

Positive and Negative Local Trend Association Patterns in Analysis of Associations between Time Series

Ildar Batyrshin¹ and Valery Solovyev²

¹ Mexican Petroleum Institute, Mexico, D.F.
batyr1@gmail.com

² Kazan Federal University, Kazan, Russia

Abstract. The paper introduces new time series shape association measures based on Euclidean distance. The method of analysis of associations between time series based on separate analysis of positively and negatively associated local trends is discussed. The examples of application of the proposed measures and methods to analysis of associations between historical prices of securities obtained from Google Finance are considered. An example of time series with inverse associations between them is discussed.

Keywords: Time series shape association measure, moving approximation transform, local trend associations, hierarchical clustering, Google Finance.

1 Introduction

Many time series similarity measures have been introduced in time series data mining during last two decades [1,2,8-11, 14,15]. These measures usually used in time series clustering and similarity search in time series databases. The following examples of the similarity queries over sequence databases have been mentioned in [1]:

- Identify companies with similar pattern of growth;
- Determine products with similar selling patterns;
- Discover stocks with similar movement in stock prices.

In [3-4], it was pointed out a need in the measures of associations between time series that additionally to similarity between time series could measure inverse relationships between them. Additionally to considered above examples such measures could be used for finding sequences of competitive companies or products with inverse patterns when the rising patterns of one sequence correspond to falling patterns of another one. In [3], the measures of local trend associations (LTA) based on Moving Approximation Transform (MAT) have been introduced and examples of their application to analysis of possible relationships between elements of economic and financial systems have been considered. In [5], the method of construction of positive and negative local trend association patterns for any pair of time series based on MAT with window size $k=2$ has been proposed. An application of this method to analysis of associations between well production data in petroleum reservoirs has been

considered. In [6,7], the general methods for constructing association measures based on distance and similarity measures have been introduced. These methods generate sample Pearson’s correlation coefficient as a particular case. Based on these results in this paper the new measures of association between time series are introduced and the method of analysis of positive and negative association patterns in time series proposed in [5] is generalized on sliding window with size $k \geq 2$. The methods are demonstrated on example of time series of end-of-day prices of securities downloaded from Google Finance [13].

The paper is organized as follows. The definition and examples of time series shape association measures are considered in Section 2. Section 3 introduces new time series shape association measures. In Sections 4 and 5 these measures and local trend association patterns based on MAT are used in clustering and analysis of associations of time series from Google Finance. The last section contains conclusions.

2 Time Series Shape Association Measures

A time series of length n , ($n \geq 1$), is a sequence of real values $x = (x_1, \dots, x_n)$. Denote X a set of such time series. Suppose p, q ($p \neq 0$) are real values. Denote $q(n)$ a constant time series of the length n with all elements equal to q . We will write $x = const$ if $x = q(n)$ for some q , and $x \neq const$ if $x_i \neq x_j$ for some $i \neq j$ from $\{1, \dots, n\}$. Denote $px+q = (px_1+q, \dots, px_n+q)$, $x+y = (x_1+y_1, \dots, x_n+y_n)$.

Definition [4]. A time series shape association measure is a function $A: X \times X \rightarrow [-1, 1]$ satisfying for all $x, y \in X$ the properties:

- A1. $A(x, y) = A(y, x)$, (symmetry)
- A2. $A(x, x) = 1$, (reflexivity)
- A3. $A(x, -x) = -1$, (inverse reflexivity)
- A4. $A(x, -y) = -A(x, y)$, $x, y \neq const$, (inverse relationship)
- A5. $A(x, y+q) = A(x, y)$, for any $q \geq 0$. (translation invariance)

A shape association measure is referred to as scale invariant if it satisfies:

- A6. $A(x, py) = A(x, y)$, if $p > 0$. (scale invariance)

A sample Pearson correlation coefficient:

$$corr(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{1}$$

is an example of a translation and scale invariant time series shape association measure. The correlation coefficient is often considered as a measure of linear relationship between variables. A slightly modified example of time series from [5] given in

Table 1 and Fig. 1 shows that the correlation coefficient is not so good for measuring time series shape associations. It is reasonable to suppose that there are positive association between time series x and y and negative association between x and $-y$ but we have for them $corr(x,y) = corr(x,-y) = 0$.

Table 1. Example of three synthetic time series with $corr(x,y) = corr(x,-y) = 0$

i	1	2	3	4	5	6	7	8	9	10
x	100	80	50	60	90	150	200	250	180	140
y	200	120	10	20	40	50	60	70	40	20
$-y$	-200	-120	-10	-20	-40	-50	-60	-70	-40	-20

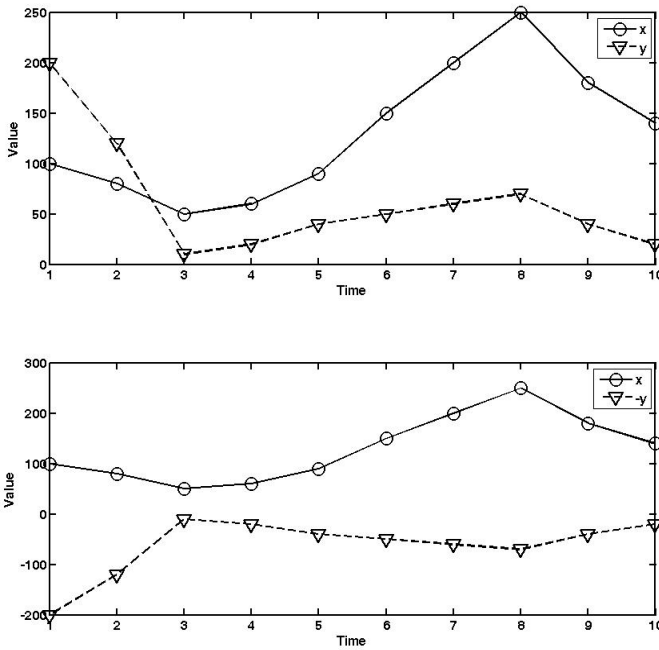


Fig. 1. Example of three synthetic time series with $corr(x,y) = corr(x,-y) = 0$

In [3], it was considered a measure of local trend associations (LTA) based on Moving Approximation Transform (MAT). MAT transforms time series $x = (x_1, \dots, x_n)$ into a sequence $MAT_k(x) = (a_1, \dots, a_{n-k+1})$ of slope values (local trends) of simple linear regressions $f_i = a_i t + b_i$, of time series values (x_i, \dots, x_{i+k-1}) , $i \in \{1, \dots, n-k+1\}$ in sliding window of size k . Suppose $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ are two time series and $MAT_k(x) = (a_{x1}, \dots, a_{xm})$, $MAT_k(y) = (a_{y1}, \dots, a_{ym})$, $k \in \{2, \dots, n-1\}$, $m = n-k+1$, are their MATs. The function (2) is called a measure of local trend associations:

$$lta_k(x,y) = \frac{\sum_{i=1}^m a_{xi} \cdot a_{yi}}{\sqrt{\sum_{i=1}^m a_{xi}^2 \cdot \sum_{j=1}^m a_{yj}^2}}, \quad k \in \{2, \dots, n-1\}. \tag{2}$$

It is easy to see that this measure for all $k \in \{2, \dots, n-1\}$ satisfies A1-A6. This measure evaluates associations between local trends and depends on the size of sliding window (see Table 2). For window size $k= 10$, the (global) trends have different signs: $MAT_{10}(x) = 0.7697$, $MAT_{10}(y) = -0.5582$, $MAT_{10}(-y) = 0.5582$ with association values $lta_{10}(x,y) = -1$, $lta_{10}(x,-y) = 1$. In spite of the global trends of x and y have different signs, i.e. x is “increasing” and y in “decreasing” with negative global trend association $lta_{10}(x,y) = -1$, the local trend associations between time series x,y (Table 2) have positive association values for small windows showing similarity in shapes of these time series and similar dynamics of them.

Table 2. Local trend associations for time series from Table 1 for all window sizes k

K	2	3	4	5	6	7	8	9	10
$lta_k(x,y)$	0.534	0.517	0.452	0.340	0.162	-0.094	-0.505	-0.895	-1
$lta_k(x,-y)$	-0.534	-0.517	-0.452	-0.340	-0.162	0.094	0.505	0.895	1

In [5], it was proposed the following shape association measure:

$$A_{UDk}(x,y) = \frac{\sum_{i=1}^m A_{xi} \cdot A_{yi}}{\sqrt{\sum_{i=1}^m A_{xi}^2 \cdot \sum_{j=1}^m A_{yj}^2}}, \quad k \in \{2, \dots, n-1\}. \tag{3}$$

such that in (2) all slope values are replaced by their sign values $A_i = \text{sign}(a_i) \in \{-1, 0, 1\}$. For example from Table 1 this measure takes value $A_{UDk}(x,y) = 1$ for window size $k=2$ that corresponds to our perceptions that time series x and y are positively associated because both of them synchronously move up and down. For time series x and $-y$ we have $A_{UD2}(x,-y) = -1$ that reflects inverse dynamics of these time series. This measure also depends on window size like original lta_k measure.

3 New Time Series Shape Association Measures

The general procedure of construction of time series shape association measures introduced in [6,7] contains several components: 1) a time series standardization $F(x)$, 2) a dissimilarity (distance) measure $D(F(x),F(y))$, 3) a transformation of D into a similarity measure S , 4) a transformation of S into an association measure A . Each component should satisfy some conditions to define shape association measure A satisfying properties A1-A5 and, perhaps, A6. For example, the sample Pearson’s

correlation coefficient (1) can be constructed using the standardization (4), the Euclidean distance (5), the similarity measure (6) and the formula (7) for an association measure:

$$F(x)_i = \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2}} \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j. \quad (4)$$

$$D(x, y) = \sqrt[2]{\sum_{i=1}^n |F(x)_i - F(y)_i|^2}, \quad (5)$$

$$S(x, y) = 1 - \frac{1}{4} D^2(x, y) = 1 - \frac{1}{4} \sum_{i=1}^n |F(x)_i - F(y)_i|^2, \quad (6)$$

$$A_{S,L}(x, y) = S(x, y) - S(x, -y). \quad (7)$$

The papers [6,7] have considered odd standardizations $F(x)$ satisfying the property: $F(-x) = -F(x)$. We introduce here shape association measures based on standardization

$$F(x)_i = \frac{x_i - \text{MIN}(x)}{\text{MAX}(x) - \text{MIN}(x)}, \text{ where } \text{MIN}(x) = \min\{x_1, \dots, x_n\}, \text{ MAX}(x) = \max\{x_1, \dots, x_n\}, \quad (8)$$

that satisfies the property: $F(x) + F(-x) = 1$. Note that standardization (8) is widely used in time series data mining. We use the similarity measure $S(x, y)$ based on Euclidean distance (9) and two association measures (10), (11), where $x, y \neq \text{const}$, (see [6,7]):

$$S(x, y) = 1 - \frac{1}{n} D^2(x, y) = 1 - \frac{1}{n} \sum_{i=1}^n |F(x)_i - F(y)_i|^2. \quad (9)$$

$$A_{S,M}(x, y) = \begin{cases} S(x, y), & \text{if } S(x, y) > S(x, -y) \\ -S(x, y), & \text{if } S(x, y) < S(x, -y) \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$A_{S,P}(x, y) = (S(x, y) - S(x, -y)) / (1 - \min(S(x, y), S(x, -y))). \quad (11)$$

The resulting association measures can be presented as follows:

$$A_{S,M}(x, y) = \text{sign}(\sum_{i=1}^n (2F(x)_i - 1)(2F(y)_i - 1)) \cdot \left[1 - \frac{1}{n} \min\{\sum_{i=1}^n |F(x)_i - F(y)_i|^2, \sum_{i=1}^n |F(x)_i + F(y)_i - 1|^2\} \right], \quad (12)$$

$$A_{S,P}(x, y) = \frac{\sum_{i=1}^n (2F(x)_i - 1)(2F(y)_i - 1)}{\max(\sum_{i=1}^n |F(x)_i - F(y)_i|^2, \sum_{i=1}^n |F(x)_i + F(y)_i - 1|^2)}. \quad (13)$$

From the construction it follows that these measures are translation and scale invariant association measures. Note that these measures and the correlation coefficient are not defined for constant time series due to division by 0 in (4) and (8).

For benchmark example given in Table 1 and Fig. 1 we have $A_{S,M}(x, y) = 0.806$, $A_{S,M}(x, -y) = -0.806$, $A_{S,P}(x, y) = 0.313$, $A_{S,P}(x, -y) = -0.313$. Note that we have for this example $\text{corr}(x, y) = 0$, $\text{corr}(x, -y) = 0$. Hence, from this point of view the new time series shape association measures give better choice than correlation coefficient.

4 Time Series Clustering Based on Association Measures

Consider an example of time series of end-of-the-day prices of securities downloaded from Google Finance [13] and denoted here as follows: 1-XRX, 2-HPQ, 3-ERIC, 4-NOK, 5-BBRY, 6-AAPL, 7-IBM, 8-LNVGY. Each time series contains 251 data measured during period 19.02.2013 – 14.02.2014 and smoothed by moving average with window size 5 (see Fig. 2). Using different association measures $A(x,y)$ we calculated associations between these time series and converted them to similarity measures: $S(x,y) = abs(A(x,y))$. Applying single linkage clustering [15] to these similarity measures we obtained the hierarchical clusterings of time series.

For association measure $A_{S,M}$ it was constructed the clustering: $\{\{\{C1, C2\} C3\}, C4\}, C5\}$ with the following two clusters joined on high level of similarity:

- C1: $\{\{1-XRX$ (Xerox), 6-AAPL (Apple)), 4-NOK (Nokia)), securities with “rising” price, and
- C2: $\{5-BBRY$ (BlackBerry), 7-IBM (IBM)), securities with “falling” price.

The clusters C1 and C2 are joined together because they have high negative associations defining high similarity between them.

- C3: $\{8-LNVGY$ (Lenovo)) is joined with these two clusters because it has high positive associations with cluster C1.
- C4: $\{2-HPQ$ (Hewlett Packard)) and C5: $\{3-ERIC$ (Ericsson)) are joined with other clusters on lower levels of similarity. C4 has positive association with 1-XRX and C5 has negative association with 2-HPQ.

For association measure $A_{S,P}$ we obtained the similar clustering. It is surprising that for correlation coefficient we obtained almost the same clustering with small difference in cluster C1: $\{\{6-AAPL$ (Apple), 4-NOK (Nokia)), 1-XRX (Xerox))

Based on the analysis of association values we can propose a hypothesis that the companies of the cluster C2 are competitive with the companies from the cluster C1 because the time series from these two clusters have negative association values. This analysis is based on association measures defined by metrics. Applying single linkage clustering to similarity measure $S(x,y) = abs(lta_{100}(x,y))$ we have obtained the following clustering of companies: $\{\{D1, D2\}, D3\}$, where

D1: $\{\{6-AAPL, 4-NOK\}, 8-LNVGY\}$,

D2: $\{\{5-BBRY, 7-IBM\}, 1-XRX\}$, where 1-XRX has high negative associations with other two time series from this cluster. D2 is joined with D1 because 1-XRX has high positive association with 6-AAPL.

D3: $\{2-HPQ, 3-ERIC\}$. These time series have sufficiently high negative association. D3 joined with D1 and D2 on lower level of similarity due to similarity between 2-HPQ and 1-XRX.

Local trend association measures give possibility to find associations between “increasing” and “decreasing” trend patterns important in many applications of time

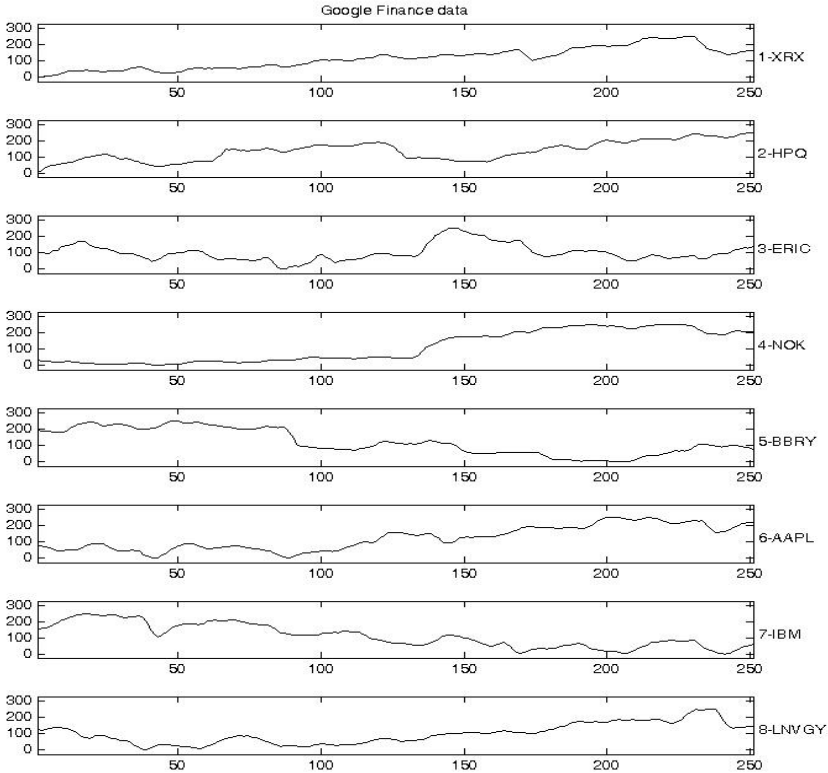


Fig. 2. Google Finance data after smoothing by moving average ($w=5$)

series analysis while distance based measures $A_{S,P}$, $A_{S,M}$ and correlation coefficient $corr$ evaluate associations between “high” and “small” values of time series with respect to their “mean” values. In the following section we propose the method of analysis of positively and negatively associated trend patterns of time series.

5 Analysis of Positively and Negatively Associated Trend Patterns in Time Series

In [5], it was proposed the method of construction of positively and negatively associated patterns of time series based on linear regressions constructed by MAT for window size $k=2$. Two linear regressions computed in the same window for time series x and y are considered as positively associated if they have the slope values with the same sign, i.e. both are positive or both are negative. These linear regressions are considered as negatively associated if they have opposite signs. In last case when one time series is increasing another one is decreasing in the same window and vice versa. The maximal sequences of positively (negatively) associated linear regressions in

consecutive windows of size 2 have been considered as positively (negatively) associated patterns for considered two time series. This method was used in [5] in analysis of associations between monthly well production data in oilfields.

An application of this technique for highly oscillating data, for example for time series of daily prices of securities in stock market, requires to smooth time series data and/or to consider local trends for windows larger than $k=2$ that also smooth data fluctuations. Here we propose the method of visualization of positively and negatively associated patterns in time series when MAT applied for windows with size greater than 2. By changing window size in MAT it is possible to do more detailed analysis of local trend associations. Separate analysis of positively and negatively associated local trend patterns for window size $k=30$ is presented in Fig. 3 in comparative analysis of 5-BBRY and 6-AAPL time series. The local trend association for these time series is negative and equals to -0.7433 . Positively associated and negatively associated local trends (linear regressions) of these time series are presented separately on the top and on the down of Fig. 3 correspondingly. It is interesting, that the Up-Down association measure (3) for these local trends of time series takes value zero: $A_{UDk}(x,y) = 0$, because exactly one half of local trends of time series are positively associated and another half is negatively associated. The positively associated local trends are located mostly on the left side of the domain (see two upper charts in Fig. 3). The negatively associated local trends are located mostly on the right side of the domain, see two lower charts in Fig. 3. The right part of these charts in presented also in Fig. 4. From Fig. 4 we see that only 10 local trends (11%) in this part of chart are positively associated and 79 local trends (89%) are negatively associated. An analysis of Figures 3 and 4 gives possibility to generate a hypothesis that BLACKBERRY and APPLE companies are competitive companies due to the large number of negatively associated local trends of considered time series.

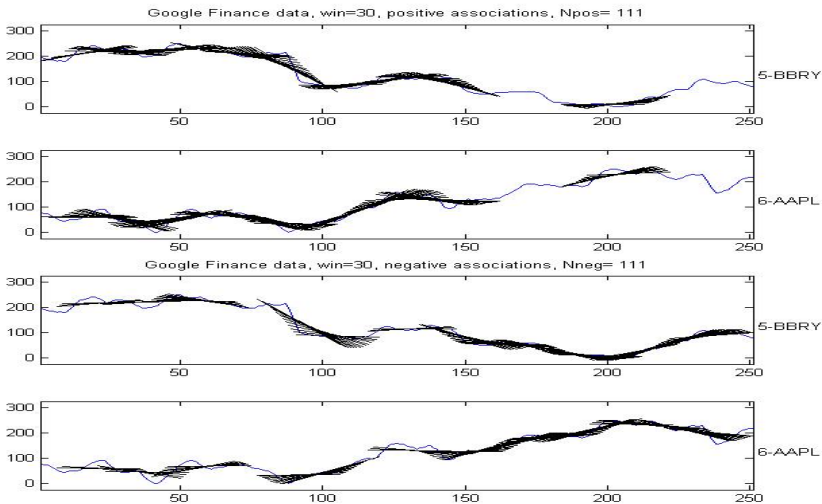


Fig. 3. Positively and negatively associated moving approximations of Google Finance data in sliding window of size $k = 30$

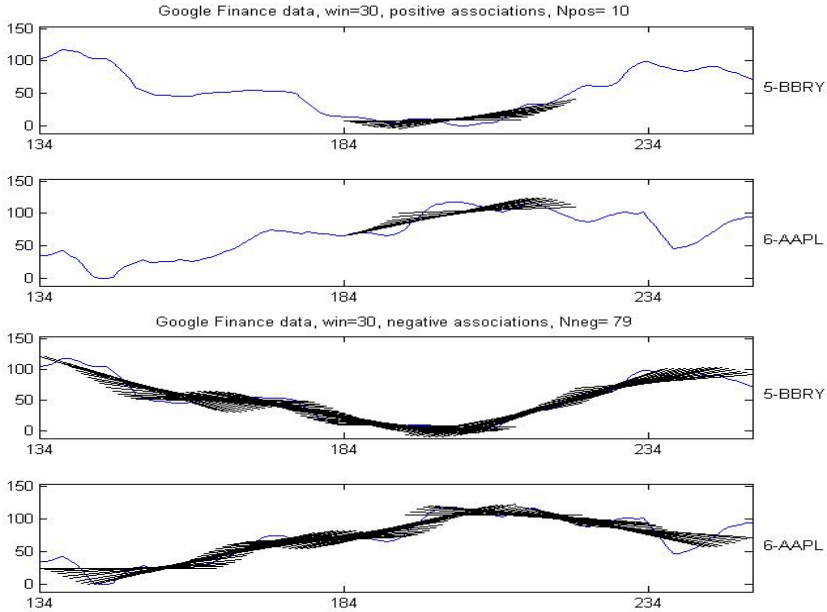


Fig. 4. The right part of Fig. 3 with high number of negatively associated local trends

6 Conclusion

The new time series shape association measures based on Euclidean distance and on popular type of data standardization are introduced and compared on synthetic benchmark example and on analysis of time series from Google Finance. The new method of analysis of positive and negative associations between time series based on separate analysis of positively and negatively associated local trends is considered. The proposed methods give possibility to generate hypothesis about possible direct and inverse relationships between complex system elements described by time series. This approach can be used for analysis of complex dynamic systems like financial, economic, industrial, ecological, meteorological, social etc. for generation of hypothesis about possible associations between system elements. Such association analysis of complex dynamic systems can be considered as a part of data-driven analysis of complex system including additional methods of data analysis.

Acknowledgments. This work was partially supported by IMP projects D.00507 and H.0016.

References

1. Agrawal, R., Faloutsos, C., Swami, A.: Efficient Similarity Search in Sequence Databases. In: Lomet, D.B. (ed.) FODO 1993. LNCS, vol. 730, pp. 69–84. Springer, Heidelberg (1993)
2. Agrawal, R., Lin, K.-I., Sawhney, H.S., Shim, K.: Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. In: 21st International Conference on Very Large Databases, pp. 490–501. Morgan Kaufmann, San Francisco (1995)
3. Batyrshin, I., Herrera-Avelar, R., Sheremetov, L., Panova, A.: Moving Approximation Transform and Local Trend Associations in Time Series Data Bases. In: Batyrshin, I., Kacprzyk, J., Sheremetov, L., Zadeh, L.A. (eds.) Perception-based Data Mining and Decision Making in Economics and Finance. SCI, vol. 36, pp. 55–83. Springer, Heidelberg (2007)
4. Batyrshin, I., Sheremetov, L., Velasco-Hernandez, J.X.: On Axiomatic Definition of Time Series Shape Association Measures. In: Workshop on Operations Research and Data Mining, ORADM 2012, Cancun, pp. 117–127 (2012)
5. Batyrshin, I.: Up and Down Trend Associations in Analysis of Time Series Shape Association Patterns. In: Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Olvera López, J.A., Boyer, K.L. (eds.) MCPR 2012. LNCS, vol. 7329, pp. 246–254. Springer, Heidelberg (2012)
6. Batyrshin, I.: Constructing Time Series Shape Association Measures: Minkowski Distance and Data Standardization. In: BRICS CCI 2013, Brasil, Porto de Galhinas (2013)
7. Batyrshin, I.: Association Measures and Aggregation Functions. In: Castro, F., Gelbukh, A., González, M. (eds.) MICAI 2013, Part II. LNCS, vol. 8266, pp. 194–203. Springer, Heidelberg (2013)
8. Das, G., Gunopulos, D.: Time Series Similarity and Indexing. In: Handbook on Data Mining, pp. 279–304. Lawrence Erlbaum Associates (2003)
9. Fu, T.-C.: A Review on Time Series Data Mining. Engineering Applications of Artificial Intelligence 24, 164–181 (2011)
10. Goldin, D.Q., Kanellakis, P.C.: On Similarity Queries for Time-Series Data: Constraint Specification and Implementation. In: Montanari, U., Rossi, F. (eds.) CP 1995. LNCS, vol. 976, pp. 137–153. Springer, Heidelberg (1995)
11. Liao, T.W.: Clustering of Time Series Data – A Survey. Pattern Recognition 38, 1857–1874 (2005)
12. Rafiei, D., Mendelzon, A.O.: Querying Time Series Data Based on Similarity. IEEE Transactions on Knowledge and Data Engineering 12, 675–693 (2000)
13. Google Finance, Historical Prices, <http://www.google.com/finance> (February 19, 2013–February 14, 2014)
14. Buza, K., Nanopoulos, A., Schmidt-Thieme, L.: Fusion of Similarity Measures for Time Series Classification. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part II. LNCS, vol. 6679, pp. 253–261. Springer, Heidelberg (2011)
15. Everitt, B.S., Landau, S., Leese, M., Stahl, D.: Cluster Analysis, 5th edn. John Wiley & Sons, Ltd., Chichester (2011)