

### 11.3. СЕНТИМЕНТНЫЙ АНАЛИЗ ВЛИЯНИЯ НОВОСТЕЙ НА ФОНДОВЫЕ РЫНКИ

Перминов Г.И., к.т.н., доцент кафедры  
бизнес-аналитики факультета бизнес-информатики;  
Ахмедова С.С., студент факультета  
бизнес-информатики

*Национальный исследовательский  
университет Высшая школа экономики*

[Перейти на Главное МЕНЮ](#)  
[Вернуться к СОДЕРЖАНИЮ](#)

В этом исследовании, рассматривается фондовый рынок, как игровое поле с большим количеством игроков. Мнение трейдера может меняться, когда он узнает, например, что компания выпустила новую линейку продуктов или когда компания меняет свою бизнес-стратегию, которая увеличивает прибыль. Новости является решающим источником такой информации, и добавление новой информации является основным фактором в мотивации движений на рынке. В работе исследуется возможность использования для анализа новостей и степени влияния на фондовые рынки сентимент-анализа с построением байесовского классификатора с корректировкой результатов по анализу таблицы сопряженности, также известной как матрица ошибок или случайностей. Использование предлагаемого метода анализа ленты новостей для выработки стратегии трейдера на реальном массиве показал работоспособность предлагаемого метода.

#### ВВЕДЕНИЕ

В этом исследовании рассматривается фондовый рынок как игровое поле с большим количеством игроков. Это модель рынка как аукциона, где игроки продают и делают предложения по некоторым активам. Стоимость актива в значительной степени определяется его текущими финансами и его потенциалом роста, но, к сожалению, информация такого рода не предоставляется по запросу. Чем больше информации есть у игроков, тем лучше они могут судить о возможностях актива и определить его текущее значение и потенциал для будущего роста этого показателя. Мнение трейдера может меняться, когда он узнает, например, что компания выпустила новую линейку продуктов или когда компания меняет свою бизнес-стратегию, которая увеличивает прибыль. Новости является решающим источником такой информации, и добавление новой информации является основным фактором в мотивации движений на рынке. Поэтому вопрос, в какой степени могут новостные данные влиять на рынки, является актуальной задачей.

В работе исследуется возможность использования для анализа новостей и степени влияния на фондовые рынки сентимент-анализа с построением байесовского классификатора с корректировкой результатов по анализу таблицы сопряженности, также известной как матрица ошибок или случайностей. Использование предлагаемого метода анализа ленты новостей для выработки стратегии трейдера на реальном массиве показал работоспособность предлагаемого метода.

#### Вопросы исследования

В работе исследуется технология и преобладающие методы анализа новостной ленты, в котором качественные, текстовые новостные данные превращается в количественные показатели настроения участников рынка. В связи с волатильностью на фондовом рынке колебания цен на основе настроений и репортажи являются общими.

Трейдеры опираются на самую разнообразную публично – доступную информацию для формирования своих решений. В рамках этой работы, анализируются публично доступные новости с использованием ком-

пьютеров для выработки рекомендаций для трейдеров для торговли на фондовых рынках.

Трейдеры могут использовать мощные компьютеры для скоростного чтения новостей, редакционных статей, веб-сайтов компаний, блогов и даже сообщений Twitter, а затем позволять делать машинные выводы относительно того, что все это означает для рынка.

В работе исследуется возможность применения для анализа новостей байесовского классификатора.

#### Методы исследования

Используются методы машинного обучения для обработки текстового ввода новостей, чтобы определить количественные оценки настроений – сентиментов. Рассматриваются различные типы доступных новостей и то, как они обрабатываются для формирования вклада в финансовые модели. Рассматриваются приложения использующие новости, для прогнозирования аномальных возвратов (доходов), для торговых стратегий, для диагностических целей, а также использование новостей для управления рисками.

По мере роста вычислительной мощности появляются технологии, которые позволяют эффективно извлекать, агрегировать и классифицировать большие объемы новостей. Такая технология может быть применена для количественного построения модели, как для высокочастотной, так и низкочастотной торговли на фондовых рынках.

Автоматизированный анализ новостей может стать ключевым компонентом движущим алгоритмическими торговыми подразделениями для создания стратегии и выполнения планов. Трейдеры, которые используют эту технологию, смогут сократить время, необходимое для реагирования на последние или экстренные новости (т.е. снизить время задержки в несколько раз). Технология новостной аналитики (**NA**) также может быть использована для оказания помощи в работе обычных менеджеров количественного фонда для мониторинга рыночных настроений в конкретных компаниях, брендах и секторах.

#### Результаты исследований

Основной идеей технологий **NA** является автоматизация человеческого мышления и рассуждения. Трейдеры, спекулянты и частные инвесторы выделяют свои ожидания относительно доходности активов, а также размера позиций и уровня неопределенности (волатильности), перед принятием инвестиционного решения. Они внимательно штудируют недавние экономические и финансовые новости, чтобы получить наиболее четкую картину текущей ситуации. Используя свои знания о том, как рынки вели себя в прошлом в различных ситуациях, люди будут косвенно сопоставлять текущую ситуацию с прошлыми, наиболее похожими на нынешнюю.

Новостная аналитика стремится ввести технологии для автоматизации или полуавтоматизации этого подхода. Благодаря автоматизации процесса суждения, человеческое звено принимающее решение, может действовать на больших, следовательно, более диверсифицированных наборах активов. Эти решения также принимаются более оперативно (снижение латентности). Автоматизация или полуавтоматизация процесса суждения человека расширяет пределы инвестиционного процесса [3] обозначает этот процесс, как интеллектуальная амплификация (**IA**).

Как показано на рис. 1, новостные данные являются дополнительным источником информации, которая может использоваться для улучшения (традиционного) инвестиционного анализа. Тем не менее, важно помнить, что **НА** в области финансов является многопрофильным полем, которое опирается на финансовую экономику, финансовый инжиниринг, поведенческую структуру финансов и искусственный интеллект (в частности, обработки естественного языка).

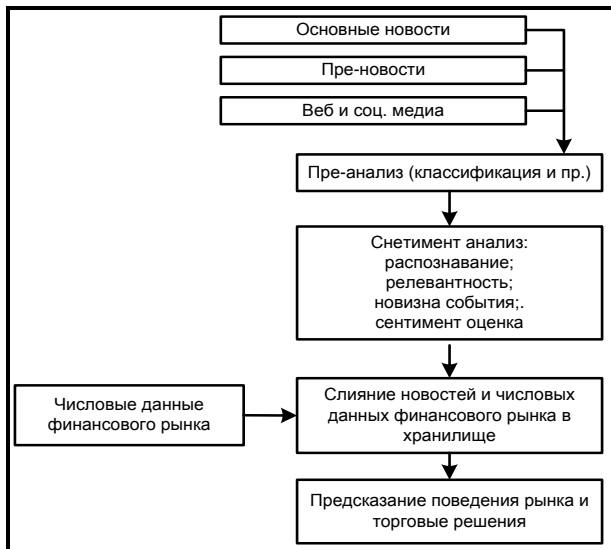


Рис. 1. Информационные потоки и архитектура моделирования

Сравнение точности байесовской модели в сравнении с обычным счетчиком новостей показало, что точность увеличилась на 9% (табл. 1).

Таблица 1

**СРАВНИТЕЛЬНАЯ ТОЧНОСТЬ КЛАССИФИКАТОРОВ**

Классификатор	Точность %
Счетчик слов	61,99
Наивная байесовская модель	70,63

Здесь под точностью подразумевается пропорция случаев в тестовом наборе, которые были классифицированы корректно.

**Состояние вопроса**

W.S. Chan [10] показал, что цены на акции с запаздыванием реагируют на поступающую информацию, из чего можно предположить, что новости могут обеспечить реальный способ для анализа финансовых рынков. Задача заключается в изучении взаимосвязи между данными фондового рынка и лингвистическими данными средств массовой информации, и проиллюстрировать, в какой степени они могут внести свой вклад в разработку инвестиционных стратегий.

Так же Ryan Timmons и Kari Lee [22] утверждают, что финансовые статьи и новости играют важную роль в оказании влияния на движение цен на акции, и есть некоторое отставание между тем, когда новость получает освещение в СМИ, и тем, когда рынок получает колебание, чтобы отразить эту информацию. Они построили систему классификации текстов для категоризации новостных статей, содержащих ссылки на спис-

ки компаний и предсказания воздействия новостей на котировки их акций.

В последнее время появился растущий объем научной литературы, которая утверждает, что масс-медиа влияет на настроения инвесторов, следовательно, цены на активы, и поэтому волатильность цен активов и риски напрямую от нее зависят [2, 8, 10, 12, 17, 22, 23]. Трейдеры и другие участники рынка усваивают и анализируют новости максимально быстро, пересматривая и восстанавливая равновесие своих позиций и активов соответственно.

Хотя актуальность использования новостей широко признана, вопрос о том, как включить их данные эффективно, в количественные модели и в целом в процесс принятия решений инвестиционного бизнеса, является открытым. При изучении того, каким образом новости оказывают свое влияние на рынок, Бэмбер, Баррон и Стобер [8] установили, что объем торгов увеличивается в дни с информационными релизами.

**Источники новостных данных**

Рассмотрим различные источники новостных и информационных потоков, которые могут быть применены для обновлений (количественных) инвесторской базы убеждений и знаний.

**1. Новости**

Новости относится к массовой информации и включают новости, изданные авторитетными источниками. Они транслируются через газеты, радио и телевидение. Они также предоставлены в распоряжение трейдеров сервисами ленты новостей. Электронные версии газет также постепенно растут в объеме и количестве.

**2. Предварительные новости**

Эти относятся к исходным данным, которые исследуются журналистами, прежде чем они пишут статьи новостей. Они происходят от первичных источников информации, таких как комиссии по ценным бумагам и биржам, их отчеты и заявки, судебные документы и сведения из государственных учреждений. Она также включает в себя потоки запланированных объявлений, таких как макроэкономические новости, выпуски отраслевой статистики, доклады о доходах компании и других корпоративных новостей.

**3. Web и социальная медиа**

Финансовые новости можно разделить на группы по синхронности и регулярности, т.е. ожидаемые (по четкому расписанию) и управляемые событиями асинхронные новости (незапланированные или неожиданные новости). Основные новости, слухи или информация полученная из социальной медиа, как правило, поступает в асинхронной, неструктурированной текстовой форме. Значительная часть предварительных новостей прибывает в заранее запланированное время и в целом в структурированном виде. Календарные (новости) или новостные сводки чаще всего имеют четко определенный числовой и текстовый контент и могут быть классифицированы как структурированные данные. К ним относятся объявления макроэкономические и о прибыли. Макроэкономические новости, в частности, экономические показатели крупнейших экономических сфер, широко используется в автоматизированной торговле. Они имеет влияние на крупнейшие и наиболее ликвидные рынки, такие как валютные и

фьючерсные. Фирмы часто осваивают сложные и быстрые торговые стратегии. Поскольку эти новостные события, как правило, хорошо документированы, то тщательная обработка тестирования стратегий является вполне возможной.

Чтобы эффективно применять новости в сфере управления активами и торговых решений, необходимо уметь идентифицировать новости, которые являются релевантными и актуальными. Это особенно важно для «внутридневных» приложений, где алгоритмы должны быстро реагировать на достоверную информацию. Задача состоит в том, чтобы идентифицировать информационное событие; т.е. различить статьи, которые ссылаются на старые новости (ранее опубликованные), от подлинно новых новостей. Как и следовало ожидать, P.C. Tetlock, M. Saar-Tsechansky и S. Macskassy [26] подтверждают, что рынок сильно реагирует именно на новые новости. Они же провели исследование событий, которое иллюстрирует влияние новостей на кумулятивные аномальные возвраты (CARs).

### Применение сентимент-анализа

В последние годы появилось большое количество работ по сентимент-анализу [5, 7, 9, 13, 14, 15, 19, 20].

В сентимент-анализе используются индикаторы настроений. Поэтому все работы можно сгруппировать в два направления:

- определение способов порождения и отбора признаков, описывающих временной ряд;
- включение отобранных признаков в модели финансовых временных рядов, по которым можно предсказывать их поведение.

#### 1. Определение признаков в виде положительных и отрицательных слов

Эти методы помогают для разделения событий на положительные и отрицательные для выбора портфеля, и выделения тех событий, которые имеют хорошие корреляции (с поправкой на риск) избыточные прибыли.

Самым общим методом сентимент-анализа является общий датчик информации. Он описывает подход к анализу настроений, который представляет каждую новость в виде двух чисел: количества положительных и количества отрицательных слов, упоминаемых в ней. Эти слова выбираются из словаря «General Inquirer» [3, 18], в котором перечислены 1915 положительных слов и 2291 негативных. Все слова, которых нет в этих списках, игнорируются. В своих работах P.C. Tetlock, M. Saar-Tsechansky и S. Macskassy [26] указывают на высокую корреляцию пропорций количества негативных и позитивных слов в статье с доходностью акций.

Pablo Daniel Azar [21] рассмотрел ряд статей и при помощи специально созданного словаря вел подсчет в них позитивных и негативных слов (рис. 2).

Каждая точка на рис. 2 представляет собой новость. Ось  $x$  – нормированная доля негативных слов (переменная  $neg$ ), ось  $y$  – нормированная доля положительных слов (переменная  $pos$ ). Круглые точки демонстрируют очень высокую, аномальную прибыль. Прямоугольные точки имеют очень низкий, аномальный доход. Аномальной доходностью тут называем только отклонения, связанные с новостными данными, в сущности, только то значение, которое не зависит от рисков в рыночных движениях и другой информации, не связанной с пресс мероприятиями. Это значение ежедневной доходности акций, которое не может быть

объяснено с помощью традиционных методов. Судя по этому графику, трудно придумать хорошее правило, чтобы отделить круглые точки от прямоугольных.

В работе Романенко А.А. [2] для прогнозирования поведения целевого ряда используется разметка временных рядов. Предлагается алгоритм порождения признаков из размеченных временных рядов и генетический алгоритм отбора признаков на размеченных временных рядах.

Аналогичный подход предлагается Кононенко Д.С. [1]. Алгоритм основан на разметке временного ряда на сегменты роста и падения и построения классификатора с помощью логистической регрессии. Отбор лучшей модели также производился генетическим методом.

#### 2. Построение модели поведения временного ряда с помощью новостей

Использование индикаторов настроений в финансовых новостях в основном строится на подсчете количества положительных и отрицательных слов. Но, как следует из рис. 2, такая мера индикаторов настроений не обеспечивает объяснение большей части дисперсии в аномальных отклонениях от доходности акций.

Метод подсчета положительных и отрицательных слов ограничен, потому что он отражает сложность новостей с использованием только двух атрибутов: числа положительных слов и числа отрицательных слов в тексте.

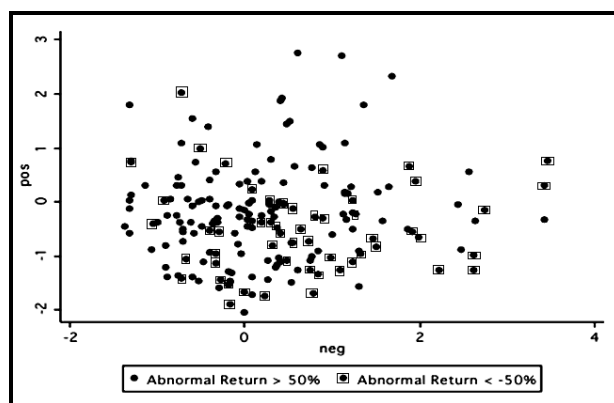


Рис. 2. Переменные  $pos$  и  $neg$  для различных статей

#### Постановка проблемы

Таким образом, рассматривая различные представления текстов, необходимо получить лучшее правило для классификации положительных и негативных новостей. В данной работе исследуется возможность перейти к разделению положительных и негативных новостей с помощью метода «наивный байесовский классификатор».

Для краткости описываем, как алгоритм принимает вход  $(x_1, \dots, x_n)$  и выводит классификации  $y$ . Для реализации алгоритмов использовалось программное обеспечение с открытым исходным кодом Weka 3.5.8 [6, 16]. Наивная байесовская модель, как и многие другие вероятностные модели, рассматривают атрибуты и классификацию экземпляра как случайные величины. Эта модель мира предполагает, что переменные связаны друг с другом, но только вероятностным образом. Если в тексте упоминается слово «катастрофа», существует высокая вероятность того, что общее настрое-

ние будет фиксировано как негативное. Но есть также небольшая вероятность того, что фон, в данном случае контекст, позитивен («катастрофа была предотвращена»). Что еще более важно, наличие или отсутствие данных слов позволяет нам обновить наши представления о настроениях текста. Пусть имеется пять атрибутов  $X_1, \dots, X_5$ , соответствующих наличию или отсутствию пяти различных слов. Обозначим соответствующую классификацию буквой  $C$ . Вероятностные методы имеют три ключевых компонента.

Предварительное убеждение о классификации текста. Оно представлено посредством распределения вероятностей  $P(C)$ .

Модель, которая сообщает нам вероятность наблюдения случая  $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ ; учитывая, что класс  $C = c$ . Эта модель выражается в виде распределения основной вероятности  $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | C = c)$ .

Можно использовать правило Байеса для обновления наших представлений, о классификации текстового сообщения используя слова, которые оно содержит:

$$P(C | X_1, X_2, \dots, X_n) = \frac{P(X_1, X_2, \dots, X_n | C)P(C)}{P(X_1, X_2, \dots, X_n)}$$

Прогнозируемый класс  $C$  – это класс с наибольшей вероятностью возникновения приведенных атрибутов.

Наивный Байес зарекомендовал себя как очень полезный алгоритм даже в ситуациях, когда предположение о независимости на определенных условиях не полностью оправдывается. Текстовая категоризация является одной из таких ситуаций.

Сравним точность байесовской модели в сравнении с обычным счетчиком новостей. Под точностью мы будем подразумевать пропорцию случаев в тестовом наборе, которые были классифицированы корректно. Так, например, если в тестовой выборке 100 случаев, 76 из которых были верно распределены по категориям, то его точность составит 76%.

Проблема указанной выше структуры в том, что нам необходимо абсурдно большое количество параметров, если мы не накладываем никаких других условий. Если атрибуты  $X_1, \dots, X_5$  являются бинарными, у нас есть:

2 в 5-ой степени = 32 возможных вариаций.

Каждый случай  $x_1, \dots, x_5$  соответствует двум параметрами структуры модели.

Первый задающий:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | C = 0)$$

Второй задающий вероятности:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | C = 1)$$

Число примеров, которым мы должны обучить эти переменные растет, по крайней мере, линейно с увеличением числа параметров. Таким образом, если бы мы имели 266 переменных, нам бы потребовалось больше примеров, чем атомов во Вселенной, чтобы узнать параметры. Это невозможно. Нам нужны предположения, которые упростят модель. Предположение Байеса состоит в том, что все атрибуты являются независимыми в данной классификации. Таким образом, мы можем расписать:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | C = 1)$$

как

$$P(X_1 = x_1 | C = 1) * P(X_2 = x_2 | C = 1) * \dots$$

$$* \dots * P(X_n = x_n | C = 1)$$

Используя это предположение, можем утверждать, что число параметров линейно по числу переменных  $n$ . Для переменной  $X_i$  нам необходимо знать значения:

$$P(X_i = x_i | C = c),$$

где  $x_i, c \in \{0, 1\}$ .

Так как  $x_i$  и  $c$  принимают только бинарные значения, то необходимо знать только четыре параметра.

В общей сложности это дает 20 параметров. Если бы мы имели 266 переменных, мы должны были бы знать  $266 * 4 = 1\,064$  параметров. Это намного меньше, чем 10 в степени 80 параметров, которые мы бы имели, не сделай мы байесовское предположение.

Наивный Байес зарекомендовал себя как очень полезный алгоритм даже в ситуациях, когда предположение о независимости на определенных условиях не полностью оправдывается. Текстовая категоризация является одной из таких ситуаций.

Следует отметить, что, когда мы используем термин «слово», мы имеем в виду корень слова. Таким образом, слова negative («отрицательный») и negation («отрицание») сводятся к стволу neg. Аналогично и с другими словами.

В модели, которая используется далее, – обычный наивный байесовский метод был дополнен так называемой таблицей сопряженности, также известной как матрица ошибок или случайностей. Она часто используется в методах классификации и представляет собой матрицу, в которой каждый столбец описывает предсказанный класс, а каждая строка – фактический.

Используя матрицу неточностей, можно отметить, откуда происходит большая часть ошибок прогнозирования: классификация положительных статей как отрицательных или негативных статей как позитивных. Это является общей проблемой обработки естественного языка. Для преодоления этой проблемы в алгоритме разработана система, которая при нахождении слова  $N$  из нашего лексикона будет проверять, есть ли любое из сохраненных в кэше слов-отрицаний вблизи  $N$ . Если оно будет найдено, алгоритм увеличивает значение.

### Практическая реализация предлагаемого метода

Практическая реализация вышеизложенных методов производилась на примере бесплатного сервиса ATH Sentiment Index [17]. Значения рассчитываются на онлайн сервисе, в который автоматически подгружаются новостные статьи известных источников.

Индекс ATH Sentiment предназначен для измерения текущего глобального сентимента. Индекс представляет собой отношение номинально положительных новостных статей в общем количестве опубликованных новостных статей, оцененных как положительно, так и отрицательно в пределах временного окна рассмотрения. Нейтральные статьи оцениваются, как ни положительные, ни отрицательные и не являются частью индекса.

Выбор, какую именно информацию использовать для анализа, является важным шагом любой задачи текстового анализа, которая требует глубокого понимания предмета. Мы сканируем ленту новостей из наиболее важных международных источников финансовых новостей, балансируя между необходимостью не упустить каких-либо серьезных событий и в то же время, избегая многой неактуальной информации. Каждый день сотни новостных статей из Европы, Китая, Индии, Японии и США анали-

зируются и вносятся в историческую базу. Фрагмент ленты новостей из сайта <http://www.banki.ru/news> ниже:

- 21 ноября 2013 г. 15:09. Греция будет принимать бюджет-2014 без одобрения кредиторов (лента новостей);
- 20 ноября 2013 г. 15:15. Фонды США сделали многомиллиардную ставку на акции банков еврозоны (мониторинг банковской прессы);
- 10:49. Бернанке обещает сохранять мягкую политику так долго, как потребуется (мониторинг банковской прессы);
- 08:10. Бернанке: ФРС остается приверженной стимуливанию экономики (лента новостей);
- 18 ноября 2013 г. 17:18. Европейский регулятор недоволен слишком малым числом ликвидированных банков (лента новостей);
- 12:29. Алексей Улюкаев: «Ничто больше не упадет с неба, за все нужно будет бороться» (мониторинг банковской прессы);
- 09:12. Результаты стресс-тестов банков еврозоны будут готовы осенью 2014 г. (лента новостей), и т.д.

Каждая новость сначала разбирается на отдельные слова и комбинации слов, например глагол + предлог или прилагательное + существительное. Исходя из содержания каждой статьи, наш алгоритм для автоматической маркировки вычисляет вероятности для оценки статьи как положительной и отрицательной соответственно. Если обе эти вероятности ниже определенного экспериментально порога, новость считается нейтральной. Наиболее релевантные слова для сентимент-анализа были взяты из работы [21].

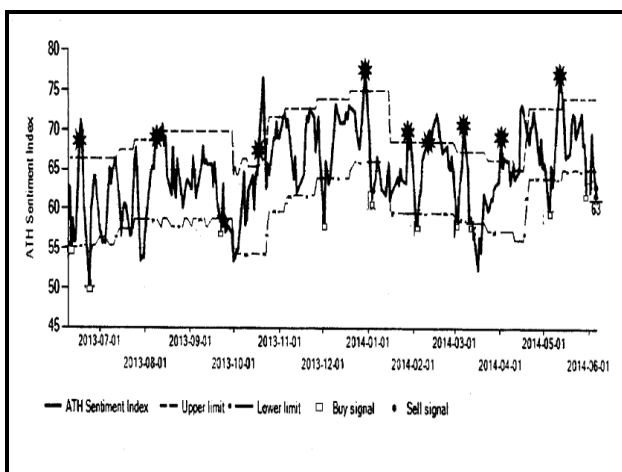


Рис. 3. График индекса ATH

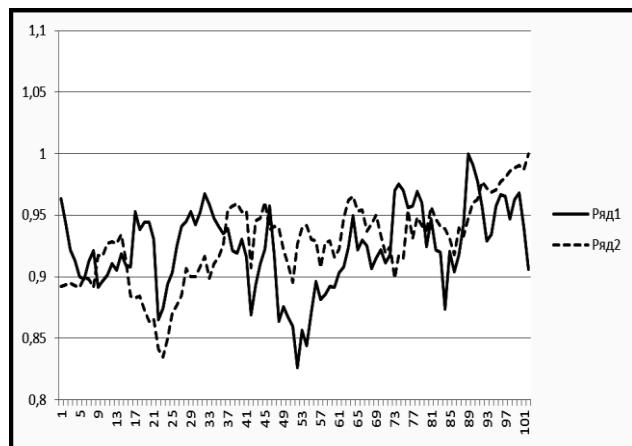


Рис. 4. График нормализованных значений сентимент-индекса ATH и индекса OMX30

Когда настроения в виде индекса **ATH** превышает верхний предел в первый раз, это считается сигналом к продаже. Когда настроения падает ниже нижнего предела – это сигнал к покупке (рис. 3).

Для того чтобы оценки правильности работы алгоритма, график колебаний индекса **ATH** сопоставлялся графику индекса OMX Stockholm30 (рис. 4). На графике рядом 1 обозначены преобразованные данные нашей новостной ленты, а рядом 2 – значения индекса OMX Stockholm30.

В табл. 2 приведен фрагмент использования предлагаемого метода для анализа ленты новостей для выработки стратегии трейдера на реальном массиве.

Таблица 2

**ФРАГМЕНТ ИСПОЛЬЗОВАНИЯ ПРЕДЛАГАЕМОГО МЕТОДА АНАЛИЗА ЛЕНТЫ НОВОСТЕЙ ДЛЯ ВЫРАБОТКИ СТРАТЕГИИ ТРЕЙДЕРА НА РЕАЛЬНОМ МАССИВЕ**

Дата входа	Вход	Дата выхода	Значение на входе	Значение на выходе	Результат, %
2014-05-06	Покупка	2014-05-13	1 358,25	1 368,25	0,7
2014-04-03	Продажа	2014-05-06	1 352,5	1 357	-0,3
2014-03-12	Покупка	2014-04-03	1 305	1 352,5	3,6
2014-03-10	Продажа	2014-03-12	1 311,5	1 305	0,5
2014-03-04	Покупка	2014-03-10	1 323	1 328	0,4
2014-02-12	Продажа	2014-03-04	1 309	1 323	-1,1 %
2014-02-04	Покупка	2014-02-12	1 249,5	1 309	4,8
2014-01-29	Продажа	2014-02-04	1 314	1 267,5	3,6
2014-01-07	Покупка	2014-01-29	1 319	1 314,5	-0,3
2014-01-02	Продажа	2014-01-07	1 319,25	1 319	0
2013-12-03	Покупка	2014-01-02	1 304,5	1 328,25	1,8%
2013-10-21	Продажа	2013-12-03	1 288,75	1 303	-1,1
Общий итог					8,8

Анализ результатов использования предлагаемого метода анализа ленты новостей для выработки стратегии трейдера на реальном массиве (табл. 2) показал работоспособность предлагаемого метода. Также стала ясна необходимости разработки собственного инструмента по классификации новостей с использованием русскоязычного словаря слов-индикаторов.

**Выводы**

Точность предложенных до сегодняшнего дня моделей по анализу новостей и их влиянию на финансовые рынки требует внедрения новых методик.

В работе сделана попытка извлечения сентиментов, соответствующих финансовым изменениям из источников новостей, и использования общих настроений рынка в качестве рыночного предсказателя. В качестве оценки влияния новостей на поведение рынка был использован байесовский классификатор с корректировкой результатов по анализу таблицы сопряженности, также известной как матрица ошибок или случайностей. Использование предлагаемого метода анализа ленты новостей для выработки стратегии трейдера на реальном массиве показал работоспособность предлагаемого метода.

**Литература**

1. Кононенко Д.С. Прогнозирование событий [Текст] / Д.С. Кононенко // Машинное обучение и анализ данных. – 2010. – Т. 1 ; №1. – С. 113-115.
2. Романенко А.А. Событийное моделирование и прогноз финансовых временных рядов [Текст] / А.А. Романенко // Машинное обучение и анализ данных. – 2011. – Т. 1 ; №2. – С. 236-243.
3. Acharya V.V., Pedersen L.H. Asset pricing with liquidity risk // Journal of financial economics. 2005. Vol. 77. Pp. 375-410.

4. An energizing list of positive words. <http://www.the-benefits-of-positive-thinking.com/list-of-positive-words.html>.
5. Antweiler W., Frank M.Z. Is all that talk just noise? The information content of internet stock message boards // Journal of finance. 2004. Vol. 3. Pp. 1259-1294.
6. Baker and Stein. Market liquidity as a sentiment indicator, 2004.
7. Bamber L.S., Barron O.E., Stober T.L. Trading volume and different aspects of disagreement coincident with earnings announcements // The accounting review. 1997. Vol. 72. Pp. 575-597.
8. Bautin M., Vijayarenu L., Skiena S. International sentiment analysis for news and blogs. In Second Int.Conf. on Weblogs and Social Media. 2008. (ICWSM 2008).
9. Brown R. Incorporating news analytics into algorithmic trading strategies: increasing the signal-to-noise ratio. In The handbook of news analytics in finance. 2011. Chapter 14. John Wiley & Sons.
10. Cahan R., Jussa J., Luo Y. Breaking news: how to use sentiment to pick stocks // MacQuarie research report. 2009.
11. Chan W. S. Stock price reaction to news and no-news: Drift and reversal after headlines // Journal of financial economics. 2003. Vol. 70. Pp. 223-260.
12. Fung G., Yu J., Hongjun Lu. The predicting power of textual information on financial markets // IEEE intelligent informatics bulletin. 2005. Vol. 5 ; No. 1.
13. Gautam M. Automated analysis of news to compute market sentiment: its impact on liquidity and trading review. 2011.
14. Godbole N., Srinivasaiah M., Skiena S. Large-scale sentiment analysis for news and blogs. In proceedings of the first international conference on weblogs and social media. 2007. Pp. 219-222.
15. <http://sourceforge.net/projects/weka/files/weka-3-5/3.5.8/>
16. <http://www.avoidingtheherd.com/>
17. Leinweber D. Nerds on Wall Street. Wiley Financial 2009.
18. Lloyd L., Kechagias D., Skiena S. Lydia: a system for large-scale news analysis. In Proc. 12th string processing and information retrieval. 2005. Vol. LNCS 3772. Pp. 161-166.
19. Mittermayer M., Knolmayer G.F. Text mining system for market response to news: a survey // Working paper. 2006. №184.
20. Negative feeling words. [http://eqi.org/fw\\_neg.htm](http://eqi.org/fw_neg.htm).
21. Pablo D.A. Sentiment analysis in financial news a thesis presented by. Harvard College, Cambridge, Massachusetts. 2009. Vol. 1.
22. Pang B., Lee L. Opinion mining and sentiment analysis. Foundations and trends in information retrieval. 2008. Vol. 2 ; №1-2. Pp. 1-135.
23. Timmons R., Lee K. Predicting the stock market with news articles. 2007. CS224n. Final report.
24. Schumaker R., Hsinchun Chen. A quantitative stock prediction system based on financial news // Information processing and management. 2009. Vol. 45. Iss. 5. Pp. 571-583.
25. Team A. Machine readable news and algorithmic trading. 2010. // Reuters and market news international, white paper.
26. Tetlock P.C., Saar-Tsechansky M., Macskassy S. More than words: quantifying language to measure firms' fundamentals // In proceedings of 9-th annual texas finance festival. 2007.

## Ключевые слова

Сентиментный анализ; влияния новостей на фондовые рынки; байесовский классификатор; индекс ATH Sentiment.

## РЕЦЕНЗИЯ

Актуальность исследования. Новости является решающим источником изменения цен на рынках. Попытка ввести новостную аналитику в виде технологии для автоматизации или полуавтоматизации построения модели поведения цен на рынках может привести к большей оперативности принятия решений трейдером. Поэтому вопрос, в какой степени могут новостные данные влиять на рынки, является актуальной задачей.

Научная новизна и практическая значимость. В настоящее время работы по влиянию ленты новостей на поведение рынка можно сгруппировать в два направления:

- определение способов порождения и отбора наиболее значимых признаков, описывающих временной ряд;
- включение отобранных признаков в модели финансовых временных рядов, по которым можно предсказывать их поведение.

Рецензируемая работа принадлежит второму направлению. Но, в отличие от доминирующего подхода к получению модели в виде использования индикаторов настроений в финансовых новостях с подсчетом количества положительных и отрицательных слов, авторы делают попытку использования расчета индекс ATH Sentiment, предназначенного для измерения текущего глобального сентимента.

Заключение: рецензируемая статья отвечает требованиям, предъявляемым к научным публикациям, и может быть рекомендована к публикации.

*Кирсанов А.П., д.т.н., профессор кафедры «Бизнес-аналитика» Национального исследовательского университета «Высшая школа экономики»*

[Перейти на Главное МЕНЮ](#)  
[Вернуться к СОДЕРЖАНИЮ](#)

*Перминов Геннадий Иванович*

*Ахмедова Саида Сабир*