2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

Mining Exceptional Activity Patterns in Microstructure Data

Yuming Ou, Longbing Cao, Chao Luo and Li Liu
Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia
{yuming, lbcao, chaoluo, liliu}@it.uts.edu.au

Abstract

Market Surveillance plays an important role in maintaining market integrity, transparency and fairnesss. The existing trading pattern analysis only focuses on interday data which discloses explicit and high-level market dynamics. In the mean time, the existing market surveillance systems are facing challenges of misuse, mis-disclosure and misdealing of information, announcement and order in one market or crossing multiple markets. Therefore, there is a crucial need to develop workable methods for smart surveillance. To deal with such issues, we propose an innovative methodology — microstructure activity pattern analysis. Based on this methodology, a case study in identifying exceptional microstructure activity patterns is carried out. The experiments on real-life stock data show that microstructure activity pattern analysis opens a new and effective means for crucially understanding and analysing market dynamics. The resulting findings such as exceptional microstructure activity patterns can greatly enhance the learning, detection, adaption and decision-making capability of market surveillance.

1. Introduction

In many types of markets such as capital and electricity markets, *market surveillance* is essential to design market models and business rules, as well as maintain the market integrity, transparency and fairness [1, 2]. The existing market surveillance systems usually rely on surveillance rules for alerting of suspect findings in the market. Most of the surveillance rules are predefined and based on business rules, while some of the surveillance rules may come from statistics and reporting results which can capture more sophisticated abnormal trading behaviour and market movement. These rules play an important role in filtering obvious offences against market business rules, regulation rules, and explicitly exceptional market movements.

However, these existing surveillance systems are facing challenges of diversified, dynamic, distributed and cyber-based misuse, mis-disclosure and misdealing of information, announcement and order in one market or crossing multiple markets. Such challenges cannot be handled by the existing systems and techniques usually used in exchanges.

In addition, the current price movement and trading pattern analysis mainly focus on interday data such as closing prices. The resulting analytical results are not workable for real-time market surveillance because they cannot catch and filter the microstructure behaviour every second of every day. There is a crucial need to develop breakthrough methodologies and techniques to discover hidden knowledge in the *market microstructure data* under the increasing financial and trading globalization.

In this paper, to deal with the above issues, we propose an innovative methodology — *microstructure activity pattern analysis* [3], which studies the investor's behaviours by following and involving market microstructure theories. Investor's behaviours are recorded in *market microstructure data* [3] consisting of investor's actions and interactions with other investors in one market or crossing multiple markets, as well as their embodiment in market dynamics. *Microstructure activity pattern analysis* aims to identify the investor's behaviour patterns hidden in *microstructure data*.

The reminder of this paper is organized as follows. Section 2 briefly introduces the market microstructure data. In Section 3 we propose our innovative methodology. A case study in identifying *exceptional microstructure activity patterns* is carried out in Section 4. We conclude this paper and present our future work in Section 5.

2. Market microstructure and data

978-0-7695-3496-1/08 $25.00 © 2008 IEEE
DOI 10.1109/WIIAT.2008.160

Market microstructure [4, 5] is "a branch of finance concerned with the details of how exchange occurs in markets" [6]. "The major thrust of market microstructure research examines the ways in which the working processes of a market affects determinants of transaction costs, prices, quotes, volume, and trading behavior." [6] One of main issues studied by market microstructure is the behaviours of market maker and investor.

Transactional data recording the investor behaviour in markets obeying market microstructure theory present a unique structure. We call such data *market microstructure data* [3]. In the stock market, *market microstructure data* is intraday data mainly including orderbook, trading records and market data. It presents the following major characteristics [3] that are not usually seen in many other applications:

- *rich semantic,*
- *time frame* and *gradient,*
- *granularity dynamics,* and
- *heterogeneous.*

The above characteristics of *market microstructure data* bring challenges to the existing pattern mining approaches. Therefore, it is unreasonable to simply conduct the traditional data mining techniques on it.

3. Framework of microstructure activity pattern analysis

Figure 1 shows the proposed framework of microstructure activity pattern analysis. There are three steps in the framework:

Modeling Behaviour (MB). The aim of this step is to model the investors' behaviors. Due to the characteristics of *market microstructure data* mentioned above and the fact that *market microstructure data* only implicitly reflects investor's behaviours, it is necessary to first extract the investor's activities from *market microstructure data* and represent them by a proper model which is favourable to pattern mining. Domain knowledge is involved in this step. This step can be formulised as follows:

$$D \xrightarrow{MB} A \qquad (1)$$

where $D = \{d_1, d_2, \ldots, d_n\}$ is *market microstructure data* and $A = \{a_1, a_2, \ldots, a_m\}$ are investors' activities extracted from D.

Mining General Patterns (MGP). In this step, data mining techniques are conducted on the behaviour model to identify patterns in terms of technical interestingness. Many categories of activity patterns [3, 7] can be discovered, such as impact-targeted patterns [8] and exceptional patterns. However, the identified patterns are generally interesting patterns and may be uninteresting to business people. We call them general patterns. This step can be formulised as follows:

$$A \xrightarrow{MGP} P \qquad (2)$$

where $P = \{p_1, p_2, \ldots, p_j\}$ are general patterns.

Generating Actionable Patterns (GAP). The last step is to generate actionable patterns from general patterns. Specific business interestingness measures are used to prune the general patterns. The distilled patterns are operational for business people because both technical and business concerns are considered. This step can be formulised as follows:

$$P \xrightarrow{GAP} \tilde{P} \qquad (3)$$

where $\tilde{P} = \{\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_k\}$ are actionable patterns.

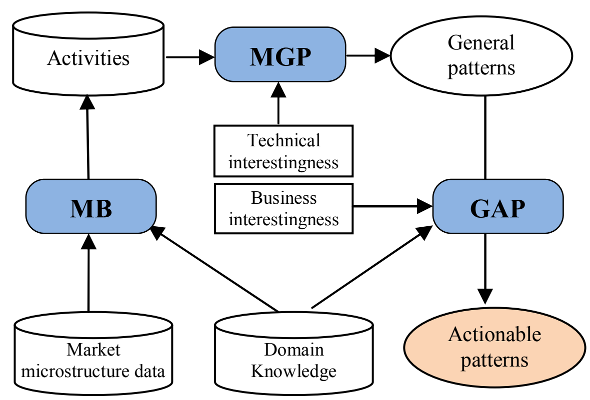

Figure 1. Framework of microstructure activity pattern analysis

4. Case study: identifying exceptional microstructure activity patterns

In this section, we carry out a case study following the framework proposed in Section 3 to identify exceptional microstructure activity patterns. In the stock market, the exceptional microstructure activity patterns are very useful for smart surveillance because they are against average people's exceptions and may reflect the abnormal trading behaviours. Once the market surveillance system monitors or detects *exceptional microstructure activity patterns* taking place in the market, then it can alert surveillance officers to the abnormal market movements.

4.1. Modeling behaviour (MB)

Firstly we need to model the trading behaviors. Considering the fact that every order follows market microstructure theory and indicates information about order holder's intention, the representation of such order sequences should reflect them accordingly. In stock markets, although the values of a particular order attribute vary from order to order, they actually reflect a trader's intentions, and cater for the particular stage of their lifecycles. For instance, for a single time point, an order may present in one of the following states (s) in its lifecycle: $s \in$ {*new, traded partly, traded entirely, deleted, outstanding*}. Further, even for the same values of a particular order attribute, they may indicate divided circumstances that reflect investor's varying motivation and behaviour.

Therefore, the proper representation of an order should reflect order holder's intention, actions and the order's lifecycle. For this purpose, we propose a vector-based order representation: A five-dimension vector $O(d, \delta, \rho, \varphi, \varepsilon)$ is defined to represent an order. Among the five dimensions, dimension $d \in \{B, S\}$ reflects the trade direction of order, dimension $\delta \in \{\delta_H, \delta_M, \delta_L\}$ stands for the probability that the order can be traded, dimension $\rho \in \{\rho_S, \rho_M, \rho_L\}$ measures the size of order, dimension $\varphi \in \{\varphi_0, \varphi_1, \varphi_N\}$ represents how many trades the order leads to and dimension $\varepsilon \in \{\varepsilon_0, \varepsilon_1, \varepsilon_{-1}\}$ reflects the balance of order at the time of market close. The proposed order vector O encloses plenty of semantics: (1) indicating the direction, probability and size of an order to be traded, (2) reflecting an order's movements during its lifecycle.

For intraday microstructure data, an order at most lasts for one day since its generation. In addition, orders placed by different investors indicate different intentions, beliefs and desires. Thus it is reasonable to construct order sequences in terms of trading day and order investor. A microstructure order sequence Ω is the sequence of orders in vectors for a trader within a trading day:

$$\Omega = \{O_1(d^1, \delta^1, \rho^1, \varphi^1, \varepsilon^1), O_2(d^2, \delta^2, \rho^2, \varphi^2, \varepsilon^2), \ldots, O_j(d^j, \delta^j, \rho^j, \varphi^j, \varepsilon^j), \ldots\} \quad \textbf{(4)}$$

where O_j is an order vector. It systematically reflects an investor's intention, order lifecycles and trading activities in the market.

4.2. Mining general patterns (MGP)

In this step, data mining techniques are conducted on the microstructure sequences to discover *exceptional microstructure activity patterns*.

At the beginning, the *target data set* and *benchmark data set* are selected by a sliding time window (assuming size $m + 1$). As shown in the Figure 2, all data on the current *target trading day* comes into the *target data set*, the data on the m previous days before the *target trading day* are called *benchmark day 1, 2, …, m* respectively. All data drawn from the *benchmark day j* fits into *benchmark data set j*. Through sliding the time window in the available data set, we can generate *target data set* and *benchmark data sets* for a *target trading day*.

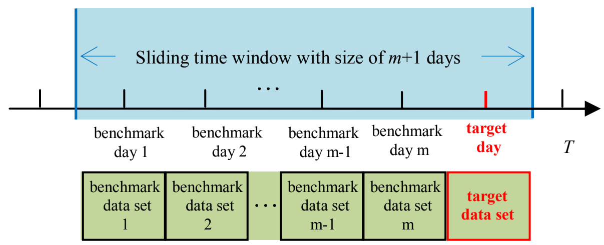

Figure 2. Target and benchmark data sets

Two technical interestingness measures are used to identify *exceptional microstructure activity patterns* in *target data set*. They are defined as follows:

Definition 1*. Intentional Interestingness* (I_i): I_i quantifies the intentional interestingness of a pattern as defined in the following formula:

$$I_i = Sup_t \times \frac{|\Omega|}{AvgL_t} \quad \textbf{(5)}$$

where Sup_t is the support of sequence in the *target data set* of day t, | | is the length of sequence and $AvgL_t$ is the weighted average length of sequences in the *target data set*. This metric reflects that investors tend to use a series of orders to deploy their intentions.

Definition 2*. Exceptional Interestingness (Ie): Ie* quantifies the exceptional interestingness of a pattern as defined in the following formula:

$$I_e = \frac{\frac{Sup_t}{AvgL_t} \times \sum_{j=1}^{m} \omega_j}{\sum_j^m (\frac{SupB_j}{AvgLB_j} \times \omega_j)} \quad \textbf{(6)}$$

where $SupB_j$ is the support of sequence in the *benchmark data j*, $AvgLB_j$ is the weighted average length of sequences in the *benchmark data j*, ω_j is the weight for the *benchmark data j*, and m is the number of *benchmark days*. I_e reflects how exceptional a pattern presents on *target day* than on *benchmark days*.

Consequently, a sequence is an *exceptional microstructure activity pattern*, if it satisfies the conditions: $I_i \geq I_{i0}$ and $I_e \geq I_{e0}$, where I_{i0} and I_{e0} are the thresholds given by users or domain experts.

4.3. Generating actionable patterns (GAP)

We use two business interestingness measures to prune the general patterns. They are *return (R)* and *abnormal return (AR)* of security. In the stock markets, *return* reflects the gain or loss of a single security over a specific period while *abnormal return* indicates the difference between the actual *return* of a security and the *expected return* estimated by the *capital asset pricing model*.

An exceptional pattern is interesting to business people, if it satisfies the conditions: $|R| \geq R_0$ and $|AR| \geq AR_0$, where R_0 and AR_0 are the thresholds given by users or domain experts.

4.4. Experimental results

Our approach has been tested on a real-life stock data set. The data consists of *240* trading days from *2005* to *2006* for a security.

Table 1 shows the samples of the *exceptional microstructure activity patterns* discovered by the approach. These patterns reflect the traders' exceptional intentions on the corresponding day. For example, on May 24, 2005, the I_i and I_e for the pattern $\{(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0), (S, \delta_S, \rho_S, \varphi_1, \varepsilon_0)\}$ are *0.054* and *11.2* respectively. This indicates a strong intention and exception of trading activities conducted on that day. Furthermore, the absolute *return* and *abnormal return* on *24/05/2005* are as high as *6.82%* and *6.38%* respectively which are really interesting to business people. Consequently, these results from both technical and business sides present business people strong indicators showing that there likely was abnormal trading behaviour on that day.

Table 1. Exceptional microstructure pattern samples

Date	Exceptional Microstructure Patterns	I_i	I_e	| R | (%)	|AR| (%)
14/01/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}), (B, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.025	7.6	2.68	1.38
18/01/2005	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}), (S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.026	10.8	2.25	1.85
20/01/2005	$\{(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1}),(S, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.038	5.4	2.98	1.56
28/01/2005	$\{(B, \delta_H, \rho_S, \varphi_1, \varepsilon_0),(B, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.030	8.2	2.63	1.97
24/05/2005	$\{(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.054	11.2	6.82	6.38
16/06/2005	$\{(B, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.025	6.1	2.64	2.47
01/07/2005	$\{(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H, \rho_S, \varphi_0, \varepsilon_{-1}), (B, \delta_H, \rho_S, \varphi_1, \varepsilon_0)\}$	0.028	54.0	2.21	1.28
13/07/2005	$\{(B, \delta_M, \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.028	5.6	9.55	9.12
15/09/2005	$\{(B, \delta_H , \rho_S, \varphi_0, \varepsilon_{-1}),(B, \delta_H , \rho_S, \varphi_0, \varepsilon_{-1})\}$	0.035	8.8	1.43	3.49
05/12/2005	$\{(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0),(S, \delta_M, \rho_S, \varphi_1, \varepsilon_0)\}$	0.035	8.6	1.85	3.41

5. Conclusions and Future Work

In this paper, a breakthrough methodology of *microstructure activity patterns analysis* has been proposed. Following the proposed methodology, we have carried out a case study in mining *exceptional activity patterns* in *market microstructure data*. Substantial experiments on real-life stock data have shown that market *microstructure pattern activity analysis* opens a new and effective means for crucially understanding and analyzing market dynamics. The resulting findings such as *exceptional microstructure activity patterns* can greatly enhance the learning, detection, adaption and decision-making capability of market surveillance.

In the future, we plan to study the issues under *microstructure activity pattern analysis*, for instance, discovery of patterns other than exceptional microstructure patterns. We expect to build up a complete and powerful methodology — *microstructure activity patterns analysis*, to enhance the existing market surveillance systems.

6. Acknowledgement

The project is partially supported by Australian Research Council Discovery grants DP0773412, LP0775041 and DP0667060.

7. References

[1] O'Hara, M., "Designing markets for developing countries", *International Review of Finance*, 2001, vol. 2, pp.205-215

[2] Dehdashti, E., "Monitoring and surveillance of wholesale electricity markets: roles, responsibilities and challenges", *Power Engineering Society General Meeting*, 2005, IEEE, vol.3, pp.3035-3041

[3] Cao, L. and Ou, Y., "Market Microstructure Patterns Powering Trading and Surveillance Agents", *Journal of Universal Computer Sciences*, 2008 (to appear).

[4] Madhavan, A., "Market microstructure: A survey", *Journal of Financial Markets*, 2000, vol. 3, pp. 205- 258

[5] Harris, L., Trading and Exchanges: *Market Microstructure for Practitioners*, Oxford University Press, 2003.

[6] *http://en.wikipedia.org/wiki/Market_microstructure*

[7] Cao, L., Zhao and Y., Zhang, C. and Zhang, H., "Activity Mining: from Activities to Actions", *International Journal of Information Technology & Decision Making*, 2008, 7(2), pp. 259 - 273

[8] Cao, L., Zhao, Y. and Zhang, C., "Mining Impact-Targeted Activity Patterns in Imbalanced Data", *IEEE Trans. on Knowledge and Data Engineering*, 2008 (to appear)