

ПРОБЛЕМЫ КЛАССИФИКАЦИИ: МЕТОД КОМИТЕТОВ

УДК: 519.8

Никонов О.И. д.ф.-м. н., профессор
директор департамента БИММ

Уральский федеральный университет, ВШЭМ
Институт математики и механики УрО РАН

Чернавин Ф.П. аспирант
кафедра анализа систем и принятия решений
Уральский федеральный университет, ВШЭМ

М.А. Медведева к.ф.-м.н.
кафедра анализа систем и принятия решений
Уральский федеральный университет, ВШЭМ

Аннотация. В докладе мы рассматриваем частный случай нейронных сетей – сети, построенные методом комитетов. Описанный подход, был впервые предложен Эйблору и Кэйлором [1]. Дальнейшее развитие метод комитетов получил в Екатеринбурге в Институте Математики и механики УрО РАН в школе по распознаванию образов. В настоящее время подходы, базирующиеся на методе комитетов базируются на результатах Вл.Д. Мазурова и М.Ю. Хачая [2,3]. Мы описываем обобщенную конструкцию р-комитета и ее приложения к простейшему случаю распознавания значений бинарной случайной величины, отвечающей поведению курса одной валюты относительно другой, принимающей, следовательно, одно из двух значений «рост» или «уменьшение». Другое возможное приложение описываемого подхода к практическим задачам приведено в работе [4].

Ключевые слова: метод комитетов, прогнозирование, валютный рынок, бинарные опционы.

Abstract. In the paper we consider a particular case of a neural network - a committee method. The described approach was first introduced by the paper of Ablow, C.M. and Kaylor, D.J [1]. Then further development of the committee machine methods have been carried out by Yekaterinburg pattern recognition school in Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences. Nowadays theory of committee machine structures is based on results of Mazurov, Vl.D. and Hachay M.Yu. [2,3]. In the presentation we describe the generalized p-committee construction and its application to the simplest case with two possible solutions: the course of one currency to another either grows, or falls. Another issues related to committee methods are considered in the paper [4].

Keywords: committee method, forecasting, currency market, binary options.

Введение

Курс евро к доллару на рынке FOREX рассматривается как бинарная случайная величина, которая имеет два возможных исхода: рост или уменьшение. В рамках данной работы предполагается, что курс валюты является отражением общего состояния экономики и перспектив ее развития,

также нами предполагается, что фондовые индексы тоже являются отражением состояния экономики. Нами предполагается, что существует взаимозависимость между изменением курса валюты и уровней фондовых индексов.

Целью работы является показать применимость метода комитетов при прогнозировании движения курса евро к доллару на рынке FOREX на 10 минутном интервале исходя из изменений уровней фондовых индексов и курса евро относительно средних значений за 30 минут.

Понятие комитета впервые появилось в работах по распознаванию образов: в совместной статье Эйблоу и Кейлора [1] было введено понятие комитетного решения для системы строгих однородных линейных неравенств. В общем виде метод комитетов можно представить как частный случай нейронной сети. Дальнейшее развитие метода комитетов осуществлялось в екатеринбургской школе распознавания образов Института математики и механики УрО РАН. Современная теория комитетных конструкций опирается на результаты, полученные Вл. Д. Мазуровым и М.Ю. Хачаем [2,3].

Результаты работы могут быть интересны для краткосрочных инвесторов, использующих в своей работе стратегии внутридневной торговли для получения прибыли на разнице курсов валют. Они могут выступать в качестве дополнительного источника прогнозных данных о краткосрочных тенденциях валютного рынка и в качестве основы для алгоритмов высокочастотной торговли.

Метод комитетов

Методом комитетов называется композиция из линейных дискриминантов, разделяющая множество точек в пространстве факторов на 2 класса. За счет одновременного использования нескольких линейных дискриминантов, метод комитетов позволяет учитывать нелинейные связи переменных, что повышает качество классификации.

Математически p -комитет можно сформулировать следующим образом. Пусть заданы множество $M \subset R^n$ и класс функций $F \subset \{R^n \rightarrow R\}$. Известно, что $M = K_1 \cup K_2$, причем множества K_1 и K_2 заданы своими конечными подмножествами $A \subset K_1$, $B \subset K_2$. Задачей построения комитета называется задача нахождения набора функций $f = (f_1, \dots, f_q), f \in F$ такого, что:

$$\begin{cases} |\{p | f_p(a) > 0\}| > p, & \forall a \in A, \\ |\{p | f_p(b) < 0\}| \geq q - p, & \forall b \in B. \end{cases} \quad (1)$$

где $p = [0..q - 1]$, q – число членов комитета.

Рассмотрим задачу построения комитета через решение задачи частично-целочисленного линейного программирования.

Пусть $x_j \in K_1$, где $j \in J$, $x_i \in K_2$, где $i \in I$, J -число элементов первого множества, I - число элементов второго множества.

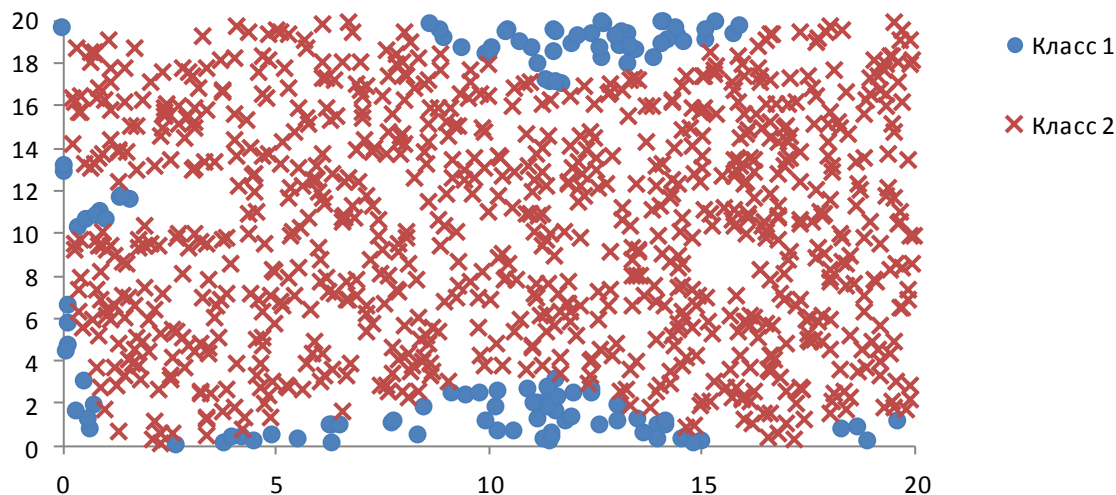


Рисунок. 1. Множество точек, разделенное на 2 класса

Поскольку $M \subset R^n$, где n является размерностью пространства (число параметров), то договоримся обозначать $x \in M$, как $x_{z,j} \in K_1, x_{z,i} \in K_2$, где $z \in Z$ (через Z обозначается множество $\{1,2,\dots,n\}$).

Назовем t -ым членом комитета последовательность $(w_1^t, w_2^t, \dots, w_z^t)$. Соответственно введем множество "невязок" $V = (v_1^t, v_2^t, \dots, v_j^t)$ и $V' = (v_1'^t, v_2'^t, \dots, v_i'^t)$, где t – индекс, указывающий на принадлежность к члену комитета.

Для построения p -комитета для всех j, i должна быть справедлива система неравенств:

$$\begin{cases} \sum_t v_j^t \leq p \\ \sum_t v_i'^t < q - p \end{cases}, \quad (2)$$

где $p = [0 \dots q - 1]$.

Задача построения p -комитета сводится к решению следующей системы неравенств.

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i^t < 0, \forall i \in I. \\ \sum_t v_j^t \leq p \\ \sum_t v_i^t < q - p \end{array} \right. \quad (3)$$

Приведем графический пример классификации.

Пусть имеется следующее множество точек, разделенных на 2 класса в двумерном пространстве.

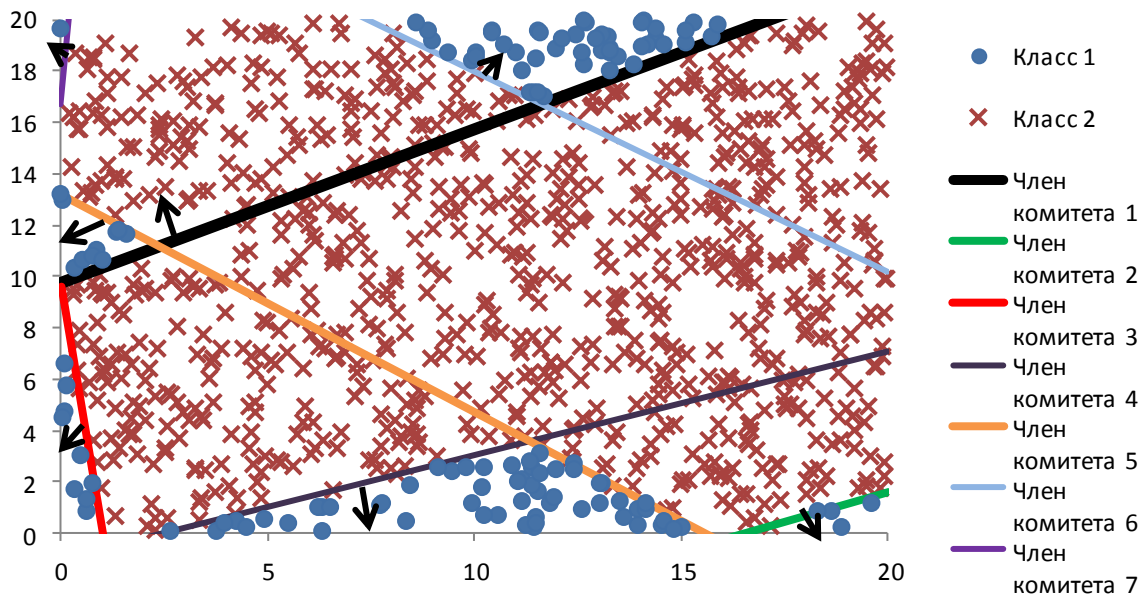


Рисунок 2. Разделение множества р-комитетом из 7 членов (p=7).

Попробуем разделить множество р-комитетом из 7 членов, в котором р принимает значение, отличное от q/2, 0 и q-1.

Стрелочками на рис. 2 показаны направления голосования членов комитета за принадлежность наблюдения 1 классу.

Заметим, что для практических задач не всегда целесообразно использовать комитетные решения, имеющие на обучающей выборке 100% качество классификации, поскольку в исходных данных часто имеются ошибки и шумы, в связи с чем несмотря на 100% качество классификации на обучающей выборке, качество решения на тестовой выборке значительно снижается, на что указывается в работе Вл. Д. Мазурова [2].

Введем дополнительные множества невязок $Y = (y_1, y_2, \dots, y_j)$ и $Y' = (y'_1, y'_2, \dots, y'_i)$, $Y, Y' \in \{0,1\}$.

Запишем задачу построения р-комитета как задачу частично-целочисленного программирования:

$$\left\{ \begin{array}{l} \sum_z w_z^t * x_{z,j} - c - L * v_j^t > 0, \forall j \in J \\ \sum_z w_z^t * x_{z,i} - c + L * v_i^t < 0, \forall i \in I. \\ \sum_t v_j^t \leq p + L * y_j \\ \sum_t v_i^t < q - p + L * y'_j \end{array} \right. \quad (14)$$

$$\min \sum_j y_j + \sum_i y'_i$$

Задачей является нахождение комитета, имеющего минимальное число ошибок на обучающей выборке.

Применение метода комитетов к прогнозированию движения курса

Прогнозирование курса евро к доллару осуществлялось в период, когда открыты только европейские фондовые рынки, то есть в период с 11:00 до 17:30 МСК. В данный период на курс евро к доллару не оказывают влияния новости об изменении американских фондовых индексов. По мнению авторов, на направление движения изучаемой величины влияют изменения уровней открытия, закрытия, минимального и максимального уровней относительно средних значений за 30 минут следующих фондовых индексов: CAC 40 – фондовый индекс Франции, DAX 30 – фондовый индекс ФРГ и FTSE 100 – фондовый индекс Британии, также нами учитывается изменение самого курса евро к доллару относительно среднего за последние 30 минут. Данные для прогнозирования динамики курса евро получены с сайта MFD.RU.

Заметим, что хотя Британия не входит в зону евро, экономики Британии и стран евро зоны являются взаимозависимыми, и, следовательно, изменение уровня индекса FTSE 100 также оказывает влияние на курс евро.

На основании обучающей выборки значений данных факторов были построены модели прогноза будущего движения цены закрытия. Обучающая выборка состоит из 1 000 наблюдений значений факторов за период с 28.09.2014 по 09.10.2014.

На обучающей выборке был построен комитет из 7 членов. Решающим правилом, в случае использования метода комитетов, является следующее соотношение: прогнозируется рост курса евро к доллару в случае, если комитетное решение за рост голосовало 4 и более членов, в противном случае прогнозируется падение.

Приведем результаты комитетного решения на обучающей выборке.

Таблица 1. Результаты комитетного решения на обучающей выборке.

Голосование комитета	Общее число наблюдений	Число верных решений	Число неверных решений	Доля правильных решений
«За» рост	518	362	156	69.9%
«Против» роста	482	330	152	68.5%
Итого:	999	692	307	69.3%

Как видно из таблицы 1 точность комитетного решения на обучающей выборке составила 69,3%, при этом точность в случае голосования за рост составила 69,9%, в случае голосования против роста – 68,5%, в целом комитет на обучающей выборке с одинаковой точностью прогнозирует как рост курса, так и снижение курса.

Полученная модель была применена для прогнозирования движения курса на тестовой выборке, состоящей из 1 271 наблюдений за период с 09.10.2014 по 27.11.2014. Периоды, за которые были получены значения изучаемых выборок, не пересекаются.

Результаты комитетного решения на тестовой выборке приведены в таблице 2.

Таблица 2. Результаты комитетного решения на тестовой выборке.

Голосование комитета	Общее число наблюдений	Наблюдения со случаями роста	Наблюдения со случаев снижения	Доля правильных решений
«За» рост	608	320	288	52,6%
«Против» роста	663	338	325	51,0%
Итого:	1271	658	613	51,8%

Согласно таблице 2 точность комитетного решения на тестовой выборке снизилась до 51,8%, при том, что на обучающей выборке точность составляла 69,2%.

Точность комитетного решения на тестовой выборке в первые 4 торговых дня (с 09.10.2014 по 14.10.2014) составила 62,9%, точность решения в период с 15.10.2014 по 28.11.2014 составила 50,4%.

По мнению авторов, точность комитетного решения на тестовой выборке снижается с увеличением времени, что может быть связано с изменением влияния изменения факторов на курс евро.

Предположим, что комитет достигает максимальной точности в течении 4 дней следующих за обучающей выборкой. Для подтверждения данного предположения построим комитет из 7 членов на обучающей выборке в период с 04.09.2014 по 14.10.2014 (1000 наблюдений).

Результаты комитетного решения, построенного на обучающей выборке за период с 04.09.2014 по 14.10.2014, на тестовой выборке приведены в таблице 3.

Таблица 3. Результаты комитетного решения на тестовой выборке.

Голосование комитета	Общее число наблюдений	Наблюдения со случаями роста	Наблюдения со случаев снижения	Доля правильных решений
«За» рост	659	322	337	48,9%
«Против» роста	469	241	228	51,4%
Итого:	1128	563	565	49,9%

Точность комитетного решения на тестовой выборке в первые 4 торговых дня (с 15.10.2014 по 20.10.2014) составила 56,3%, точность решения в период с 15.10.2014 по 28.11.2014 составила 49,0%. Заметим, что точность комитета, обученного ранее на обучающей выборке с 09.10.2014 по 27.11.2014, в период с 15.10.2014 по 20.10.2014 составила 49,3%. То есть, изменение обучающей выборки позволила увеличить точность комитетного решения, следовательно, предположение о том, что точность комитетного решения снижается с увеличением времени является допустимым.

Формирование торговой стратегии

Метод комитетов позволяет предсказать только падение или рост курса, в связи с чем при обычной «игре» на рынке FOREX применение данного метода возможно для уточнения решений иных методов анализа, позволяющих определить курс валюты на следующие 10 минут. По мнению авторов, метод комитетов представляет интерес при торговле бинарными опционами. Бинарный опцион - опцион, который в зависимости от выполнения оговоренного условия в оговоренное время либо обеспечивает фиксированный размер дохода (премию), либо не приносит ничего. Так как опцион покупают заранее по фиксированной цене, общий итог либо положительный (в размере разности между премией и ценой опциона), либо отрицательный (на величину стоимости опциона). Как правило, размер (модуль) положительного результата меньше, чем отрицательного. Обычно речь идёт о том, будет ли биржевая цена на базовый актив выше (или ниже) определённого уровня. Фиксированная выплата производится в случае выигрыша опциона, независимо от степени изменения цены (насколько она выше или ниже). Бинарные опционы позволяют точно знать размер выплаты и возможных рисков ещё до заключения контракта, что обеспечивает возможность проще управлять большим количеством торговых операций.

Поведем условную торговлю бинарными опционами в периоды с 09.10.2014 по 14.10.2014 и с 15.10.2014 по 20.10.2014 используя метод комитетов. Будем считать, что торговля осуществляется 100 условными единицами, в случае, если решение комитета верно, то выплачивается 200 условных единиц, у случае проигрыша теряется 100 условных единиц, по данным правилам, в случае если вероятность роста или снижения равна 50%, тогда математическое ожидание выигрыша равна 0. Рассмотрим возможный выигрыш, в случае применения комитетного решения. Так, в период с 09.10.2014 по 14.10.2014 общее число 10 минутных интервалов составило 143 из них верных решений комитета – 90,

ошибочных решений – 53, следовательно, совокупный выигрыш составит 7 400 условных единиц. В период с 15.10.2014 по 20.10.2014 общее число 10 минутных интервалов составило 144 из них верных решений комитета – 81, ошибочных решений - 63 следовательно, совокупный выигрыш составит 7 400 условных единиц.

Заключение

Использование комитетов при принятии инвестиционных решений при игре на изменениях валютного курса допустимо, точность комитетного решения в течении 4 торговых дней, следующих после обучающей выборки, превышает 55%. Заметим, что для практического применения метода комитетов для инвестиционных решений необходимо периодически осуществлять перестройку комитетного решения. В случае, торговли на фондовом рынке FOREX применение метода комитетов возможно для уточнения решений иных методов анализа, позволяющих определить курс валюты на следующие 10 минут. Отдельное применение метода комитетов возможно, при «игре» на бинарных опционах, когда требуется предсказать только направление курса. Хотелось бы отметить, что для построения комитета нами использованы только показатели индексов DAX 30, CAC 40, FTSE 100 и курс евро к доллару, добавление большего числа индексов, таких как Euronext 100, Euro Stoxx 50 и других, может привести к увеличению точности комитетного решения. В дальнейшем авторами планируется проведение исследование метода комитетов на иных временных интервалах с использованием большего числа параметров.

Список литературных источников

1. Мазуров Вл. Д. Метод комитетов в задачах классификации и оптимизации. М., Наука, 1990.
2. Мазуров Вл. Д., Хачай М.Ю. Конструкции метода комитетов // Известия Уральского Университета. Серия Математика и механика. 1999, Выпуск 2 (14), С.77-109.
3. Mazurov, Vl.D., Nachay M.Yu The committee machine constructions // News of the Ural University. Series Mathematic-mechanic. 1999. Issue 2 (14). p. 77-109.
4. Никонов О.И., Чернавин Ф. П. Построение рейтинговых групп заемщиков – физических лиц с применением метода комитетов // Деньги и кредит , №11, 2014, С. 52-55.