

Использование методов интеллектуального анализа данных для обработки данных колебания курса валют

И.Р. Шарапов¹ Д.А. Савельев^{1,2}

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. Методы интеллектуального анализа данных возможно использовать во многих отраслях, в частности, в экономике. В данной работе исследуются особенности применения методов интеллектуального анализа данных для классификации валют по степени стабильности в целях упрощения прогноза колебаний валютного курса.

1. Введение

В реальной экономике деньги как товар имеют свою цену – курс. Валютный курс зависит от традиционных ценообразующих элементов: спроса и предложения, которые, в свою очередь, подвержены влиянию множества факторов [1]. Основные факторы, влияющие на валютный курс – это решения центральных банков государств и международных организаций в области кредитно-денежной политики относительно выпускаемой ими валюты. То есть внутреннее положение государств напрямую влияет на колебание курса валют. Любое событие или мероприятие может весомо повлиять на ценность той или иной валюты [1].

В современном мире трудно классифицировать события как влияющие и не влияющие на колебание курса валют. Провести зависимость между событием и изменением положения валюты возможно только при подробном анализе подобных случаев, которые уже произошли [2]. Реализовать множество статистических вычислений поможет программа Rstudio [3-4]. По сути, чтобы предсказать будущее с весомой вероятностью, нужно проанализировать прошлое [5]. Для этого актуально использовать интеллектуальный анализ данных. Также это называется анализом временных рядов [5]. Исследуя статистику прошедших событий с помощью инструментов интеллектуального анализа, специалист, разбирающийся в предметной области, может составить прогноз с определенной вероятностью [6].

Использование данной разработки актуально в сфере экономики, международного ведения бизнеса, инвестиций. Существует множество отраслей, где будет полезно использование анализа колебаний валют.

Помимо данного метода используются: теория паритета покупательной способности, принцип относительной экономической стабильности, построение эконометрической модели. Но данные методы являются разработкой специалистов экономической направленности и в интеллектуальному анализу данных относятся лишь косвенно.

Целью данной работы является просмотр наиболее точных инструментов для анализа и для проведения и оценки зависимости с помощью RStudio.

2. Прогнозирования курса доллара с применением анализа временных рядов

На протяжении многих лет мировая экономика прочно привязана к такой мировой валюте как доллар. Наиболее актуальное использование знаний экспертов может предсказать поведение курса этой валюты на незначительный срок с довольно высокой вероятностью. В данном разделе для прогнозирования курса доллара будет проводиться с помощью интеллектуального анализа данных. Временной отрезок будет взят из недалекого прошлого для сравнения результата прогнозирования с реальной картиной.

Во многих областях деятельности необходимы периодические замеры каких-либо показателей, для выявления градации, деградации или отсутствия изменения значения показателя. Между этими замерами проходит определенный интервал времени, который с каждым последующим замером остается неизменным. Его называют интервалом выборки (sampling interval). А образующийся в результате выборки ряд данных значений измерений называют временным рядом (time series).

Во временных рядах выделяют неслучайную (детерминированную) и случайную компоненты. Неслучайная компонента дает возможность проверить влияние тех или иных факторов на изменение значения измерения. Математическая модель неслучайной компоненты используется для прогнозирования поведения временного ряда.

Для составления временного ряда, проведения анализа и составления прогноза используется RStudio. Данные о состоянии курса доллара рассматривались в период с 01.06.2011 по 18.07.2011. Для построения модели временного ряда общего числа курсов доллара используется метод ARIMA (авторегрессия интегрированного скользящего среднего).

Модель ARIMA наиболее подходящий для данной ситуации инструмент построения временного ряда, так как данная модель позволяет составлять прогноз, зависящий как от текущего и прошлого значения зависимой переменной, так и от текущих и прошлых значений величины случайного возмущения. В нашем случае курс доллара имеет и значение зависимой переменной, и значение случайного возмущения. Подбор ARIMA модели для ряда наблюдений за курсом доллара состоит из 4 этапов:

- идентификация модели;
- оценивание модели;
- тестирование модели;
- использование модели для прогнозирования.

Аналогичные инструменты для анализа временных рядов:

- ECM - модель корректировки отклонений;
- VAR - векторная авторегрессионная модель;
- VECM - векторная модель корректировки отклонений.

В отличие от всех вышеперечисленных методов модель ARIMA на практике показывает наиболее точный и наиболее вероятный прогноз и меньше подвержен вероятности ошибки прогноза. Все эти плюсы обосновываются статической составляющей рассматриваемого множества значений. В случае изменения данных в процессе их обработки, данная модель не сможет произвести адекватного прогноза. В этом случае рекомендуется использовать метод ECM.

Каждый шаг будет описан запросом на языке R по следующему алгоритму:

1. Преобразование данных о курсе доллара во временной ряд, с учетом разделения десятичных значений запятой и предположительным периодом измерения курса - раз в 7 дней. Фрагмент кода в RStudio, преобразование данных о курсе доллара во временной ряд:

```
dollar1 <- read.table("dollar.txt", dec=",")$V3  
dollar <- ts(dollar1[1:56], frequency=7)
```

Информационный критерий Акаике (AIC) — критерий, применяющийся исключительно для выбора из нескольких статистических моделей. Разработан в 1971 как «an information criterion» («(некий) информационный критерий»).

2. Проверка коэффициентов AIC каждого значения курса доллара с помощью модели ARIMA изменяя третью переменную order(m). Результат работы кода представлен на рисунке 1. Проверка коэффициентов AIC:

```
for(m in 1:7){
  mm <- paste("model0", m, sep="")
  assign(mm, arima(dollar, order=c(0,0,m)))
  cat(paste(mm, ":", AIC(get(mm)), "n", sep=""))
}

model01: -67.8129312683548
model02: -80.9252194390681
model03: -82.7498433251648
model04: -82.4022042909309
model05: -84.5913013237983
model06: -83.0836200480987
model07: -82.2056122336345
```

Рисунок 1. Модели ARIMA, 2 шаг.

Наибольшее значение коэффициент принимает при модели *model05*. Значение m=5.

3. Проверка коэффициентов каждого значения курса доллара с помощью модели ARIMA, изменяя первую переменную *order(m)* (рисунок 4). Результат работы кода показан на рисунке 2. Фрагмент кода в RStudio, проверка коэффициентов каждого значения:

```
for(m in 0:5){
  mm <- paste("model", m,"05", sep="")
  assign(mm, arima(dollar, order=c(m,0,5)))
  cat(paste(mm, ":", AIC(get(mm)), "n", sep=""))
}

model005: -84.5913013237983
model105: -82.772885907062
model205: -81.8467316861655
model305: -79.9942752423242
model405: -83.3710277055226
model505: -80.7835758225641
```

Рисунок 2. Модели ARIMA, 3 шаг.

Наибольшее значение коэффициент принимает при модели *model005*. Значение m=0.

4. Проверка коэффициентов каждого значения курса доллара с помощью модели ARIMA, изменяя первую переменную *order(m)* (рисунок 6). Результат работы кода показан на рисунке 3. Фрагмент кода в RStudio, изменение первой переменной *order(m)*:

```
for(m in 0:5){
  mm <- paste("model0", m,"5", sep="")
  assign(mm, arima(dollar, order=c(0,m,5)))
  cat(paste(mm, ":", AIC(get(mm)), "n", sep=""))
}

model005: -84.5913013237983
model015: -78.0533071948326
model025: -69.4383206787167
model035: -58.0629434485615
model045: -36.173674031108
model055: -15.1052746809435
```

Рисунок 3. Модели ARIMA, 4 шаг.

Наибольшее значение коэффициент принимает при модели *model005*. Значение m=0. Модель *model005* имеет минимальный AIC, поэтому она будет использоваться для прогнозирования.

5. Построим график предсказания, добавив туда реальные значения курса доллара с 27 июля по 9 августа распределяя одно значение на одну неделю (рисунок 4). Фрагмент кода в RStudio, построение графика предсказания:

```
plot(dollar, xlim = c(1,11), ylim = c(27.3, 28.5))
```

```
lines(predict(model005, n.ahead = 14, se.fit =TRUE) $pred, lty=2)
lines(ts(dollar1[56:70], start = 9, frequency = 7))
```

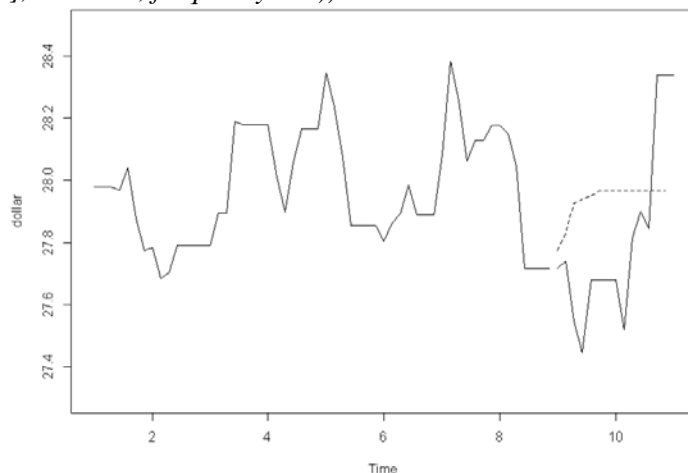


Рисунок 4. Прогноз изменения курса доллара (пунктирная линия), ось абсцисс - номера недели начиная с 1 июля 2011 года.

Лучшим результатом работы будет сравнение фактического значения курса доллара с спрогнозированным. Прогноз не определил резкого возрастания доллара 29 июля, но показал возрастание курса в последующие две недели (рисунок 10).

09.08.2011	28.5210	+0.1828
06.08.2011	28.3382	+0.4950
05.08.2011	27.8432	-0.0564
04.08.2011	27.8996	+0.0842
03.08.2011	27.8154	+0.2950
02.08.2011	27.5204	-0.1592
30.07.2011	27.6796	+0.0889
29.07.2011	27.5907	+0.1468
28.07.2011	27.4439	-0.1021
27.07.2011	27.5460	-0.1953

Рисунок 5. Фактическое значение курса доллара в период с 27 июля до 9 августа 2011.

3. Классификация валюты по отношению к мировому рынку

Существенной характеристикой любой валюты в мире является ее стабильность на мировом рынке. Множество факторов влияют на стабильность валюты. Для примера можно привести сравнение резервной и доминирующей валюты. Каждая из них имеет свое место в мировом обороте, но показатель стабильности очень размыт. Валюты по степени стабильности классифицируют на: твердую валюту, мягкую валюту, среднюю валюту.

Соответственно, твердая валюта означает стабильное положение курса на протяжении долгого времени. Мягкая - отсутствие стабильности. Средняя - показатель стабильности нельзя отнести ни к мягкой ни к твердой валюте. В данном разделе будет взят набор данных с числовым показателем влияния валюты на мировой рынок, числовым показателем изменения курса валюты за месяц, процент пользования валютой на мировом рынке и будет проведена классификация этих значений по вышеперечисленным классам валют. Валюты может быть мягкой, твердой или средней. Для актуальности информации, данные будут взяты с временного периода финансового кризиса 2008 года. Название валюты до классификации рассматриваться не будет.

Результаты работы программы представлены на рисунка 6-7. В результате анализа данных мы получили: $\epsilon = 0.233945$.

В результате данные были разделены на обучающую выборку и контрольную выборку, затем была произведена классификация объектов каждой выборки, и вероятность ошибочной классификации составила 23%.

В данном случае на рисунке 6 и на рисунке 7 белым цветом отображены объекты валюты, которые являются мягкими, фиолетовым цветом выделены средние и синим твердые. Зеленым цветом выделены валюты не относящиеся ни к одной из вышеперечисленных групп. Проведя классификацию по своим классам (рис. 1) и по классам объектов заданными программой (рис. 2), в результате получаются несколько разные классы некоторых объектов. Данные случаи исследуются специалистами в этой области, но также стоит отметить, что анализ подобного рода наглядно показывает отношение всех объектов к какому либо классу, в зависимости от его характеристик. Следовательно, взяв один из объектов, найдя его на графике результата классификации, по его цвету возможно сказать о его стабильности, с вероятностью ошибки 23%.

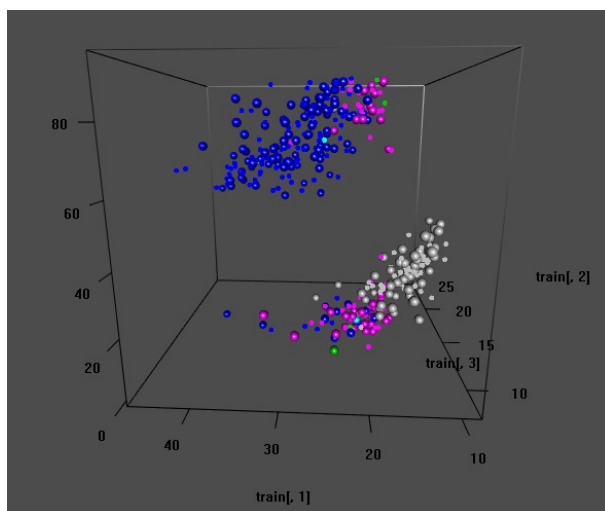


Рисунок 6. Объекты из контрольной выборки со своими настоящими классами.

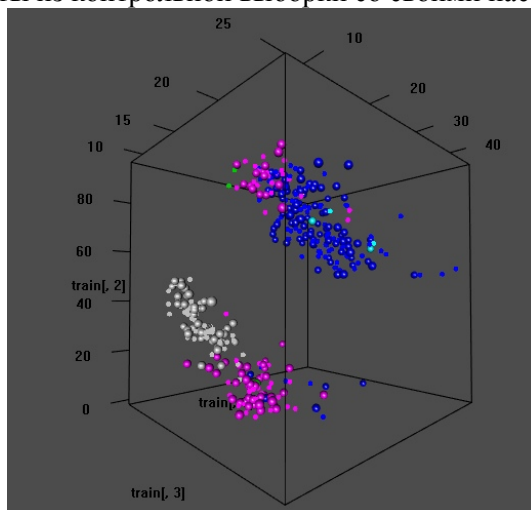


Рисунок 7. Объекты из контрольной выборки с классами, назначенными классификатором.

4. Заключение

Реализация деятельности в отрасли изучения валюты напрямую связана со статистикой. Возможность учесть все факторы, исключить человеческий фактор, трезво оценить значимость каждого фактора и многие другие функции возможно реализовать с помощью интеллектуального анализа данных. В данной работе был отображен минимум возможностей

этого инструмента. В будущем наиболее благоприятным для изучения предметом может стать совмещение использования этих инструментов. Гибриды подобного рода могут автоматизировать работу специалистов, при этом повысив качество выполнения работы за счет отсутствия человеческого фактора.

5. Литература

- [1] Алексеев, И.А. Современный экономический словарь: справочник / И.А. Алексеев, Е.С. Григорьев. – М.: Инфра-М, 2005. – 1096 с.
- [2] Чубукова, И.А. Data Mining: учебное пособие. – Казань: Издательство университета информационных технологий БИНОМ, 2006. – 382 с.
- [3] Michalski, S. Machine Learning / S. Michalski, G. Carbonell, M. Mitchell // Opt. Tioga Publishing Comany. – 1983. – P. 270-275.
- [4] Verzani, J. Getting started with RStudio // O'Relly Media, 2011. – 99 p.
- [5] Зиновьев, А.Ю. Визуализация многомерных данных: учебное пособие // Красноярск: Издательство Красноярского государственного технического университета, 2000. – 180 с.
- [6] Mitchell, T. Machine Learning. – London: McGraw-Hill., 1997. – 255 p.

The predict currency fluctuations using data mining techniques

I.R. Sharapov¹, D.A. Savelyev^{1,2}

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Abstract. Data mining methods can be used in many industries, in particular, in the economy. This paper explores the features of the application of data mining methods for classifying currencies according to the degree of stability in order to simplify the forecast of exchange rate fluctuations.