

УДК 004.421.4 : 339.13.017

ОПРЕДЕЛЕНИЕ ЦЕНЫ ЛИМИТНОГО ОРДЕРА С ВЫСОКОЙ ВЕРОЯТНОСТЬЮ ИСПОЛНЕНИЯ

И.О. МОРОЗОВ, С.А. ПОТТОСИНА

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 31 мая 2013

Рассматривается алгоритм определения цены посылаемого на биржу лимитного ордера с фиксированной ценой для повышения вероятности его исполнения и уменьшения рисков, связанных с рыночными ордерами. Метод основан на анализе обработанных сообщений первого уровня, поступающих с цифровых торговых платформ в режиме реального времени. Приведены практические результаты торговли внутри дня.

Ключевые слова: лимитный ордер, совершенные сделки, спрос, предложение.

Введение

В связи с высокой волатильностью рынка ценных бумаг в преддверии наступающей рецессии и в свете последних так называемых «флеш крэшей» [1], становится еще более актуальной задача определения цены рыночного инструмента в режиме торговли внутри дня. Для совершения сделок на цифровых торговых платформах используются различные виды ордеров, основными из которых являются рыночный и лимитный ордера. Рыночный ордер – это заявка на покупку или продажу акций по рыночной цене. Он исполняется моментально и включает риск получить акции по цене, которая отличается от ожидаемой. Это может привести к потерям капитала в сделке. Используя лимитный ордер, т. е. заказ на покупку или продажу определенного количества акций по указанной цене или лучше, можно избежать этого риска. Такой ордер исполняется по любой лучшей доступной цене в указанном диапазоне и включает риск экспирации ордера до момента его исполнения, если доступной цены на рынке не будет [2].

В данной работе рассматривается алгоритм, представляющий собой комбинацию системы обработки входных данных и системы принятия решений. Задачей данного алгоритма является определение цены посылаемого на биржу лимитного ордера для повышения вероятности его исполнения по данной цене. Система обработки входных данных принимает торговые сообщения, поступающие с фондовых бирж США, таких как Нью-Йоркская фондовая биржа (NYSE), автоматизированные котировки Национальной ассоциации дилеров по ценным бумагам (NASDAQ) и Американская Фондовая Биржа (AMEX). Эти данные содержат различного рода аномалии, шумы, возникающие в процессе передачи и сделки между брокерами. Такие данные называют «сырыми» и требуют фильтрации, перед тем как попадут на следующий этап работы торговой системы. Анализируя обработанные данные, система принятия решений определяют цену лимитного ордера.

Виды данных

Цифровые торговые платформы отправляют подписчикам данные двух уровней. Первый уровень (1st Level data):

– информация о совершенных сделках;

– информация о верхних в очереди ценах спроса и предложения, так называемые лучшие спрос и предложение или просто квота.

Второй уровень (2nd Level data):

– все ордера, поступившие на биржу, а также их изменения, исполнение или закрытие.

Выделенные категории кардинально различаются по структуре, поэтому требуют различных методов анализа.

Второй уровень данных (order book) в данном исследовании не участвовал в связи с высокой стоимостью оборудования для обработки в режиме реального времени и ценой подписки на получение.

Совершенные сделки

Когда цена спроса совпадает с ценой предложения, происходит сделка, которая фиксируется торговой платформой и отправляется в сообщении всем подписчикам. После преобразования из бинарного формата эти данные имеют вид, представленный в табл. 1, где «Время» – время в секундах от начала дня по UTC, «Символ» – торговый инструмент, участвовавший в сделке, «Цена» – цена, по которой была совершена сделка, «Объем» – количество акций, участвовавших в сделке, «Кумулятивный объем» – суммарное количество акции участвовавших в сделках по данному торговому инструменту до текущего момента за текущее календарное число. В зависимости от времени, система принимает от 200 до 3000 сообщений о сделках за 1 с по 100 исследуемым ликвидным инструментам.

Таблица 1. Представление информации о сделке

Время	Символ	Цена	Объем	Кумулятивный объем
61803,068	VZ	37,650	100	3631374
61803,178	FXI	38,852	177	9735605
61803,568	FCX	44,965	100	5872764

Для фильтрации шумов и аномалий в данных о ценах сделок был выбран критерий Пирса, т.к. он имеет ряд преимуществ по отношению к критерию Шовене [3]:

- имеет строгую формулировку;
- не делает никаких предположений касательно исключаемых данных;
- позволяет исключать несколько аномалий за одну итерацию;
- легко реализуем с использованием таблицы критериев.

Алгоритм представляет собой рекурсивную оконную фильтрацию на основе оценки максимально возможного отклонения данных

$$R = \frac{|x_i - x_m|_{\max}}{\sigma}, \quad (1)$$

где σ – стандартное отклонение, x_i – текущее значение ряда, x_m – среднее значение ряда.

Размер окна равен 25 значениям. Соответственно, набор значений критерия Пирса [3] для данного окна в зависимости от числа исключений: 2,307 для одного, 2,019 – двух, 1,840 – трех, 1,709 – четырех, 1,604 – пяти, 1,517 – шести, 1,442 – семи, 1,375 – восьми, 1,315 – девяти. Если рассчитанное значение R для нового значения цены превышает табличное значение критерия, то оно помечается как аномальное и исключается из анализа [3].

На роль алгоритма рекурсивного расчета дисперсии в окне длины n были рассмотрены два варианта с линейной вычислительной сложностью: простой рекурсивный алгоритм [4]:

$$\sum_n x = \sum_{n-1} x - x_1 + x_n, \quad (2)$$

$$\sum_n x^2 = \sum_{n-1} x^2 - x_1^2 + x_n^2, \sigma^2 = \frac{\sum_n x^2 - (\sum_n x)^2}{n-1}. \quad (3)$$

и онлайн алгоритм [3]:

$$\bar{x}_n = \bar{x}_{n-1} + \frac{x_n - \bar{x}_{n-1}}{n}, \quad (4)$$

$$\sigma^2 = \frac{(n-2)\sigma_{n-1}^2 + (x_n - \bar{x}_n)(x_n - \bar{x}_{n-1})}{n-1}, \quad n > 1, \quad (5)$$

где n – размер окна.

Для анализа быстродействия алгоритмы были реализованы на языке программирования C++.

Таблица 2. Результаты оценки быстродействия алгоритмов

Алгоритм расчета дисперсии	Объем данных, количество сделок	Время на обработку, мс
Простой рекурсивный	$1 \cdot 10^7$	29
	$1 \cdot 10^9$	2894
Онлайн	$1 \cdot 10^7$	13
	$1 \cdot 10^9$	1347

Из анализа быстродействия алгоритмов (табл. 2) видно, что время обработки растет линейно в зависимости от размера входных данных. Для включения в конечную систему был принят алгоритм с наименьшим временем обработки – онлайн алгоритм.

Для исключения нулевых значений дисперсии при расчете критерия Пирса (1) в алгоритм поступают только те сделки, цена которых отличаются хотя бы на 1 цент от цены предыдущей сделки.

Лучшие спрос и предложение

Набор лимитных ордеров, каждый из которых характеризуется ценой и объемом, формирует очередь заказов. На рис. 1 представлены ордера с лучшими ценами, которые находятся вверху очереди и называются лучшими спросом и предложением.

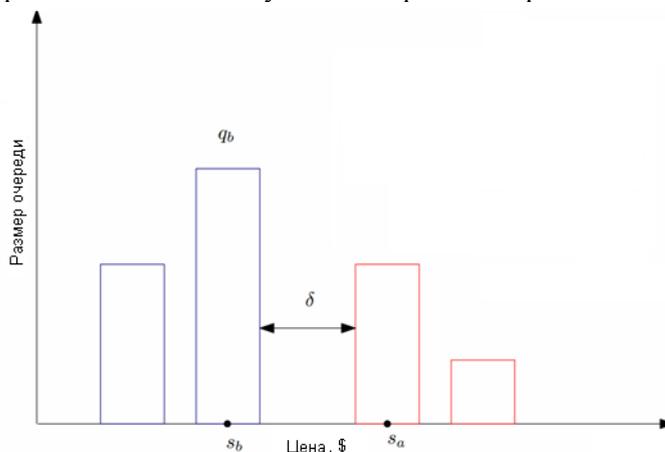


Рис. 1. Схематическое изображение очереди заказов: s_b – цена лучшего предложения, s_a – цена лучшего спроса, q_b – объем лучшего предложения, q_a – объем лучшего спроса, δ – спред

При изменении цены или объема лучшего спроса или предложения по торгуемому финансовому инструменту, торговая платформа отправляет подписчикам обновленную информацию о цене – квоте. После преобразования данные имеют вид, представленный в таблице.

Таблица 3. Представление квоты

Время	Символ	Цена лучшего предложения	Объем лучшего предложения	Цена лучшего спроса	Объем лучшего спроса
61803,068	VZ	37,65	100	37,66	10
61803,178	FXI	38,85	20	38,87	400
61803,568	FCX	44,96	100	44,97	200

Разность между ценой спроса и предложения называется спредом. На исследуемых платформах минимальный спред равен 1 центу США. Соответственно, цена спроса всегда должна быть больше цены предложения.

В связи с широким применением на рынке ордеров типа «все или ничего» или «немедленно или отменить» цены спроса и предложения имеет моментальные единичные скачки. Также поступающие в алгоритм квоты имеют высокую плотность, достигающую 15000–20000 сообщений в секунду по 200 исследуемым торговым инструментам. Распределение первой разности последовательности цен спроса представлено на рисунке.

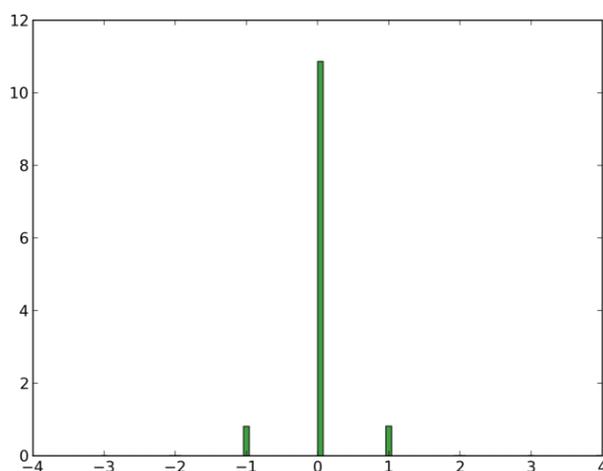


Рис. 2. Распределение изменений цены спроса

С учетом распределения изменений цены спроса и предложения и свойства инертности цены в основу алгоритма фильтрации цен спроса и предложения был положен рекурсивный медианный фильтр. В алгоритме используется три сортированные последовательности, две из которых имеют внутреннюю организацию в виде бинарных деревьев – *lower_median* и *bigger_median* и одна – в виде очереди – *real_values*. Размер окна $window = 2n + 1$, где $n \in N$. Алгоритм имеет следующий порядок выполнения:

1. Получить новый элемент из входной очереди – *key*.
 2. Если размер *lower_median* равен $(window + 1) / 2$ и последний элемент в последовательности *lower_median* меньше либо равен *key*, то добавить *key* в *bigger_median*, иначе добавить в *lower_median*.
 3. Добавить *key* в *real_values*.
 4. Если размер *real_values* больше, чем *window*, то достать первый элемент из *real_values* и удалить его в одном из *lower_median* или *bigger_median*.
 5. Пока размер *lower_median* больше, чем $(window + 1) / 2$, переместить последний элемент из *lower_median* в *bigger_median*.
 6. Пока размер *bigger_median* больше, чем $window / 2$, переместить последний элемент из *bigger_median* в *lower_median*.
 7. Для получения медианного значения, если размер *real_values* равен *window*, то достать из *lower_median* последний элемент, иначе достать из *real_values* последний элемент.
- Использование бинарных деревьев для операций вставки в сортированную последовательность позволяет получить вычислительную сложность $O(n)$. Полученный алгоритм выигрывает по быстродействию у стандартных алгоритмов, основанных на алгоритмах сортировки, вычислительная сложность которых колеблется от $O(n^2)$ до $O(n \log(n))$ [3].
- Графики цен до и после фильтрации представлены на рисунке.

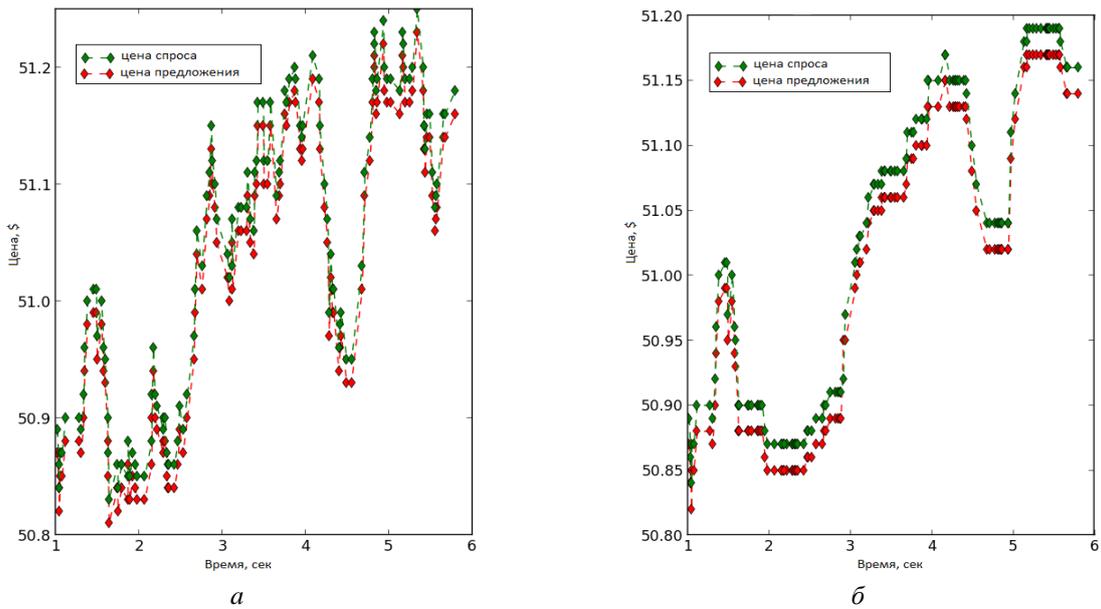


Рис. 3. Графики цен спроса и предложения: *a* – до фильтрации; *б* – после фильтрации

Определение цены лимитного ордера

Разработанный алгоритм основан на сравнении текущих отфильтрованных цен совершенных сделок и отфильтрованных цен спроса и предложения и использует следующие ключевые моменты.

1. Если медианный спред меньше текущего спреда, то:
 - если цена последней совершенной сделки лежит вне диапазона цен спроса и предложения, то использовать текущие спрос и предложение;
 - если сделка была на покупку, то использовать ее цену как цену спроса, а цену предложения рассчитать по формуле $s_b = s_a - \delta$;
 - если сделка была совершена на продажу, то использовать ее как цену предложения, а цену спроса рассчитать по формуле $s_a = s_b + \delta$;
2. Если медианный спред больше текущего спреда, то использовать текущие цены спроса и предложения.

Результаты практического применения алгоритма в торговле представлены в табл. 4. На рис. 4 показана гистограмма вероятности того, что ордер будет выполнен по цене равной (0 центов на гистограмме) или лучшей, чем заявлена в ордере.

Таблица 4. Практические результаты работы алгоритма

Количество лимитных ордеров			Вероятность исполнения, %	Среднее время исполнения, мс
Всего отправлено	Полностью исполнено	Частично исполнено		
346	312 (90,17%)	10 (2,89%)	93,06	5923

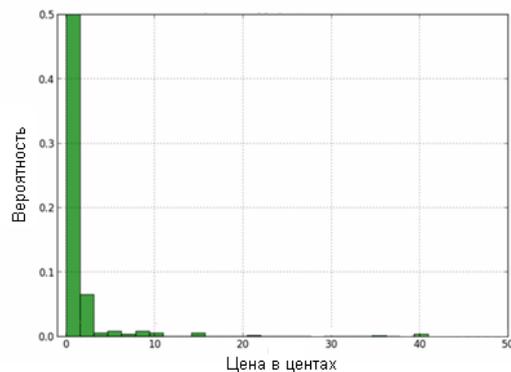


Рис. 4. Гистограмма вероятности исполнения по цене равной или лучшей, чем заявлена в ордере

Заключение

Рассмотренные методы фильтрации имеют линейную вычислительную сложность и позволяют без задержек обрабатывать большие объемы сообщений по совершенным сделкам, ценам спроса и предложения, поступающих с фондовых бирж. Алгоритм определения цены лимитного ордера в режиме реального времени имеет высокую эффективность (93,06 %), что подтверждается практическими результатами исследования.

PRICE ESTIMATION OF LIMIT ORDER WITH HIGH EXECUTION PROBABILITY

I.O. MOROZOV, S.A. POTTOSINA

Abstract

Limit order price estimation algorithm is considered in the article. Algorithm increases execution probability. It is based on the processing of the first level capital market data such as tick trades and quotes. It is shown and proved by practical results that algorithm allows successfully detect price of limit order for good position entry and exit.

Список литературы

1. *Lauricella T.* // The Wall Street Journal. Archived from the original on 21 October 2010. Retrieved 28 October 2010.
2. *Larry Harris.* Trading & Exchanges. Oxford, 2003.
3. *Ross S. M.* // Journal of Engineering Technology. New Haven, 2003.
4. *Knuth D.E.* The Art of Computer Programming, volume 2: Seminumerical Algorithms. Boston, 1998.