

Финансовые временные ряды: прогнозирование и распознавание нарушений динамики

Демин А. В.¹, Витяев Е. Е.²

¹Институт систем информатики СО РАН, пр. Лаврентьева, д. 6, г. Новосибирск, 630090, Россия.

²Институт математики СО РАН, пр. ак. Коптюга, д. 4, г. Новосибирск, 630090, Россия.

vityaev@math.nsc.ru, alexandredemin@yandex.ru

Аннотация. В настоящее время серьезной проблемой при прогнозировании финансовых временных рядов, а также других нестационарных процессов, является определение применимости предиктора к текущей динамике ряда. Для решения этой проблемы в работе предлагается новая технология прогнозирования, суть которой заключается в том, что параллельно с обнаружением закономерностей, предсказывающих будущую динамику ряда, также обнаруживаются высоковероятные закономерности, описывающие поведение ряда на обучающем интервале. По нарушениям этих закономерностей далее определяется, насколько текущая ситуация соответствует обучающему интервалу и можно ли в данной ситуации делать прогноз. Преимущества новой технологии демонстрируются на примере предсказания курсов акций ГМК «Норильский Никель». Приводится описание прототипа программного комплекса «FinDiscovery», реализующего предложенную технологию.

Ключевые слова: обнаружение закономерностей, предсказание, извлечение знаний, интеллектуальный анализ данных

1 Введение

Задача прогнозирования финансовых временных рядов была и остается актуальной, поскольку предсказание является необходимым элементом любой инвестиционной деятельности, ведь сама идея инвестирования – вложения денег с целью получения дохода в будущем – основывается на идеи прогнозирования будущего. В последнее время, когда стали доступны мощные средства сбора и обработки информации, задача прогнозирования финансовых временных рядов также становится и одной из самых популярных задач для практического применения различных Data Mining методов [1]. Широкое применение Data Mining методов в данной области обусловлено наличием в большинстве временных рядов сложных закономерностей, которые не обнаруживаются линейными методами.

Однако практическое применение Data Mining методов для прогноза финансовых временных рядов сталкивается с проблемой нестационарности этих рядов. Закономерности развития финансовых рядов постоянно меняются, причем, как показывает практика, эти изменения могут происходить очень быстро. В результате успех или неудача при использовании прогнозирующих систем, обученных на исторических данных, во многом зависит от того, будет ли будущая динамика ряда соответствовать той динамике, которая была на обучающем множестве, или нет.

Зачастую для преодоления нестационарности финансовых данных предлагается регулярно переобучать прогнозирующие системы по мере поступления новых данных. Однако очевидным недостатком данного метода является его запаздывание: к тому моменту, когда в обучающем множестве накопится достаточное количество новых данных, чтобы новая динамика ряда нашла отражение в закономерностях системы, эта динамика может либо уже закончиться, либо система может успеть дать неприемлемо большое количество ложных прогнозов. Таким обра-

зом, возникает проблема того, как определить, применима ли прогнозирующая система к текущей ситуации или нет.

Для решения этой проблемы нами предлагается использовать реляционный подход к извлечению знаний [1-2]. В рамках данного подхода нами была разработана технология предсказания финансовых временных рядов, использующая механизм проверки применимости полученных прогнозов к текущей ситуации. Технология заключается в том, что параллельно с обнаружением закономерностей, предсказывающих будущие значения ряда, обнаруживать также высоковероятные закономерности, описывающие динамику ряда на обучающем интервале, чтобы в дальнейшем, по нарушениям этих закономерностей, определять, соответствует ли текущая ситуация обучающему интервалу или нет. Мы исходим из предположения, что высоковероятные закономерности, описывающие динамику ряда, определяют его нормальное состояние, а нарушение этих закономерностей – аномальное: резкую смену тренда при выходе каких-либо новостей, обвал на рынке и т.д. Если текущая ситуация не соответствует обучающему интервалу, т.е. если произошло нарушение в динамике ряда, то следует воздержаться от применения закономерностей, предсказывающих будущие значения ряда, поскольку при их обнаружении такие ситуации не были учтены и велика вероятность получения ложного прогноза.

Важной особенностью предлагаемого подхода является возможность обнаружения таких признаков нарушения динамики, которые не встречались ранее на обучающей выборке. Для обнаружения признаков нарушения динамики не требуется наличия обучающего материала, т.е. известных примеров нарушений в динамике ряда, как это требуется существующими Data Mining методами.

2 Метод обнаружения вероятностных закономерностей

Предлагаемый нами метод позволяет для заданного класса гипотез Th обнаружить на данных множество вероятностных закономерностей в виде логических правил вида

$$A_i \& \dots \& A_n \rightarrow A_0, \quad (1)$$

где, $A_i \in U(Th)$, $i = 0, \dots, n$ – литеры, $U(Th)$ – множество всех литер.

В работах [3-5] нами был предложен язык описания классов гипотез, который позволил разработать интерактивную систему задания классов гипотез, проверяемых на данных, и реализовать её в системе «Discovery».

Алгоритм поиска закономерностей основан на методологии семантического вероятностного вывода [1-2], который позволяет находить все вероятностные закономерности вида (1) с максимальной вероятностью предсказывающие целевую литеру A_0 . Отличительной особенностью семантического вероятностного вывода является использование понятия вероятностной закономерности [1-2], которое звучит следующим образом. Правило $A_i \& \dots \& A_n \rightarrow A_0$ является вероятностной закономерностью, если для любого правила $A_{i1} \& \dots \& A_{ik} \rightarrow A_0$, такого что $\{A_{i1}, \dots, A_{ik}\} \subset \{A_i, \dots, A_n\}$, условная вероятность $0 < p(A_0 | A_{i1} \& \dots \& A_{ik}) < p(A_0 | A_i \& \dots \& A_n)$.

Для реализации семантического вероятностного вывода был разработан алгоритм направленного перебора правил [3,5], позволяющий существенно сократить пространство поиска. Суть алгоритма заключается в последовательном наращивании условной части правил с проверкой выполнимости условий вероятностной закономерности. Сокращение перебора осуществляется за счет использования эвристики, которая заключается в том, что, начиная с момента, когда длина условной части правил достигает некоторой заданной величины, алгоритм начинает последовательно наращивать условные части только тех правил, которые являются вероятностными закономерностями.

Вероятностные неравенства, входящие в определение вероятностной закономерности, проверяются на данных с использованием статистических критериев [4].

3 Метод обнаружения нарушений динамики временного ряда

Для обнаружения нарушений динамики временного ряда нами был разработан следующий метод, основанный на идее обнаружения аномальных событий [6-7].

1. Сформулировать вид гипотез Th , описывающий нормальную динамику временного ряда.

2. На анализируемом временном периоде обнаружить множество вероятностных закономерностей REG для заданного вида гипотез Th , предсказывающих все возможные литеры из $U(Th)$. Для этого необходимо последовательно брать каждую литеру из $U(Th)$ в качестве целевой и при помощи описанного выше метод находить множество предсказывающих её закономерностей.

3. Среди обнаруженного множества закономерностей REG отобрать множество высоковероятных закономерностей REG' , описывающих нормальную динамику временного ряда. Для этого необходимо зафиксировать некоторый порог T , $0 < T \leq 1$, к примеру, $T = 0.9$. Если $p(R) > T$, где R – правило из REG , то добавить правило R в множество REG' .

4. Сформировать множество закономерностей REG'' , описывающих нарушения динамики временного. Для этого необходимо для каждого правила $R = A_1 \& \dots \& A_m \rightarrow A_n$ из REG' взять его отрицание $\neg R = A_1 \& \dots \& A_m \rightarrow \neg A_n$. Конъюнкция $A_1 \& \dots \& A_m \& \neg A_n$ будет являться признаком нарушения динамики ряда, а правило $R' = A_1 \& \dots \& A_m \& \neg A_n \rightarrow Abnormal$ будет прогнозировать нарушение динамики с вероятностью $p(R)$. Полученное правило R' необходимо добавить в множество REG'' .

4 Технология прогнозирования финансовых временных рядов

Используя описанные выше методы, нами предлагается следующая технология прогнозирования финансовых временных рядов.

1. Сформулировать вид гипотез Th_1 . На обучающем временном периоде D обнаружить множество вероятностных закономерностей REG_1 заданного вида гипотез Th_1 , предсказывающих дальнейшую динамику временного ряда, к примеру, увеличение или уменьшение цены, либо достижение ценой определенного уровня и т.п. Полученное множество закономерностей REG_1 будет составлять основу прогнозирующей системы.

2. Сформулировать вид гипотез Th_2 , описывающий динамику временного ряда на обучающем периоде D . На том же обучающем временном периоде D обнаружить множество закономерностей REG_2 , заданного вида гипотез Th_2 , прогнозирующих нарушения динамики временного ряда.

3. При прогнозировании использовать предсказания, полученные по закономерностям REG_1 только в том случае, если степень отклонения в текущей динамике ряда меньше заданного порога, в противном случае считать, что текущая динамика ряда не соответствует обучающему периоду, и прогноз отсутствует. Степень отклонения в динамике ряда рассчитывается по прогнозам закономерностей REG_2 следующим образом $\delta = (\sum_{R \in REG_2^+} p(R)) / (\sum_{R \in REG_2} p(R))$, где

$REG_2^+ \subseteq REG_2$ – множество закономерностей из REG_2 , условная часть которых выполнена в текущей ситуации.

5 Пример разработки торговой системы

Продemonстрируем преимущество описанной технологии предсказания финансовых временных рядов на примере разработке торговой системы, предсказывающей курс акций ГМК «Норильский Никель». На рисунке 1 приведен график, отражающий динамику цен закрытия часа акций ГМК «Норильский Никель» на фондовой бирже ММВБ за период с 23 августа 2006 г. по 27 февраля 2009 г. Приведенные данные интересны тем, что достаточно продолжительный возрастающей тренд заканчивается резким обвалом курса в ноябре 2007 г. с последующей неоднократной сменой динамики ряда.

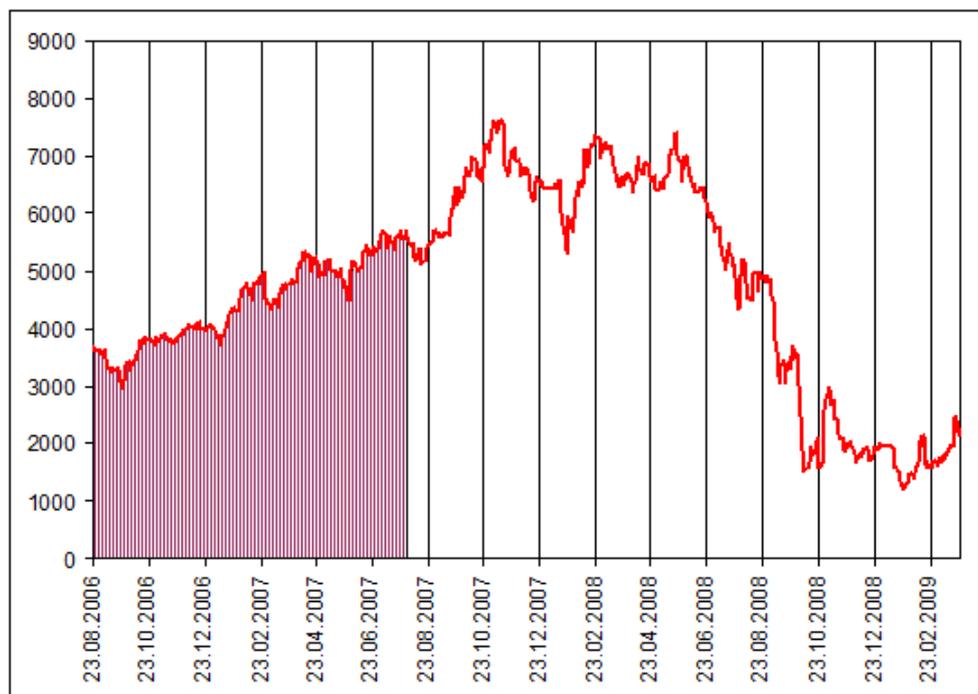


Рис.1. Динамика цен ГК «Норильский Никель».

На примере данного финансового инструмента покажем, каким образом использование технологии обнаружения признаков нарушения динамики ряда позволяет уменьшить количество ложных прогнозов. Для этого мы специально выбрали обучающее множество, включающее период возрастающего тренда с 23 августа 2006 г. по 31 июля 2007 г. (выделенный на Рис. 1 штриховкой). Построим на данном обучающем периоде две торговые системы: первую – без использования указанной технологии, а вторую – с использованием. Затем посмотрим, как поведут себя обе системы после 31 июля 2007 г. при резкой смене динамики временного ряда.

Основой обеих торговых систем являются вероятностные закономерности, обнаруженные на обучающем периоде и предсказывающие рост или падение стоимости акций на пять часов вперед. Для обнаружения закономерностей нами был использован простой класс гипотез Th , сравнивающий между собой цены закрытия последних пяти часов с учетом текущего времени и дня недели.

Первая торговая система основывается на следующем простом правиле открытия позиций. Если закономерности прогнозируют рост цены, то предлагается открыть позицию на покупку и держать ее пять часов. Если закономерности прогнозируют падение цены, то надо открыть позицию на продажу и держать ее пять часов. Если при этом ранее была открыта противоположная позиция, то надо её предварительно закрыть, а если была уже открыта аналогичная позиция, то надо продлить срок её удержания ещё на пять часов.

Результаты работы первой торговой системы на периоде с 1 августа 2007 г. по 27 февраля 2009 г. приведены на Рис. 2 и в Таблице 1. На Рис. 2 приведен график доходности системы в сравнении с доходностью при пассивном инвестировании (стратегия «купил и держи»), а в Таблице 1 представлены показатели финансовой эффективности системы. При детальном анализе графика доходности можно заметить, что система хорошо работает на периодах роста цены, т.е. когда динамика ряда соответствует динамике обучающего периода, и терпит убытки на периодах падения и бокового движения цен.

Для второй торговой системы мы, при помощи метода обнаружения нарушений динамики, для указанного обучающего периода нашли множество правил, прогнозирующих нарушение динамики ряда. Для обнаружения этих правил мы, для простоты, использовали тот же самый класс гипотез Th . Далее, в соответствии с описанной выше технологией прогнозирования, мы использовали прогнозы о росте или падении цен для открытия позиций только в тех случаях, когда степень отклонения динамики ряда была меньше заданного порога. В остальном правила открытия и закрытия позиций были такими же, как и у первой системы.

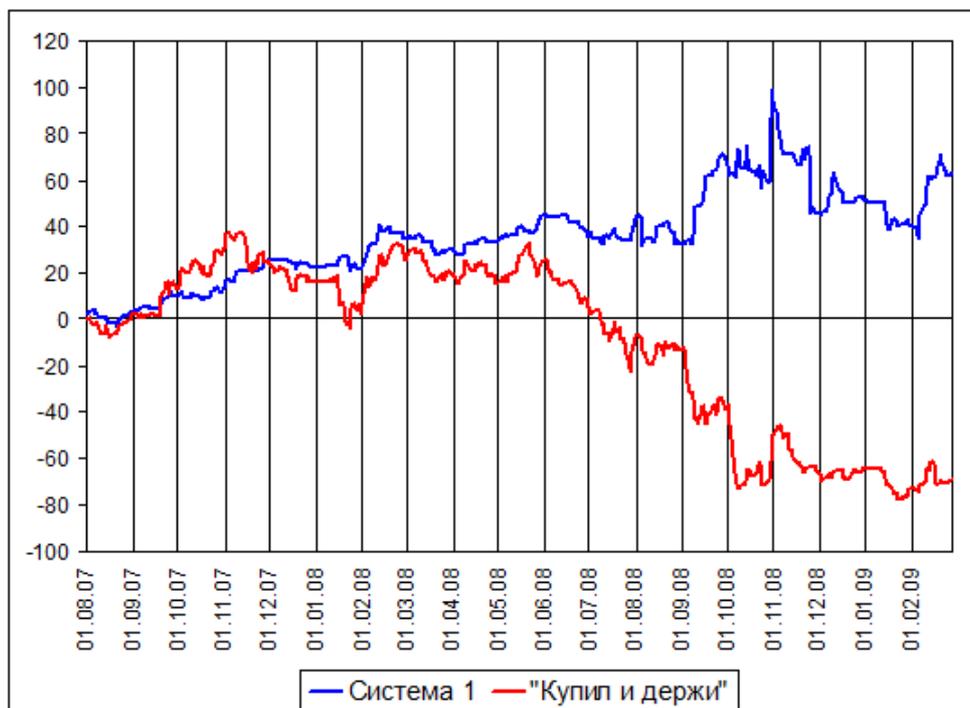


Рис.2. Доходность первой системы в сравнении со стратегией «купил и держи».

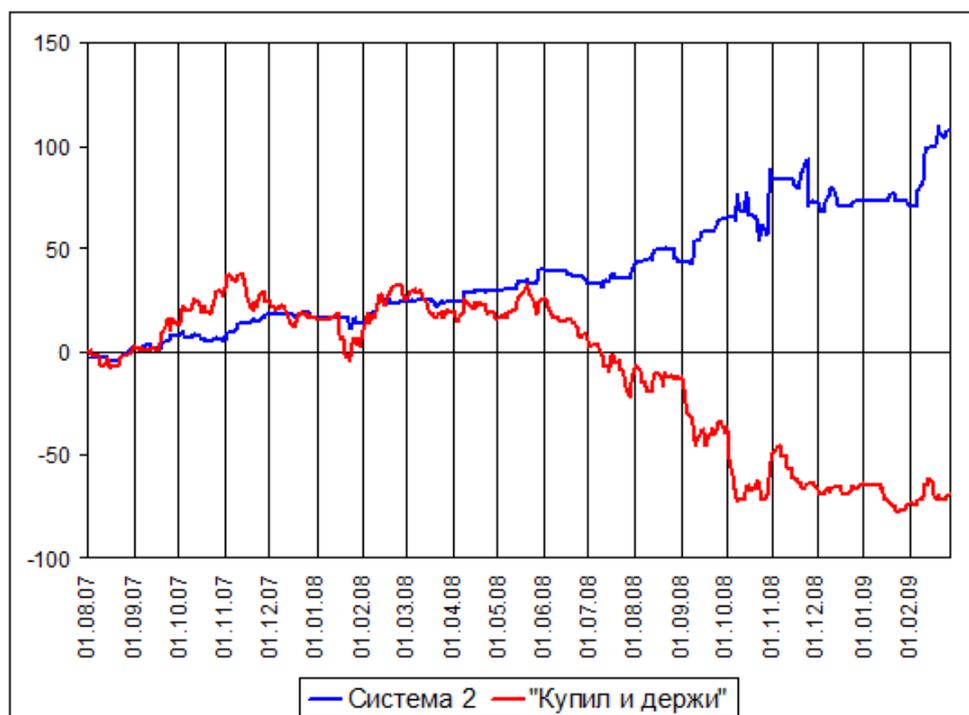


Рис.3. Доходность второй системы в сравнении со стратегией «купил и держи».

Результаты работы второй торговой системы на периоде с 1 августа 2007 г. по 27 февраля 2009 г. приведены на Рис. 3 и в Таблице 1. Как видно по результатам работы, использование технологии обнаружения признаков нарушения динамики ряда позволило отфильтровать ложные сигналы, что привело существенному росту процента прибыльных сделок и сократило мак-

симальную просадку счета почти в два раза. График доходности системы также стал более плавным и стабильным.

Таблица 1. Сравнение финансовых показателей торговых систем.

<i>Показатель</i>	<i>Система 1</i>	<i>Система 2</i>
Доходность	63.44 %	108.27 %
Процент прибыльных сделок	51.91 %	59.56 %
Максимальная просадка счета	-41.79 %	-25.69 %

Приведем несколько примеров обнаруженных закономерностей, соответствующих нормальной динамике ряда на обучающем периоде, и полученных на их основе правил, предсказывающих нарушение этой динамики.

1. Закономерность нормальной динамики:

$$(Hour(t) = 14) \& (C(t-5) < C(t)) \rightarrow (C(t-5) < C(t-1)), p = 0.975 .$$

Правило, предсказывающее нарушение:

$$(Hour(t) = 14) \& (C(t-5) < C(t)) \& \neg(C(t-5) < C(t-1)) \rightarrow Abnormal, p = 0.975 .$$

2. Закономерность нормальной динамики:

$$(C(t-5) < C(t-2)) \& (C(t-1) < C(t)) \rightarrow (C(t-5) < C(t)), p = 0.952 .$$

Правило, предсказывающее нарушение:

$$(C(t-5) < C(t-2)) \& (C(t-1) < C(t)) \& \neg(C(t-5) < C(t)) \rightarrow Abnormal, p = 0.952 .$$

3. Закономерность нормальной динамики:

$$(DayOfWeek(t) = 2) \& (C(t-5) < C(t)) \rightarrow (C(t-5) < C(t-2)), p = 0.836 .$$

Правило, предсказывающее нарушение:

$$(DayOfWeek(t) = 2) \& (C(t-5) < C(t)) \& \neg(C(t-5) < C(t-2)) \rightarrow Abnormal, p = 0.836 .$$

Здесь t – текущий момент времени, $C(t)$ – значение временного ряда в момент времени t , $Hour(t)$ – текущий час, $DayOfWeek(t)$ – текущий день недели (1 – понедельник, 2 – вторник и т.д.), p – условная вероятность правила.

В полученных закономерностях можно проследить следующую тенденцию в поведении ряда: если ряд растет на некотором промежутке времени, то он растет на нем достаточно непрерывно. Эта особенность поведения полностью соответствует растущему тренду, который наблюдается на всем обучающем периоде. В тоже время максимальная частота нарушения этих закономерностей наблюдается на отрезках бокового тренда, т.е. как раз тогда, когда поведение ряда существенно отличается от поведения на обучающем интервале.

6 Программный комплекс «FinDiscovery»

Проведенные эксперименты, а также полученные ранее результаты применения реляционного подхода в финансах [1-3] свидетельствуют о возможностях успешного использования предложенной технологии прогнозирования финансовых временных рядов для нужд реальной торговли. Возможности современного интернет-трейдинга позволяют в режиме реального времени получать данные о ходе торгов на мировых биржах. Это делает перспективным разработку программной системы, реализующей предложенную технологию и позволяющей в режиме реального времени получать прогнозы о движении цен и рекомендации о совершении сделок.

В настоящий момент нами разработан прототип программного комплекса «FinDiscovery», который является адаптацией системы извлечения знаний «Discovery» [3,5], специально предназначенной для анализа финансовых временных рядов (Рис. 4). Программный комплекс

«FinDiscovery» дает возможность пользователю в интерактивном режиме задавать класс гипотез для обнаружения закономерностей на финансовых данных, осуществлять поиск закономерностей для заданного класса и использовать найденные закономерности для разработки торговых систем. Разработанные пользователем торговые системы могут быть использованы при торговле в реальном времени. Для этого в программном комплексе имеется возможность подключения к торговому терминалу QUIK, что позволяет в режиме реального времени получать данные о ходе торгов на биржах ММВБ и РТС. Полученные сигналы об открытии или закрытии позиций могут быть либо разосланы в виде рекомендаций клиентам торговой системы, либо отправлены непосредственно в терминал QUIK в виде команды на совершение сделки. Во втором случае программный комплекс превращается в полностью автоматизированную торговую систему, совершающую сделки без участия человека. Подобные автоматизированные торговые системы обладают рядом очевидных преимуществ над обычными участниками рынка: неутомляемость, неподверженность эмоциям, более высокая скорость реагирования и принятия решений.

Index	<DATE>	<TIME>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>	Prediction	TextSignals	Signals
0	19990112	150000								
1	19990115	120000								
2	19990118	110000								
3	19990118	120000								
4	19990118	140000								
5	19990118	150000								
6	19990118	160000								
7	19990119	130000								
8	19990121	110000								
9	19990125	130000								
10	19990126	150000								
11	19990128	160000								
12	19990129	110000	16.1	16.1	16	16	2100	0	0	0
13	19990201	160000	17.9	18.8	17.9	18.8	4200	0	0	0
14	19990202	110000	18.91	18.91	18.91	18.91	1000	0	Close Long	0
15	19990202	120000	19.5	19.8	19.5	19.8	1600	0	0	0

MTS	Real-time	Send signals	Signal filter	Position	Position rate	Month rate
GAZP	No	No	None	Short	0.2048 %	5.2602 %
SMKN	No	No	None	Short	0.2742 %	0.1899 %
PLZL	No	No	None	Long	0.8357 %	7.5334 %
ROSN	No	No	None	Short	0.8264 %	0.0222 %
SBER	No	No	None	Close	0 %	8.3152 %
SNGS	No	No	None	Short	0.4486 %	4.8661 %
TATN	No	No	None	Long	0.1977 %	8.7377 %

Рис.4. Программный комплекс «FinDiscovery».

7 Заключение

В данной работе мы на примере финансовых временных рядов показали, каким образом использование технологии обнаружения нарушений динамики временного ряда позволяет улучшить качество прогнозирования нестационарных процессов. Однако имеются и другие возможности для эффективного использования указанной технологии, а именно:

1. В качестве предвестника существенного изменения динамики ряда. Данная возможность имеет большое значение для определения начала кризисных явлений на финансовых рынках. Если частота и количество нарушений начинает существенно возрастать, то это может свидетельствовать об изменении закономерности развития процесса и служить предвестником критических явлений на финансовых рынках. Из-за различной информированности участников рынка реакция на события, способные существенно изменить динамику рынка, приходит не сразу и имеет лавинообразный характер. Используя предложенную технологию можно отследить начало этих изменений и раньше на них отреагировать.

2. Для выбора предиктора среди набора предикторов, который в наибольшей степени соответствует текущей динамики ряда. Предположим, что у нас имеется набор предикторов, обученных на различных участках временного ряда. Эти участки могут быть специально подобраны таким образом, чтобы каждый из них содержал вполне определенную динамику ряда. Для

прогноза выбирается тот предиктор, обучающий интервал которого более всех остальных соответствует текущей динамике ряда. Для определения соответствия можно, к примеру, использовать приведенный выше показатель степени отклонения в динамике ряда. Очевидно, что от выбранного таким образом предиктора можно ожидать более точного прогноза. В результате можно добиться существенно лучшего качества прогнозирования по сравнению с каждым отдельным предиктором из заданного набора.

8 Благодарности

Работа поддержана грантом РФФИ 08-07-00272-а; интеграционными проектами СО РАН №№ 47, 111, 119, а также Советом по грантам Президента РФ и государственной поддержке ведущих научных школ (проект НШ-335.2008.1).

Литература

- [1] Kovalerchuk B., Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid methods. – Kluwer Academic Publishers, 2000. – p.308.
- [2] Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. – Новосибирск: НГУ, 2006. – 293 с.
- [3] Демин А.В., Витяев Е.Е. Реализация универсальной системы извлечения знаний «Discovery» и ее применение в задачах финансового прогнозирования // Информационные технологии работы со знаниями: обнаружение, поиск, управление. – Новосибирск, 2008. – Вып. 175: Вычислительные системы. – С. 3-47.
- [4] Демин А.В., Витяев Е.Е. Метод предсказания в языке первого порядка // Информационные технологии работы со знаниями: обнаружение, поиск, управление. – Новосибирск, 2008. – Вып. 175: Вычислительные системы. – С. 57-88.
- [5] Демин А.В., Витяев Е.Е., Полоз Т.Л. Реализация универсальной системы извлечения знаний «Discovery» и ее применение в задачах медицинской диагностики // Труды Всероссийская конференция с международным участием «Знания – Онтологии – Теории», Том 1, Новосибирск, 2007. – С. 63–70.
- [6] Витяев Е.Е., Ковалерчук Б.К., Федотов А.М., Барахнин В.Б., Белов С.Д., Дурдин Д.С., Демин А.В. Обнаружение закономерностей и распознавание аномальных событий в потоке данных сетевого трафика // Вестник НГУ, Информационные технологии. – 2008. – Т. 6. – Вып. 2. – С. 57-68.
- [7] Kovalerchuk B., Vityaev E., Holtfreter R. Correlation of Complex Evidence in Forensic Accounting Using Data Mining // Journal of Forensic Accounting 1524-5586. – Vol.VIII. – R.T. Edwards, Inc., 2007. – pp. 53-88