Enhanced Symbolic Description in Analyzing Patterns and Volatility on the Forex Market

3 authors:

Krzysztof Kania
University of Economics in Katowice
20 PUBLICATIONS   **24** CITATIONS

Przemysław Juszczuk
University of Economics in Katowice
37 PUBLICATIONS   **88** CITATIONS

Jan Kozak
University of Economics in Katowice
43 PUBLICATIONS   **203** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project — Modelling of phase transformations and element speciation in pyrometallurgical slags of different chemistry in atmospheric pressure and temperatures in the range of 800-1500 deg. C View project

Project — GenEvANN View project

World Scientific
www.worldscientific.com

Enhanced Symbolic Description in Analyzing Patterns and Volatility on the Forex Market

Krzysztof Kania, Przemysław Juszczuk* and Jan Kozak

Department of Knowledge Engineering
Faculty of Informatics and Communication
University of Economics in Katowice
Ul. 1 Maja 50, 40-287 Katowice, Poland
*przemyslaw.juszczuk@ue.katowice.pl

In this paper, we propose a novel approach for transforming financial time-series values into symbolic representation based on value changes. Such approach seems to have few advantages over the existing approaches; one of the most obvious is noise reduction in the data and another one is possibility to find patterns which are universal for investigating different currency pairs. To achieve the goal we introduce a preprocessing method that allows initial data transformation. We also define a text-based similarity measure which can be used as an alternative for methods allowing to find the exact patterns in the historical data. To effectively evaluate our method, we present a concept that allows to predict the potential price movement direction and compare it with the actual price direction observed in the historical data. Such a method gives an opportunity not only to indicate the different price patterns based on the symbolic representation but also at the same time evaluate the predictive power of such patterns. The proposed approach is experimentally verified on 10 different currency pairs, each covering approximately a period of 10 years.

Keywords: Symbolic representation; data prediction; pattern recognition.

1. Introduction

Many different theories suggest that the past behavior of the price includes enough information to efficiently predict the future price movement. The common statement "history repeats itself" allows to assume that identifying some price patterns in the past and further observing these patterns in the future may lead to a similar market behavior. In the efficient market hypothesis introduced originally by Fama in Ref. 1, the above patterns-related statements were undermined. Another author, Malkiel,

in his work[2] suggested that the observed data seem to strongly support the random-walk hypothesis. The basic idea behind these works was the assumption that it is unlikely to derive an efficient trading system capable of achieving promising results related to the price prediction.

Nowadays, these above-mentioned works seem to be questioned in some way and works related to the Adaptive Markets Hypothesis[3] can be interpreted as the counter-theory to those proposed by Fama and Malkiel in recent years. Hence, focusing more on the Forex market, there exist even newer works suggesting that the trading systems adapted to the Forex market can actually be effective.[4] Taking into account these observations leads to the conclusion that deriving an effective trading system capable of achieving profits on the market is possible and could be done on the basis of some set of rules. Such an approach can be understood as the rule-based trading system, where a set of rules is derived on the basis of three completely different philosophies:

- fundamental analysis based on macro- and micro-economic, financial and other qualitative and quantitative factors that may affect the instrument's value;
- sentiment analysis based on investigating the overall attitudes, emotions, opinions and other subjective impressions that may have an influence on the market; and
- technical analysis based on the assumption that history repeats itself, so in similar situations markets behave similarly, and on finding the unknown underlying patterns developing over time, that may be useful in prediction of the market movement.

Thus the main idea behind this concept is to derive the trading system based on the set of rules. Examples of such system can be found in Ref. 5. However, moving towards the remaining two, sentiment analysis and fundamental analysis, can be a lot more difficult and includes often the aspects like working with the unstructured data or text mining. Examples of such system are described in Ref. 6.

Numerous studies were focused on deriving some efficient rules on the basis of the time-series data. The method proposed in this paper comes under the technical analysis approach as it uses values taken from the time series itself and is based on the description created from these values.

Finding patterns in the financial data is among one of the most popular trading concepts available for investors. Concepts related to identifying patterns, like the flag, "head and shoulders" and triangle formation, directly on the chart are still discussed and addressed in multiple publications.[7–9] More complex patterns involve using additional techniques like Fibonacci retracement levels based on the Fibonacci numbers.[10,11] Most of these techniques are more or less based on older publications related to discovering some patterns in price like Gartley harmonic formation first introduced in Ref. 12 and further extended in modern research.[13]

Japan candle formations are a valuable source of information related to patterns as well. However, in the case of candle formations, most of the works are focused

directly on identifying some similarities between the present price and patterns observed in historical data. There are a few papers covering the aspect of finding approximate patterns among the candle formations. This was done e.g. with the use of self-organizing maps.[14] The presented concept is based on the description, so besides the basic information, additional data related to patterns is introduced.

In this paper, we aim to connect the concepts related to finding patterns on the chart with the extension of the presented data by the additional simple symbolic pattern. The main advantage of such action is a possibility to find many patterns instead of complex price sequences occurring rarely and only for selected instruments. We can assume that patterns found for the single instrument can occur as well in other cases. We investigate this statement on the basis of different currency pairs from the Forex market.

From the point of view of investors who rely on technical analysis, symbolic description created with our method has one more advantage as well — it gives a clear and obvious interpretation visualization (picture) of the market. It makes possible to judge the situation on the market very quickly. For most of the investors, the crucial question is the direction of changes in a short or long perspective (depending on investment horizon).

To achieve the goal, we first introduce a simple preprocessing mechanism that allows to transform the present price chart into a sequence of symbols describing the relative price change between not only two neighboring values. Next, we define a text-based distance conceptually similar to the Hamming distance, which we use not only to determine the differences between two chart fragments but also to calculate the degree of difference. Finally, we verify the usability of patterns found in the experiments by observing the price change after the pattern has occurred and comparing it with the same observations for different currency pairs. This is basically seeking the confirmation that patterns found for the single currency pair are also usable for different currency pairs.

The paper is organized as follows: first we discuss the concept of the symbolic aggregation in data; next, we present the methodology for transforming the initial data into the symbolic data and we define the conditions for attributing whether the selected pattern has a rising or falling trend. In Sec. 4 we describe in detail the way to evaluate the single pattern in terms of its predictive capabilities. The last part of the paper covers the details of the experiments and a short discussion.

2. Symbolic Aggregation and Distance in Financial Data

Efficient and accurate similarity searching in a large time series is an important but nontrivial problem. Many dimensionality reduction techniques have been proposed for the effective representation of time-series data in order to realize such similarity searching, including Singular Value Decomposition (SVD), the Discrete Fourier Transform (DFT), the Adaptive Piecewise Constant Approximation (APCA), the recently proposed Symbolic Aggregate Approximation (SAX) or the Extended

approach based on SAX (ESAX). The SVD method is based on the reduction of the matrix dimensionality used e.g. in image processing.[15] Thus the rule-based trading system consists of a number of technical indicators that could be used to generate the rules, there is a possibility to reduce its number on the basis of statistical methods. The Principal Component Analysis (PCA) could be effectively used here to remove market indicators, for which the information derived to the system is not significant. At the same time, calculating the correlation among different indicators could lead to similar assumptions and eventually, to reduce the complexity of the trading system.

Among one of the most interesting approaches towards the time-series analysis is the use of the Discrete Fourier Transformation. The original idea in this case is to find the magnitude and location of the tones to build the actual signal. By increasing the number of tones, it is possible to more precisely recreate the original price behavior on the chart. The idea behind this concept is close to that used for the moving average indicator, where the price is smoothed on the basis of some predefined number of historical values. The idea of using the Fourier transformation in the time-series analysis was used e.g. in Ref. 16.

The best-known proposition for time-series transformation to the symbolic description is the SAX approach originally proposed in Ref. 17. It allows a very good dimensionality reduction together with distance measures that can be used for further data mining tasks like classification or pattern finding. Its usability was proved in many areas.[18] However, the SAX algorithm uses fixed parameters of the word size and alphabet size. The algorithm does not clearly show how to define word size and alphabet size using a time-series dataset.

Sometimes, dimensionality reduction is not the main problem. Financial time-series data has its own characteristics over other time-series data. One of its special characteristics is that it is typically characterized by a few critical points and multiresolution consideration is always necessary for long-term and short-term analyses. The second one is that financial time-series data is continuous, large and potentially unbound. Moreover, especially the technical analytical methods for critical or extreme points, which the original SAX cannot handle, are very important to discover.[19] So, other propositions were developed like ESAX or TSX.[20] On the other hand, the symbolic notation also facilitates quick assessment of the current market situation and also enables automatic semantic analysis and description using the concepts used by investors such as trend, channel, formation, turning points, etc. With the growing computing power of computers, the benefits from reducing the computational complexity and the analysis time are not so important.

3. Proposed Methodology

The proposed method differs from the other methods for constructing the symbolic description. It is dedicated to the analysis of financial data; while the main purpose of methods such as SAX is to limit the multidimensionality and the ability to study a

time series based on a reduced set of data, the proposed method is used to extend the description of each value in a row, which — in our opinion — will help in better understanding of the context and broadening the scope of analyses. Whereas in other methods a set of original values is replaced with one symbol from the proposed alphabet, in our approach each value in time series is described by a string of a length of n symbols, where n depends on the analysis horizon.

In analyzing the phenomena from financial markets such as prices, share prices or stock market indicators, the values themselves are not as important as their changes. Relative changes are particularly useful when we want to compare values from different ranges. For instance, in stock market analysis a change of price from 3 to 2.5 (absolute change: -0.5, relative change: -16.67%) is more significant and important than a change from 300 to 315 (absolute change: 15.0, relative change: 5.00%). Hence to overcome the limitations of other methods and to reduce a loss of these important points, in this work we propose a novel construction of symbolic description that aims at considering that investors are interested mainly in how values change, finding patterns that predict future changes and preserving the context of the particular values in time series.

The first step of the proposed method is introducing a dictionary which assigns a predefined symbol for each relative price change. We experimented with the dictionary initially defined by the domain expert. We consider integer values from the range $\langle -5, 5 \rangle$ as symbols.

All data are price values observed at the end of the day. We define the symbolic representation of the time series as follows:

$$
\begin{aligned}
d &= (f(x_i, x_{i-1}), f(x_i, x_{i-2}), \ldots, f(x_i, x_{i-n})) \\
&= (f(\Delta x_{-1}), f(\Delta x_{-2}), \ldots, f(\Delta x_{-n})),
\end{aligned} \tag{1}
$$

where by $f(x_i, x_{i-j})$ we describe the price change between the ith and jth readings. Thus any price change defined as $f(\Delta x_{-2})$ will be transformed into a single element of the symbolic representation. Such an element is described as follows:

$$
f(\Delta(x_i)) = \begin{cases} k & \text{if } (2 \cdot k + 1) \cdot s \leq \Delta x_i, \\ \ldots \\ 1 & \text{if } s \leq \Delta x_i < 3 \cdot s, \\ 0 & \text{if } |\Delta x_i| \leq s, \text{where } s = \overline{\Delta x_{-1}}, \\ -1 & \text{if } -s \leq \Delta x_i < 3 \cdot s, \\ \ldots \\ -k & \text{if } \Delta x_i < -(2 \cdot k + 1) \cdot s, \end{cases} \tag{2}
$$

where k are some discrete values (or symbols) used for the proposed representation; and by s we define some average price change used for further calculations. Finally, any price change given by Δx_i is in some range given by s, or its multiplication. Thus the greater the price change, the higher the number (symbol) is set for the given price.

Such an approach eliminates the problem of comparing the patterns generated with data from different price ranges, and greatly reduces the number of symbols introduced in the patterns.

Contrary to the previously described methods, we aim at enlarging the description in order to not to make it shorter to obtain deeper analysis and wider context. So we propose enhanced two-dimensional description while the proposed earlier descriptions are one-dimensional.

We compare two different symbolic representations on the basis of modified Hamming distance, where not only information about the difference in the given position is calculated, but also a size of this difference does matter. We make use of the following definition:

$$\Theta(a, b) = \sum_{i=1}^{n} |a_i - b_i|, \tag{3}$$

where a and b are symbolic representations of the time series, while n is their length. To estimate the strength of any found patterns we define the support which corresponds to the strength of the given pattern.

$$\text{Supp} = \{|X| : x \in X, y \in Y, \Theta(x, y) < k\}, \tag{4}$$

where k is the acceptable value of $\Theta()$ function defined in (3). The whole concept is presented in Fig. 1.

The BUY or SELL signal, in this case, can be understood in a twofold manner. The concept of pattern finding for the symbolic representation can be a part of the

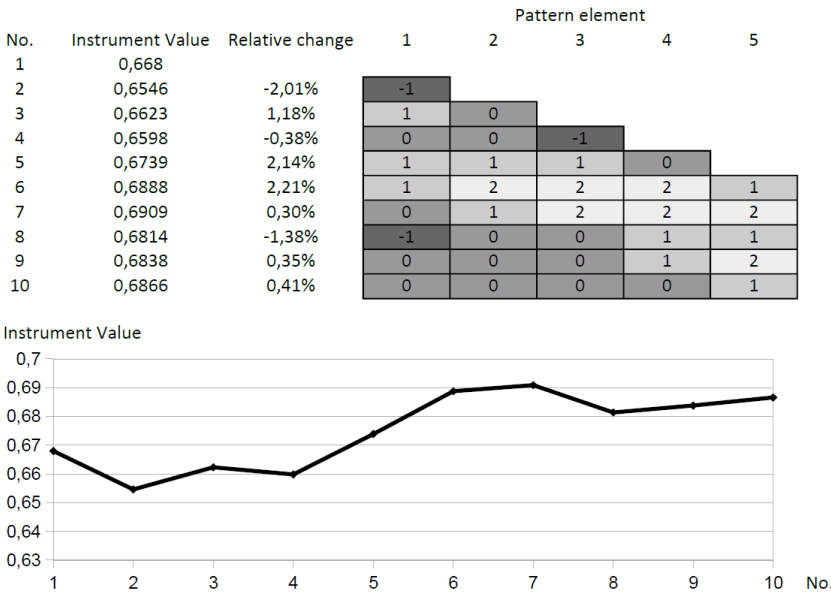| No.	Instrument Value	Relative change	1	2	3	4	5
1	0,668						
2	0,6546	-2,01%	-1				
3	0,6623	1,18%	1	0			
4	0,6598	-0,38%	0	0	-1		
5	0,6739	2,14%	1	1	1	0	
6	0,6888	2,21%	1	2	2	2	1
7	0,6909	0,30%	0	1	2	2	2
8	0,6814	-1,38%	-1	0	0	1	1
9	0,6838	0,35%	0	0	0	1	2
10	0,6866	0,41%	0	0	0	0	1

Fig. 1.   Symbolic representation of last sessions for the example instrument.

fully automatic trading system. Thus, eventual pattern match leads to the immediate opening of the transaction. One should note that details about the potential risk or profit related to the single transaction are out of the scope of this paper. We investigate if the predicted price direction corresponds to the actual change. The alternative, for the fully automatic trading system, can be a decision support system, for which the signal is derived to the decision maker for further investigation.

The general idea behind the pattern finding and the symbolic representation can be interpreted as some set of rules present in the trading system. While the trading systems based on market indicators (technical or fundamental) generate signals on the basis of rules related to these indicators, the trading system using the proposed symbolic representation uses rules connected with these patterns.

4. Pattern Evaluation

Finding a set of patterns that can be used in the trading system is a difficult task. However, an additional problem is related to the evaluation of the proposed solution in terms of the system efficiency. In literature, we can observe various aspects related to the evaluation of the generated signals. The most prominent examples are based on the account balance observed during the whole analyzed period. Such a method is easy in terms of implementation and interpretation. At the same time it allows fast verification of the proposed approach, since the positive balance of the account means positive results. However, there were publications like Ref. 21, where initially the effective system had some drawbacks when the additional transaction costs were included in the analysis.

The second most popular approach moves towards the analysis of all signals generated by the system. Calculating the number of good as well as the bad signals is relatively easy to estimate values like the accuracy.[22]

In this work we focused on patterns with a length of n, however, every considered pattern should not include any missing values. The whole available data was divided chronologically into two separate parts (learning and testing). Such an approach was necessary to further evaluate the capability of the method to find patterns and estimate their predictive capabilities.

A string of symbols from the learning set was analyzed towards finding the patterns with the given length n (we included only patterns which gave us the clear BUY or SELL signals). For every pattern found we also preserved the information about the future price direction. In this case, it was the symbol observed directly after the pattern. For example, consider the pattern with a length of 5:

$$[2, 2, 3, 4, 3]3,$$

its successor is the symbol 3 which is treated as a decision.

As a result, a set of exactly same patterns with different decisions could occur. In such a situation, the result equal to the majority of votes is selected (in the rare situation, in which the same number of observations had different decisions, the

random decision is selected). One should know that symbols 5, 4, 3 and 2 are considered as the BUY decision, while the -2, -3, -4 and -5 are the SELL decision. Remaining cases 1, 0 and -1 are WAIT decision.

In such a situation, all found patterns are written as

$$[2, 2, 3, 4, 3]5 : 6, 3 : 12, 2 : 10, 1 : 3, 0 : 6, -1 : 1, -2 : 1,$$

which is interpreted as follows: in the learning set, the pattern $[2, 2, 3, 4, 3]$ occurred 39 times, where for six times its successor was the symbol 5, 12 times its successor was the symbol 3 and so on. For such a pattern, an exact rule is defined. The premise of the rule is the whole pattern $[2, 2, 3, 4, 3]$, while its decision is BUY, because 28 occurrences of this pattern in the learning set indicated the BUY (symbols 5, 4, 3 and 2), 10 occurrences indicated the WAIT (1, 0 and -1) and finally, one symbol for the SELL. Thus the rule could be written as

$$[2, 2, 3, 4, 3] \Rightarrow \text{BUY}.$$

Such a set of rules was examined on the testing set, which was not used in the process of rules generation. During the tests, all patterns equal to the rule premise were selected. Further, the decision fitting the rule premise was taken. If the decision assigned to the rules was exactly the same, which for the BUY means the symbol greater than or equal to 2, we consider this situation as the correct classification. In any other case, we treat such observation as the incorrect classification.

5. Experiments

We experimented with 10 different currency pairs: AUDUSD, EURCHF, EURGBP, EURUSD, GBPUSD, USDJPY, NZDJPY, NZDUSD, USDCAD and USDCHF. Each covers 2,500 day sessions, which is equal to approximately 8 years: July 2007–July 2017. We investigated the patterns with lengths of five and 10 elements which correspond to five and 10 daily sessions, respectively. For all results, we used the notation $|x|y$, where x corresponds to the length of the pattern and y is the absolute value of the minimal symbol value necessary to identify the pattern. For example, $|5|2$ means all patterns of length 5 and minimal symbol value 2 or higher, for the rising trend, and -2 or lower for the falling trend. Thus, symbols -1, 0 and 1 observed in the pattern are considered as a side trend. We stick to the assumption that patterns related to side trend (consisting only of the above symbols) are out of the interest for the potential decision makers and were omitted from the analysis. The whole experiments were focused on two different aspects of the financial data. In Sec. 5.1, we analyzed the volatility of selected currency pairs. Our goal was to investigate how volatile are the analyzed data, measured by measures such as standard deviation. The volatile data are much more difficult to predict, so this can be interpreted also as "a forecast difficulty" measure. We also analyzed the return values possible to achieve during the span of 10 years. The second part of the

experiments was directly related to the generation of patterns. We divided all experiments in Sec. 5.2 into two groups:

- first group included results obtained for the exact patterns, where every found string should be identical to the pattern and include at least single element equal to or greater than 2 (or equal to or smaller than −2);
- second group included results for the string, where a minimal deviation from the original pattern (defined as in the previous section) was acceptable, where every found string should include at least four elements equal to or greater than 2 in the case of patterns of length 5 and at least eight elements equal to or greater than 2 in the case of longer patterns.

The following concepts were analyzed:

- what is the number of patterns found for every considered currency pairs?
- what is the size of the support for the considered currency pairs?
- what is the approximate effectiveness of prediction for all found patterns?

In the last problem, we investigated all patterns found for the given currency pair. Eventually, the predictive power of these patterns was investigated in the case of the nine remaining currency pairs. The main question was: how effective are the patterns found for the single currency pair in the case of all remaining currency pairs? This problem is investigated in Sec. 5.2. Thus the whole dataset including 10 different currency pairs was divided in such a way that every currency pair was used as a training set (which was 10% of all data) while the remaining currency pairs were used as a test set. The whole procedure was repeated 10 times for all considered currency pairs.

5.1. *Volatility analysis for selected currency pairs*

In this subsection, we performed some basic analysis of the selected currency pairs. The analysis covered the whole period of 10 years with daily changes of instrument value. Thus the length of the time series was over 2,500 values. In Table 1 we can see the selected statistics: average, standard deviation, minimum and maximum values along with the range. The last column shows the interquartile range used to estimate how many readings fell into the nearest surrounding of the average value.

We expect that the results should be uniformly distributed near the 0.0 value. However, in the case of USDJPY, this value is slightly moved towards the negative values. Standard deviation values are rather similar. It could be expected that in case of GBPUSD the standard deviation will be visibly higher since this instrument is considered to be very volatile. What is interesting is that the same observation is seen for the NZDUSD currency pair.

In most of the cases, minimum and maximum values observed for currency pairs are not exceeding the thresholds of −5% and 5%. During the past years, we observed

Table 1. Basic statistics for the analyzed data.

	Average	Std. dev.	Min.	Max.	Range	Interquartile range
AUDUSD	0.01	0.54	−3.22	3.17	6.39	0.58
EURCHF	−0.01	0.44	−3.03	3.41	6.44	0.33
EURGBP	0.01	0.56	−3.12	3.03	6.15	0.64
EURUSD	0.01	0.76	−3.04	3.24	6.28	0.73
GBPUSD	0.00	1.02	−5.32	6.41	11.73	0.97
USDJPY	−0.10	0.62	−3.21	3.35	6.56	0.81
NZDJPY	−0.02	1.09	−7.92	3.54	11.46	1.05
NZDUSD	0.00	0.91	−6.38	4.72	11.10	1.03
USDCAD	0.01	0.65	−3.64	3.49	7.13	0.70
USDCHF	0.00	0.68	−4.14	5.01	9.15	0.75

some stagnation for the EURCHF currency pairs leading to the side trend near the price value equal to 1.2000. This may be the reason for the very small interquartile range observed for this currency pair — which is 0.33.

To better capture the differences between currency pairs for the lowest and for the highest achieved returns we generated histograms for the selected currency pairs. In Fig. 2 we can find details about the achieved returns for six selected currency pairs. To improve readability of charts we omitted all return values in the range of $\langle-0.5, 0.5\rangle$. We can observe that despite the average values oscillating near 0.0 (see Table 1) for some currency pairs, there exists some pressure for positive (or negative) values. This can be seen e.g. for EURGBP currency pair, where positive relative changes above $+1\%$ occur more frequently than in the case of negative values. A similar situation is observed for USDCAD, while NZDJPY seems to be more negative-oriented.

Presented results showed that even among different currency pairs we can observe slightly unbalanced difficulty, where some instruments potentially could lead to higher profits. The same observation is confirmed in Table 1 where the observed results are not uniform for all currency pairs. However, in the next part of the experiments we show that despite these differences, the quality of generated patterns seems to be not affected by the difficulty of the currency pair.

5.2. *Patterns and the size of the support*

Results related to the exact patterns (columns 2 and 3 of Table 1) confirm our observation, while a number of patterns found in the case of approximate patterns do not differ significantly among shorter and longer patterns. This is an interesting observation, because an approximate pattern |5|3 means that string with the length of five elements seems to be a much stronger trend confirmation than the exact pattern with the same length. Exact longer patterns indicate that the proposed solution gave a small number of patterns, and in the case of more difficult version |10|3 we can see that there was only one currency pair with a single pattern (see Table 2), which has no use for the potential decision maker.
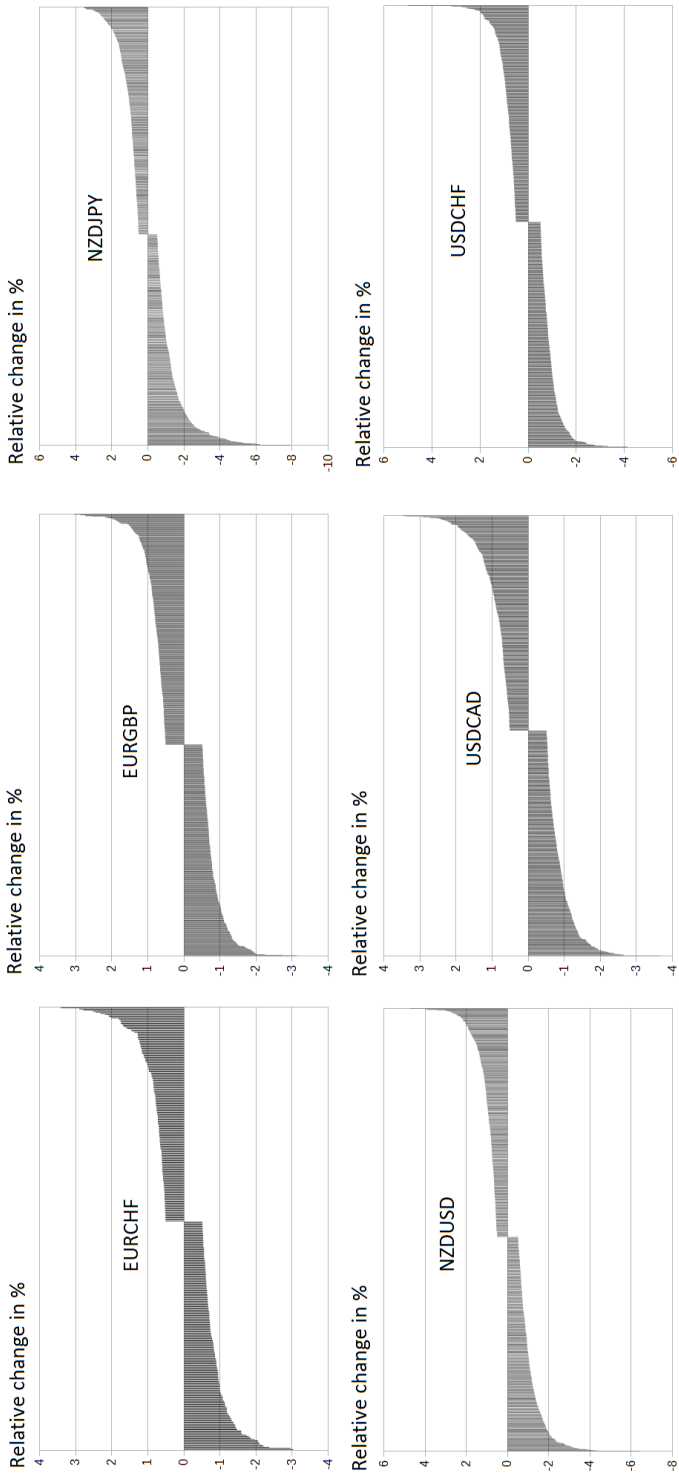
Fig. 2.   Distributions of returns for selected currency pairs.

Table 2.   The numbers of patterns observed in the case of exact and approximate solutions for all currency pairs. Exact solutions: Every found string should include at least a single element equal to or greater than 2 (or equal to or lower than −2). Approximate solutions: Every found string should include at least four elements equal to or greater than 2 in the case of patterns of length 5 and at least eight elements equal to or greater than 2 in the case of patterns of length 10.

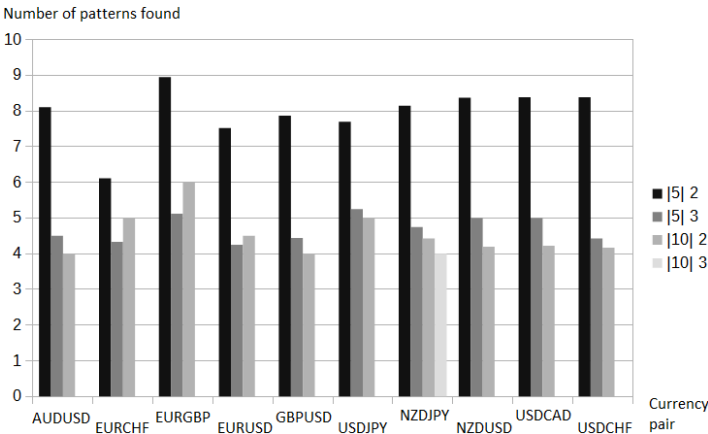| Currency pair	Exact solutions				Approx. solutions			
	$\|5\|2$	$\|5\|3$	$\|10\|2$	$\|10\|3$	$\|5\|2$	$\|5\|3$	$\|10\|2$	$\|10\|3$
AUDUSD	48	6	2	0	14	9	2	2
EURCHF	25	3	1	0	18	15	6	10
EURGBP	55	8	1	0	14	8	3	1
EURUSD	52	8	2	0	15	8	2	2
GBPUSD	53	9	3	0	14	7	2	6
USDJPY	53	4	1	0	15	10	2	0
NZDJPY	40	4	7	1	18	14	6	6
NZDUSD	54	9	5	0	13	11	0	2
USDCAD	47	7	9	0	13	8	1	0
USDCHF	55	7	6	0	13	14	1	0

Fig. 3.   The average sizes of the support for different patterns — the exact method.

We also investigated the average support of every found pattern. By the support, we understand a number of historical observations corresponding to this pattern. Support sizes in the case of exact solutions can be found in Fig. 3. We can see that in most cases (except EURCHF which is the second element of the histogram) the support size for every considered currency pair is similar. Support size for the $\|5\|2$ pattern is higher in all the remaining cases, which was expected. However, in the case of patterns $\|5\|3$ and $\|10\|2$ we observe a similar support size. Finally, the most difficult patterns identified by $\|10\|3$ were observed only in the case of single currency pair — NZDJPY.

Similar experiments were conducted for the approximate patterns, which can be seen in Fig. 4. Here we can see some interesting facts. Support size in the case of $\|5\|2$ is
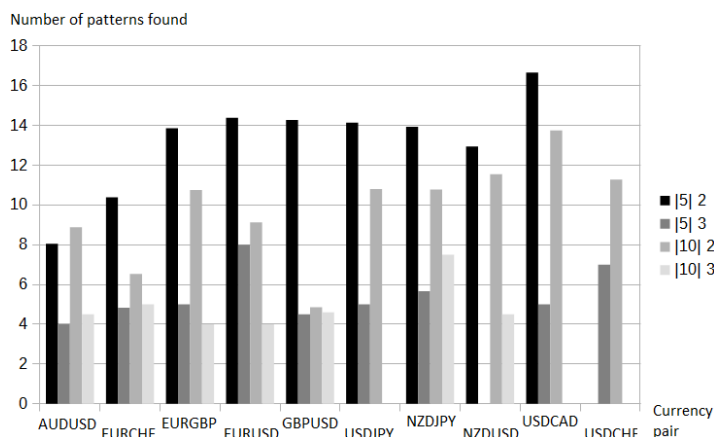
Fig. 4.   The average sizes of the support for different patterns — the approximate method.

not as concise as before. In the case of USDCHF currency pair, not even single support of the pattern was found, which is surprising. Differences between patterns |5|3 and |10|2 are also more visible. However, size of the support in the most difficult cases including patterns described as |10|3 was around 4 for six currency pairs while in one case it was slightly below 8. That clearly shows that the approach introduced in this paper based on the proposed text-based measure allows increasing the number of patterns useful for the decision maker.

Further, in Sec. 5.3, we investigate the predictive capabilities of different patterns in both the exact and approximate approaches.

5.3. *Predictive capabilities of patterns*

The efficiency of the found patterns was verified on the basis of six different cases. For every pattern, we estimated the direction of the price movement on the basis of the method introduced in Sec. 4. In this analysis, we selected only those patterns which carried the most information (to be more specific, the patterns with the largest size of support), which were |5|2.

We investigated the efficiency of patterns related to rising and falling trends for the exact and approximate patterns. Moreover, we selected only those patterns which carried the most information (to be more specific, patterns with the largest size of support), which were |5|2. We analyzed the falling trend, the rising trend and the side trend. For all these cases we investigated the efficiency of the exact and approximate methods.

We used the classical accuracy measure to estimate the number of patterns, for which the predicted price direction was exactly the same as the actual price direction. To be more specific, for patterns indicating the BUY direction we calculated a number of situations corresponding to this pattern, for which an actual price increase was found.
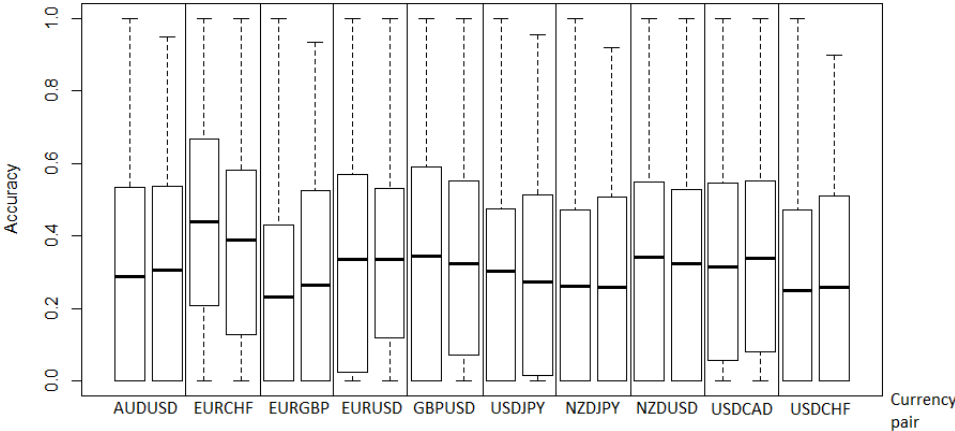
Fig. 5.   Comparison of accuracy for predicting the falling trend — the exact |5|2 pattern.

All presented charts include the information presented as boxplots for the exact patterns, where the first figure covers the patterns related to the falling trend, while the second figure represents results for the rising trend. All boxplots were divided into 10 separate fragments. Each fragment includes two boxes — first is related to the test data, while the second one is related to the training data.

In Figs. 5 and 6 we can see the results for the exact method — falling trend (Fig. 5) and rising trend (Fig. 6). For both cases, the estimated median of results is almost always below the accuracy value of 0.4. Only for one case for the EURCHF currency pair, it was possible to achieve a higher median value for the testing set. Focusing only on the right boxplots suggests that the average accuracy of the found patterns for the testing set is not very high. However, we should remember that for these results we take into account all found patterns, thus many of them could lead
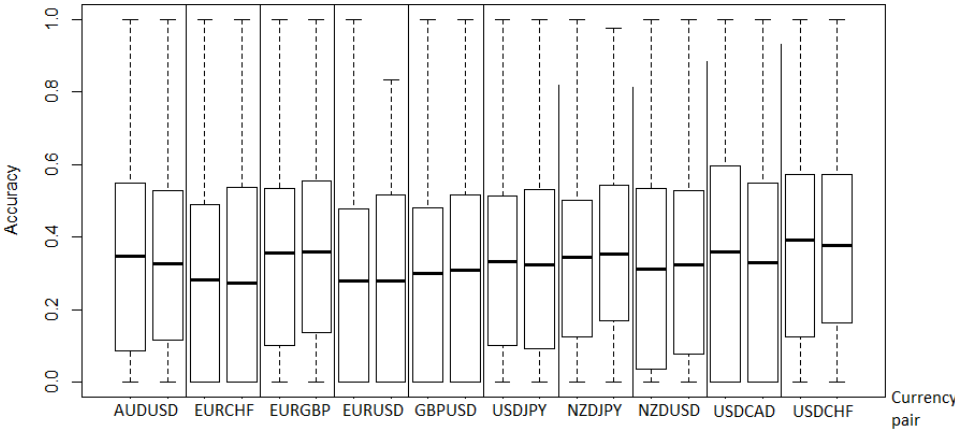

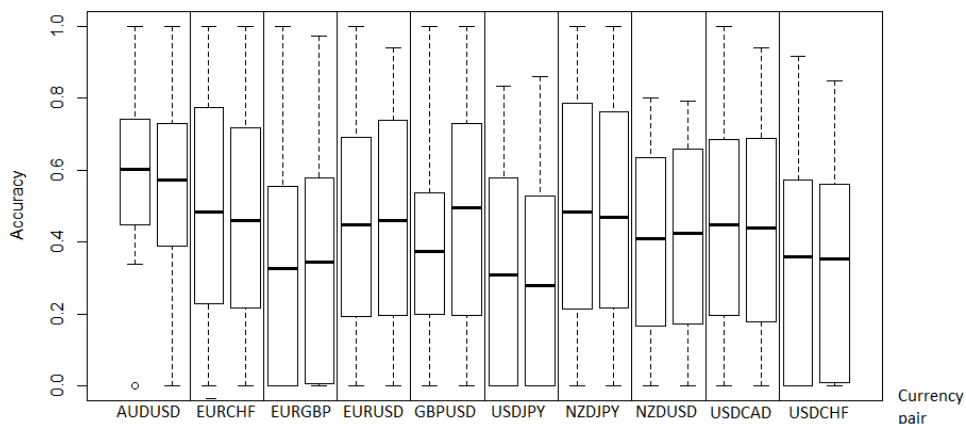Fig. 6.   Comparison of accuracy for predicting the rising trend — the exact |5|2 pattern.

Fig. 7.   Comparison of accuracy for predicting the falling trend — the approximate |5|2 pattern.

to poor predictive capability. The most important information observed on these charts is related to the highest achieved accuracy, which for almost all currency pairs is close to 1.0. Thus there are patterns in the set which exhibit very good predictive capabilities and could be used in the trading system. However, the main drawback of this approach is that we actually cannot say which of these patterns present such good efficiency. On the other hand, it is obvious that all patterns cannot be used in the system at the same time.

Assumptions made at the beginning of Sec. 5 lead to the situation in which the number of exact patterns visibly exceeds the number of approximate patterns. The accuracy values are presented in Fig. 7 (for the falling trend) and in Fig. 8 (for the rising trend). In both cases, the results are not as uniform as in case of exact patterns. The median value for both figures is mostly in the range of $\langle 0.2, 0.6 \rangle$. At the same time, the highest accuracy achieved for different currency pairs is not as high as

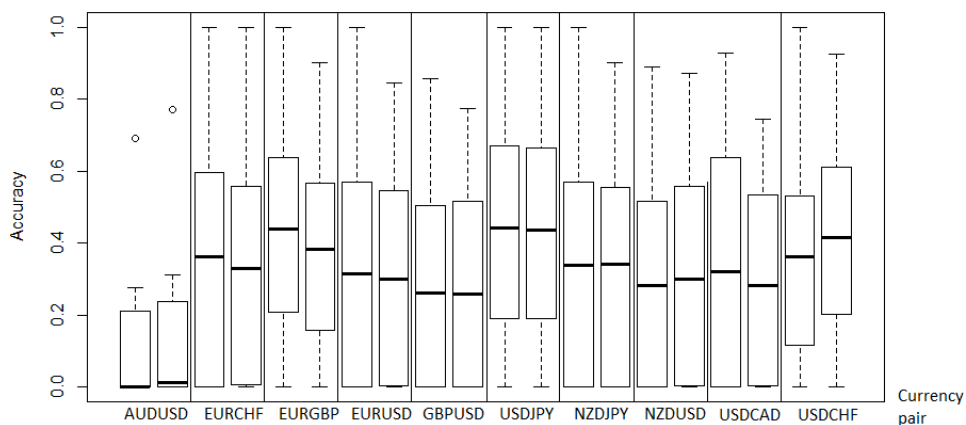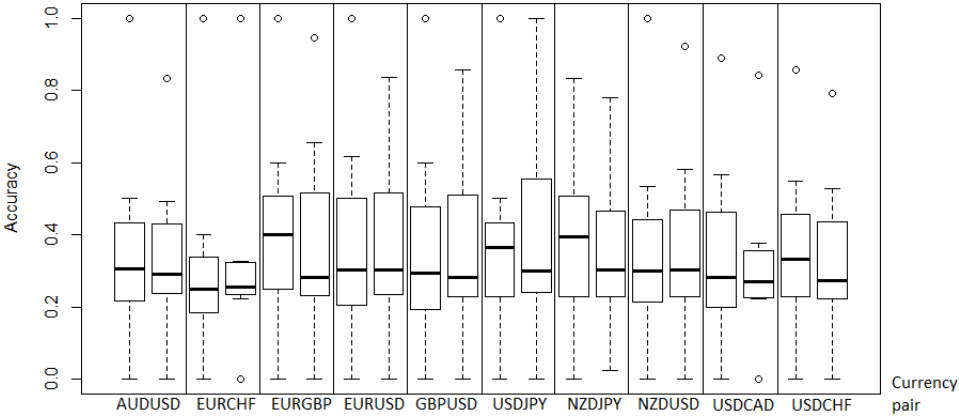Fig. 8.   Comparison of accuracy for predicting the rising trend — the approximate |5|2 pattern.

Fig. 9.   Comparison of accuracy for predicting the side trend — the exact |5|2 pattern.

it was seen for the exact patterns. It may be related to the fact that the number of patterns, in this case, is relatively small, thus it is more difficult to find a pattern with the accuracy close to 1. Differences between the first and the third quartiles (the size of the boxplots) are dependent on the selected currency pair, while the size of the exact pattern of boxplots was relatively uniform. It is worth noting that for all analyzed currency pairs there exists at least one pattern for which the observed accuracy was equal to 0 for both the learning (left boxplot) and testing (right box-plot) sets.

The last part of the experiments involved the analysis of the side trend and number of patterns fitting this observation. In Fig. 9 we can see the results for the exact pattern, while Fig. 10 shows the results for the approximate pattern. In both cases, we can see that for the most observations it was not possible to find at least one pattern with the highest accuracy of prediction of 100%. Especially in the second
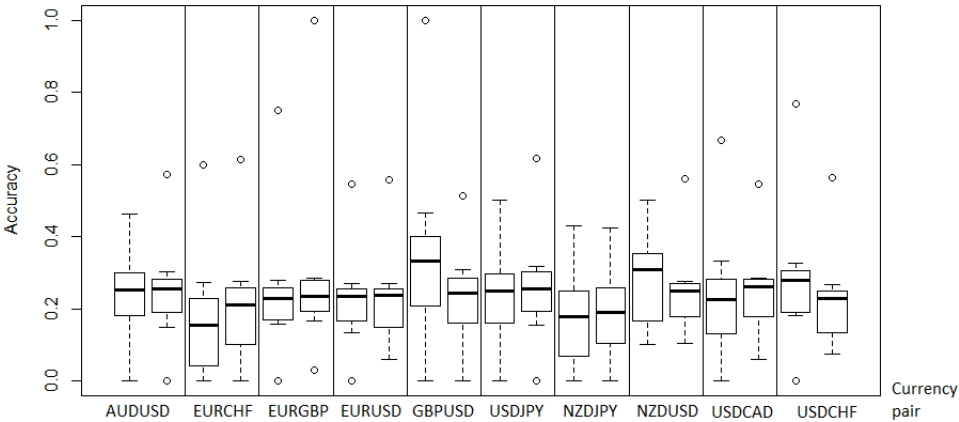

Fig. 10.   Comparison of accuracy for predicting the side trend — the approximate |5|2 pattern.

chart the observed accuracy is rather small. Moreover, almost all maximal values are shown on the chart by the dots. That means that these results are considered as outliers, thus for the most patterns the achieved accuracy was relatively small.

We can see that in the case of rising and falling trends there exist some patterns which can give very good accuracy of prediction. On the other hand, the same results cannot be confirmed for the side trend. Thus the most important problem is to indicate the time span of the rising (or falling) trend. Such identification of trend change could bring great improvement in the results.

5.4. *Case analysis*

Below, we discuss the example of using single pattern for generating the signals for the trading system. Let us assume that some currency pair curr and 20 continuous readings are given. The symbolic representation for the currency pair is obtained on the basis of the method discussed at the end of Sec. 3 and it can be seen in Fig. 11.

The left part of the figure presents the symbolic representation. We analyze the pattern with the length equal to 5, thus we omit the first four readings. The right part of the figure represents the decisions related to readings. For each reading, we give the symbolic representation of the reading following the pattern, while the second column is related to the decision. We assumed that the symbol equal to or greater than 2 means the BUY signal, symbol equal to or lower than -2 is considered as the SELL signal while the remaining values are considered as the WAIT. The example pattern can be seen at the bottom of Fig. 11.

No.	Pattern element					Decision	
	1	2	3	4	5		
1	-1					0	WAIT
2	1	0				0	WAIT
3	0	0	-1			1	WAIT
4	1	1	1	0		1	WAIT
5	0	2	2	2	2	2	BUY
6	0	1	2	2	2	1	WAIT
7	-1	0	0	1	1	2	BUY
8	0	0	0	1	2	0	WAIT
9	0	0	0	0	1	-1	WAIT
10	-1	-2	-2	-1	-1	-2	SELL
11	0	0	-1	0	1	-2	SELL
12	0	2	1	1	1	1	WAIT
13	1	2	2	2	2	2	BUY
14	3	2	3	3	2	-2	BUY
15	3	3	3	3	3	3	BUY
16	2	1	2	2	1	1	WAIT
17	1	1	2	2	1	2	BUY
18	0	0	1	1	1	0	WAIT
19	0	2	1	2	2	-2	BUY
20	0	0	-1	1	1	1	WAIT
Pattern found							
	0	2	2	2	1		

Fig. 11. Symbolic representation for 20 successive readings and the pattern example.

No.	Pattern element					Decision	
	1	2	3	4	5		
1	-1					0	WAIT
2	1	0				0	WAIT
3	0	0	-1			1	WAIT
4	1	1	1	0		1	WAIT
5	0	2	2	2	2	2	BUY
6	0	1	2	2	2	1	WAIT
7	-1	0	0	1	1	2	BUY
8	0	0	0	1	2	0	WAIT
9	0	0	0	0	1	-1	WAIT
10	-1	-2	-2	-1	-1	-2	SELL
11	0	0	-1	0	1	-2	SELL
12	0	2	1	1	1	1	WAIT
13	1	2	2	2	2	2	BUY
14	3	2	3	3	2	-2	SELL
15	3	3	3	3	3	3	BUY
16	2	1	2	2	1	1	WAIT
17	1	1	2	2	1	2	BUY
18	0	0	1	1	1	0	WAIT
19	0	2	1	2	2	-2	SELL
20	0	0	-1	1	1	1	WAIT

	Pattern found						
	0	2	2	2	1		

No.	1	2	3	4	5	Θ	Decision
5	1	2	2	2	1	1	BUY
13	1	2	2	2	2	2	BUY
19	0	2	1	2	2	1	SELL

Fig. 12.   Making the decision for all found symbolic representations corresponding to the analyzed pattern.

In Fig. 12 we select only those readings for which the actual symbolic representation matches the analyzed pattern. We assumed the Θ value as equal to or lower than 2, which means that all symbolic representations of the readings which differ by not more than 2 are accepted. We observe three such situations present in the reading number: 5, 13 and 19. We consider the BUY pattern, thus we expect that the decision after analyzing the reading is equal to the BUY decision.

We can see that such decision is observed in two out of three readings, thus the overall accuracy for this particular example is equal to 66%.

6. Conclusions

In this paper, we described a novel approach for presenting the financial data related to the Forex market with the use of symbols predefined by the domain expert. We defined a text-based similarity measure based on the concept used in the SAX measure allowing to calculate the distance between two series of symbols. We do not target to reduce the complexity of the data. We rather investigate the possibilities to extend the existing methods of visualization by adding a simple mechanism that allows to estimate the general trend direction. Moreover, this concept to represent the data can be extended in such a way that the number of symbols used in the description can be larger than the initial 11 symbols proposed in this paper. It could lead to deriving more complex patterns, which could bring great benefit for making decisions on the markets.

The experimental part of this work covered three different aspects. The first part was related to estimating the difficulty of the data. We understand difficulty as the volatility of the selected currency pair on the market. We did not investigate the number of trend changes during the whole analyzed period, however, we prepared charts showing the relative changes during two successive readings for the selected currency pairs.

The second aspect of experiments covered the problem of finding and preparing the set of rules based on the symbolic representation. We aimed to find the exact and approximate patterns of different length. We calculated a number of patterns found for every considered currency pair and we measured the support of every pattern.

Finally, we divided data into 10 sets: each set including the training set consisting of a single currency pair and the test set including nine remaining currency pairs. The last fragment of experiments covered the aspect of analyzing the predictive capabilities of generated symbolic patterns. We investigated the efficiency of patterns in the case of rising, falling and side trends. For the first two cases, results confirmed that there exist patterns with very high accuracy (even equal to 100%). For the side trend, patterns generated in the second part of the experiments failed to achieve 100% accuracy even for the selected currency pairs. Thus any decisions made during the side trend could lead to increasing the risk.

Initial results indicate that the proposed approach could be effectively used as a preprocessing mechanism allowing to eliminate the noise in the data. Symbolic presentation of the financial data greatly simplifies the process of finding the patterns in the data. However, the main problem is still to identify the moment of a trend change and identify currency pairs being in the side trend. Thus, automatic identification of the side trend could greatly improve any techniques based on the technical analysis.

References

1. E. Fama, The behavior of stock-market prices, *J. Bus.* **38**(1) (1965) 34–105.
2. B. G. Malkiel, *A Random Walk Down Wall Street* (W. W. Norton & Company, 1985).
3. A. W. Lo, Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis, *J. Invest. Consult.* **7**(2) (2005) 21–44.
4. A. Shmilovici, Y. Kahiri, I. Ben-Gal and S. Hauser, Measuring the efficiency of the intraday Forex market with a universal data compression algorithm, *Comput. Econ.* **33**(2) (2009) 131–154.
5. M. A. H. Dempster, T. W. Payne and Y. Romahi, Computational learning techniques for intraday FX trading using popular technical indicators, *IEEE Trans. Neural Netw.* **12**(4) (2001) 744–754.
6. Y. Shynkevich, T. M. McGinnity, S. A. Coleman and A. Belatreche, Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, *Decis. Support Syst.* **85** (2016) 74–83.
7. R. Cervello-Royo, F. Guijarro and K. Michniuk, Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data, *Expert Syst. Appl.* **42**(14) (2015) 5963–5975.

8.   R. Arvalo, J. Garcia, F. Guijarro and A. Peris, A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting, *Expert Syst. Appl.* **81** (2017) 177–192.

9.   T. Hartle, Triangles and trends, *Tech. Anal. Stock Commod.* **18**(2) (2000) 24–30.

10. L. Pesavento, *Fibonacci Rations with Pattern Recognition* (Traders Press, Inc., Greenville, 1997).

11. V. Gaucan, How to use Fibonacci retracement to predict Forex market, *J. Knowl. Manage. Econ. Inf. Technol.* **2** (2011) 1–14.

12. H. Gartley, *Profits in the Stock Market* (Traders Press, New York, 1935).

13. R. Teseo, The Gartley setup, *Tech. Anal. Stock Commod.* **19**(1) (2001) 40–47.

14. P. Cisowski and M. Zaton, Neural pattern recognition with self-organizing maps for efficient processing of Forex market data streams, in *Proc. Int. Conf. Artificial Intelligence and Soft Computing* (2010), pp. 307–314.

15. G. Golub and C. Reinsch, Singular value decomposition and least squares solutions, *J. Numer. Math.* **14** (1970) 403–420.

16. M. Y. Chen and B. T. Chen, Online fuzzy time series analysis based on entropy discretization and a fast Fourier transform, *Appl. Soft Comput.* **14** (2014) 156–166.

17. J. Lin, E. Keogh, S. Lonardi and B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, in *Proc. 8th ACM SIGMOD Workshop Research Issues in Data Mining and Knowledge Discovery* (2003), pp. 2–11.

18. J. Lin, E. Keogh, L. Wei and S. Lonardi, Experiencing SAX: A novel symbolic representation of time series, *Data Min. Knowl. Discov.* **15**(2) (2007) 107–144.

19. B. Lkhagva, Y. Suzuki and K. Kawagoe, New time series data representation ESAX for financial applications, in *Proc. 22nd Int. Conf. Data Engineering Workshops* (2006).

20. B. Lkhagva, Y. Suzuki and K. Kawagoe, Extended SAX: Extension of symbolic aggregate approximation for financial time series data representation, in *Proc. 17th Data Engineering Workshop of IEICE* (2006).

21. H. Zhua and Z. Q. Jiang, Profitability of simple technical trading rules of Chinese stock exchange indexes, *Physica A* **439**(1) (2015) 75–84.

22. C. Evans, K. Pappas and F. Xhafa, Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation, *Math. Comput. Model.* **58**(5–6) (2013) 1249–1266.