656	0.5007	0.5789	-3.6186
14 0	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-02	5.4e+13	0.6637	0.002	54891.7344	0.5024	0.6173	-8.4406
14 1	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-01	6e+13	0.6865	0.0019	54802.0664	0.5063	0.6748	-9.5436
14 2	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.50e-01	7.36e+13	0.7159	0.0026	54547.8008	0.5129	0.7254	-11.683
14 3	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-03	2.79e+13	0.7244	0.0004	41286.5508	0.4876	0.5264	-3.8742
14 4	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-02	2.83e+13	0.7249	0.0005	41204.6992	0.4884	0.5335	-4.8996

14 5	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 5.00e-02	2.89e+13	0.7432	0.0022	41200.3984	0.4901	0.5694	-9.8166
14 6	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.00e-01	3.21e+13	0.7589	0.0027	41254.25	0.4924	0.6298	-10.2086
14 7	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 1.50e-01	4.03e+13	0.7792	0.0036	41166.25	0.4971	0.6801	-11.6648
14 8	5	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 0.00e+00	4.62e+14	0.2771	0.0002	153559.0	0.5924	0.7968	-1.6138
14 9	10	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 0.00e+00	1.19e+14	0.513	0.0003	81471.7031	0.5257	0.6489	-2.5434
15 0	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 0.00e+00	5.26e+13	0.6423	0.0004	55025.9336	0.4996	0.5694	-2.4057

15 1	20	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[DecorrPhi]; ; 0.00e+00	2.78e+13	0.7231	0.0003	41288.0508	0.4874	0.5226	-3.3932
---------	----	--	----	----------------------------	----------	--------	--------	------------	--------	--------	---------

Таблица 2. Результаты экспериментов по выбору значения относительного коэффициента сглаживания фоновой темы. Итерации 1-10.

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность
15 2	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[SmoothPhi_back] : ; 1.00e-02	4.59e+13	0.6223	0.0003	51356.375	0.4972	0.5558	-3.2510
15 3	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[SmoothPhi_back] : ; 5.00e-02	4.59e+13	0.6228	0.0003	50991.25	0.4976	0.5512	-3.0265
15 4	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[SmoothPhi_back] : ; 1.00e-01	4.59e+13	0.6228	0.0003	50789.625	0.4973	0.5476	-3.0555
15 5	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[SmoothPhi_back] : ; 1.50e-01	4.59e+13	0.6227	0.0003	50658.25	0.4969	0.545	-3.0276

15 6	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[SmoothPhi_back] : ; 2.00e-01	4.59e+13	0.6223	0.0003	50571.812 5	0.4964	0.5427	-3.0751
---------	----	--	----	----------------------------------	----------	--------	--------	----------------	--------	--------	---------

Таблица 3. Результаты экспериментов по выбору значений коэффициентов разреживания для матриц Φ и Θ . Итерации 11-20.

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность Θ	Размер ядра	Контрастность	Чистота	Когерентность
158	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 1.00e-01; - 2.10e-01	5.12e+13	0.8165	0.7875	54221.0664	0.5221	0.6154	-2.1779
159	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 2.00e-01; - 4.70e-01	5.25e+13	0.8198	0.7863	54050.5352	0.5258	0.6194	-2.1998
160	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 3.00e-01; - 8.00e-01	5.24e+13	0.8218	0.825	53943.5352	0.5282	0.6193	-2.2609
161	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 4.00e-01; 1.25e+00	5.17e+13	0.8227	0.8473	53914.7344	0.5291	0.6161	-2.1628
162	15	{u'text': 0.5, u'bigrams': 1.0,	20	[SparsePhi, SparseTheta]: -	5.11e+13	0.8228	0.8648	53918.0664	0.529	0.6093	-2.0612

		u'references': 1.0, u'authors': 1.0}		5.00e-01; - 1.87e+00							
167	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: -6.00e-01; ; - 2.81e+00	5.07e+13	0.8218	0.8823	53983.4648	0.5276	0.5989	-1.9052
168	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 7.00e-01; ; - 4.37e+00	5.06e+13	0.8194	0.8983	54134.2656	0.5242	0.5836	-2.0455
169	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 8.00e-01; ; - 7.50e+00	5.12e+13	0.8151	0.9137	54353.332	0.5186	0.5599	-1.9709

Таблица 4. Результаты тестирования стратегии для числа тем в окрестности 15: 13, 14, 16, 17 тем

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность
17 4	16	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 6.00e-01; ; - 2.81e+00	4.43e+13	0.8286	0.8913	50773.75	0.5215	0.5835	-2.5001
17 5	17	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 6.00e-01; ; - 2.81e+00	3.88e+13	0.8346	0.8997	48000.293	0.5149	0.5669	-2.6296
17 6	14	{u'text': 0.5, u'bigrams': 1.0,	20	[SparsePhi, SparseTheta]: -	5.88e+13	0.8131	0.8731	57642.285 2	0.5329	0.6087	-2.3225

		u'references': 1.0, u'authors': 1.0}		6.00e-01; ; - 2.81e+00							
17 7	13	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 6.00e-01; ; - 2.81e+00	6.69e+13	0.805	0.8622	61662.539 1	0.5417	0.629	-1.8164

Таблица 5. Примеры документов по темам

d1. Ограниченная рациональность, теория перспектив

1. Moshe_Levy, Haim_Levy (2002). Prospect Theory: Much Ado About Nothing?
2. Ralph_Hertwig, Anastasia_Kozyreva (2020). Bounded rationality: A vision of rational choice in the real world
3. I_Myuller (1974). Employment of Heuristics in Forecasting
4. Robert_Libby, Peter_C._Fishburn (1977). Behavioral Models of Risk Taking in Business Decisions: A Survey and Evaluation
5. David_Peel (2013). Subjective skewness of return as an explanation of the optimal choice between gambles in cumulative prospect theory
6. Nicola_Lacetera, Devin_G_Pope, Justin_R._Sydnor (2010). Heuristic Thinking and Limited Attention in the Car Market
7. Hanqing_Jin, Xun_Yu_Zhou (2011). Greed, Leverage, and Potential Losses: A Prospect Theory Perspective
8. Enrico_G._De_Giorgi, Thorsten_Hens (2006). Making prospect theory fit for finance
9. Krzysztof_Kontek (2009). Are People Really Risk Seeking for Losses?
10. Krzysztof_Kontek (2010). Multi-Outcome Lotteries: Prospect Theory vs. Relative Utility

d2. Про-экологическое поведение

1. Philippa_Howden-Chapman, Michael_Keall, Kate_Whitwell, Ralph_Chapman (2019). Evaluating natural experiments to measure the co-benefits of urban policy interventions to reduce carbon emissions in New Zealand.
2. JongRoul_Woo, Sesil_Lim, Yong-Gil_Lee, Sung-Yoon_Huh (2018). Financial Feasibility and Social Acceptance for Reducing Nuclear Power Plants: A Contingent Valuation Study
3. Meir_Russ (2016). The probable foundations of sustainabilism: Information, energy and entropy based definition of capital, Homo Sustainabiliticus and the need for a "new gold"
4. David_Adams, Steven_Tiesdell (2010). Planners as Market Actors: Rethinking State–Market Relations in Land and Property
5. Sharachandra_Lele, Oliver_Springate-Baginski, Roan_P._Lakerveld, Debal_Deb, Prasad_Dash (2013). Ecosystem Services: Origins, Contributions, Pitfalls, and Alternatives
6. Jakob_Wachsmuth (2014). Cross-sectoral integration in regional adaptation to climate change via participatory scenario development
7. Reza_Dashti, Saeed_Afsharnia, Farid_Ghaderi (2010). AGA (Asset Governance Assessment) for analyzing affect of subsidy on MC (Marginal Cost) in electrici

8. Thomas_Bauwens (2013). What Roles for Energy Cooperatives in the Diffusion of Distributed Generation Technologies?
9. Akhmad_Hidayatno, Arry_Rahmawan_Destyanto, Christin_Arauna_Hulu (2019). Industry 4.0 Technology Implementation Impact to Industrial Sustainable Energy in Indonesia: A Model Conceptualization
10. Carlos_Andrés_Uribe (2014). The Dark Side of Social Capital Re-examined from a Policy Analysis Perspective: Networks of Trust and Corruption

d3. Ограниченная рациональность, методы подталкивания и поведение потребителя

1. Kalle_Grill (2014). Expanding the Nudge: Designing Choice Contexts and Choice Contents
2. Zhijian_Zhang, Xueyuan_Wang (2020). Nudging to promote household waste source separation: Mechanisms and spillover effects
3. Jeffrey_R._Brown, Jeffrey_R._Kling, Sendhil_Mullainathan, Marian_V._Wrobel (2013). Framing Lifetime Income
4. David_G._Taylor, Iryna_Pentina, Monideepa_Tarafdar (2017). "Is More Less, or Is Less More?: Social Media's Role in Increasing (and Reducing) Information Overload from News Sources " Book Chapter in The Dark Side of Social Media: A Consumer Psychology Perspective, by Angeline Close Scheinbaum
5. Karoline_Kuenzel, Moritz_Loock, Claudio_Cometta (2014). How Punishment and Reward Increase Customer Acceptance of Demand Response in the Energy Industry
6. Wilfred_Amaloss, Chuan_He (2016). Does Informative Advertising Increase Market Price? An Experimental Investigation
7. Dai_Li-na (2010). The Habit Formation,Uncertainty,Liquidity Constraint and the Inhabitant Consumption:A Study Based on the Provincial Level Panel Data
8. David_Hagmann, Emily_H._Ho, George_Loewenstein (2019). Nudging out support for a carbon tax
9. Nan_Hu, Paul_A._Pavlou, Jie_Jennifer_Zhang (2017). On Self-Selection Biases in Online Product Reviews
10. Elmar_Nubbemeyer (2011). A Reconsideration of Full-Cost Pricing

d4. Общие вопросы поведенческой экономики

1. Khondkar_E._Karim (2016). Advances in Accounting Behavioral Research
2. Herbert_A._Simon (1984). Models of Bounded Rationality: Economic Analysis and Public Policy
3. WU_Ying-yu (2012). A Study on the Formation Path of Humanism Capital Value——Based on Bounded Rationality Man Hypothesis
4. Nathan_Berg (2014). The consistency and ecological rationality approaches to normative bounded rationality
5. Bruno_S._Frey, Adrian_Furnham, Alan_Lewis (1986). The Economic Mind: The Social Psychology of Economic Behaviour.
6. Gerd_Gigerenzer (2018). The heuristics revolution: Rethinking the role of uncertainty in finance
7. Todd_J._Zywicki (2018). Future Imperfect: The Behavioral Economics of Behavioral Law & Economics
8. Marina_Ryzhkova (2018). Convergence Of Behavioral Economics And Orthodox Economic Theory
9. Daniel_McFadden (2014). The new science of pleasure: consumer choice behavior and the measurement of well-being
10. Stephen_O._Murray, Joseph_H._Rankin (1980). Use diffusion: An extension and critique

d5. Поведенческие факторы в корпоративном управлении

1. Koen_Vliek (2015). Effectuation heuristics for the incumbent firm
2. Angela_T._Hall, Stacy_Hickox, Jennifer_Kuan, Connie_Sung (2017). Barriers to Employment: Individual and Organizational Perspectives
3. Olivier_Meier, Aurélie_Sannajust (2014). Public to Private transactions and cognitive biases: A European study

4. Amit_Jain, Raymond-Alain_Thietart (2013). Knowledge based transactions and decision framing in Information Technology Outsourcing
5. Mateja_Drnovsek, Pankaj_C._Patel, Melissa_S._Cardon (2012). ENTREPRENEUR'S PSYCHOLOGICAL CAPITAL AND VENTURE GROWTH: TESTING THE GOAL MEDIATED RELATIONSHIPS
6. Laura_Cousens, Kathy_Babiak, Cheri_L._Bradish (2006). Beyond Sponsorship: Re-Framing Corporate-Sport Relationships
7. Sangeeta_Nadkarni, Bruce_W._Stening (1989). Human Resource Management in Remote Communities
8. HU_Chun-hu (2013). People-Oriented, Behavioral Values and Humanistic Management Accounting Mechanism
9. Anna_Plotnikova, Krsto_Pandza, Herald_Sales-Cavalcante (2020). How Strategy Professionals Develop and Sustain an Online Strategy Community – The Lessons from Ericsson
10. Zhao_Dan-hu (2014). Research on Supermarket's Participation in Agriculture-supermarket Jointing under the Bounded Rationality Perspective

d6. Экспериментальная экономика

1. Alexander_S._Kritikos, Friedel_Bolle (2005). Utility-Based Altruism: Evidence from Experiments
2. Iris_Bohnet, Stephan_Meier (2005). Deciding to Distrust
3. Harald_David_Stein (2010). Allocation Rules with Outside Option in Cooperation Games with Time-Inconsistency
4. Karl_Schurter, Bart_J._Wilson (2009). Justice and Fairness in the Dictator Game
5. Wolfgang_Breuer, Christiane_Helduser, Philipp_Schade (2016). Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game
6. Giovanni_Ponti (2000). Cycles of Learning in the Centipede Game
7. Werner_Güth (2000). Experimental game theory
8. Alan_G._Sanfey (2009). Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses
9. Catherine_C._Eckel, Angela_C._M._de_Oliveira, Philip_J._Grossman (2008). Gender and Negotiation in the Small: Are Women (Perceived to Be) More Cooperative than Men?
10. Julio_J._Rotemberg (2006). Minimally Acceptable Altruism and the Ultimatum Game

d7. Влияние фрейминга и других факторов на покупательную способность; электронная коммерция

1. Theerasak_Thanasankit (2003). E-commerce and cultural values
2. K._Caleb_Chan, Karen_D._Loch (1994). Towards building a theory of IT-enabled value-adding partnership: factors contributing to the effectiveness of electronic interface systems
3. Rong_He, Heqing_Li, Zeng_Lian, Jie_Zheng (2020). The effect of culture on consumption: A behavioral approach
4. Yang_Fan (2004). Importance of Commodity Package in Sale
5. Gregory_K._Dow (1984). Myopia, amnesia, and consistent intertemporal choice
6. Kenneth_D._Brown (2005). Framing production: technology, culture and change in the British bicycle industry
7. Ulrich_Witt (2011). The Dynamics of Consumer Behavior and the Transition to Sustainable Consumption Patterns
8. Kong_Rong (2009). Study on Trust Game of Micro-Credit in China
9. A._V._Muthukrishnan, Seshan_Ramaswami (1999). The Framing Effects of Bundling
10. Ulrich_Witt (2011). The Dynamics of Consumer Behavior and the Transition to Sustainable Consumption Patterns

d8. Влияние маркетинговой стратегии на потребителя

1. Andrew_Crane (1997). The dynamics of marketing ethical products: A cultural perspective
2. Zhang_Lu (2012). On Packaging Design and Marketing Strategy
3. Jing_Zhang, Josée_M._M._Bloemer (2008). The Impact of Value Congruence on Consumer-Service Brand Relationships
4. Jiankang_Mu (2017). The Study on Activation Strategy of Time-honored Brand
5. H._C._Purohit (2011). Consumer Values and Ethics as a Determinant of Buying Decision: An Analytical Study
6. Giuseppe_Aliperti, Francesco_Rizzi, Marco_Frey (2018). Cause-related marketing for disaster risk reduction in the tourism industry: A comparative analysis of prevention- and recovery-related campaigns
7. Jing_Jian_Xiao, Barbara_O'Neill (2018). Mental accounting and behavioural hierarchy: Understanding consumer budgeting behaviour
8. Shuto_Ogura, Kosuke_Tanaka, Shintaro_Terabe, Hideki_Yaginuma, Nang_Kang (2019). FRAMING EFFECT OF INFORMATION PROVISION FOR SHOPPING BEHAVIOR CHANGE
9. Ludmila_Bandeira_Lima_B Barros, Martin_de_La_Martinière_Petroll, Cláudio_Damacena, Marc_Knoppe (2019). Store atmosphere and impulse: a cross-cultural study
10. Hyun_Joung_Lee, Peter_J._Boyle, Jong-Ho_Lee (2011). The Effects of Decision Control and the Number of Alternatives in Purchase Decision Processes

d9. Поведенческая теория фирмы

1. Rashmi (2015). CORPORATE SOCIAL RESPONSIBILITY: INITIATIVE TAKEN BY RELIANCE INDUSTRIES LIMITED IN INDIA
2. HE_Qi-yan (2012). Psycho-analysis on the Dimission of Hotel Staff and the Countermeasures——The Case of Hotel Management Specialty Graduates of Higher Vocational Schools
3. Andrea_Zangiacomi, Marco_Sacco, Elena_Pessot, Alberto_De_Zan, Massimiliano_Bertetti (2018). A Perspective for the Implementation of a Path Towards the Factory of the Future: The Italian Case
4. Lin-Hua_Lu, Poh-Kam_Wong (2019). Performance feedback, financial slack and the innovation behavior of firms
5. Tomasz_Obloj (2011). Incentive Life-Cycles: The Role of Learning, Ability, and Cognitive Biases in the Division of Value within Firms
6. Jiao_Huang, Muriel_Tichit, Monique_Poulot, Ségolène_Darly, Shuangcheng_Li, Caroline_Petit, Christine_Aubry (2015). Comparative review of multifunctionality and ecosystem services in sustainable agriculture.
7. Kevin_Mole (1999). UK support for small firms: an examination of business advisors' heuristics
8. Mehmet_N._Aydin, Mark_E._Bakker (2008). Analyzing IT maintenance outsourcing decision from a knowledge management perspective
9. Vinita_Sinha, K._S._Subramanian, Sonali_Bhattacharya, Kaushik_Choudhary (2012). The contemporary framework on social media analytics as an emerging tool for behavior informatics, HR analytics and business process
10. Wageeh_A._Nafei (2015). The Effects of Psychological Capital on Employee Attitudes and Employee Performance: A Study on Teaching Hospitals in Egypt

d11. Поведенческая экономика и налогообложение

11. Simon_James (2006). Taxation and the Contribution of Behavioral Economics
12. Otto_Eibl, Jozef_Zagraban (2016). Do Negative News Create Tax Evaders? Evidence from the Tax Compliance Experiment
13. Robin_Boadway (2012). From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views

14. Steuart_Henderson_Britt (1940). Rules of Evidence- An Empirical Study in Psychology and Law
15. Erich_Kirchler, Valerie_Braithwaite (2007). The Economic Psychology of Tax Behaviour: References
16. George_W._Borden, Dale_J._Menkhaus, Glen_D._Whipple, Ray_A._Field, Elizabeth_Hoffman (1990). Factors Influencing Bids For Vacuum Packaged Retail Beef: An Experimental Economics Approach
17. a*tefura_Gabriela (2011). Tax Compliance Behaviour
18. Nahariah_Jaffar (2009). Fraud Detection: The Moderating Role of Fraud Risk Level
19. Benjamin_A._Templin (2010). Social Security Reform: Should the Retirement Age Be Increased?
20. Christian_A._Vossler, Michael_McKee (2015). Behavioral effects of tax withholding on tax compliance: Implications for information initiatives

d12. Поведенческие финансы

1. Meir_Statman (2019). A Second Generation Behavioral Finance
2. Kwang-Hoon_Ji (2007). The Behaviors of The Korean Stock Markets Following Large Price Changes: An Empirical Study on the Behavioral Finance Hypothesis
3. Zhang_Wei, Zhang_Xiao-tao (2005). Research on Asymmetry of Volatility of Shanghai Stock Market—Comparison of GJR-and VS-GARCH
4. Liu_Jun (2002). Behavioral Finacial Theory and Investment Fund Manager's Choice
5. Elian_Wiener (2005). The psychology of money : personal finance
6. George_Bakos, George_Petrakos (2017). Overreaction Analysis In State-Owned Equities
7. Mohamed_Djarraya (1980). Behavioral Models of Dividend Policy and Their Implications to Financial Management
8. Frank_T._Magiera (2007). The Psychology of Ethics in the Finance and Investment Industry
9. Idárraga_Calderón, Cindy_Alejandra (2018). Behavioral Finance. How psychological factors can influence the stock market
10. Gong_Chun-xia (2013). Analysis of the Cash Dividend Policy of Chinese Public Companies from the Perspective of Corporate Finance Behavior

d13. Экспериментальные подходы к анализу готовности платить

1. Choongbeom_Choi, Anna_S._Mattila (2014). The effects of promotion framing on consumers' price perceptions
2. Jin_H._Paik, Martin_Scholl, Rinat_A._Sergeev, Steven_Randazzo, Karim_R._Lakhani (2020). Innovation Contests for High-Tech Procurement
3. Jesse_A._Matheson (2012). Prices and social behavior: A study of adult smoking in Canadian Aboriginal communities
4. Faculteit_Economie (2006). The Impact of Uncertainty on Incentive Framing Effectiveness in a Multidimensional Task Environment
5. Adewale_Oparinde, Ekin_Birol, Abdoul_Murekezi, Lister_Katsvairo, Michael_T._Diressie, Louis_Butare (2016). Radio Messaging Frequency, Information Framing, and Consumer Willingness to Pay for Biofortified Iron Beans: Evidence from Revealed Preference Elicitation in Rural Rwanda
6. Paul_S_Biederman, Jun_Lai (2007). Travel and Tourism: An Industry Primer
7. Clemens_F._Köhler, Els_Breugelmans, Benedict_G._C._Dellaert, Ko_de_Ruyter (2007). Choice for Choice: The effects of consumer choice task framing on the decision to use a recommendation agent
8. Patrali_Chatterjee (2011). Framing Online Promotions: Shipping Price Inflation and Deal Value Perceptions
9. Anne_Odile_Peschel, Carola_Grebitus, Bodo_Steiner, Michele_M._Veeman (2015). A Behavioral Approach to Understanding Green Consumerism Using Latent Class Choice Analysis
10. Sean_R._McMahon, Laura_A._Orr (2017). Pop psychology? Searching for evidence, real or perceived, in bestselling business books

d14. Игровые подходы и теория перспектив в решении различных проблем

1. Xiaoran_Yu, Guanglong_Dong, Changyu_Liu (2020). The Tripartite Regulation Game of Carbon Financial Products Based on the Prospect Theory
2. Zhonghai_Liu, Kuangqi_Li, Ning_Zhang, Lianyou_Li, Yi_Fan (2017). Are Farmers' Behaviors Rational When They Pay Less for Social Endowment Insurance? Evidence from Chinese Rural Survey Data
3. Anna_Radzewicz, Radosaw_Winiewski, Justyna_Brzezicka, Oksana_Kurji-Wysocka, Marek_Walacik (2012). Rationality of actors and impulsive action in the property market
4. Niklas_Etzell (2016). The framing of bundling coupons: an experimental study (Available on Internet)
5. Ya-Lan_Li, Qi-Qing_Song (2016). Evolution Model Analysis of Group Buying Games
6. Yu-Jui_Huang, Zhenhua_Wang (2020). Optimal Equilibria for Multi-dimensional Time-inconsistent Stopping Problems
7. A._C._Burton, K._S._Carson, Susan_M._Chilton, W._George_Hutchinson (2005). Resolving questions about bias in real and hypothetical referenda
8. Alexander_W._Cappelen, Ottar_Mæstad, Bertil_Tungodden (2010). Demand for Childhood Vaccination – Insights from Behavioral Economics
9. Toshiaki_Koide (2000). "Dislike-for-inflation psychology" and endogenous price fluctuations reconsidered

d15. Политика в отношении современных проблем

1. David_S._Wall (2018). How Big Data Feeds Big Crime
2. Giancarlo_Fortino, Lidia_Fotia, Fabrizio_Messina, Domenico_Rosaci, Giuseppe_M._L._Sarné (2020). Trust and Reputation in the Internet of Things: State-of-the-Art and Research Challenges
3. Lillian_C_Borrone (2005). Part 1: 2005 Thomas B. Deen Distinguished Lecture: Sparking the Globalized Trade and Transportation Connection: Supplying Freight System Responses to Global Trade Demands
4. Amber_Roth (2016). States of emergency: How jurisdictions are re-framing homelessness as a crisis
5. Lesley_Frank (2015). The Breastfeeding Paradox
6. Odysseas_Kopsidas (2017). An Application of 'Willingness to Pay' Method as a Quantifier for Environmental Impact Assessment
7. Dayna_Roslynn_Albert (1998). Creating spaces for dieticians to promote food security, can we move beyond the charity model?
8. David_Waite, Bruce_Whyte, Jill_Muirie (2020). From an agreeable policy label to a practical policy framework: inclusive growth in city-regions
9. M._B._de_Ondarza, J._M._Tricarico (2017). Review: Advantages and limitations of dairy efficiency measures and the effects of nutrition and feeding management interventions
10. LI_Xiang-mei (2008). Thinking on Security Information in the Administration of Entertainment Places

Таблица 6. Результаты анализа чувствительности для различных начальных инициализаций модели

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перспексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность

84	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.28e+17	0.8190000000000001	0.8713	54144.8008	0.5225	0.5876	-1.6582
85	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.21e+17	0.8204	0.8682	53907.8672	0.5263	0.5982	-2.7488
86	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.27e+17	0.8178	0.8716	54203.0664	0.5209	0.5812	-1.8428
87	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.25e+17	0.8155	0.8718	54339.8008	0.5173	0.5752	-2.4951
88	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.33e+17	0.8164	0.8721	54266.332	0.519	0.5798	-2.8328
89	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e-01; -1.87e+00	2.24e+17	0.8207	0.8703	53986.8008	0.5257	0.5939	-3.0041

		: 1.0, u'authors': 2.0}									
9 0	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e- 01; -1.87e+00	2.21e+17	0.8174	0.8729	54070.535 2	0.5207	0.579 2	-2.5849
9 1	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e- 01; -1.87e+00	2.09e+17	0.8187	0.8734	54123.332	0.5215	0.586 3	-3.2852
9 2	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e- 01; -1.87e+00	2.17e+17	0.82	0.8713	53953.734 4	0.5246	0.592 3	-2.3434
9 3	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	20	[SparsePhi, SparseTheta]: -5.00e- 01; -1.87e+00	2.3e+17	0.8212	0.8667	54056.265 6	0.5258	0.598	-2.2601

Таблица 7. Примеры ключевых биграмм для модели, построенной по другой начальной инициализации.

Столбцы 1-15 – номера тем для результатов новой модели. Строка «Соответствие теме» содержит номер темы исходной модели, которой соответствует тема новой модели

Номер темы	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Соответствие теме	11	-	4	-	5	9	15	12	-	3*	8	2	14	13	6
	framing,	framing,	behavioral_economics,	investor_sentiment,	decision_making	framing,	framing,	behavioral_finance,	framing,	nudge,	framing,	supply_chain,	framing,	framing,	ultimatum_game,
	prospect_theory,	credit_card,	framing,	stock_return,	framing	corporate_governance,	behavioral_economics,	prospect_theory,	decision_making,	framing,	consumer_behavior,	bounded_rationality,	trust_game,	public_good,	prospect_theory,
	decision_making,	economic_growth,	decision_making,	framing,	behavioral_economics,	decision_making,	decision_making,	stock_market,	experimental_approach,	nudging,	e_commerce,	energy_efficiency,	public_good,	dictator_game,	experimental_economics,
	bounded_rationality,	information_security,	bounded_rationality,	stock_price,	bounded_rationality,	social_responsibility,	ecosystem_service,	investment_decision,	climate_change,	climate_change,	decision_making,	decision_making,	decision_making,	willingness_pay,	game_theory,
	loss_aversion,	behavioral_economics,	economic_theory,	real_estate,	business_model,	corporate_social,	public_service,	financial_market,	human_resource,	decision_making,	customer_satisfaction,	climate_change,	experimental_economics,	prosocial,	expected_utility,
	transaction_cost,	economic_development,	endowment_effect,	stock_market	human_resource,	family_firm,	land_use,	decision_making,	resource_management,	public_health,	purchase_intention,	framing,	experimental_approach,	contingent_valuation,	bounded_rationality,
	behavioral_theory,	long_term,	neoclassical_economics,	trading_volume,	social_behavior	corporate_finance,	long_term,	loss_aversion,	long_term,	food_waste,	real_estate,	behavioral_economics,	dictator_game,	behavioral_economics,	decision_making,
	tax_compliance,	local_government,	financial_crisis,	short_term,	experimental_economics,	cognitive_biases,	sustainable_development,	individual_investor,	air_pollution,	food_security,	consumer_psychology,	renewable_energy,	bounded_rationality,	decision_making,	utility_theory,
	theory_firm,	climate_change,	behavioral_economics,	decision_making,	moral_hazard,	listed_company,	social_science,	asset_pricing,	e_r,	health_care,	marketing_strategy,	low_carbon,	trust_trustworthiness,	social_behavior,	utility_function,

risk_taking,	decision_making,	law_economics,	exchange_rate,	public_policy,	earnings_management,	nudge,	risk_ave	renewable_energ	behav-	product_ser-	en-	so-	field_ex	nash_eq
tax_eva-	social_security,	ratio-	earnings_announ-	resource_manage-	finan-	hu-	dispo-	developing_cou-	food_sys-	ser-	electric-	so-	fram-	trust_ga-
refer-	developing_coun-	experi-	bounded_rational	earnings_manage-	bounded	politi-	effi-	social_capital,	food_saf-	pur-	evolu-	third_pa-	choice_	labora-
fram-	low_income,	ecologi-	long_term,	experimental_app	behav-	so-	stock_pr	social_norm,	food_ch	con-	nash_equ	pub-	so-	so-
firm_per-	labor_market,	pub-	abnormal_return,	long_term,	fam-	pub-	fi-	welfare_state,	food_su	con-	chain_ma	que_le,	experi-	dicta-
gain_los-	public_sector,	so-	momentum_strat	social_psycholog	firm_per	experi-	stock_e	e_n,	sustaina-	so-	en-	so-	so-	behav-

Примечание. * Тема 10 частично схожа с темой 3 исходной модели, однако акцентирует внимание на вопросах подталкивания в вопросах питания и продовольствия

Таблица 8. Результаты тестирования состава модальностей

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность
179	13	{u'bigrams': 1.0}	20	[SparsePhi, SparseTheta]: -6.00e-01; ; -2.81e+00	8190.0	0.7805	0.9057	63522.6172	0.5052	0.5281	-2.0101
181	13	{u'text': 0.5, u'bigrams': 1.0}	20	[SparsePhi, SparseTheta]: -6.00e-01; ; -2.81e+00	329000.0	0.8196	0.8488	60385.7695	0.5654	0.6787	-2.2311
183	13	{u'text': 0.5, u'bigrams': 1.0,	20	[SparsePhi, SparseTheta]: -6.00e-01; ; -2.81e+00	2030000000.0	0.816	0.8513	60800.7695	0.5587	0.6669	-2.5371

		u'authors': 1.0}									
18 5	13	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	20	[SparsePhi, SparseTheta]: - 6.00e-01; ; - 2.81e+00	6.69e+13	0.805	0.8622	61662.539 1	0.5417	0.629	-1.8164

Таблица 9. Результаты тестирования весов модальностей

	Число тем	Модальности	Число итераций	Стратегия регуляризации	Перплексия	Разреженность Φ	Разреженность θ	Размер ядра	Контрастность	Чистота	Когерентность
103	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[]: ;	5.26e+13	0.6423	0.0004	55025.867 2	0.4996	0.5694	-2.4057
104	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 0.5, u'authors': 1.5}	10	[]: ;	2.07e+13	0.6305	0.0003	54901.601 6	0.502	0.5796	-2.8385
105	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.5}	10	[]: ;	3.59e+15	0.6459	0.0003	55076.867 2	0.4985	0.5655	-2.44
106	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 0.5,	10	[]: ;	1.4e+15	0.6351	0.0002	54991.601 6	0.5003	0.5744	-2.9979

		{u'authors': 2.0}									
107	15	{u'text': 0.5, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	10	[];	2.41e+17	0.6497	0.0003	55121.668	0.4973	0.5611	-2.472
108	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 0.5, u'authors': 1.0}	10	[];	1.16e+13	0.5468	0.0007	53550.535 2	0.5239	0.636	-2.5666
109	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.0}	10	[];	2e+15	0.5666	0.0006	53880.199 2	0.5189	0.6226	-2.5961
110	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 0.5, u'authors': 1.5}	10	[];	8.51e+14	0.5505	0.0007	53660.668	0.5224	0.6331	-2.6132
111	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 1.0, u'authors': 1.5}	10	[];	1.44e+17	0.5702	0.0006	53961.535 2	0.5176	0.6196	-2.5245
112	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 0.5,	10	[];	6.14e+16	0.5539999999999999	0.0007	53750.933 6	0.5209	0.6297	-2.6115

		u'authors': 2.0}									
113	15	{u'text': 1.0, u'bigrams': 1.0, u'references': 1.0, u'authors': 2.0}	10	[]: ;	1.02e+19	0.5735	0.0006	54052.265 6	0.5161	0.6163	-2.5948

Приложение 4. Скрипт предобработки данных

```
from collections import Counter
import nltk
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import numpy as np
from itertools import chain
stop_file='stop_words.txt'

tokenizer = RegexpTokenizer(u'[a-zA-Z][\s\p{L}]+') # only words
lemmatizer = WordNetLemmatizer()

# merge two set of stop words
stop_words = set(stopwords.words('english'))
esw = open(data_dir+stop_file, "r").read().decode('utf-8').split()
esw=set(esw)
full_stop_words=set(chain(stop_words,esw))
# Список одиночных слов, которые необходимо добавить к биграммам:
white_unigrams=set(['nudge', 'nudging', 'framing'])

# Токенизация, лемматизация, удаление стоп слов
def text_preprocess(plain_text):
    intermediate = tokenizer.tokenize(plain_text.lower())
    lemmatized = [lemmatizer.lemmatize(w) for w in intermediate]
    preprocessed_text = [i for i in lemmatized if i not in full_stop_words]
    return preprocessed_text

# Подсчет числа униграмм (одиночных слов)
def get_counted_unigrams(preprocessed_text):
    counts_uni = Counter(preprocessed_text) # счетчик частот униграмм
    unigrams_count_line="" # unigrams in line with counts
    unigram_tags_count_line=""
    unigram_tags=[]
    for char in counts_uni.keys():
        unigrams_count_line=unigrams_count_line+char+' '+str(counts_uni[char])+' '
        if (char in white_unigrams): #& (counts_uni[char]>=3):
            unigram_tags.append(char)
            unigram_tags_count_line=unigram_tags_count_line+char+' '+str(counts_uni[char])+' '
    uniword='|||'.join(preprocessed_text)
    return uniword, unigrams_count_line, unigram_tags, unigram_tags_count_line

# Список биграмм, которые необходимо исключить из анализа:
stop_list=set(['et_al', 'e_g', 'de_los', 'de_la', 'case_study', 'p_p',
    'main_result', 'study_provides', 'research_question', 'better_understanding', 'per_cent',
    'per_caput', 'de_l', 'en_el', 'en_la', 'n_de', 'f_r', 'sur_le', 'dans_le', 'r_e',
    'result_show',
    'decision_maker',
    'result_suggest',
    'united_state',
    'making_process',
    'result_indicate',
    'research_ha',
    'future_research',
    'ha_become',
    'methodology_approach',
    'important_role',
    'cognitive_dissonance'
    ])

# Подсчет числа биграмм
def get_counted_bigrams(preprocessed_text, unigram_tags, unigram_tags_count_line):
    bigramm_cutoff_value=1
    bigrams_count_line="" # bigrams in line with counts
    biword_list = [' '.join(b) for b in nltk.bigrams(preprocessed_text)]
    biword = "|||".join(biword_list+unigram_tags)
    fdist = nltk.FreqDist(biword_list)
    bigrams_count_line="" # bigrams in line with counts
    #bigrams_tags=""
    for word, word_count in fdist.items():
        if word not in stop_list:
            bigrams_count_line=bigrams_count_line+ word +' '+str(word_count) +' '
    line=bigrams_count_line+unigram_tags_count_line
    return biword, line

# Формирование файла в формате .vpw
import io
vpw_name='Ec_Bu_be_vpw.txt'
table['title']=table['title'].astype("string")
```

```

table['authors_name']=table['authors_name'].astype("string")
table['List_outCitations']=table['List_outCitations'].astype("string")
table['title'] = table['title'].replace(np.nan, "", regex=True)
table['paperAbstract']=table['paperAbstract'].astype("string")
table['paperAbstract'] = table['paperAbstract'].replace(np.nan, "", regex=True)
table['bigrams']="
bigrams_list=[]

with io.open(data_dir+vpw_name, 'w', encoding='utf8') as vpw_lemmas_file: #file to write
    for index, row in table.iterrows():
        title_abstract=row['title']+ ' '+ row['paperAbstract']
        preprocessed_text=text_preprocess(title_abstract)
        if len(preprocessed_text)>1:
            uniword, titleAndAbstract_unigrams_count_line, unigram_tags, unigram_tags_count_line = get_counted_unigrams(preprocessed_text)
            biword, titleAndAbstract_bigrams_count_line= get_counted_bigrams(preprocessed_text, unigram_tags, unigram_tags_count_line)
            bigrams_list.append(biword)
            List_outCitations=[cit.strip() for cit in row['List_outCitations'].split(',') if len(cit.strip())>0] # len(cit.strip())>0 - т.к. есть пустые "
            new_line=row['_id'] + '|authors ' + row['authors_name'].decode('utf8') + '|text ' + titleAndAbstract_unigrams_count_line + '|bigrams ' + titleAndAbstract_bigrams_count_line + '|references ' + ','.join(List_outCitations)
            vpw_lemmas_file.write(unicode(new_line))
            vpw_lemmas_file.write(unicode('\n'))
table['bigrams']=bigrams_list
table.to_csv(data_dir+'Ec_Bu__be_p1_p2_filtered_preprocessed2.csv', encoding='utf-8', index=False)

```

Примечание: # - комментарий к выполнению кода

Приложение 5. Скрипт для вычисления когерентности тем

```

import itertools
import numpy as np
import pandas as pd

# Загрузка таблицы с исходными преобработанными данными
data_dir='./data/'
file_name='Ec_Bu__be_p1_p2_filtered_preprocessed2.csv'
#header=['_id', 'authors', 'fieldsOfStudy', 'id', 'inCitations', 'journalName', 'outCitations', 'paperAbstract', 'title', 'year']
table_prep = pd.read_csv(data_dir+file_name, delimiter=',', quotechar='"', error_bad_lines=False, engine='python')
table_prep['bigrams'] = table_prep['bigrams'].apply(lambda x: frozenset([word.strip() for word in x.split("||")]))

def CoherenceCalculating(topics_directory):
    # открываем и считываем все слова всех тем - имя директории (topics_directory) зависит от числа строк в таблице результатов
    topic_files = [i for i in os.listdir(topics_directory) if (i != 'coherence.txt') & (i.endswith(".txt")) & (i.startswith("d"))]
    print len(topic_files)
    # словарь, ключи которого - названия топиков (d8_tokens, b10_tokens и т.п.),
    # элементы - кортеж: (словарь слов->частоты, словарь пар->частоты(+PMI), связность)
    topics_dict = {}
    # topics_coherence = {} # формируем пустой словарь для связности

    # Формируем словарь из тем с элементами:
    for topic_file in topic_files:
        # считываем топ-слова темы
        topic_words = []
        with open(topics_directory+topic_file, 'r') as f:
            topic_words = [line.split(',')[0] for line in f] # в файле и слово, и значение, берем только слово [0]
        words = dict.fromkeys(topic_words, 0)
        # составляем всевозможные пары комбинаций (сразу конвертируем в наборы)
        pairs = dict.fromkeys([frozenset(p) for p in itertools.combinations(topic_words, 2)], 0)
        coherence = 0.0
        # имя топика = имя файла (без .txt)
        topics_dict[topic_file[:-4]] = (words, pairs, coherence) # [i for i in itertools.combinations(topic_words, 2)])

rows = table_prep['bigrams']
# iterate over rows and then over topics and calculate counts of words & pairs
for row in rows:
    for topic in topics_dict:
        # unpack tuple
        words, pairs, coherence = topics_dict[topic]

        # increment word
        for word in words:
            words[word] += int (word in row)

        # increment pair
        for pair in pairs:

```



```

    pairs[pair] += int(pair.issubset(row))

# re-pack
topics_dict[topic] = (words, pairs, coherence)

D = len(rows)

# calculate PMI coherence by topic
for topic in topics_dict:
    # unpack tuple
    words, pairs, coherence = topics_dict[topic]
    # calc PMI and csum them for all pairs
    for pair in pairs:
        PMI = calc_PMI(pair, pairs, words, D)
        coherence += PMI
    # now save tuple of count and PMI for pair
    pairs[pair] = pairs[pair], PMI

# normalize PMI by pairs count
coherence = coherence/float(len(pairs)) # = k(k-1)/2
# re-pack
topics_dict[topic] = (words, pairs, coherence)

# save results and cumulate coherence
coherence = 0.0
with open(topics_directory+"coherence.txt", 'w') as f:
    for topic in topics_dict:
        coherence += topics_dict[topic][2]
        f.write(topic + " " +str(topics_dict[topic][2])+ "\n")

    coherence = coherence/float(len(topics_dict))
    f.write("Итоговая средняя когерентность = " + str(coherence))
return coherence, topics_dict

# =====
# Вычисление Pointwise Mutual Information, D - vol of corpus
def calc_PMI(pair, pairs, words, D, set_zero_count_by=1e-6):
    word = tuple(pair)
    # if smth is rare
    N12 = pairs[pair]
    if pairs[pair] == 0:
        N12 = set_zero_count_by*set_zero_count_by
    N1 = words[word[0]]
    if N1 == 0:
        N1 = set_zero_count_by
    N2 = words[word[1]]
    if words[word[1]] == 0:
        N2 = set_zero_count_by
    return np.log(float(D*N12)/float(N1*N2))

```

Приложение 6. Скрипт для построения тематической модели

```

import artm

# -- Вводим основные (domain - d) и фоновые (Background - b) темы.

def SetTopicsNum(d, b):
    background_topics = []
    domain_topics = []
    all_topics = []

    for i in range(1, d+b+1):
        if i <= d:
            topic_name = "d" + str(i)
            domain_topics.append(topic_name)
        else:
            topic_name = "b" + str(i)
            background_topics.append(topic_name)
        all_topics.append(topic_name)
    return all_topics, domain_topics, background_topics

# ---
#-- Инициализуем модель
batch_vectorizer = artm.BatchVectorizer(data_path='collection_batches3',
                                       data_format='batches')
main_dictionary = artm.Dictionary()
main_dictionary.load_text(dictionary_path='collection_batches3/main_dictionary.txt')

```

```

#Вводим контролируемые метрики модели.
def SetScores(model):

    model.scores.add(artm.PerplexityScore(name='Perplexity_Score',
        dictionary=main_dictionary))
    model.scores.add(artm.SparsityPhiScore(name='SparsityPhiScore_bigrams',
        class_id='bigrams'))
    model.scores.add(artm.SparsityThetaScore(name='SparsityThetaScore'))
    model.scores.add(artm.TopTokensScore(name='Bigrams_Top10_Tokens',
        num_tokens=15,
        dictionary = main_dictionary,
        class_id='bigrams'))
    model.scores.add(artm.TopTokensScore(name='Text_Top10_Tokens',
        num_tokens=15,
        dictionary = main_dictionary,
        class_id='text'))
    model.scores.add(artm.TopTokensScore(name='References_Top10',
        num_tokens=10,
        dictionary = main_dictionary,
        class_id='references'))
    model.scores.add(artm.TopTokensScore(name='Authors_Top15',
        num_tokens=15,
        dictionary = main_dictionary,
        class_id='authors'))
    model.scores.add(artm.TopicKernelScore(name='DomainTopicKernelScore',
        probability_mass_threshold=0.25,
        class_id='bigrams' ))

```

Приложение 7. Скрипт для выгрузки результатов на веб-сайт

```

# === Формирование перечня ключевых слов ===
import os
def GetBigrams(d):
    modality_folder='bigrams\'
    b_row=""
    bigrams_files = [i for i in os.listdir(results_dir+modality_folder)]
    bigrams_file_sorted=sorted(bigrams_files, key=lambda x: int(re.search(r'\d+(?=\.)', x).group()))
    bigrams_file=bigrams_file_sorted[d-1]
    b_list = open(results_dir+modality_folder+bigrams_file, "r", encoding='utf-8').read().split("\n")[:-1] #удаляем последнюю пустую строку
    for ind, bigram in enumerate(b_list):
        bigram_single=bigram.split(',')[0].replace('_', ' ')
        b_row=b_row+bigram_single+', ' #игнорируем вероятности. можно добавить условие на включение в список в зависимости от вероятности
    b_row_=""<p>"""+b_row[:-2]+"""</p>""
    return b_row_

# === Формирование ссылок на авторов. берем txt файл с авторами - преобразуем в html фрагмент
def GetAuthorsLinks(d): #d - номер темы - по нему смотрим, какой фрагмент файла брать
    modality_folder='authors\'
    authors_files = [i for i in os.listdir(results_dir+modality_folder)]
    authors_file_sorted=sorted(authors_files, key=lambda x: int(re.search(r'\d+(?=\.)', x).group()))
    authors_file=authors_file_sorted[d-1]
    a_list = open(results_dir+modality_folder+authors_file, "r", encoding='utf-8').read().split("\n")[:-1]
    authors_link_list=[]
    for ind, author in enumerate(a_list):
        single_author_list=author.split(',')
        authors_link=""<p><a href=https://www.semanticscholar.org/author/"+single_author_list[1].strip()+"" target="_blank">""+single_author_list[0].replace('_', ' ')+""</a></p>""
        authors_link_list.append(authors_link)
    return authors_link_list

import semanticscholar as sch
# === Преобразование из файлов ссылок с кодом (в папке references_codes) в файлы ссылок с названиями (кладутся в references_titles)
# ===Запускается единожды

def CreateCitDoc():
    topic_folder1='references_codes\'
    topic_folder2='references_titles\'
    ref_files = [i for i in os.listdir(results_dir+topic_folder1)]
    citation_general_list=[]
    for file_ref in ref_files:
        with open(results_dir+topic_folder2+file_ref, 'w+', encoding='utf-8') as f:

            cit_list = open(results_dir+topic_folder1+file_ref, "r", encoding='utf-8').read().split("\n")[:-1]

```

```

for ind, cit in enumerate(cit_list):
    citation=""
    try:
        paper = sch.paper(cit, timeout=60)
        authors_list=[]
        for author in paper['authors']:
            authors_list.append(author['name'])
        authors=(', '.join(authors_list))
        citation=authors + " (" + str(paper['year']) + "). " + paper['title']
    except Exception:
        pass
    f.write(citation + "|" + cit+"\n")
    print (citation + "|" + cit+"\n")
CreateCitDoc()

# === Берет файлы results_dir+references_titles\ и формирует из них ссылки для html

def GetCitationsLinks(d):
    modality_folder="references_titles\"
    ref_files = [i for i in os.listdir(results_dir+modality_folder)]
    ref_files_sorted=sorted(ref_files, key=lambda x: int(re.search(r'\d+(?=\.)', x).group()))
    ref_file=ref_files_sorted[d-1]
    ref_list = open(results_dir+modality_folder+ref_file, "r", encoding='utf-8').read().split("\n")[:-1]
    citation_general_list=[]
    for ind, cit in enumerate(ref_list):
        single_cit_list=cit.split('|')
        if len(single_cit_list[0])>0:
            citation_href="<p><a href=https://www.semanticscholar.org/paper/" + single_cit_list[1] + "" target="_blank">"" + single_cit_list[0] + ""</a></p>"
            citation_general_list.append(citation_href)
    return citation_general_list

# Получаем название темы по ее номеру
def GetTopicTitle(d):
    titles_list = open(results_dir+Topics_titles.txt', "r", encoding='utf-8').read().split("\n")
    keys=[]
    title_names=[]
    for title in titles_list:
        title_name=title.split(' ')
        keys.append(title_name[0])
        title_names.append(title_name[1])
    title_dictionary = dict(zip(keys, title_names))
    title_ = title_dictionary.get(str(d))
    return title_

# === Формирование общего html для темы - склеивание всех частей ===

def CreateHtmlForTopic(d):
    html_dir=results_dir+'html\
    with open(html_dir+'d'+str(d)+'.html', 'w+', encoding='utf-8') as f:
        title=GetTopicTitle(d)
        message_begin = ""<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN""http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" >
    <head>
    <meta charset="utf-8">
    <link rel="stylesheet" type="text/css" href="STYLE.css">
    <title>""+str(title)+""
    </title>
    </head>
    <body background-color=#f2f2f2><div>
    ""
    f.write(message_begin)

    # ===== TITLE =====
    if title is not None:
        f.write("""<h1>""+title+""</h1>""")

    # ===== BIGRAMS =====
    bigrams=GetBigrams(d)
    f.write(bigrams)
    # =====
    f.write("""<br>""")
    f.write( ""<table><tr><td width=55%> """)
    # ===== REFERENCES =====
    f.write("""<h2>""+'Ключевые публикации по теме:' + ""</h2>""")
    cit_link_list=GetCitationsLinks(d)
    for cit in cit_link_list:
        f.write(cit)
    f.write("""<br>""")

```

```

f.write( ""</td> """)
f.write( ""<td width=8%></td> """)
# ==== AUTHORS =====
f.write( ""<td> """)
f.write( ""<h2>""+'Основные авторы по теме:'+"""</h2>""")
authors_link_list=GetAuthorsLinks(d)
for authors_link in authors_link_list:
    f.write(authors_link)
# =====
f.write( ""</td></tr> <tr><td colspan=3><p><a href="index.html">На главную</a></p></td></tr></table>""")
f.write( ""</div>""")
f.write( ""</body>
</html>""")
# ==== ВЫЗОВ ГЛАВНОЙ ПРОЦЕДУРЫ ====
for i in range(15):
    print (i+1)
    CreateHtmlForTopic(i+1)

```