

МЕТОД СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА ЭКОНОМИЧЕСКИХ ОБЪЕКТОВ И СИСТЕМ

Филькин Михаил Евгеньевич, старший научный сотрудник ЦЭМИ РАН, Москва, mfilkin@mail.ru

Доклад на Второй конференции «Системные проблемы отечественной мезоэкономики, микроэкономики, экономики предприятий» Отделения «Моделирование производственных объектов и комплексов» (Отделения № 5) ЦЭМИ РАН (Москва) посвящен адаптации инструментария оценки параметров скрытых марковских цепей для оценки неизмеримых параметров экономических объектов и систем. Последовательно описаны этапы применения алгоритма для анализа скрытых переменных экономической модели. Для специфического решения экономических задач применены алгоритмы Витерби, Баума–Уэлша и прямого-обратного хода. Предложенный нами метод может быть использован для анализа широкого круга экономических систем, обладающих марковским свойством, а именно таких, где текущее состояние системы по вероятностному распределению зависит от ее предыдущего состояния. Приведена ссылка на пример использования алгоритма для эмпирической оценки уровня конкуренции на реальном розничном рынке.

Ключевые слова: эконометрический метод, конкуренция, скрытые марковские цепи, алгоритм Витерби, алгоритм Баума–Уэлша, розничный рынок.

Классификация JEL: C13.

В задачи стратегического планирования и анализа реальных экономических систем часто входит необходимость оценивания тех или иных параметров исследуемых объектов: рынков, предприятий, региональных экономик. Для этих целей исследователь может использовать мощный эконометрический инструментарий, разрабатываемый не одно столетие и получивший новое ускорение в последние десятилетия в связи с развитием информационных и вычислительных технологий. В зависимости от специфики задачи оценка реальных параметров экономических систем может быть получена, к примеру, методами регрессионного, вариационного, дисперсионного или других видов анализа. В нашем докладе представлен, на наш взгляд, перспективный метод оценивания параметров реальных экономических систем вероятностным методом скрытых марковских моделей. Этот метод, базовые основы которого были разработаны в 1960–1970-е гг., в настоящее время активно используется в различных областях науки – от криптографии до анализа структуры ДНК и распознавания речи. Использование метода в экономических задачах обычно сводится к прогнозированию движения финансовых рынков, однако его общность и доказанная итерационная сходимость позволяет применять ряд алгоритмов для широкого круга проблем экономики.

Существует ряд экономических объектов или систем, которые можно описать в терминах *скрытых состояний* и *видимых переменных*. В общем виде формализация

выглядит следующим образом. Пусть в некоторой экономической системе есть скрытые переменные, которые невозможно наблюдать непосредственно. Например, такой переменной может быть состояние розничного рынка: конкурентное – неконкурентное. Исследователь, изучая тот или иной рынок подчас не может непосредственно выявить сговор, тайные картельные соглашения или неявные договоренности о разделе рынка или ценообразовании. Поэтому состояние рынка [конкуренция – сговор] можно считать примером *невидимой* переменной. Однако на этом рынке есть *видимые* переменные, – например, розничные цены товаров, либо объемы продаж этих же товаров. *Невидимые* переменные тем или иным образом влияют на уровень и динамику *видимых*. Задача исследователя: собрать данные по *видимым* переменным и на основе их анализа определить возможные состояния *невидимых*. Этой цели служит предлагаемый метод скрытых марковских цепей. Рассмотрим алгоритм работы этого метода на простом модельном примере.

Представим, что у вас есть друг, он живет в Люксембурге, где гипотетически возможны два состояния погоды: дождь или солнце. Нам известно, что наш друг каждый день может делать только одно из трех дел: ходить по магазинам, делать уборку в доме или писать книгу. Каждое утро друг выглядывает из окна и решает, что ему сегодня делать. Его решение зависит от того, какую погоду он наблюдает, выглянув в окно. Предположим, что выбор решения о том, как ему поступить, имеет вероятностный характер. Так, если друг увидит, что на улице идет дождь, то с вероятностью 10% он пойдет в магазин, с вероятностью 30% сделает уборку, с вероятностью 60% будет писать книгу. Если же друг обнаружит, что на улице светит солнце, значения вероятностей для каждого решения существенно изменятся: с вероятностью 80% он пойдет в магазин, с вероятностью 15% останется дома, чтобы заняться работой над книгой, с вероятностью 5% займется уборкой. Таким образом, мы можем составить дискретную матрицу исходов дня нашего друга для двух (скрытых для нас) состояний погоды (табл. 1).

Таблица 1

Распределение вероятных способов поведения для двух состояний погоды, %

Погода	Магазин	Книга	Уборка
Дождь	10	60	30
Солнце	80	15	5

Для нас состояния погоды – «Дождь» и «Солнце» – являются скрытыми. Однако представим, что мы звоним каждый вечер другу и спрашиваем, что он делал сегодня. У нас для каждого дня под номером t возможны три вида наблюдений O_t : «магазин», «книга», «уборка».

На основании многолетних наблюдений мы выяснили, что если сегодня в Люксембурге солнечный день, то завтра с вероятностью 80% также будет солнце, а с вероятностью 20% пойдет дождь. В свою очередь, если сегодня дождь, то шанс того,

что завтра будет солнечный день, составляет 60, а 40% – вероятности продолжения дождливой погоды. Таким образом, у нас имеется матрица, составленная на основе наблюдений, переходов между *скрытыми* состояниями погоды (табл. 2).

Таблица 2

Матрица переходов между скрытыми состояниями, %

Состояния		Завтра	
		Дождь	Солнце
Сегодня	Дождь	40	60
	Солнце	20	80

Итак, наша задача поиска скрытых состояний заключается в том, что на основании множества наблюдений O_t ($t = 1, \dots, T$) и матрицы переходов A выявить наиболее вероятное состояние погоды в каждый день t в Люксембурге.

Скрытые марковские модели позволяют также предсказать вероятность следующего состояния (а, следовательно, и вероятность поведения друга) на следующий день $T + 1$ по набору наблюдений, предшествующих моменту T . Алгоритм для вычисления апостериорных вероятностей последовательности состояний при наличии последовательности наблюдений называется *алгоритмом прямого-обратного хода*. Он подробно описан в (Лоренс, Рабинер, 1986). Это первая из задач, которая может возникнуть при поиске скрытых состояний.

Однако легко представить ситуацию, когда нам неизвестна матрица переходов между состояниями погоды. Более того, нам может быть неизвестно распределение вероятных способов поведения нашего друга при каждом из состояний. Нам доступно только знание числа *скрытых* состояний и совокупность наблюдений O_t . Для оценки вероятности проявления каждого из *скрытых* состояний используется комбинация алгоритма Витерби (разработана и описана в его работе 1967 г.) и алгоритма Баума–Уэлша (описанного в совместной с Петри работе (Баум, Петри, 1966)). С помощью алгоритма Витерби ведется поиск наиболее подходящего набора состояний, который в контексте цепей Маркова получает наиболее вероятную последовательность произошедших событий. В свою очередь, с помощью алгоритма Баума–Уэлша вычисляются такие неизвестные параметры скрытой марковской модели, как вероятности перехода между состояниями и распределение вероятностей появления отдельных наблюдений.

Итак, решение задачи поиска *скрытых* состояний можно представить в виде последовательности этапов.

Этап 1. Для каждой последовательности наблюдений *видимых* состояний с помощью матрицы перехода вычисляется вероятность получения именно этой последовательности наблюдений.

Этап 2. Вычисляются такие параметры модели, которые наиболее точно описывают имеющуюся последовательность наблюдений.

Этап 3. С помощью обновленных параметров модели повторяются этапы 1 и 2 до необходимого уровня сходимости итерации.

Этап 4. Полученные на предыдущих этапах параметры модели применяются для поиска последовательности *скрытых* состояний, наилучшим образом (в терминах максимизации вероятности) описывающих данную последовательность наблюдений.

Этап 5. Смещается окно наблюдений и для нового набора повторяются процедуры расчета и вычисления последовательности *скрытых* состояний, заданной этапами 1–4. Делаем это для всего доступного эмпирического ряда. Итогом будет состояние скрытой переменной, заданная в динамике с учетом переходных состояний. Кроме того, на этом этапе будет посчитана матрица переходов между состояниями, которая имеет прогнозную силу, если в качестве начального состояния выбрать данные сегодняшнего дня. Такие результаты – еще одно дополнительное преимущество использования метода скрытых марковских цепей для анализа и прогнозирования состояния экономических объектов или систем.

В качестве примера для исследования рынка предложенным способом был выбран розничный рынок нефтепродуктов Санкт-Петербурга и Ленинградской области в период 1999–2007 гг. Результаты исследования опубликованы в 2017 г. в работе (Филькин, 2017). Выбор этого рынка и этого периода времени был обусловлен несколькими причинами. Во-первых, в эти годы происходила ускоренная автомобилизация домашних хозяйств, флагманами которых исторически являлись крупные города. Таким образом, ситуация на других региональных рынках нефтепродуктов в настоящий момент может иметь характерные черты, присущие Санкт-Петербургу, в качестве одного из своего рода *эталонов*. Во-вторых, период охватывает время до создания Петербургской топливной биржи (АО «Санкт-Петербургская международная товарно-сырьевая биржа»), которая изменила систему экономических взаимоотношений между оптовыми и розничными продавцами на топливном рынке. В-третьих, имеется возможность сравнить выводы нового метода исследования с существующими разработками (Панфилов, 2008; Филькин, 2010). Наконец, для этого периода истории данного рынка существует подробная база данных розничных цен, которая дает возможность исследовать вопрос на уровне отдельных АЗС. Особенностью метода было применение непрерывной функции распределения для *невидимой* переменной. Расчеты показали высокий уровень конкурентной борьбы между участниками рынка в политике ценообразования. При этом уровень конкуренции имеет сезонный характер, в среднем повышаясь в летний период, когда характерен повышенный спрос на бензин; вследствие этого продавцы могут иметь более серьезный стимул для кооперации по поводу цены.

Метод скрытых марковских цепей может быть использован для решения широкого круга других задач стратегического планирования и изучения эмпирических рынков. Выбор числа возможных *скрытых* состояний, а также характер их функций

распределения является экзогенным фактором модели и должен определяться исходя из практических соображений.

СПИСОК ЛИТЕРАТУРЫ

- Панфилов М.В.* Есть ли сговор на российском рынке розничной продажи нефтепродуктов? // Магистерская диссертация РЭШ, 2008.
- Филькин М.Е.* Метод скрытых марковских цепей для оценки уровня конкурентности. // Конкурентоспособность в глобальном мире: экономика, наука, технологии. 2017. № 8 (X. 4). С. 127–132.
- Филькин М.Е.* Эмпирический анализ конкуренции на розничном рынке бензина // Экономическая наука современной России. 2010. № 3 (50) С.78–92.
- Baum L.E., Petrie T.* Statistical Inference for Probabilistic Functions of Finite State Markov Chains // The Annals of Mathematical Statistics. 1966. № 37 (6). P. 1554–1563.
- Viterbi A.J.* Error bounds for convolutional codes and an asymptotically optimum decoding algorithm // IEEE Transactions on Information Theory. 1967. № 13 (2). P. 260–269.
- Viterbi A.J.* Error bounds for convolutional codes and an asymptotically optimum decoding algorithm // IEEE Transactions on Information Theory. 1967. № 13 (2). P. 260–269.
- Lawrence R. Rabiner, Juang B.H.* An introduction to hidden Markov models // IEEE ASSP Magazine. 1986. № 4–15.