An Empirical Study of Similarity Search in Stock Data

Lay-Ki Soon **Sang Ho Lee**

School of Information Technology, Soongsil University
1-1 Sangdo-dong, Dongjak-gu, Seoul 156-743, Korea

laykisoon@gmail.com, shlee199@gmail.com

Abstract

Using certain artificial intelligence techniques, stock data mining has given encouraging results in both trend analysis and similarity search. However, representing stock data effectively is a key issue in ensuring the success of a data mining process. In this paper, we aim to compare the performance of numeric and symbolic data representation of a stock dataset in terms of similarity search. Given the properly normalized dataset, our empirical study suggests that the results produced by numeric stock data are more consistent as compared to symbolic stock data.

Keywords: financial data mining, similarity search, data normalization, computational finance.

1 Introduction

Stock data mining plays an important role in realizing the vision of autonomous financial market analysis or computational finance. The Efficient Market Theory asserts that it is impossible to infer a consistent and global forecasting model to the stock market by using any information that the market already knows. Stock data mining does not accept nor reject this theory; instead it aims to discover subtle short term conditional patterns and trends in wide range of financial data (Kovalerchuk and Vityaev 2000).

Being one of the key applications in time series data mining, stock data mining generally focuses on trend modeling and forecasting (Han and Kamber 2006). In this paper, we have analyzed the stock dataset by performing similarity search. By identifying stocks that share similar behavior, we can gain insight into the underlying pattern, which is helpful in further analysis, such as stock market forecasting. Nevertheless, one of the main challenges in stock data mining is to find the effective knowledge representation of the stock dataset (Kovalerchuk and Vityaev 2000, Kovalerchuk et al. 2000, Roddick and Spiliopoulou 2002). As such, the main objective of this empirical study is to explore the suitability and compare the performance of numeric and symbolic stock dataset in similarity search. Our experimental results demonstrate notable difference between the results produced by these two types of data representation.

This paper is organized as follows. Section 2 briefly discusses some related works which motivate our experiments. Section 3 presents the stock dataset and data pre-processing steps, particularly data transformation or normalization. Section 4 and 5 explain the experiments and the results of similarity search on numeric and symbolic stock dataset respectively. Lastly, we conclude this paper in Section 6.

2 Motivation and Related Works

Financial data are conventionally represented in numeric format for data mining purpose. However, recent works have demonstrated promising results of representing financial data symbolically. For an instance, Kovalerchuk et al. (2002) argues that symbolic relational data mining is more suitable in incorporating background knowledge. Their proposed methodology outperforms numeric financial data in generating IF-Then rules. In (Ting et al. 2006), sequential and non-sequential association rule mining (ARM) were used to perform intra and inter-stock pattern mining, where each stock is represented symbolically based on its performance with respect to a user-defined threshold. Similarly, we have also obtained encouraging results in finding correlations among the stocks with respect to an index by representing the stock dataset symbolically (Soon and Lee 2007). Although similarity search is the main concern of our empirical study, these related works have motivated us to compare the performance of numeric and symbolic representation of our stock dataset in this context.

Given a time series dataset, similarity search can be performed using either complete or subsequence matching (Keogh 2006). Most of the research works pertaining to similarity search in time series data focused on finding suitable distance measures or dimension reduction (Han and Kamber 2006, Keogh 2006). Several distance measures are widely applied in this context, such as Euclidean distance, Dynamic Time Warping (DTW) and Longest Common Subsequence. Since almost 80% of the published works used Euclidean distance, we have decided to use it for the similarity search on our numeric dataset. As for the symbolic dataset, the similarity is evaluated using the measure applied on categorical variables in cluster analysis, as described in Section 5. Complete matching is performed on both numeric and symbolic datasets.

This work was supported by Seoul R&BD Program (10581cooperateOrg93112).

3 Stock Dataset and Data Pre-Processing

3.1 Stock Dataset

For the purpose of this empirical study, we have used the stocks dataset of Malaysia Exchange, dated from December 31, 2001 till December 31, 2006, which amounts to 1233 days (KLSE Daily 2007). The trading of the stock market within a day is recorded in a single text file. Figure 1 shows a snippet of a file dated December 7, 2006. Each line represents the trading information of an index or a stock. Table 1 describes the information presented in the first line of Figure 1.

```
COMPOSITE,12/07/06,1100.3100,1102.7900,1092.5400,1098.2600,12191
TRADING,12/07/06,152.2500,152.6800,151.1900,151.9800,1926
SECBOARD,12/07/06,93.5100,93.6600,91.9000,92.7400,814
IND_PRODUCTS,12/07/06,91.2300,91.2800,90.3900,90.7000,776
FINANCE,12/07/06,8806.3100,8835.3100,8733.5200,8787.2600,684
PROPERTY,12/07/06,677.2100,677.2100,669.3200,673.5400,677
CONSTRTN,12/07/06,203.8500,204.5700,200.6300,202.6300,500
PLANTATN,12/07/06,4221.3900,4273.1900,4217.3700,4253.6000,393
CONSUMER,12/07/06,267.6500,267.7900,263.7500,264.7700,294
MINING,12/07/06,493.3900,493.3900,471.3700,475.7700,0
INDUSTL,12/07/06,2238.8100,2257.0200,2238.8100,2250.5200,0
MESDAQ,12/07/06,122.0900,122.4500,118.5500,119.6700,302297700
TECHNOLOGY,12/07/06,28.9900,29.0400,28.3400,28.5100,6718000
0012,12/07/06,0.255,0.255,0.250,0.250,5627
```

Figure 1: Snippet of our stock dataset.

Information	Description
COMPOSITE	Kuala Lumpur Composite Index
12/07/06	Dated December 7, 2006
1100.3100	Opening / Previous value (Last value of December 6, 2006).
1102.7900	Highest value of the day.
1092.5400	Lowest value of the day.
1098.2600	Closing / Last value of the day.
12191	Trading volume.

Table 1: Information recorded at first line of Figure 1.

Similar to other real-world datasets, our stock dataset requires comprehensive data cleaning. Generally, the existing noise in our stock dataset can be grouped into three main categories, which are duplicate records, inconsistent stock codes and missing values. Details of the data cleaning process performed on our stock dataset can be referred at (Soon and Lee 2007).

3.2 Data Pre-Processing for Numeric Representation

After the comprehensive data cleaning, we have then transformed the stock dataset into numeric representation, as shown in Figure 2. We have used six parameters for the numeric representation, which include the opening, closing, highest, lowest values, trading volume and the price changes (closing – opening) of the stocks and indices. One file has been created for each parameter. Figure 2 illustrates the file recording the opening values of stocks and indices. Each line denotes the opening values of a particular stock or index from December 31, 2001 till December 31, 2006, separated by comma.

Missing values are observed when certain stocks or indices record null values on certain days. For the purpose of this empirical study, only stocks and indices which have at most 10 null values are selected. As a result, only the records of 11 indices and 789 stocks were used in the similarity search. The existing null values in all the selected records were filled by using the average of its first left and first right non-null values.

```
7061,1.7009,1.7936,1.8095,1.826,1.8383
5657,-0.77904,-0.86301,-0.93246,-0.985
7120,1.681,1.777,1.8076,1.8061,1.8045,
5924,0.19104,0.1437,0.11364,0.10622,0.
9148,0.28108,0.29373,0.2517,0.22829,0.
5185,-1.3989,-1.5662,-1.6562,-1.7276,-
5185X,-0.078845,-0.10955,-0.1042,-0.06
5185Y,-0.038806,-0.01619,0.11185,0.266
7090,0.95123,0.84384,0.8299,0.80857,0.
4952,-0.90873,-1.0161,-0.99205,-1.0094
6696,-1.2787,-1.1561,-1.32,-1.3853,-1.
2305,0.87135,0.84059,0.79005,0.77678,0
9547,0.51139,0.49728,0.51611,0.53285,0
```

Figure 2: File recording the open values of stocks or indices.

Assuming L_o lists 800 lines of opening values for the selected stocks and indices while C contains all the stock codes and indices, each line l_{o_i} is

$$l_{o_i} = \{c_i, v_i, v_2, \ldots, v_n\} \tag{1}$$

where $l_{o_i} \in L_o$, $c_i \in C$, $n = 1233$ and $v_1, v_2, \ldots, v_n$ are the opening values for stock c_i. For example, stock 7061 in the first line of Figure 2 records an opening value of 1.7009 on December 31, 2001, 1.7936 on January 2, 2002 and so on so forth.

As mentioned above, Euclidean distance will be applied in our similarity search on numeric stock dataset. Since Euclidean distance is very sensitive to some unnecessary distortions, our numeric dataset needs to be normalized or transformed. The data transformation steps performed include offset translation, amplitude scaling, removing of linear trend, and removing of noise using moving average smoothing method.

Offset translation is done by removing/adding a certain offset value from/to a sequence of values. In our case, using mean as the offset value, we have removed every data series, say l_{o_i} for the opening values of c_i from its mean, which transforms l_{o_i} to

$$offset\_l_{o_i} = l_{o_i} - mean(l_{oi})$$

As for amplitude scaling, the standard deviation *stddev* of l_{o_i} is applied where l_{o_i} is transformed to

$$amp\_l_{o_i} = \frac{offset\_l_{o_i}}{stddev(l_{o_i})}$$

In other words, $amp\_l_{o_i}$ is z-normalized after undergoing offset translation and amplitude scaling. Subsequently, in order to remove the linear trend from our dataset, a best fitting line to the data sequences is subtracted from $amp\_l_{o_i}$. The main purpose of removing linear trend is to remove the influence of time from the time series. Lastly, the noise from our data series is removed by averaging each data points in the time series with its neighbors. In our case, we have used 5-point moving average method.

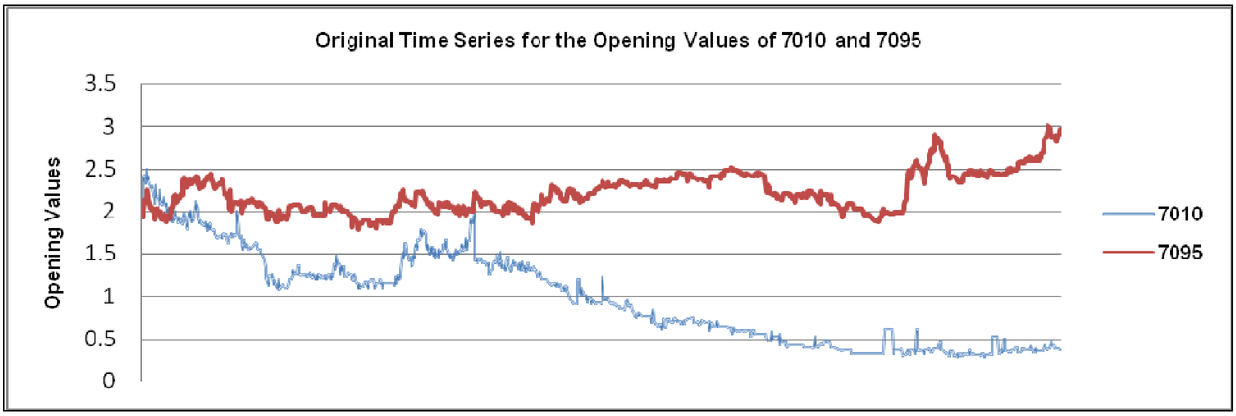
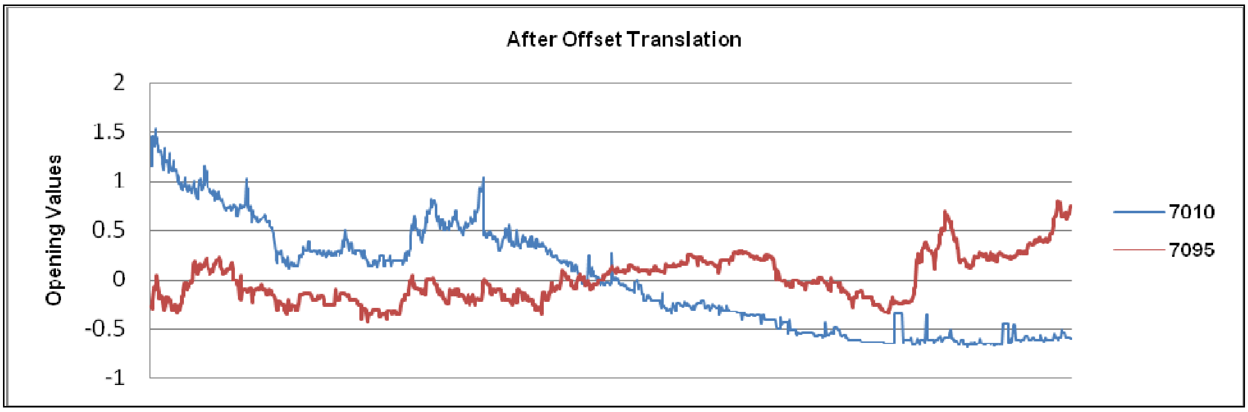
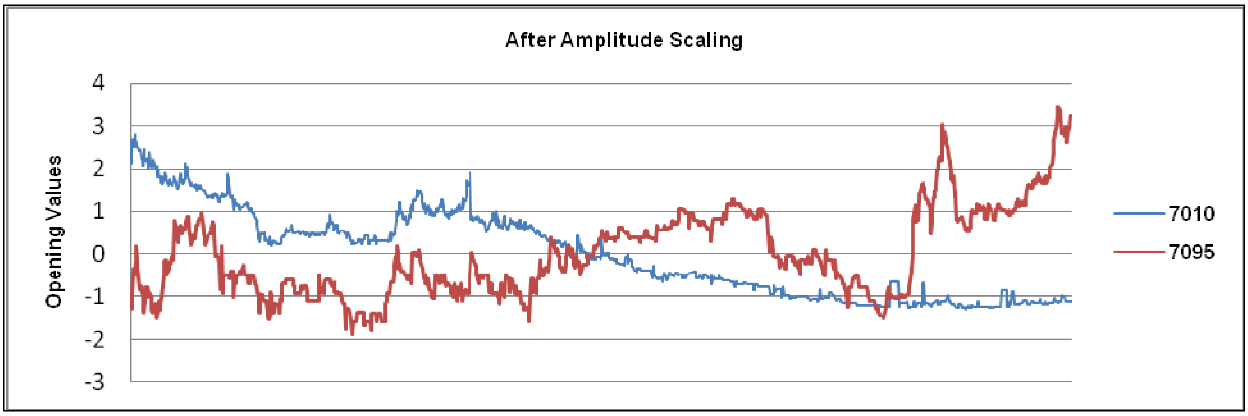


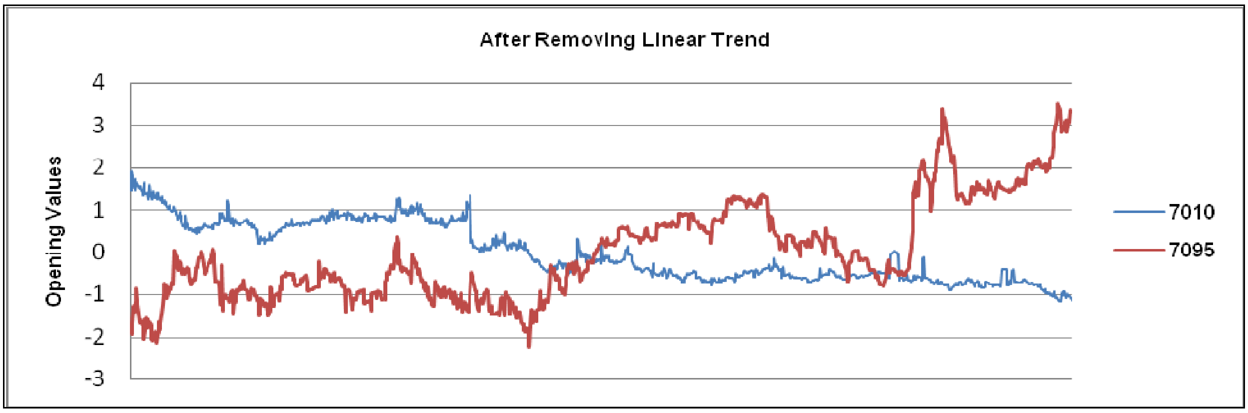
Figure 3: The original time series for the opening values of stocks 7010 and 7095.



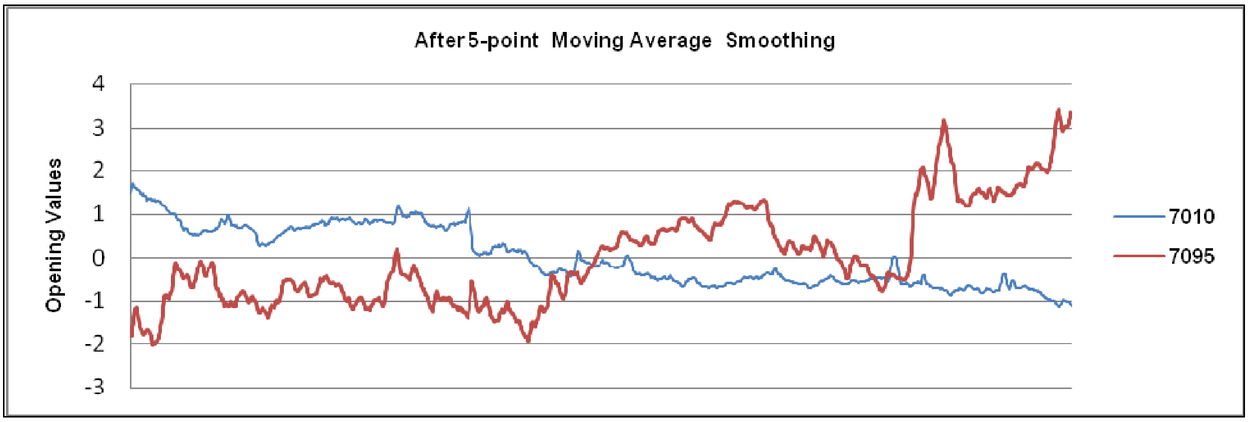
(a) After offset translation



(b) After amplitude scaling



(c) After removing linear trend



(d) After 5-point moving average smoothing

Figure 4: The time series for the opening values of stocks 7010 and 7095 after each data transformation.

Both removal of linear trend and noise are performed using Matlab. To better explain the effects of these data transformations, Figure 3 shows the original time series for the opening values of stocks 7010 and 7095 prior to any transformation, while Figure 4 illustrates the changes after each data transformation process. The final version of time series for the opening values of both 7010 and 7095 are shown in Figure 4(d). Our numeric stock dataset is now ready for similarity search.

3.3 Data Pre-Processing for Symbolic Representation

For the symbolic data representation of our stock dataset, let C represents all the stocks and indices, the performance of a particular stock or index c_i is recorded in one line l_i such that

$$l_i = \{c_i, p_1, p_2, \ldots, p_n\} \tag{2}$$

where $c_i \in C$, $n = 1233$ and $p_i \in \{UP, DOWN, SAME\}$. *UP*, *DOWN* and *SAME* denote positive, negative and unchanged performance of a stock or index respectively. For example, the performance of COMPOSITE in Table 1 is DOWN. After the conversion, we have only one file where each row is dedicated to the performance of a specific stock or index, as shown in Figure 5.

```
COMPOSITE,up,down,up,up,down,up,down,
MINING,down,up,up,up,down,same,up,dow
PLANTATN,down,down,up,up,down,up,down
PROPERTY,down,up,up,up,down,up,down,d
FINANCE,down,up,up,up,down,up,down,do
TRADING,up,down,up,down,down,up,up,do
CONSTRTN,up,down,up,up,up,up,down,dow
IND_PRODUCTS,up,down,up,up,up,up,down
CONSUMER,up,down,up,up,up,up,down,dow
5959,up,same,up,up,down,up,up,down,up
5959WA,same,same,up,up,up,up,up,down,
7054,up,up,up,up,up,up,down,down,up,d
```

Figure 5: Symbolic representation for stock dataset.

4 Similarity Search in Numeric Representation

Let nl_{o_x} and nl_{o_y} be the two normalized or transformed time series for the opening values of stocks c_x and c_y, Euclidean distance D is measured by

$$D(nl_{o_x}, nl_{o_y}) = \sqrt{\sum_{i=1}^{1233}(nl_{o_{x_i}} - nl_{o_{y_i}})^2}$$

Although we can use squared Euclidean distance to optimize the processing time, it does not make significant difference given the volume of our dataset. Likewise, similarity search on time series for closing, highest, lowest values, trading volume and price changes are performed using the same formula.

Table 2 to 7 list the top 10 most similar pairs of stocks or indices in terms of opening, closing, highest, lowest values, trading volume and price changes. *R* ranks the stock or index pairs according to their respective distance values *D*, whereas the relationship between the stocks or indices in each pair is described by *Remarks* (Malaysia Exchange 2007).

R	Stock/Index Pairs		D	Remarks
1	3816	3816F	2.51	Sister companies
2	Composite	Trading	3.78	Indices
3	2453	2569	3.86	Same industry (Plantation)
4	1295	1295F	4.05	Sister companies
5	Plantation	1961	4.26	1961 belongs to Plantation index.
6	5398X	5398WC	5.05	Sister companies
7	2577	2577PA	5.33	Sister companies
8	2445	6823	5.57	Same industry (Plantation)
9	7061	7017W	5.86	Industrial Products -Warrants/Loans
10	Composite	Finance	5.89	Indices

Table 2: Top 10 most similar stocks/indices by opening value.

R	Stock/Index Paris		D	Remarks
1	3816	3816F	2.54	Sister companies
2	Composite	Trading	3.77	Indices
3	2453	2569	3.87	Same industry (Plantation)
4	1295	1295F	4.03	Sister companies
5	Plantation	1961	4.57	1961 belongs to Plantation index.
6	5398X	5398WC	5.09	Sister companies
7	2577	2577PA	5.38	Sister companies
8	2445	6823	5.66	Same industry (Plantation)
9	7061	7017W	5.8	Industrial Products - Warrants/Loans
10	Composite	Finance	5.82	Indices

Table 3: Top 10 most similar stocks/indices by closing value.

R	Stock/Index Pairs		D	Remarks
1	3816	3816F	2.51	Sister companies
2	Composite	Trading	3.79	Indices
3	2453	2569	3.84	Same industry (Plantation)
4	1295	1295F	4.08	Sister companies
5	Plantation	1961	4.33	1961 belongs to Plantation index.
6	5398X	5398WC	5.12	Sister companies
7	2577	2577PA	5.46	Sister companies
8	7061	7017W	5.56	Industrial Products - Warrants/Loans
9	2445	6823	5.7	Same industry (Plantation)
10	9903	9903W	5.79	Sister Companies

Table 4: Top 10 most similar stocks/indices by highest value.

Obviously, the top 5 most similar pairs of stocks or indices in Table 2 to 5 are consistent. These stocks are highly similar mainly due to the fact that they are either sister companies or they belong to the same industry. Besides, indices COMPOSITE and TRADING also performed similarly. Rows in these tables where similar stocks are neither sisters companies nor belonging to the same industry are highlighted.

Note that although 7061 and 7017W belong to different industries, both of them behaved similarly by opening, closing, highest and lowest values within December 31, 2001 to December 31, 2006. Figure 6 illustrates the time series of the opening values of these two stocks, where the time series are obviously close to each other. Likewise, stocks 7017W and 4103 demonstrate highly similar trend in lowest values in spite of belonging to different industries, as illustrated in Figure 7.

R	Stock/Index Pairs		D	Remarks
1	3816	3816F	2.44	Sister companies
2	Composite	Trading	3.7	Indices
3	2453	2569	3.91	Same industry (Plantation)
4	1295	1295F	3.91	Sister companies
5	Plantation	1961	4.32	1961 belongs to Plantation index
6	2577	2577PA	4.66	Sister companies
7	5398X	5398WC	4.99	Sister companies
8	2445	6823	5.46	Same industry (Plantation)
9	7017W	4103	5.61	Warrants/Loans – Trading/Services
10	7061	7017W	5.81	Industrial Products - Warrants/Loans

Table 5: Top 10 most similar stocks/indices by lowest value.

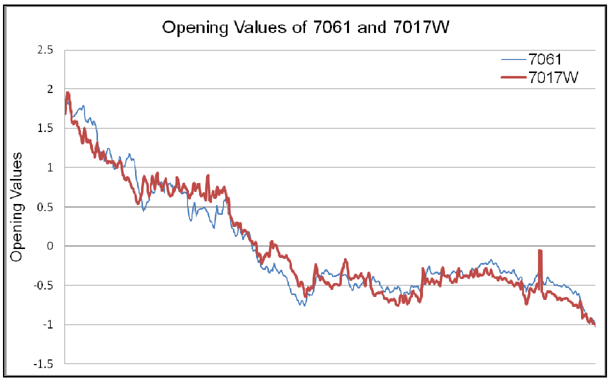


Figure 6: The time series for the opening values of stocks 7061 and 7017W.

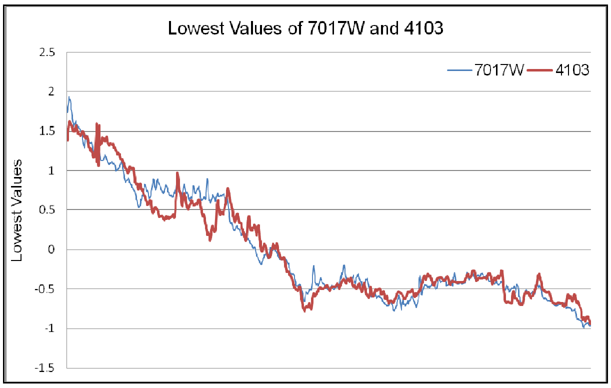


Figure 7: The time series for the lowest values of stocks 7017W and 4103.

R	Stock/Index Pairs		D	Remarks
1	1163	1252	0.64	Same industry (Finance)
2	7044	7044W	6.22	Sister companies
3	9903	9903W	7.3	Sister companies
4	Trading	Industrial Products	8.82	Indices
5	7005	7927	9.1	Industrial Products - Trading/Services
6	1945	1945WB	9.24	Sister companies
7	9474	9474W	10.28	Sister companies
8	Property	Industrial Products	11.85	Indices
9	1481	1481WA	12.4	Sister companies
10	6246	9628	12.82	Properties – Construction

Table 6: Top 10 most similar stocks/indices by trading volume.

R	Stock/Index Pairs		D	Remarks
1	1163	1252	0.42	Same industry (Finance)
2	1252	4448P	1.5	Finance – Industrial Products
3	1163	4448P	1.7	Finance – Industrial Products
4	Composite	Trading	1.76	Indices
5	5916	1252	2.08	Industrial Products – Finance
6	1163	5916	2.1	Finance – Industrial Products
7	5916	4448P	2.33	Same industry (Industrial Products)
8	Composite	Finance	2.6	Indices
9	1252	7081	2.62	Finance – Trading/Services
10	4847	9911	2.64	Same industry (Trading/Services)

Table 7: Top 10 most similar stocks/indices by price changes.

On the other hand, the similarity search on our numeric stock data with regard to trading volume produces slightly different result (Table 6) as compared to Table 2 to 5. For an instance, the most similar pair by trading volume does not appear in Table 2 to 5. Besides, despite recording a distance of 9.1 according to the trading volume (fifth most similar pair), 7005 and 7927 record distance values of more than 19 by opening, closing, highest and lowest values.

Referring to Table 7 which shows the top 10 most similar pairs of stocks or indices by price changes, only 3 pairs of them belong to the same industry. In fact, none of the top 100 most similar pairs by price changes are sister companies. Nevertheless, the first 3 pairs in Table 7 obviously demostrate one of the important properties of distance measures, which is triangular inequality, where

$$D(1163,1252) \leq D(1163,4448P) + D(1252,4448P)$$

In short, the results of the similarity search on numeric stock data with regard to the opening, closing, highest and lowest values are relatively consistent. As such, we may conclude that the opening, closing, highest and lowest values of the stock dataset are more suitable to be used as the parameters when performing similarity search, compared to the trading volume and price changes of the stocks or indices.

5 Similarity Search in Symbolic Representation

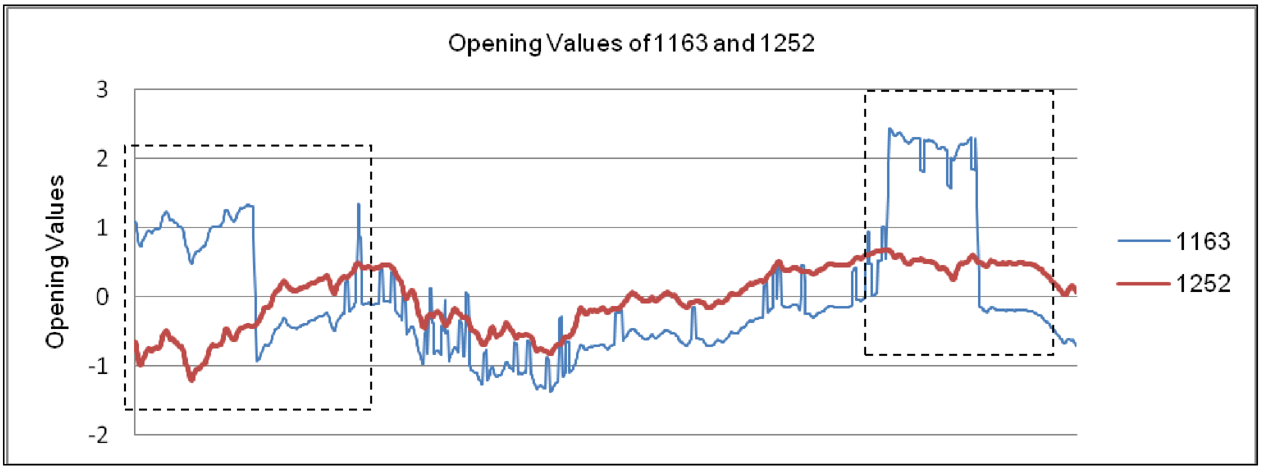
For the symbolic representation of stock dataset in format (2), the similarity search among stocks or indices c_i and c_j are evaluated by using the following formula

$$D(c_i, c_j) = \frac{d - m}{d}$$

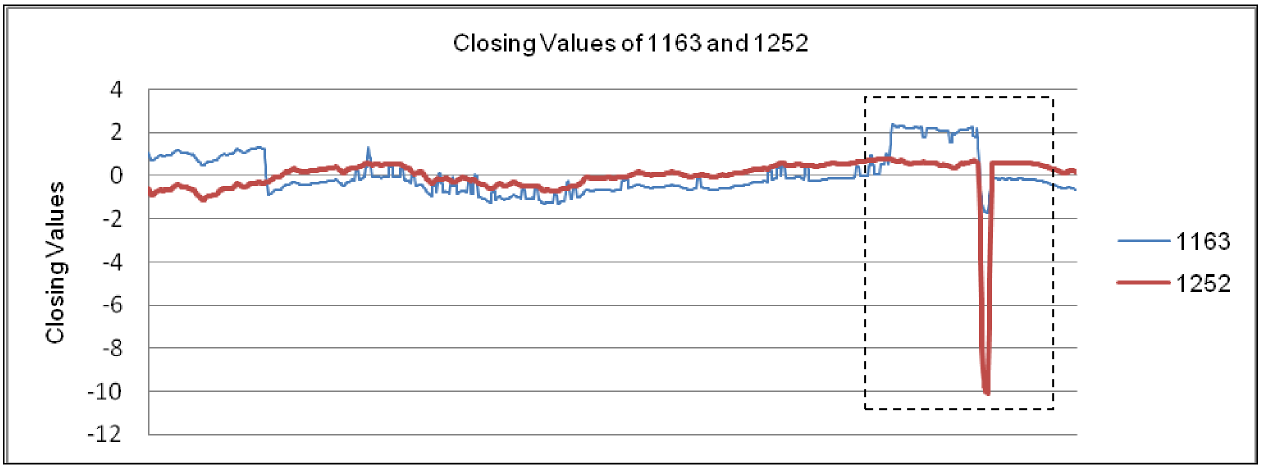
where d is the total number of variables and m is the number of matching symbols (*UP*, *DOWN* and *SAME*) occurred between c_i and c_j.

R	Stock/Index Pairs		D	Remarks
1	1163	1252	0	Same industry (Finance)
2	1163	4448P	0	Finance – Industrial Products
3	1252	4448P	0	Finance – Industrial Products
4	5185X	1163	0.01	Warrants/Loans - Finance
5	5185X	1252	0.01	Warrants/Loans - Finance
6	5185X	4448P	0.01	Warrants/Loans – Industrial Products
7	4057LB	1163	0.03	Warrants/Loans - Finance
8	4057LB	1252	0.03	Warrants/Loans - Finance
9	4057LB	4448P	0.03	Warrants/Loans – Industrial Products
10	5185X	4057LB	0.04	Same industry (Warrants/Loans)

Table 8: Top 10 most similar stocks/indices using symbolic stock dataset



(a) Opening values.



(b) Closing values.

Figure 8: The time series of stocks 1163 and 1252.

As aforementioned, the measure used on symbolic dataset is widely used in calculating dissimilarity or similarity among categorical variables in cluster analysis. In our case, *d* is 1233 since our stock dataset are recorded in 1233 days. Table 8 shows the top 10 most similar pairs of stocks or indices produced by similarity search using the symbolic stock data.

Since the symbolic representation of the stock dataset is mainly based on the price changes of the stocks and indices, Table 7 is the most suitable to be compared with the result listed in Table 8. Note that both Table 7 and 8 share the same first three most similar pairs of stocks. However, the subsequent pairs in both tables are totally different. The main reason is because symbolic representation merely captures the naive changes of the price (positive, negative or same) without considering the actual magnitude of price changes.

On the other hand, if we compare Table 8 with Table 2 to 6, there is also an obvious difference where out of the top 10 most similar pairs in Table 8, only the first pair appears in Table 6. Besides, only the first and 10th pairs in Table 8 are from the same industry. Note that the first 3 pairs record 0-distant among themselves. However, taking the first pair as an example, the similarity search on the numeric representation of this pair produces distance values of 32.48 by opening values, 40.68 by closing, highest and lowest values, and 0.64 by trading volume.

Although 1163 and 1252 appear as the most similar pair in Table 8 where the symbolic representation (2) is formed using their opening and closing values; a closer look at the time series for the opening and closing values of these two stocks show significant differences between them, as illustrated in Figure 8(a) and 8(b) respectively. The dotted-lined-boxes highlight significant differences in the sub-sequences.

Based on our observation on the result generated by the similarity search on symbolic representation of the stock dataset, we may conclude that the results are naively generated by considering only the superficial behavior of the stock dataset.

6 Conclusion and Future Works

In this paper, we have conducted an empirical study to compare the performance of numeric and symbolic representation of our stock dataset in terms of similarity search. Euclidean distance is applied to numeric dataset based on six parameters – opening, closing, highest, lowest values, trading volume and price changes. Our finding suggests that the opening, closing, highest and lowest values of the stocks and indices are able to produce consistent results in similarity search, as compared to trading volume and price changes. On the other hand, symbolic stock data are represented as {*UP*, *DOWN*, *SAME*} according to the price changes. The similarity is measured based on the number of matching symbols (*UP*, *DOWN* and *SAME*) between the time series. Compared to numeric dataset, similarity search performed on symbolic dataset generates less intuitive results because it captures only the superficial behavior of the stock dataset.

Although symbolic stock data is incomparable to numeric stock data in terms of similarity search, it generally provides easier interpretation. For an instance, classification rules generated by using symbolic data representation are easier to understand and may be sufficiently useful for those who merely want to classify stocks given their daily positive or negative performances.

This empirical study is part of our preliminary effort in stock data mining. Having completed this empirical study, we plan to explore the possibility of combining numeric and symbolic data representation in stock data mining, particularly on trend modeling. It is our hope to incorporate temporal semantics of the dataset, such as the evolution of the (causal) relationships among the stocks over a period of time using dynamic Bayesian network. In terms of similarity search, we are also interested to explore the application of DTW and Symbolic Aggregate Approximation (SAX) (Keogh and Ratanamahatana 2005, Lin et al. 2003).

7 References

Han, J. and Kamber, M. (2006): *Data Mining Concepts and Techniques*. Morgan Kauffman Publishers, Elsevier, San Francisco, CA.

Keogh, E. (2006): A Decade of Progress in Indexing and Mining Time Series Data. *Tutorial of the 32nd International Conference on Very Large Data Bases*, Seoul, Korea.

Keogh, E. and Ratanamahatana, C.A. (2005): Exact Indexing of Dynamic Time Warping. In *Knowledge and Information Systems*, Springer-Verlag, 358-386.

KLSE Daily http://www.klsedaily.com. (Subscription) Accessed 15 January 2007.

Kovalerchuk, B. and Vityaev, E. (2000): *Data Mining in Finance, Advances in Relational and Hybrid Methods*. Kluwer Academic Publishers, Massachusetts.

Kovalerchuk, B., Vityaev, E. and Yusupov, H. (2002): Symbolic Methodology in Numeric Data Mining: Relational Techniques for Financial Applications. In *The Computing Research Repository*.

Lin, J., Keogh, E., Lonardi, S. and Chiu, B. (2003): A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, San Diego, CA.

Malaysia Exchange http://www.bursamalaysia.com. Accessed 10 July 2007.

Roddick, J.F. and Spiliopoulou, M. (2002): A Survey of Temporal Knowledge Discovery of Paradigms and Methods. *IEEE Transactions on Knowledge and Data Engineering* **14**(41), 750–767.

Soon, L.-K. and Lee, S.H. (2007): Explorative Data Mining on Stock Data – Experimental Results and Findings. In *Proceedings of the 3rd International Conference on Advanced Data Mining and Applications*, Harbin, China, 562-569.

Ting, J., Fu, T. and Chung, F. (2006): Mining of Stock Data: Intra- and Inter-Stock Pattern Associative Classification. In *Proceedings of 2006 International Conference on Data Mining*, Las Vegas, USA, 30-36.