

TWITTER КАК ИНДИКАТОР В ЗАДАЧАХ ЭЛЕКТРОННОЙ ТОРГОВЛИ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Трубчик А.И.

Самаль Д.И. – к.т.н., доцент

Существует теория, что история рынка ценных бумаг циклична, поэтому для торговли используется так называемый технический анализ. Для этого руководствуются ценовыми графиком, находят типовые варианты развития событий по данному товару, а также используют технические индикаторы для подтверждения своих суждений. Эти схемы используются для определения тенденций в ценовом изменении и разворота тренда, а также для определения сигналов купить или продать. Социальная сеть Twitter может быть дополнительным индикатором для этих задач.

На рынке ценных бумаг существует много факторов, которые могут вызвать изменение цены. Узнать об этих факторах можно из новостей, например таких как Reuters, Bloomberg и других. Анализ тональности текста, как один из видов анализа текста, помогает проанализировать скрытую в тексте эмоциональную оценку слов, такие как счастье или грусть, оптимизм или пессимизм [1]. А микроблоггинговый сервис Twitter является крупнейшей социальной сетью в мире и уже опережает сайты, газеты и телеканалы по скорости создания новостей.

Из этого следует, что анализ сообщений Twitter (твитов) может повысить точность прогнозирования индикаторов фондового рынка. Анализ направлен как на нахождение корреляций (т.е. что данные связаны с событиями происходящими на фондовом рынке), так и на те случаи, когда высокая активность в Twitter сигнализирует о будущих движениях рынка. Таким образом можно выдвинуть две гипотезы:

1. Существуют статистически значимые корреляции между показателями, получаемыми из твитов и рыночных данных.

2. В паре коррелирующих между собой показателями из Twitter и фондового рынка, показатели из Twitter время от времени выступает в качестве ведущего.

Чтобы проверить это, необходимо выполнить следующие задачи: собрать "сырые" данные из Twitter, произвести классификацию эмоций твитов, собрать данные с электронной торговой площадки. Для проведения анализа был использован язык программирования R. В качестве актива выступила электронная валюта Bitcoin, а данные о торгах были получены из площадок Bitfinex и BTC-E. Твиты собирались по хештегу #BTC и ключевому слову bitcoin.

Основываясь на эмоциональной оценке были выделены следующие категории твитов: позитивный, негативный, смешанный (негативный и позитивный) и нейтральный (не негативный и не позитивный) [2].

Показатель	Назначение
total	Суммарное количество твитов
pos_num	Количество позитивных твитов
pos_ppt	Процент позитивных pos_num / total
neg ..., mix ... и neu ...	– аналогичным образом для негативных и других твитов

Таблица 1 – Список показателей, получаемых на основе данных Twitter

Можно выделить и категории сообщений: с ссылками, без ссылок, включая ретвиты и без них. Это удалит шум, который создают пользователи, когда делают ретвиты популярных сообщений или массово ссылаются на какие-либо новости по теме.

Для Bitcoin использовались данные, предоставляемые торговой площадкой: операции купли-продажи и производные от них объем, минимальная и максимальная цены за интервал времени, а также группа технических индикаторов.

Выбранный объем данных слишком большой для ручного анализа, поэтому значительное число тестов в различных комбинациях проводилось автоматически, чтобы обнаружить самые интересные зависимости между данными.

По результатам анализа наибольшая корреляция $r=0.88$ наблюдается между объемом твитов без ссылок и объемом торговых сделок Bitcoin за соответствующий период. При увеличении лага между данными коэффициент корреляции значительно уменьшается. В некоторых случаях, при высокой активности на торговых площадках, Twitter перестает быть ведомым и коэффициент может быть даже отрицательным.

Таким образом необходимо провести множество исследований и создать систему интеллектуального анализа твитов в реальном времени, прежде чем станет возможным использование данных Twitter в качестве индикатора в электронной торговле.

Список использованных источников:

1. Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction / S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, A. Sakurai // Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference, pp. 800-807.
2. Sentiment Analysis of Twitter Data / A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau // Proceedings of the Workshop on Language in Social Media. Association for Computational Linguistics, Columbia University 2011, pp. 30-38.