FORECASTING DIRECTION OF EXCHANGE RATE FLUCTUATIONS WITH TWO DIMENSIONAL PATTERNS AND CURRENCY STRENGTH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUSTAFA ONUR ÖZORHAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PILOSOPHY
IN
COMPUTER ENGINEERING

MAY 2017

Approval of the thesis:

FORECASTING DIRECTION OF EXCHANGE RATE FLUCTUATIONS WITH TWO DIMENSIONAL PATTERNS AND CURRENCY STRENGTH

submitted by **MUSTAFA ONUR ÖZORHAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver ________________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Adnan Yazıcı ________________
Head of Department, **Computer Engineering**

Prof. Dr. İsmail Hakkı Toroslu ________________
Supervisor, **Computer Engineering Department, METU**

Examining Committee Members:

Prof. Dr. Tolga Can ________________
Computer Engineering Department, METU

Prof. Dr. İsmail Hakkı Toroslu ________________
Computer Engineering Department, METU

Assoc. Prof. Dr. Cem İyigün ________________
Industrial Engineering Department, METU

Assoc. Prof. Dr. Tansel Özyer ________________
Computer Engineering Department,
TOBB University of Economics and Technology

Assist. Prof. Dr. Murat Özbayoğlu ________________
Computer Engineering Department,
TOBB University of Economics and Technology

Date: 24.05.2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: MUSTAFA ONUR ÖZORHAN

Signature:

ABSTRACT

FORECASTING DIRECTION OF EXCHANGE RATE FLUCTUATIONS WITH TWO DIMENSIONAL PATTERNS AND CURRENCY STRENGTH

Özorhan, Mustafa Onur
Ph.D., Department of Computer Engineering
Supervisor: Prof. Dr. İsmail Hakkı Toroslu

May 2017, 110 pages

The value of a country's currency is expressed in terms of other countries' currencies. That value is called an exchange rate. Many currencies are freely floating and do not have a fixed value that is pegged by the central bank of a country. The value of currencies are determined in the foreign exchange market (Forex). Forex market is an extensive trading ground for traders across the world. It is available for trade 24 hours a day, 5 days a week. The trade volume per day is in excess of 4 trillion USD. Many different bilateral currency pairs are traded in the Forex market. In the Forex market a trader can profit from predicting the direction and magnitude of price fluctuations of a currency pair. Using a leverage value, it is possible to multiply wins and losses.

Technical indicators are statistical metrics whose values are calculated from price history of financial instrument. Technical indicators are generated to represent the behavior of the price and they are used to determine the future trend of the price of the financial instrument.

Chart patterns are two-dimensional formations that appear on a financial instrument’s price-action chart. Chartists and traders use these patterns to identify the cur-rent trends of the instrument to trigger buy and sell signals.

This thesis presents a method to predict the direction and magnitude of movement of currency pairs in the foreign exchange market. The method uses clustering and classification methods with a combination of two dimensional chart patterns, processed price data and technical indicator data. The input data is adapted to each trading day with a moving time-frame. The accuracy of the prediction models are tested across several different currency pairs. The experimental results suggest that using two dimensional chart patterns mixed with processed price data and the Zigzag technical indicator improves overall performance and adapting the input data to each trading period results in increased accuracy and profits. The predictions should be applicable in real world, since trading concepts such as spreads, swap commissions and leverages are taken into account.

Keywords: forex, forecasting, support vector machines, technical indicators, zigzag, chart patterns, motifs

ÖZ

DÖVİZ KURU DALGALANMA YÖNÜNÜN İKİ BOYUTLU ÖRÜNTÜLER VE PARA BİRİMİ GÜCÜ İLE ÖNCEDEN TAHMİNLENMESİ

Özorhan, Mustafa Onur
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. İsmail Hakkı Toroslu

Mayıs 2017, 110 sayfa

Ülkelerin para birimlerinin değerleri başka ülkelerin para birimlerinin değerleri cinsinden ölçülmektedir. Bu ölçüme parite denmektedir. Günümüzde pek çok para birimi dalgalı kur rejiminde serbestçe dalgalanmaktadır ve ülke merkez bankasınca başka bir para biriminin değerine sabitlenmemektedir. Para birimlerinin değerleri yabancı para takas pazarı (Foreks) pazarında belirlenmektedir. Foreks pazarı tüm dünyadan katılımcıların işlem yaptığı bir işlem platformudur. Haftanın 5 günü, 24 saat boyunca açıktır. Günlük işlem hacmi 4 trilyon ABD dolarından daha fazladır. Foreks pazarında pek çok para birimi ikilisi işlem görmektedir. Foreks pazarında işlem yapanlar para birimi ikililerinin değerlerinde meydana gelecek hareketlerin yönünü ve büyüklüğünü önceden tespit ederek kazanç elde edebilmektedir. Bir kaldıraç değeri kullanılarak, kazançlar ve kayıplar katlanabilmektedir.

Teknik indikatörler finansal bir enstrümanın önceki değerlerinden hesaplanan istatistiksel metriklerdir. Teknik indikatörler fiyatın davranışını temsil etmek için yaratılırlar ve finansal enstrümanın gelecek fiyat trendini belirlemekte kullanılırlar.

Grafik desenleri finansal bir enstrümanın fiyat-hareket grafiğinde meydana gelen iki-boyutlu oluşumlardır. Grafikçiler ve işlem yapanlar bu desenleri kullanarak enstrümanın mevcut trendini tespit eder ve alım satım sinyalleri yaratırlar.

Bu tezde yabancı para takas pazarındaki para birimlerinin hareketlerinin büyüklük ve yönlerinin tahmini için bir yöntem önerilmektedir. Yöntem kümeleme ve sınıflandırma tekniklerinin iki boyutlu grafik desenleri, işlenmiş fiyat verisi ve teknik indikatör verisi ile birleştirilmesinden faydalanmaktadır. Girdi verisi her bir işlem gününe kayan bir pencere ile uyarlanmaktadır. Tahmin modellerinin doğrulukları çeşitli farklı para birimi ikililerinde test edilmiştir. Deneysel sonuçlar iki boyutlu grafik desenleri, işlenmiş fiyat verisi ve Zigzag teknik indikatörünün kullanımının performansı arttırdığını, girdi verisinin her bir işlem anına adapte edilmesinin doğruluğa ve karlılığa olumlu etkilerinin olduğunu göstermektedir. Tahminler gerçek dünyada uygulanabilir olacak şekilde, fiyat aralıkları, faiz oranları ve kaldıraç oranları gibi ticari işlem kavramları dikkate alınarak yapılmaktadır.

Anahtar Kelimeler: foreks, tahmin etme, destek vektör makineleri, teknik indikatörler, zigzag, grafik şablonları, motifler

To my dearest daughter Alya.

ACKNOWLEDGMENTS

I want to express my deepest gratitude to my supervisor Prof. Dr. İsmail Hakkı Toroslu for his guidance, advice and encouragements throughout the research. One simply could not wish for a better advisor. I would like to thank my co-supervisor Dr. Onur Tolga Şehitoğlu for his meticulous contributions to my studies. I also would like to thank Assoc. Prof. Dr. Tolga Can and Assoc. Prof. Dr. Cem İyigün for their suggestions and comments. I would like to thank Assoc. Prof. Dr. Tansel Özyer and Asst. Prof. Murat Özbayoğlu for being members of my thesis examining committee.

I am deeply grateful to my sweet little daughter Alya and my loving wife Esra who have been there for me at all times, my parents Nurhayat and Fatih who devoted their life to their children, for their love and support. Without them, this work could not have been completed.

I would also like to thank the Central Bank of the Republic of Turkey (TCMB) and the Scientific and Technological Research Council of Turkey (TÜBİTAK) for providing the financial and temporal means throughout this study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CCI	Commodity Channel Index
DAP	Dynamically Adjusting Parameters
GA	Genetic Algorithm
RSI	Relative Strength Index
SBT	Strength Biased Trading
SCPT	Single Currency Pair Trading
SVM	Support Vector Machine
TI-MFT	Technical Indicator Medium Frequency Trading
ZZMOP	Zigzag Motif Predictor

CHAPTER 1

INTRODUCTION

1.1 Problem Definition

An exchange rate indicates the value of a country’s currency in terms of currencies. Since 1971 Smithsonian Agreement [1] many currencies are freely floating and do not have a fixed value that is pegged by the central bank of a country.

Forex is a global and decentralized market for trading currencies. It is continuously operational except weekends. The value of exchange rates are determined based on market supply and demand. Supply and demand are further determined by political conditions, market psychology and a variety of fundamental economic factors.

The trade volume per day is in excess of 4 trillion USD. Many different bilateral currency pairs are traded in the Forex market, most popular currencies are USD, EUR, GBP, JPY and CHF. The volume generated with currency pairs including these currencies account to 90.95% of all the trades taking place [2].

New data is generated with every transaction in the Forex market. A transaction in the Forex market can take place with the meeting of the buyer and the seller at the same price point. This meeting is a non-zero-sum game, due to the presence of a market maker which provides the Forex service under transaction costs such as spreads and swaps.

In the Forex market a trader can profit greatly from predicting the direction and magnitude of price fluctuations of a currency pair. Using a leverage value, trader can also multiply his wins and losses.

Many time series analysis and forecasting techniques are employed to predict the movements in the Forex and stock markets. Most of these techniques focus on directional symmetry of the created model which means that the model does not need to correctly predict the future value of a financial instrument, rather it needs to predict the future direction of movement. If prediction and future value of the financial instrument are in the same direction, then the prediction is considered to be directionally symmetric. Although it is difficult to obtain high levels of accuracy in terms of predicted values, a directionally symmetric model can prove useful in real world trading scenarios.

Exchange rates in the Forex environment are multivariate, semi-infinite time series data. As with any time series data learning, data mining, clustering, forecasting and similarity measurement applications are possible. Multiple time series are available, and are in interaction with each other. The formations and patterns recorded in a time series might effect a chain of other time series in the future. A similar phenomena occurs in different time frames. Movements and formations in the lower, fine-grained time frames not only contribute to the higher, coarse-grained time frames but they also dictate future price movements and formations. The opposite also holds true in certain cases.

There are many approaches to forecasting the future values in a time series. Some approaches try to forecast a single value that would occur in the next available time frame, some try to visually replicate recent time series data from a historic recurrence perspective, some simply try to forecast a band of values that are possible in a probabilistic manner.

The financial market has its own parameters that affect the success of the forecasts. From a real world perspective the appropriate metric to measure a forecasting model would take profitability and drawdown into account. To secure real world applicability of a forecasting model, an accord with Forex market parameters is a necessity. The parameters of the Forex market include take profit locations, stop loss locations, determination of the lot size, trailing stop loss and take profit orders, risk mitigation with counter positions, management of swaps and spreads. Therefore success of a forecasting model depends not only on the forecasting of a financial time series data, but also on correct prediction of Forex specific trade parameters.

1.2 Motivation and Contribution

In this thesis, several different techniques and approaches are used to tackle the financial time series forecasting problem.

Our first approach is using raw technical indicator data belonging to a variety of technical indicators in different combinations to exploit the techniques currently available to technical traders in a programmatic way. This approach presents a method to trade currencies using only technical indicators in the Forex environment. Our method uses variations of Relative Strength Index (RSI) and Commodity Channel Index (CCI) indicator to enter and close trades. The parameters of the technical indicators are adapted to each trading interval with a moving window. Trading is done in the 1-minute interval and highly frequent. The accuracy of the trading models are tested against historical data from 2015 to 2016. The experimental results suggest that using adaptive parameters for RSI and CCI technical indicators results in a high prediction accuracy and trade profits. Exhaustive searching and genetic algorithms are used to determine optimal parameters. Genetic algorithms prove useful to shorten the time for parameter search.

Our second approach is using genetic algorithms and support vector machines in a combination to exploit technical indicator signals for entering and exiting trade positions. We introduce a currency strength concept to understand and make use of the interaction between different financial time series available to us through different exchange rates. This approach addresses the problem of predicting direction and magnitude of movement of currency pairs in the foreign exchange market. We use Support Vector Machines (SVM) with a novel approach for input data and trading strategy. The input data contains technical indicators generated from currency price data (i.e. open, high, low and close prices) and representation of these technical indicators as trend deterministic signals. The input data is also dynamically adapted to each trading day with genetic algorithm. A currency strength biased trading strategy is incorporated, which selects the best pair to trade from the available set of currencies and is an improvement over the previous work. The accuracy of the prediction models are tested across several different set of technical indicators and currency pair sets, spanning 5 years of historical data from 2010 to 2015. The experimental results suggest that using trend deterministic technical indicator signals mixed with raw data improves overall performance and dynamically adapting the input data to each trading period results in increased profits. Results also show that using a strength biased trading strategy among a set of currency pair increases the overall prediction accuracy and profits of the models.

Our final approach is using a motif discovery mechanism in a time series with the help of a modified technical indicator to aggregate the multivariate and noisy data at hand. The discovered motifs are clustered and several learning models are trained to predict Forex specific trade parameters to enter and exit trades. This approach presents a method to predict short term trends in financial time series data found in the foreign exchange market. Trends in the Forex market appear with similar chart patterns. We approach the chart patterns in the financial markets from a discovery of motifs in a time series perspective. Our method uses a modified Zigzag technical indicator to segment the data and discover motifs, Expectation Maximization (EM) to cluster the motifs and Support Vector Machines (SVM) to classify the motifs and predict accurate trading parameters for the identified motifs. The available input data is adapted to each trading timeframe with a sliding window. The accuracy of the prediction models are tested across several different currency pairs, spanning 5 years of historical data from 2010 to 2015. The experimental results suggest that using the Zigzag technical indicator to discover motifs that identify short term trends in financial data results in a high prediction accuracy and trade profits.

Our study therefore evaluates a variety of approaches with the same set of exchange rates in a variety of different time frames and trading intervals.

We provide different approaches to forecasting financial time series, and all of these approaches contribute to the solution in different ways. The contributions are outlines as below.

We present a new way of representing and selecting the input data for training Support Vector Machines for multivariate financial time series. The first improvement is raw price data is used in combination with trend deterministic technical indicator signals and the second improvement is, the set of data used to train the learning model is dynamically adjusted via a genetic algorithm in a moving timeframe, so that indicators that are more relevant in a timeframe have more chance to reflect on the model.

Our approach presents a new notion for exploiting inter-time series interactions. A concept that we call strength bias is used for determining which currency pair to trade in the Forex market rather than trading a single pair during the entire term. Previous approaches make their forecasts using a single exchange rate. In our approach we are assessing the strengths of multiple currencies simultaneously to determine the weakest and strongest currencies at any time.

Modified Zigzag based motif discovery is introduced, which is a new method for discovering, clustering, classifying and segmenting subsequences of financial time series. Current methodology is the observation of charts via chartists. In our work raw price data is used in combination with modified Zigzag technical indicator signals and these data are used to discover motifs to predict future time series.

In terms of real-world trading focused learning approach, the study’s contribution is employing a stop-loss, take-profit and trade-window based trading approach to trade in the Forex market rather than simply trying to forecast the direction of the trend. Previous approaches try to forecast the direction of movement in the exchange rate but not the magnitude of the desired or undesired movement. In our approach we are assessing the possible direction and magnitude of movements in the currency’s future in both directions to determine trading parameters such as stop-loss, take-profit, time-stop and lot size.

1.3 Organization

The introduction is given in this chapter. In Chapter 2, background information about the Forex market and time series is provided. As the study is organized as a part-to-whole relation, accordingly, Chapter 3 includes Financial Time Series with Pure Technical Indicators, Chapter 4 includes Predicting Financial Time Series with Technical Indicator Signal Based Learning and Chapter 5 includes Motif Discovery in Financial Time Series. In Chapter 6, Experiments and Results are discussed. Last, in Chapter 7, Conclusion is presented.

CHAPTER 2

PRELIMINARY INFORMATION ON TIME SERIES, FOREX AND TECHNICAL INDICATORS

2.1 Time Series Preliminaries

A time series is the collection of values that are obtained from sequential measurements over a specific period of time. Time series analysis tries to visualize the characteristics of data. The mining, classification and forecasting of time series faces numerous difficulties. Most frequently these difficulties arise from the high dimensionality and the large volume of the data.

2.1.1 Definitions

This section provides definitions that are used in this thesis regarding time series.

Definition 1 - A time series T is an ordered sequence of n real valued variables $T = < o_1, o_2, \dots, o_n >$ where $o_n \in R$.

The observations in a time series are collected from measurements performed at uniformly spaced *time instants* which results in a fixed *sampling rate*. The time series can be *univariate* as shown in Definition 1 or it can be *multivariate* as shown in Definition 2. A *multivariate* time series spans multiple dimensions of data within the same time range.

Definition 2 - A multivariate time series *MT* is an ordered sequence of *n* vectors with m real valued variables $MT = << o_1, o_2, \dots, o_m >_1, < o_1, o_2, \dots, o_m >_2, \dots, < o_1, o_2, \dots, o_m >_n>$ where $o_m \in R$.

Time series may have a fixed length or they might be streaming in which case time instants continuously feed and grow the series. These types of time series are referred to as *semi-infinite* time series. *Semi-infinite* time series can be processed in a streaming manner or *subsequences* of it can be considered.

Definition 3 – Given a time series $T = < o_1, o_2, \dots, o_n >$ of length n a subsequence S^m of T is a series of length $m \leq n$ consisting of contiguous time instants from T such as $S^m = < o_k, o_{k+1}, \dots, o_{k+m-1} >$ where $1 \leq k \leq n - m + 1$. $S_T^m$is the set of all subsequences of length $m \leq n$ that can be derived from time series T.

Time series mining algorithms try to represent the similarity between two time series with similarity measures. Similarity between time series is usually represented from a distance perspective.

Definition 4 – Given a time series $T_1$and T_2 the similarity measure $D(T_1, T_2) = d$ is a function that takes two time series as inputs, and returns a distance *d* representing the distance between these two time series.

Financial time series are *multivariate* and *semi-infinite*. Similarity is usually measured between subsequences extracted from a single or multiple time series using a similarity measure such as the distance measure.

2.1.2 Clustering Time Series

Clustering finds groups or *clusters* in a given data set. Clustering tries to create clusters containing data that are homogeneous, while clusters themselves are as distinct as possible from each other. Clustering minimizes intracluster variance and maximizes intercluster variance.

There are different types of time series clustering approaches. In financial time series, *subsequence clustering* is generally applied. In this approach clusters are created by extracting subsequences from a single or multiple time series.

Definition 5 – Given a time series $T = < o_1, o_2, \dots, o_n >$ of length *n*, and a similarity measure $D(T_1, T_2)$, *subsequence clustering* finds *C*, the set of clusters $c_i = \{T_j' \middle| T_j' \in$

$S_T^m\}$ where c_i is a set of subsequences that maximizes intercluster variance and intracluster cohesion.

There are several time series clustering approaches, however most clustering techniques require parameter optimization based on individual series data and are incompatible with multivariate time series. Denton, Besemann and Horr[3] propose a pattern based time series subsequence clustering approach which uses radial distribution functions. Rakthanmanon, Keogh and Lonardi[4] propose an approach which includes both single and multivariate clustering based on minimum description length. In our approach we use a pattern based approach which segments and transforms a multivariate time series with expectation maximization.

2.1.3 Classification of Time Series

Classification assigns a category to each instance in a set. While clustering tries to intrinsically categorize instances, classification may know the classes in advance and be trained on an example dataset. With this approach a classifier can first learn the distinguishing features of a class and then determine the class of an unlabeled instance.

Definition 6 – Given an uncategorized time series T classification assigns it to a class c_i from a set C where $c_i \in C$ are predefined classes.

There are various classification approaches ranging from whole series classification to singular value decomposition. One frequent fallacy is overtraining, which can be overcome using time-series reduction and data selections techniques.

2.1.4 Segmentation of Time Series

Segmentation creates an approximation of the time series by reducing the dimensionality of the data. The reduction should accurately approximate the series by retaining the essential features.

Definition 7 – Given a time series $T = < o_1, o_2, \ldots, o_n >$ of length n, segmentation constructs a model T' such that dimensionality of T' is less than the dimensionality of T such that $d(T') \leq d(T)$ and T' approximates T with an error threshold e for a reconstruction function R where $D(R(T'), T) < e$.

Segmentation should minimize the reconstruction error between the reduced representation and the original time series. There are sliding window based approaches, top-down approaches and bottom-up approaches to segmentation of time series.

2.1.5 Prediction of Time Series

Time series are usually very long and many of them can be considered *smooth*. In a *smooth* time series any subsequent value for a subsequent time instance is within a predictable range.

Definition 8 – Given a time series $T = < o_1, o_2, \ldots, o_n >$ of length n, prediction estimates the time series $P = < o_{n+1}, o_{n+2}, \ldots, o_{n+k} >$ which contains k next values that are most likely to occur where $1 \leq k$.

There are a variety of prediction approaches which use neural networks, support vector machines or self-ordering maps. The predictor tries to maximize the similarity between the forecasted time series and actual time series. In financial applications the similarity between historical time series and forecasted time series might be measured differently.

2.1.6 Motifs in Time Series

A motif [5] is a subsequence of a longer time series which appears *recurrently*. Several motifs can exist within a single series, motifs can be of varying lengths and might overlap. Exhaustively determining motifs in a time series requires subsequences to be compared against other subsequences using a similarity measure, to assure *recurrent* behavior.

Definition 9 – Given a time series $T = < o_1, o_2, \dots, o_n >$ of length n, a motif M is a set of time series subsequences of T of length m, $M = \{TS_i \mid TS_i \in S_T^m\}$, and $\forall TS_i, TS_j : D(TS_j, TS_i) < e \wedge i \neq j$ holds true for a predefined error e where $D(TS_j, TS_i)$ is the similarity measure between two time series as described in Definition 4.

Subsequence clustering rarely produces meaningful results. Thus motif discovery is used to address time series problems such as anomaly detection and time series forecasting.

2.1.7 Measuring Similarity in Time Series

Most time series mining tasks requires a notion of *similarity* or *distance* between time series. For the analysis of the time series, humans inherently use the notion of *shape* and abstract themselves from problems such as amplitude, scaling, temporal warping, noise and outliers. Prominent distance measures such as the Euclidean distance cannot reach this level of abstraction. There are several categories of approaches to measuring the similarity of time series such as shape based, edit based, feature based and structure based approaches.

A sound time series similarity measure should recognize perceptually similar objects, be consistent with human intuition, emphasize features on global and local scales and abstract itself from distortions and noise [6]. To enable these properties for a similarity measure $D(T_1, T_2)$ we define several transformations to be applied to a time series $T = < o_1, o_2, \dots, o_n >$.

Definition 10 – *Amplitude shifting* creates a series $T' = < o'_1, o'_2, \dots, o'_n >$ obtained by a linear amplitude shift of the original series T where $o'_i = o_i + k$ where $k \in \mathbb{R}$ is a constant.

Definition 11 –*Uniform amplification* creates a series $T' = < o'_1, o'_2, \dots, o'_n >$ obtained by multiplying the amplitude of the original series T where $o'_i = o_i \,.\, k$ where $k \in \mathbb{R}$ is a constant.

Definition 12 – *Uniform time scaling* creates a series $T' = < o'_1, o'_2, \dots, o'_n >$ obtained by a uniform change of the time from the original series T where $o'_i = o_{\lceil k.i \rceil}$ where $k \in \mathbb{R}$ is a constant.

2.2 Forex Preliminaries

Forex is a global and decentralized market for trading currencies. It is continuously operational except weekends. The value of exchange rates are determined based on market supply and demand. Supply and demand are further determined by political conditions, market psychology and a variety of fundamental economic factors.

New data is generated with every transaction in the Forex market. A transaction in the Forex market can take place with the meeting of the buyer and the seller at the same price point. This meeting is a *non-zero-sum game* [7], due to the presence of a *market maker* which provides the Forex service under transaction costs such as spreads and swaps. With the collection of the transactions in a given period of time, a summarizing set of data –which is called a *bar*- is produced. A bar in a time frame contains opening, closing, highest and lowest prices for the given time interval. These are the prices that traders value the most in an interval and are also called as raw price data.

With the use of leverage, profits and losses can be multiplied. Leverages generally result in borrowing costs (swaps). The margin between asking and bidding price is called the spread and is generally very small to allow traders to create medium frequency applications with lower profit margins. For a trading algorithm to be profitable, the sum of profits collected by the algorithm should be higher than sum of losses, borrowing costs and transaction costs.

In this section we provide definitions about concepts and terms commonly used in Forex environment. Detailed definitions of these concepts can be found in [8].

Definition 13 – The first currency in a currency pair is called the base currency. In EUR/USD pair, EUR is the base currency.

Definition 14 – From Forex trader's perspective currency strength expresses the future value of currency and is an indicator of many factors such as the country's economic parameters and central bank interest rates. A currency's strength can be calculated against a set of other currencies or commodities.

Definition 15 – Leverage is the use of a financial instrument through borrowed capital. It allows a trader to multiply the potential return of an investment. The borrowed capital is called margin and provided to the trader by the broker firm that operates the investment. A leverage of 1:10 indicates that whenever trader opens a position of volume 1, a transaction with volume 10 is initiated by the broker firm. Earnings and losses are reflected 10 folds to the account if the leverage is 1:10.

Definition 16 – "Being long" or "going long" on a currency pair means buying the base currency in the pair against the quote currency.

Definition 17 – Margin is the borrowed capital that is provided to a trader using leverage by his broker. It allows the trader to open bigger positions than his account balance. A broker will place rules in place to protect firm's borrowed money such as a margin call –which limits the losses of a trader in a leveraged position.

Definition 18 – Pip is the smallest amount of change that can take place in a currency's value. Historically most major currency pairs were priced to four decimal points, hence a pip is valued at 0.0001 of such a currency.

Definition 19 – Since the advancement of the Forex market and an increase in the account leverages, many major currency pairs are priced to five decimal points. The fifth decimal point is 1/10th of a pip and is called a pipette.

Definition 20 – The second currency in a currency pair is called the quote currency. In EUR/USD pair, USD is the quote currency.

Definition 21 – "Being short" or "going short" on a currency pair means buying the quote currency in the pair against the base currency.

Definition 22 – Forex market is a settlement market in which buyers and sellers name their own price. Buyer’s price is the lower price and seller’s price is the higher price. Buyer’s price is also called “Bid” and seller’s price is called “Ask”. The difference between Bid and Ask prices is called spread. In a highly traded currency, the spread is lower. A lower spread would allow the trader to turn in profits in smaller price movements. Spreads can vary due to market volatility and liquidity. In this paper a fixed spread is assumed for all currencies at all times.

Definition 23 – A stop loss is an open position parameter for a Forex trader which allows a predefined amount of pips to be lost in an open position before the position is closed. The parameter allows the trader to accept losses and move on. Since Forex market is highly active and highly leveraged, without a stop loss traders can lose entire accounts in minutes.

Definition 24 – In the Forex market, the term swap refers to an interest rate swap. It is a way for Forex dealers to limit their exposure to fluctuations of interest rates on base and quote currencies. When a base currency with a higher interest rate is longed against a quote currency with a lower interest rate, swap would be positive. In the opposite case swap would be negative. In cases where interest rates are similar, long swap could be zero and short swap could be negative. In some more special cases where interest rates are similar and close to zero, both swaps could be negative. Swaps are commonly determined by the Forex brokers based on London Interbank Offered Rate (LIBOR).

Definition 25 – A take profit is an open position parameter for a Forex trader which allows a predefined amount of pips to be won in an open position before the position is closed. The parameter allows the trader to cash-in the virtual earnings in the position before the market price changes.

Definition 26 – There are two main approaches to analyze a currency: fundamental analysis and technical analysis. Fundamental analysis analyzes the economic fundamentals regarding a currency such as economic growth or interest rates. Technical analysis is the study of market price itself. Technical analysis assumes that all the fundamental factors are priced-in the market price and additional information regarding supply and demand can be deducted from the market price action.

Definition 27 – A technical indicator is a metric derived from historical price data to determine future price or direction of a currency. Technical indicators are based on technical analysis and Dow Theory and are mathematical formulae.

Definition 28 – Technical indicators are mathematical formulae and usually provide a number as output. How that number will be interpreted depends upon the trader. Different traders have different levels and numbers that they watch. A technical indicator signal is an additional set of rules and formulas that turn current or past values of a technical indicator into a trend determination method as simple as “Buy” or “Sell”.

2.3 Technical Indicator Preliminaries

Technical indicators [9] are statistical metrics whose values are calculated from price history of financial instrument. There are two types of technical indicators: lagging indicators and leading indicators. Lagging technical indicators are generated to represent the past behavior of the price. Leading indicators try to predict future behaviors of the price.

Technical indicators capture certain properties of price movements but are available to everyone in a simple, numeric form. In their original form they are numbers computed from raw price data without any meaning. With experience from the history, traders keep track of what kind of values generated by the technical indicators can be used to generate successful buy and sell signals and trade accordingly. The rules used in this process are called technical indicator signals. A static set of technical indicators or signals cannot reflect the price changes of a financial instrument indefinitely. Therefore both technical indicators and technical indicator signals are updated continuously.

Chart patterns are two-dimensional formations that appear on a financial instrument’s price chart. Chartists and traders use these chart patterns to identify trends for the instrument to trigger buy and sell signals. There are various categories of chart patterns, such as reversal chart patterns and continuation chart patterns. Reversal chart patterns appear at the end of previous trends and are followed by opposite price action, continuation chart patterns are intermediate consolidation areas in existing trends. A sample reversal chart pattern and a continuation chart pattern is shown in Figure 1 (a) and (b) respectively.

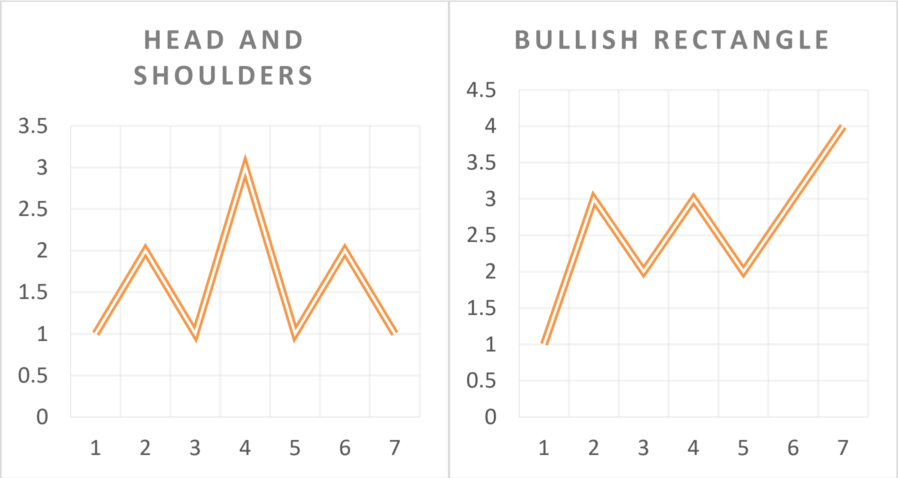

Figure 1: Sample chart patterns: (a) head and shoulders – reversal pattern, (b) bullish rectangle – continuation pattern

Financial markets, including the Forex market has three states: uptrends, downtrends and sideways trends [10]. A sample for each of these states are shown in Figure 2 for a sample Forex traded commodity (i.e. XAU/USD).

Figure 2: Trends at commodity XAU/USD: (a) uptrend, (b) downtrend and (c) sideways

As specified in [11, 12] predicting the trend of a financial instrument is not only more important but also easier than predicting the price at each time interval. We describe an example to motivate the importance of predicting the trend with the same time series from Figure 2 (a) with T1, (b) with T2 and (c) with T3.

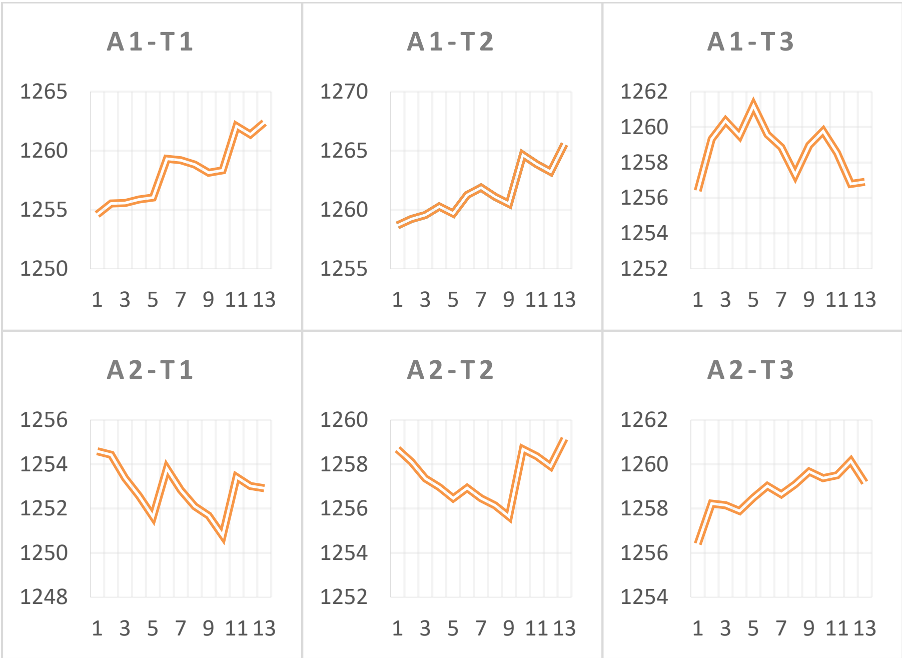

Figure 3: Account balances (a) with A1 and (b) A2

Figure 3 shows two perfect trading algorithms trading these time series; A1 is an oracle algorithm that determines the trend correctly, and keeps its positions open until the end of the trend. In T1, A1 buys at time point 1, and closes the position at time point 13. In T2, A1 sells at time point 1, and closes the position at time point 13. In T3, there is no trend, however for the sake of comparison we assume that A1 incorrectly determines that there is an uptrend and but at time point 1 and closes the position at time point 13. A1 makes a single decision at the beginning of the time series and closes its position at the end. A2 is an oracle algorithm that determines the correct price direction at each time frame. For T1, T2 and T3, A2 buys at the beginning of all positive time frames and closes the buy position at the end of the positive time frame; similarly A2 sells at the beginning of all negative time frames and closes the sell position at the end of the negative time frame - therefore each time frame is a decision time frame for A2. The resulting account balances achieved by the algorithms are presented in each of the charts.

Due to the transaction costs A2 underperforms A1 in T1 and T2. T3 represents the market conditions where there is no trend –the market is sideways. A2 outperforms A1 in this case. This example shows that determining the trends correctly can be more profitable than detecting the direction of individual price movements.

2.3.1 Zigzag Indicator

Zigzag indicator is a lagging technical indicator for financial time series data. It does not make predictions regarding the future values of a financial instrument. It is used to highlight the significant highs and lows of the instrument's historic values and eliminates the noise in the data. From a time series perspective, Zigzag performs time series segmentation as described in Definition 7. In our work we use the Zigzag technical indicator to detect legacy and novel motifs and create technical indicator signals that determine the trends in the Forex market.

Given time series $T = < t_1, t_2, \dots, t_n > (t_i \in N^+)$ Zigzag Z satisfies the following:

1. $Z =< z_1, z_2, \dots, z_n >$

2. $z_i = t_i$ if point is selected as *Zigzag point*

3. $z_i = *$ is the linear interpolation of z_{i-} and z_{i+}, the preceding and succeeding *Zigzag points* of z_i, if point is not selected as a *Zigzag point*

4. $\forall\, Zigzag\ point\ z_i\ of\ Z: (z_i > z_{i-} \wedge z_i > z_{i+}) \oplus (z_i < z_{i-} \wedge z_i < z_{i+})$

Depth and deviation are two important parameters to the Zigzag indicator, which are used to determine how much data will be filtered and how frequently the indicator will make adjustments to its previous values. The *depth* value determines the number of bars in which a bar has to be the extremum bar to be qualified as a Zigzag point. Zigzag points satisfies the following for the depth parameter:

1. $T = < t_1, t_2, \ldots, t_n > (t_i \in N^+)$

2. $Z = < z_1, z_2, \ldots, z_n >$

3. $\forall\, Zigzag\ point\ z_i \in Z\ z_i = t_j \rightarrow (\forall t_x \in \{t_{j-depth} .. t_{j+depth}\} : z_i \leq t_x) \oplus (\forall t_x \in \{t_{j-depth} .. t_{j+depth}\} : z_i \geq t_x)$

The *deviation* value of the Zigzag indicator represents the number of pip points that are required to establish a new low after a high Zigzag point, or a new high after a low Zigzag point. The new bar's value should deviate from the previous high or low value by at least the deviation amount. Zigzag points satisfies the following for the deviation parameter:

1. $T = < t_1, t_2, \ldots, t_n > (t_i \in N^+)$

2. $Z = < z_1, z_2, \ldots, z_n > \subset T$

3. $\forall\, Zigzag\ point\ z_i \in Z : z_i = t_j \rightarrow (z_{i-1} + deviation \leq z_i \wedge z_i \leq z_{i+1} + deviation) \oplus (z_{i-1} \leq z_i + deviation \wedge z_i + deviation \leq z_{i+1})$

In the stock exchange, deviation is a percentage, in the Forex market deviation is an amount of pips. Higher depth and deviation values would result in a lower number of Zigzag points hence more noise would be filtered. When selecting parameters for the Zigzag indicator, depth and deviation should be high enough to ensure noise is filtered but should be low enough to detect significant movements in instrument's price.

In the Forex environment, the values generated by the Zigzag indicator can be used in conjunction with different trading techniques such as Elliott waves [13], Fibonacci retracements [14] and chart patterns. In this work we use the Zigzag indicator to determine similarities in historic financial time series data in the form of motifs. Algorithm 1 describes the Zigzag algorithm.

Zigzag Algorithm

Require: S←start date, E ← end date, c ← currency time series, $zz_{deviation}$ ← Zigzag deviation
Ensure: Output the list of Zigzag values l_{zz} between S and E
$zz_{high} \leftarrow S$, $zz_{low} \leftarrow S$
for all $d \in < S..E >$
    case: previous Zigzag point is a high point:
    if $c[d] > c[zz_{high}]$ **then**
        $zz_{high} \leftarrow d$
        if $c[zz_{high}]-c[zz_{low}] > zz_{deviation}$ **then**
            l_{zz}.append(zz_{low})
    case: previous Zigzag point is a low point:
    if $c[d] < c[zz_{low}]$ **then**
        $zz_{low} \leftarrow d$
        if $c[zz_{high}]-c[zz_{low}] > zz_{deviation}$ **then**
            l_{zz}.append(zz_{high})
return l_{zz}

Algorithm 1: Zigzag algorithm

Figures 4 (a), (b) and (c) show the results of a variety of different Zigzag parameters being applied to the same financial time series data.

Figure 4: Zigzag with (a) depth = 10, deviation = 1% (b) depth = 20, deviation = 1% (c) depth = 20, deviation = 3%

2.3.2 RSI Indicator

The Relative Strength Index (RSI) indicator has been developed by Welles Wilder in 1978 [15]. RSI defines short term and medium term trends with respect to a strength value, which is calculated using the difference between the closing value of the current and previous bars.

The RSI indicator is an oscillator, meaning its values oscillate between predefined numbers, which are 0 and 100 in this case. RSI can be applied to any currency, stock or any other financial data. At a given time *t* and for period *n*, the RSI is calculated with the formulae given below.

$$Relative\ Strength\ Index_n(t) = 100 - \frac{100}{1 + Relative\ Strength_n(t)}$$

$$Relative\ Strength_n(t) = \frac{Average\ Gain_n\ (t)}{Average\ Loss_n\ (t)}$$

$$Average\ Gain_n(t) = \frac{(n-1) \times Average\ Gain_n\ (t-1) + Current\ Gain\ (t)}{n}$$

$$Average\ Loss_n(t) = \frac{(n-1) \times Average\ Loss_n\ (t-1) + Current\ Loss\ (t)}{n}$$

$$Current\ Gain\ (t) = Close\ (t) - Open(t)$$

$$Current\ Loss\ (t) = Open\ (t) - Close(t)$$

When the price of the financial instrument moves lower at all *n* previous time frames, RSI approaches 0, and when the price moves higher at all *n* previous time frames RSI approaches 100. As with any oscillator, specific values of the RSI indicator have specific meanings for the traders. Any RSI value less than 30 is classified as *oversold* meaning the price is too low, and any RSI value higher than 70 is classified as *overbought* meaning the price is too high. RSI indicator can stay at oversold and overbought levels for prolonged periods.

Even though traditionally RSI is applied to the closing price of the instrument, it can be applied to any price such as opening, low, high or median.

2.3.3 CCI Indicator

The Commodity Channel Index (CCI) has been developed by Donald Lambert in 1983 [16]. CCI indicator can be used to identify new trends or warns the trader of extreme market conditions. The indicator measure the current price level with respect to an average price level for a specific period of time. As the RSI indicator, CCI indicator is an oscillator.

For a given time period *t* and CCI parameter *n* the CCI is calculated with the formulae given below.

$$CCI_t(n) = \frac{Typical\ Price_t - SMA(Typical\ Price_t, n)}{Mean\ Deviation_t\ (n)\ \times\ L}$$

$$Typical\ Price_t = \frac{High_t +\ Low_t + Close_t}{3}$$

$$SMA(Typical\ Price_t, n) =\ \frac{\sum_{i=t-n+1}^{t} Typical\ Price_i}{n}$$

$$Mean\ Deviation_t(n) = \frac{\sum_{t-n+1}^{t} |Typical\ Price_t - SMA(Typical\ Price_t, n)|}{n}$$

The oscillation range of CCI varies based on the Lambert parameter L and the number of periods to average. 80% of CCI values fall between -100 and $+100$ with n=20, and L=0.015.

Similar to the RSI, a CCI number above 100 denotes a strong price action, which might further suggest a decline in the prices since the instrument is overbought. The opposite is also through for CCI numbers below -100.

2.4 Genetic Algorithm Preliminaries

A genetic algorithm is a search heuristic, which mimics the principles in biological evolution -such as natural selection, crossover and mutation [19]. The algorithm starts with a population of candidate solutions (i.e. chromosomes), and with respect to a fitness function evaluates the performance of the solution.

In our first work using pure technical indicators, we use genetic algorithms to determine the best possible parameters for various technical indicators. In this work chromosomes contain numeric genes which can take on one of several values from a predefined set. The chromosomes are used to create a trading model which better reflects the current trading interval.

In our second work in which we bring the strength bias approach, we represent the set of technical indicators and symbol parameters as a binary chromosome, and generate learning models based on that chromosome, which is in turn evolved based on model's performance as a fitness function. The binary chromosome acts as a filter to use and discard technical indicator signals for the current trading interval.

In both approaches the initial population of the genetic algorithm is initialized with a predefined number of chromosomes with randomly selected alleles. A specific number of generations of evolutions are performed with mutation and crossover operators. Also elitism is used to preserve the fittest individuals.

In both approaches re-evaluation of the fitness of same chromosomes are not necessary for the same interval thus a fitness function cache is maintained to decrease total CPU time necessary for the evolution. The fitness function of a chromosome is the total earnings achieved by the model that is trained with the technical indicator parameters indicated by the chromosome.

2.5 SVM Preliminaries

An SVM is a machine that constructs a set of hyper-planes in a multi-dimensional space for classifying and regression of data [23]. An empirical analysis of the systems used in time series forecasting shows that SVMs perform much better than the remaining models in terms of directional symmetry [22]. Even though Gaussian Process and Multilayer Perceptron is better than SVMs in terms of mean squared error and mean absolute deviation, research show that in the Forex market directional symmetry is much more important [24, 25, 26] than distance based error and deviation performance; therefore in our system we use SVMs to train on selected technical indicators.

SVM's main objective is identifying the maximum margin hyper-plane so that separation between classes of samples can be maximized. Mapping of input vectors to the high dimension feature space is performed by a kernel function. Polynomial function and radial basis function kernels are commonly used in trading systems. Due to superior performance recorded by polynomial function kernels in related work [17, 27] we use a polynomial function kernel in our system. Degree of polynomial function (d), and regularization parameter (c) are the parameters of the SVM.

To determine these parameters efficiently several different settings are experimented and the selection is made based on training and validation accuracy. Final results obtained with our algorithm using the optimal parameters are shown in section 5.

CHAPTER 3

USING PURE TECHNICAL INDICATORS FOR MEDIUM FREQUENCY TRADING IN FINANCIAL TIME SERIES

Our TI-MFT algorithm uses RSI and CCI indicators to create signals to open and close trade positions in the Forex market. In this section we take a top down approach by first explaining the overview of our algorithm, and then we proceed to the details of the trading approach.

3.1 TI-MFT Algorithm

TI-MFT algorithm has a layered architecture, the first one is the data collection layer, second layer is the calibration layer and third layer is the trading layer. The first layer is the data collection layer and it collects both tick-by-tick and 1-Minute interval data for currencies of interest. The second layer is the calibration layer. In this layer indicator parameters and trade parameters are calibrated with historical data. One year of historic data is retrieved from data collection layer and is used for training and testing with moving time frames.

The third layer is the trading layer where parameters acquired from the second layer are used to trade with tick-by-tick real time data for currencies of interest. You can find a summarization of our algorithm in pseudo code format in Algorithm 2.

In our approach we search the training space for a combination of RSI and CCI parameters to both enter and exit trades using two methods (i) exhaustively searching all combinations in the parameter space, (ii) using a genetic algorithm to determine optimal parameters. The flow of operations in the genetic algorithm is described in Figure 5.

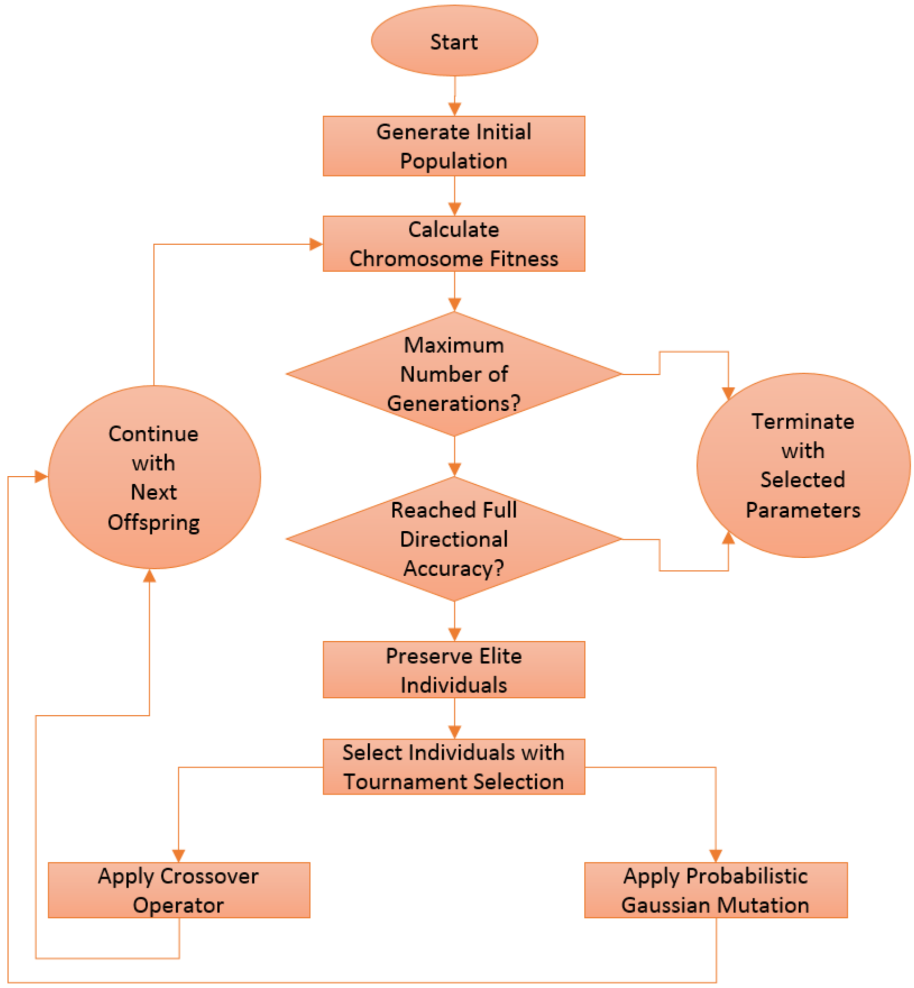

Figure 5: TI-MFT genetic algorithm flow

```
TI-MFT Trade Algorithm
Require: CL←currency list, F←future, H←history, CCIL←CCIs, RSIL←RSIs
Ensure: Trade orders
for c ← CL do
    for i←1 to H do
        for CCI in CCIL do
            for RSI in RSIL do
                signalList ← buySignal(RSI, CCI)
                signalList ← sellSignal(RSI, CCI)
        signal ← selectSignalTrigger [signalList]
        stopLoss ← selectSignalStopLoss
        takeProfit ← selectSignalTakeProfit
    for i←1 to F do
        return order = submitOrder (c, stopLoss, takeProfit)
```

Algorithm 2: An overview of our TI-MFT trade algorithm in pseudocode format

Once parameters for the trading model is determined by either of the parameter determination methods, the model can be used for trading. For a buy trade to happen a fast moving RSI value should crossover a slower moving RSI value, the same should happen for a fast moving CCI value to crossover a slower moving CCI value. When these two signals are present, a buy order can be put in place. When one of these signals is no longer present, the order is closed. A sample crossover between fast and slow moving RSI values is presented in Figure 6.

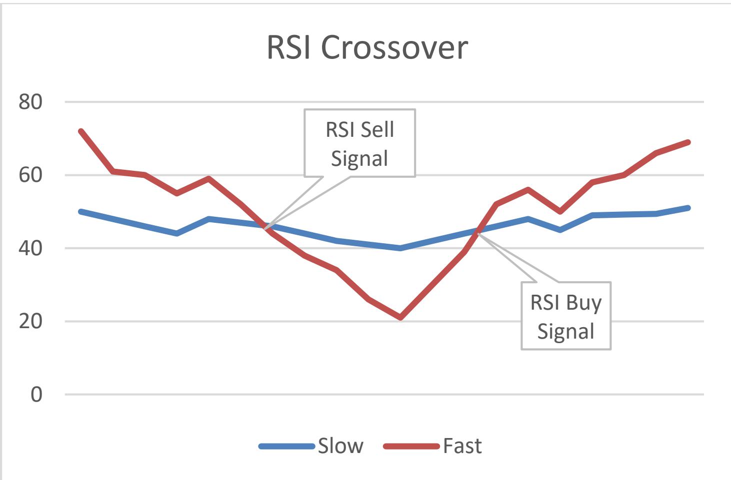

Figure 6: Fast and slow moving RSI values on the same price data

For a sell trade to happen, a fast moving CCI value should drop below a slow moving CCI value, and the same should happen for the RSI. When two of these signals are present, a sell order can be put in place. If a fast moving RSI value crosses over a slow moving RSI, or if a fast moving CCI value crosses over a slow moving CCI, one of the signals will be lost and the open trade will be closed.

For both type of the trades, a predefined take profit and stop loss set is also applied to increase trading success.

3.2 Trade Parameter Detection

Before entering a trade with a given currency, our algorithm determines the stop loss and take profit locations. This is done using the previously simulated price-action. For each type of order a separate set of values are determined to use in the future.

A trader who places a sell order expects prices to drop. The expected amount of drop is the take profit location. However, there might be a slight increase in the price before a drop occurs. Our system should be able to overlook this much of a loss at any time to reach the forecasted drop and take profit. Another possibility is that the increase in the price continues and prices does not drop in the event horizon, in which case the system should stop the losses. Hence the stop loss location should account for these two possible price movements. The stop loss location for a sell order is called “high until low” and the take profit location is called “low”.

For buy orders, a hike in the price is expected in the future, which will be the take profit location. A temporary decrease in price prior to the hike should be allowed to reach the hike forecasted. A deeper decrease should be recognized as a false signal and losses shall be suffered. The stop loss location for a sell order is called “low until high” and the take profit location is called “high”. Figure **7** outlines the price action for a sample buy signal and shows low until high and high points, which should be covered by stop loss and take profit parameters determined.

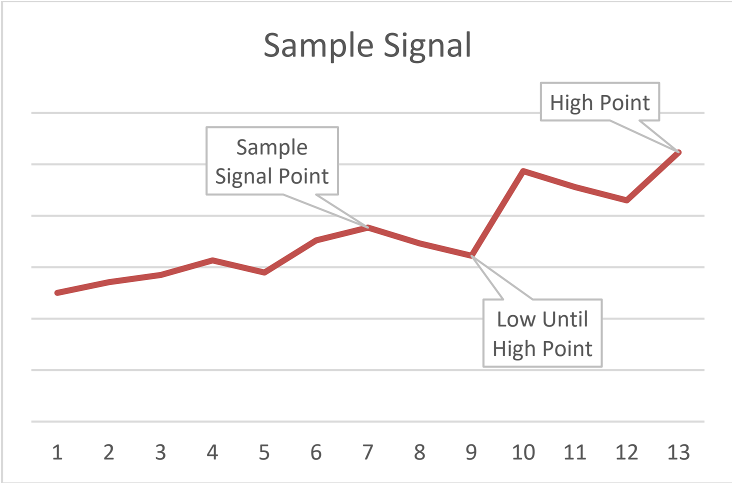

Figure 7: Sample price action on an instrument after a trade signal

The optimal values to be applied in the trade interval are determined using parameter search in the training interval by two methods: (i) all combinations of possible values are empirically tested, (ii) values are determined via a genetic algorithm.

3.3 Genetic Algorithm

In our algorithm technical indicator and trade parameters are represented as genes in a chromosome. To evaluate the fitness of a chromosome, a trading model with the parameters specified in the chromosome is generated, which is in turn simulated to determine resulting account balance for the given model. The resulting account balance is used as the fitness of the given chromosome. Table 1 illustrates a sample chromosome that is used in our system.

Table 1: A chromosome instance used in TI-MFT

Gene Name	Value
RSI_FAST_PERIOD	14
RSI_SLOW_PERIOD	21
CCI_FAST_PERIOD	30
CCI_SLOW_PERIOD	50
CCI_FAST_ACTIVE_PRICE	Median
CCI_SLOW_ACTIVE_PRICE	High
CCI_LAMBERT	0.01
STOP_LOSS	10
TAKE_PROFIT	20

CHAPTER 4

USING TREND DETERMINISTIC TECHNICAL INDICATOR SIGNALS WITH STRENGTH BIAS

The system is composed of three layers. The first one is the data preparation layer and is responsible for generating technical indicators and respective trend deterministic signals. The second layer is the trade preparation layer in which trading parameters and set of available currencies can be specified. The third layer prepares the prediction models and simulates the trades. The architecture is similar to the systems created in [17, 18] and an overview of the system is provided in Figure 8.

The first layer is the data preparation layer and it generates technical indicators and trend deterministic technical indicator signals from raw price data that has been made available to the system via the data collection infrastructure. Set of indicators and signals that are generated in this layer are explained in detail in Table 16 and 17

The second layer is the trade preparation layer. Human intervention is possible in this step, and the user can select the currencies to add to the set of available currencies or override trade parameters such as stop loss, take profit and leverage. Selection of currencies is rather important since our algorithm weighs the strengths of multiple currencies and selects the strongest one against the weakest one. The base case of selecting only two currencies changes the trading approach to single currency pair trading. The results obtained with different set of currencies will be discussed in Section 5.

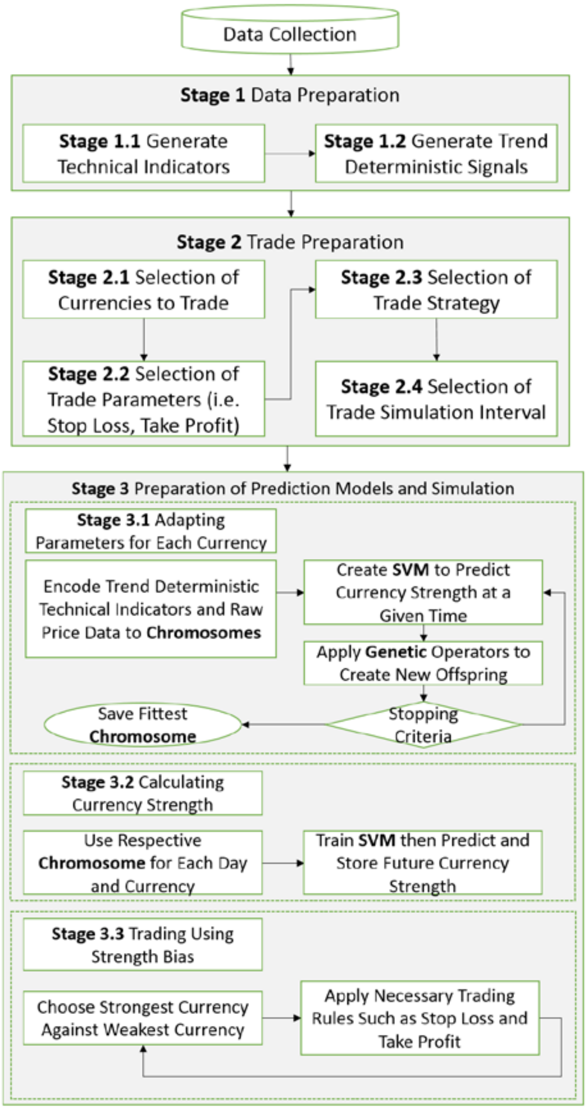

Figure 8: Overview of our system

The third layer prepares the historical prediction models and simulates the models over future data. The system trains a separate set of models for all the currencies for a trading day in the simulation interval and then selects the best performing model for each currency to predict next day's currency strength for each currency. An SVM model for a currency is fed *n* days of training data and *m* days of validation data for each day –selection of these parameters are discussed in section 4.2. The data is composed of a set of trend deterministic technical indicator signals and raw price data. Variations of the data is created to find the best performing set of inputs for a given history period using genetic algorithms [19, 20]. The SVM models try to learn which currency is strongest at a given date. A strong currency is supposed to increase and a weak currency is supposed to decrease the next day.

For each trading day, the selected SVM models assess the strength of their respective currencies. The strongest currency is bought and the weakest currency is sold. In Forex market this is a single operation in the currency pair. If the currency pair is listed as *Strong/Weak*, the system "goes long" on that pair, if it is listed as *Weak/Strong*, the system "goes short" on that pair. The trader starts with a specific amount of cash (i.e. 1000 USD) and for each Forex market operation, the trading spread applicable for the given currency pair is applied. If the position is carried overnight, applicable swap commissions are also applied. When trading term ends, the entire portfolio is cashed back to USD and earnings are reported. The swaps and spreads used by our system are taken from TrueFX [21] and are outlined in Table 2.

Table 2: Spread and swap commissions applicable

Currency Pair	Spread	Swap (Long)	Swap (Short)
EURCHF	2.5	0.5	-2
EURGBP	3.5	-1.4	0
EURUSD	1.9	-1.5	0
GBPCHF	13	3	-6.5
GBPUSD	2.1	0	-2.7
USDCHF	2.9	0	-1.5

Our system also makes use of stop loss, take profit and trailing stop loss orders. This means no single trade can result in more loss than the stop loss and more profit than the take profit amount. Trailing stop loss occurs if the price of an instrument goes in the predicted direction for a while, and then changes the direction against the prediction. Traditionally the orders are determined based on trading leverage. In our system we use a 1 to 100 trading leverage (which would mean loss of entire balance in a 1% decrease, and doubling of the balance in a 1% increase). In our system we accept 10 pips of loss as the stop loss, 10 pips of loss as the trailing stop loss and 100 pips of gains as the take profit points. Therefore our trading approach accepts a lot of losses to make a bigger win and make up for all the losses recorded.

Our system generates a set of technical indicators for each of the financial instrument selected as previously mentioned. For the experiments in this paper variations of the indicators stated in Table 5 are used.

4.1 SBT-DAP Algorithm

There are many support vector machine based time series forecasting algorithms. Selection of input data, adjustment of learning model parameters and employment of additional decision support mechanisms decide the fate of the forecasting algorithm. An empirical comparison of some popular algorithms has been made by Nesreen and Atiya in [22]. This part discusses the novel aspects of our Strength Biased Trading with Dynamically Adapting Parameters (SBT-DAP) algorithm. You can find a summarization of our algorithm in pseudo code format in Algorithm 3 and 4.

Technical indicators capture certain properties of price movements but are available to everyone. Traders keep track of signals generated by technical indicators and trade accordingly. A static set of technical indicators cannot reflect the price changes of a financial instrument indefinitely. Our program represents the technical indicator signals as a binary set of inputs for each financial instrument. Initially all the available inputs are available for all the trading instruments. Using a genetic algorithm variations of the inputs are created to train separate models. For each trading day and each trading instrument the selected inputs are used to determine the strength of the instrument in the given day.

```
SBT-DAP Prediction Model Preparation and Simulation Algorithm
Require: RPD ← raw price data, TID ← trend deterministic technical indicator
data, CL ← currency list, CPL ← currency pair list
Ensure: N models to predict N days to be forecasted
    for c in CL do
// Stage 3.1: Generate SVM's
     for i:1 to N do
          //Chromosomes represent available data to train SVM
          chromosomes = create (RPD, TID)
          // Evaluate the fitness of the given chromosome
          evaluateFitness(chromosome, RPD, TID)
          //Apply GA operations for 100 generations and retry
    chromosome.evolve()
          goto 3
          end for
// Stage 3.2: Calculate Strengths of Currencies
          for i:1 to N do
               // Predict and store currency's next day strength
               strength[c, i+1] = predict(c, SVM_i, i+1)
          end for
    end for
// Stage 3.3: Trade Using Best Pair to Buy/Sell
    for   i:1 to N do
          strongest = chooseStrongest(strength[*, i+1])
          weakest = chooseWeakest(strength[*, i+1])
          if (strongest+weakest) in CPL
               buy(strongest, weakest)
          if (weakest+strongest) in CPL
               sell(weakest, strongest)
    end for
```

Algorithm 3: SBT-DAP prediction model preparation and simulation algorithm

```
SBT-DAP Model Fitness Evaluation Algorithm
Require: RPD ← raw price data, TID ← trend deterministic technical indicator
data, chromosome ← chromosome
Ensure: Real number F representing chromosome fitness
  //Train SVM on training interval of 80 days
  SVM.train(chromosome, RPD, TID)
  //Validate SVM on validation interval of 20 days
  SVM.validate(RPD)
     balance = SVM.simulate() //Simulate on given interval
  return balance //The fitness of the chromosome
```

Algorithm 4: SBT-DAP model fitness evaluation algorithm

4.2 Genetic Algorithm

The initial population of the genetic algorithm is initialized with 50 chromosomes with randomly selected alleles. 100 generations of evolutions are performed with mutation and crossover operators. Also a 10% elitism rate is used to preserve the fittest individuals. The fitness function of a chromosome is the total earnings achieved by the model that is trained with the technical indicator parameters indicated by the chromosome. This requires a separate model to be trained by the system for each distinct chromosome. Two sample binary chromosomes are shown in Figure 9 for reference. The length of the chromosomes are 30. For the first 26 genes variations of trend deterministic technical indicator signals are used. The last 4 genes are raw price values of the currency (i.e. open, close etc.). A gene value of 0 denotes that the given raw value or signal will not be used for training the SVM while a value of 1 denotes otherwise.

Allele	MACD		RSI		SMA						WMA						WMA/SMA Crossover						CCI		SK%		Open	Close	High	Low
	5	15	5	15	5	10	20	50	100	200	5	10	20	50	100	200	5	10	20	50	100	200	5	15	5	15				
Chromosome #1	0	1	0	1	0	0	0	1	0	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1	0	0	0	1	1	1
Chromosome #2	1	0	0	0	1	1	1	0	0	1	0	0	0	1	1	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1

Figure 9: Two Sample Chromosomes

4.3 Support Vector Machines

In our algorithm, SVMs provide estimates for the financial instrument's current strength, and therefore future value. The trades made by our machine is not simply buying or selling a single financial instrument, rather it is taking position with the strongest instrument against the weakest one. Whenever the predicted strength value is positive the base currency is assumed to increase in the future, whenever the prediction is negative, the base currency is assumed to decrease in the future. The reverse applies to the quote currency. As long as actual and predicted strength values are both negative and positive, the decision is considered to be accurate in single currency pair trading.

When multiple currencies are made available to the system and strength biased currency trading is used, not only the sign but also the magnitude of predictions are important. Strength of each base and quote currency is calculated and summed to find the strongest and weakest currencies in the trading pool and strongest currency is bought against the weakest currency.

4.4 Currency Strength Calculation

Currency strength bias is a novel trading approach proposed in this work, which favors the strongest currency against the weakest currency. This section details the calculation of a currency's strength. In our system a currency is at its strongest when the price is in a local minimum, and at its weakest when the price is in a local maximum. The reason is a local minimum is followed by a local uptrend resulting in the increase of prices and a local maximum is followed by a local downtrend resulting in the decrease of prices.

While training the SVM, the classes of training data is determined using *Hill Size for Strength Tagging Training Data*, *Strength Category Granularity (SCG)* and *Strength Categorization Bucket Size (SCBS)* variables defined in Table 20. At its strongest point a currency will be tagged with (+*SCG)* and at its weakest point it will be tagged with (*–SCG). SCBS* will be used to determine how the days between the local minimum and local maximum will be tagged.

Let $C = \{c_1, c_2, \ldots, c_n\}$ denote the set of currencies which are traded against each other, where n is the total number of available currencies.

$Chr_d(c_i, c_j)$ is a chromosome that describes trend deterministic technical indicator signals that will be used to train the SVM to predict the pairwise strength of a currency $c_i \in C$ against $c_j \in C$ at day d. $SVM_d(c_i, c_j)$ is the SVM trained using $Chr_d(c_i, c_j)$ that will predict the pairwise strength of a currency $c_i \in C$ against $c_j \in C$ at day d. $Chr_d(c_i, c_j)$ and $SVM_d(c_i, c_j)$ are defined by *SBT-DAP* algorithm in Algorithm 2.

The strength $S_d(c_i)$ of a currency $c_i \in C$ at day d is defined in equation (1) where $SVM_d(c_i, c_j)$ is the pairwise strength of currency c_i relative to currency c_j at day d.

$$S_d(c_i) = \sum_{c_i \neq c_j}^{c_j \in C} SVM_d(c_i, c_j) \quad (1)$$

$SVM_d\ (c_i, c_j)$ is equal to a number x that lies in the *[-SCG, SCG]* interval where *SCG* is *Strength Categorization Granularity* defined in Table 20. $SVM_d\ (c_i, c_j)$ is valid for all $c_i \in C$ and $c_j \in C$ where $c_i \neq c_j$ and $SVM_d\ (c_i, c_j) = -SVM_d\ (c_j, c_i)$.

The currency with maximum strength $Max\_strength_d$ at day d is defined in equation (2) where total number of available currencies is n and $c_i \in C$.

$$Max\_strength_d = \max_{1 \leq i \leq n} \{S_d(c_i)\} \quad (2)$$

The currency with minimum strength $Min\_strength_d$ at day d is defined in equation (3) where total number of available currencies is n and $c_i \in C$.

$$Min\_strength_d = \min_{1 \leq i \leq n} \{S_d(c_i)\} \quad (3)$$

Currency positioning $CP_d(SC_d, WC_d)$ at day d is made with $Max\_strength_d$ which has maximum currency strength and $Min\_strength_d$ which has minimum currency strength and it is formulated below in equation (4).

$$\begin{aligned} &CP_d(Max\_strength_d, Min\_strength_d) \\ &= \begin{cases} \text{either } Max_{strength_d}/Min_{strength_d}, long(Max\_strength_d, Min\_strength_d) \\ \text{or } Min\_strength_d/Max\_strength_d, short(Min\_strength_d, Max\_strength_d) \end{cases} \quad (4) \end{aligned}$$

In Figure 10, the performance of the six currency pairs used in our work is shown between December 15, 2015 and January 15, 2016. The relative performances of the currencies change throughout the interval, but are fairly stable. For the entirety of the interval the currencies can be listed from best performing to worst performing as USD > EUR > CHF> GBP. Optimal course of action for this period would be longing the USD against the GBP, which can be attained by shorting the GBP/USD pair.

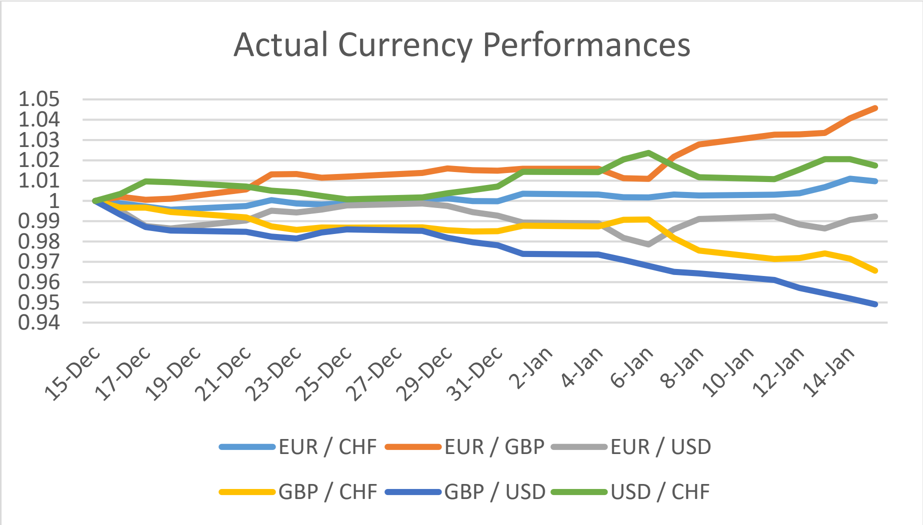

Figure 10: Price action for the currencies used in our work between 12/15/2015 and 01/15/2016

When training for the next day, our system would identify USD as the strongest currency and GBP as the weakest currency for December 15, 2015 using *Hill Size for Strength Tagging Training Data* variables. The strengths of the currencies for the days in between would be appointed using *Strength Category Granularity (SCG)* and *Strength Categorization Bucket Size (SCBS).* When classification is made for the following day of January 15, 2016, USD would have a lower strength and GBP would have a higher strength than they did back in December 15, 2015.

For the same period our system detects the pairwise strength for currencies as depicted in Figure 11 using equation (1). It can be seen from the graph that for the period starting from December 18 and ending at January 11 strength in pair GBP/USD is the lowest. For the period starting from December 21 and ending at January 6 strength in pair USD/CHF is the highest.

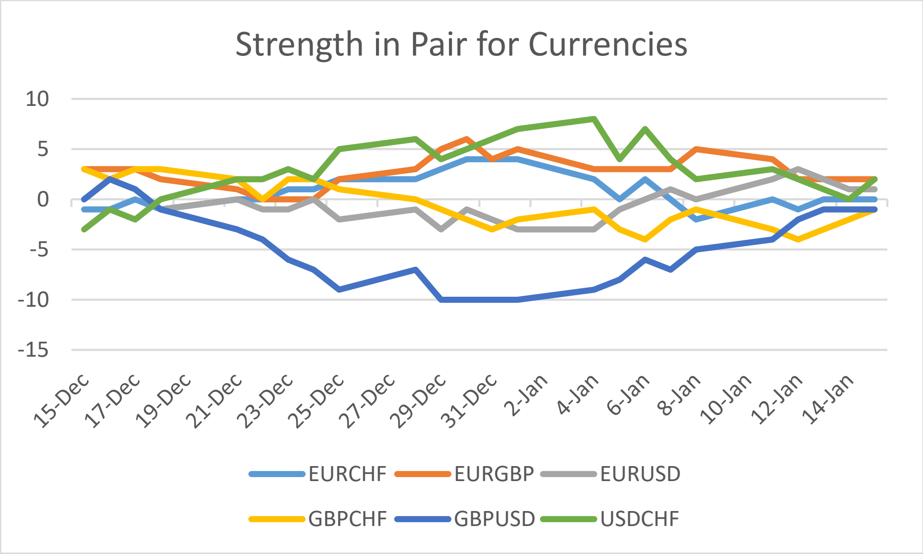

Figure 11: Strength in pair for currencies between 12/15/2015 and 01/15/2016

Using the data for each currency pair and equation (1), currency strengths are calculated for December 15, 2015 – January 15, 2015 interval. The data is represented in Figure 12. It can be seen that at the beginning the strongest currency is EUR, and as of December 18, 2015 USD is detected as the strongest currency. The strongest currency is later on changed to EUR after January 10, 2016. The weakest currency is USD in the beginning, is changed to CHF at December 18, 2015 and changed again to GBP as of December 24, 2015.

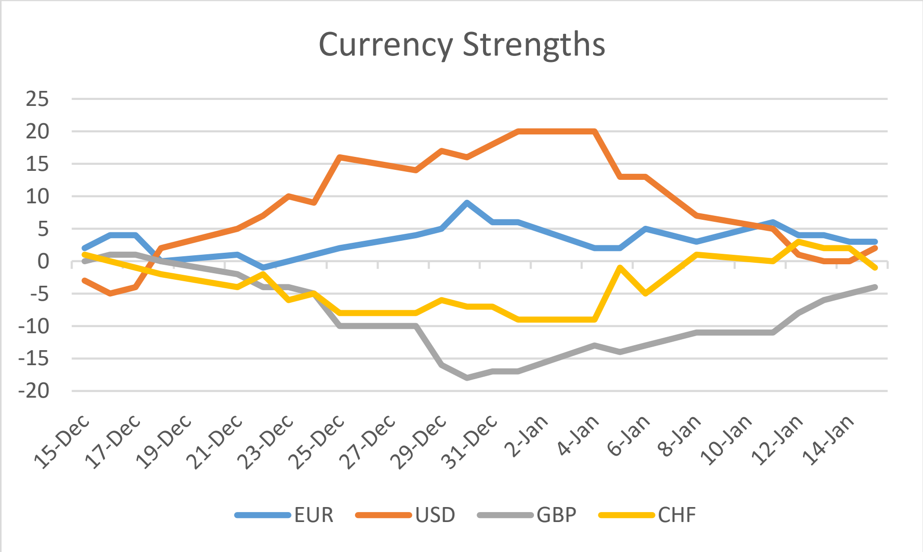

Figure 12: Currency strengths between 12/15/2015 and 01/15/2016

In this period our system takes the positions specified below as per equations (2), (3) and (4). The raw profit in this one month interval is 2.19%, with a leverage of 1:100 this could mean 219% earnings. Our system applies trading parameters stop-loss and take-profit specified in Table 3 to these trades limiting the losses from forecast errors. The result obtained by our system is slightly better with 3.17% converting to 317% earnings.

Table 3: SBT-DAP system decisions

Date	Strongest	Weakest	Decision	Result (Pips)	Raw Result (%)	System Result (%)
12/15/2015	EUR	USD	long-EURUSD	-149	-1.36	-0,09
12/18/2015	EUR	CHF	long-EURCHF	+2	+0.19	+0,19
12/21/2015	USD	CHF	long-USDCHF	-2	-0.20	-0,20
12/22/2015	USD	GBP	short-GBPUSD	+14	+0.09	+0,09
12/23/2015	USD	CHF	long-USDCHF	-18	-0.18	-0,10
12/24/2015	USD	GBP	short-GBPUSD	+354	+2.38	+2,01
01/11/2016	EUR	GBP	long-EURGBP	+95	+1.27	+1,27

CHAPTER 5

SHORT TERM TREND PREDICTION IN FINANCIAL TIME SERIES DATA WITH MOTIFS USING ZIGZAG

Our algorithm Zigzag Motif Predictor (ZZMOP) uses a modified Zigzag indicator to segment historical time series to possible motifs, clusters the motifs and defines reward/risk ratio criteria for the clusters, classifies the motifs in clusters to learn trading rules for the created clusters. The flow of these operations is shown in Figure 13.

Figure 13: Flow chart illustrating algorithmic flow of ZZMOP

Once ZZMOP constructs models required for trading, our trading algorithm uses these models to learn trading parameters and submit trade orders. The flow of these operations is shown in Figure 14.

In this section we take a top down approach by first explaining the overview of our algorithm, and then we proceed to the details of the modified Zigzag indicator, how sequence similarity applies to financial time series data and how we find signal points for detected motifs.

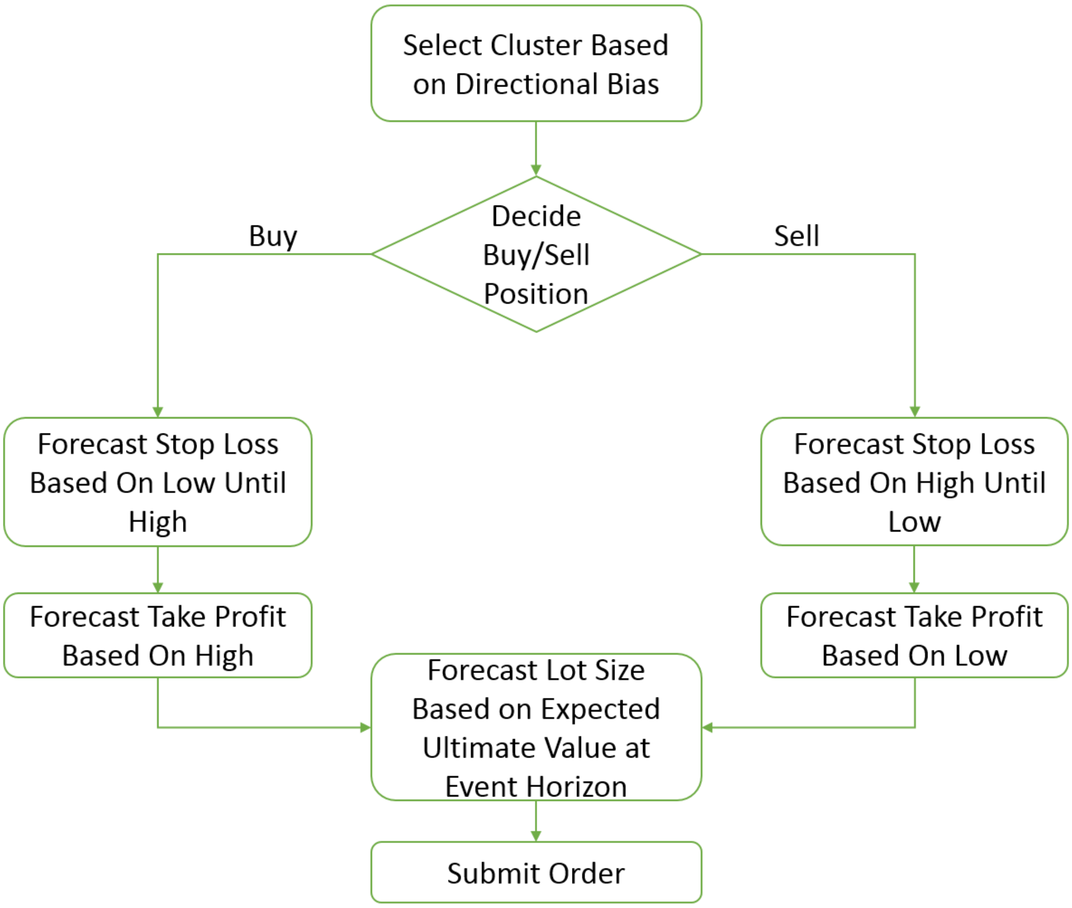

Figure 14: Flow chart illustrating trading algorithm

5.1 ZZMOP Algorithm

ZZMOP algorithm has a layered architecture, the first one is the data collection layer, second layer is the calibration layer and third layer is the model construction and trading layer. The first layer is the data collection layer and it collects both tick-by-tick and 15 Minute interval data for currencies of interest. The second layer is the calibration layer. In this layer indicator parameters, clustering parameters and data sanitization parameters are calibrated with historical data. Five years of historic data are retrieved from data collection layer and separated into disjoint sets to discover new motifs and cluster the motifs. In the third layer historical data is used to create motifs and motif clusters, and then the clusters are used for each trading period to train SVM models using the given data and previously acquired calibration results. Motifs might result in differing price movements based on their underlying characteristics. Our program represents the Zigzag technical indicator data as a set of inputs to train for the SVM to capture these characteristics. To determine SVM parameters efficiently several different settings are experimented and the selection is made based on training and validation accuracy.

In our algorithm, SVMs provide estimates for the captured motif's future behavior such as the maximum amount of decrease/increase to be expected in the event horizon, the maximum amount of inverse price-action before the expected movement and the last price at the event horizon.

Definition 29 –Given a streaming time series $T = < o_1, o_2, \dots, \dots >$, a motif M of length n with the signal point o_i, event horizon of M, E_M is a subsequence of T length n consisting of contiguous time instants from T such as $E_M = < o_{i+1}, o_{i+2}, \dots, o_{i+n} >$. A motif is characterized by the definitive Zigzag points that it contains, however due to Zigzag's nature motifs with same number of Zigzag points can be of varying lengths. The length of the event horizon for a motif is equal to the number of data points in the motif.

The trades made by our machine is not simply buying or selling a single financial instrument, rather it is taking a position with a calculated risk and reward. The amount of risk or reward is also multiplied or divided based on the position's future value. Whenever the predicted change in the value is extreme, the lot size of the order is increased accordingly and vice-versa.

The motifs recognized during the simulation by the system are clustered using the clusterer in this layer. Once clustering is done, SVM's are used to create trading parameters optimal for the associated cluster. Based on classification results trade orders are submitted to the trade engine. Trades are implicitly closed once they hit stop loss or take profit values. If a trade does not hit a stop loss or take profit until motif's event horizon, trade is closed explicitly by the trading layer. In our system we accept the stop loss and take profit values are forecasted based on SVM classifications. The order sizes are also scaled based on SVM's forecasts. You can find a summarization of our algorithm in pseudo code format in Algorithm 5 and 6.

```
ZZMOP Model Construction Algorithm
Require: CL ← currency list, RPH ← raw price history with fixed time frame
Ensure: Clusterer containing historic motif clusters
for all c ∈ CL do
    for all timeframe ∈ RPH do
        motif ← Zigzag [RPH[c, timeframe]]
        motifList.append(motif)
    clusterer [c] ← buildCluster [motifList]
    for all motif ∈ motifList do
        clusterer [c].updateClusterRewardRiskRatio(motif)
return clusterer
```

Algorithm 5: An overview of our ZZMOP model construction algorithm in pseudocode format

```
ZZMOP Trading Algorithm
Require: CL ← currency list, clusterer ← clusterers, RPF ← raw prices in the
future with fixed time frame
Ensure: Trade orders with trading parameters
for all c ∈ CL do
    for all timeframe ∈ RPF do
        motif ← Zigzag [RPF[c, timeframe]]
        motifList.append(motif)
        clusterer [c] ← updateCluster [motifList]
    for all motif ∈ clusterer[c] do
    {case: if the motif reward/risk ratio is high enough to buy:}
    if buyRewardRiskRatioHigh(cluster) then
        lotSize ← lastForecastingSVM ← train [cluster]
        stopLoss ← lowUntilForecastingSVM ← train [cluster]
        takeProfit ← highForecastingSVM ← train [cluster]
        return order = submitBuyOrder (c, lotSize, stopLoss, takeProfit)
    else if sellRewardRiskRatioHigh(cluster) then
        lotSize ← lastForecastingSVM ← train [cluster]
        stopLoss ← highUntilLowForecastingSVM ← train [cluster]
        takeProfit ← lowForecastingSVM ← train [cluster]
        return order = submitSellOrder (c, lotSize, stopLoss, takeProfit)
```

Algorithm 6: An overview of our ZZMOP trading algorithm in pseudocode format

5.2 Expressing Financial Time Series Similarity with Zigzag

The use of chart patterns in financial transactions is based on the sense of historic recurrence. Similar historic chart patterns would usually result in similar price changes in the future. Financial time series data contains various irregularities which makes the data noisy, hence measuring the similarity between the time series becomes hard. We provide three examples of these irregularities below. In Figure 15 (a) and (b) market behavior creates an extraordinary market movement against the actual trend whilst keeping the actual trend intact. These types of movements are called long squeezing or short squeezing depending on the current trend and try to eliminate open positions with stop loss orders. In Figure 15 (c) the bars around 5:00 are empty due to lack of volatility and transactions in the market.

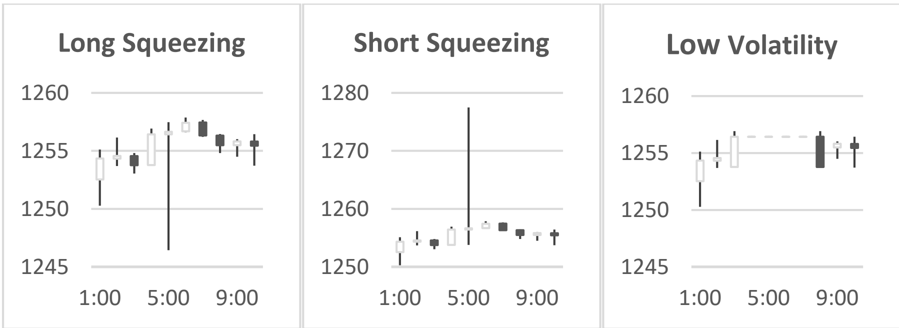

Figure 15: Examples of market noise (a) long squeezing (b) short squeezing (c) low volatility

There are two general ways of eliminating noise in financial time series data.

Averaging: Averaging models take a high variance time series and transform it into a low variance time series. For instance for a time series $T = < o_1, o_2, \dots, o_n >$, a simple averaging model would produce a new time series $T_{average} = < \frac{\sum_1^a o_{1-i}}{a}, \frac{\sum_1^a o_{2-i}}{a}, \dots, \frac{\sum_1^a o_{n-i}}{a} >$ where a is the number of past values to average. This model filters out momentary spikes in the time series, but is prone to a repetitive data with no meaning.

Summarizing: A summarizing model takes a continuous time series and transforms it into a sparse time series. For instance for a time series $T = < o_1, o_2, \dots, o_n >$, a summarizing model would produce a new time series $T_{summary} = < o_1, *, \dots, *, o_k, *, * , *, \dots\ o_n >$. This model filters out repetitive data with no meaning, but might be prone to spikes in the time series. In our approach we interpolate the blank data in between Zigzag points linearly to measure Zigzag thickness coverage.

Zigzag is a summarizing method for eliminating noise in the data. It can be used on a multivariate time series to segment it. During segmentation not only it reduces the dimensionality of the original time series but also it creates a summary of the data based on its prominent features. It has certain parameters to filter spikes in the times series and repetitive data in the time series. Due to the low number of data points it has, it is easier to measure Zigzag time series similarity with other time series. The impact of Zigzag and moving average models are shown in Figure 16. The moving average used in Figure 16 (b) is a fast moving average which averages 15 timeframes.

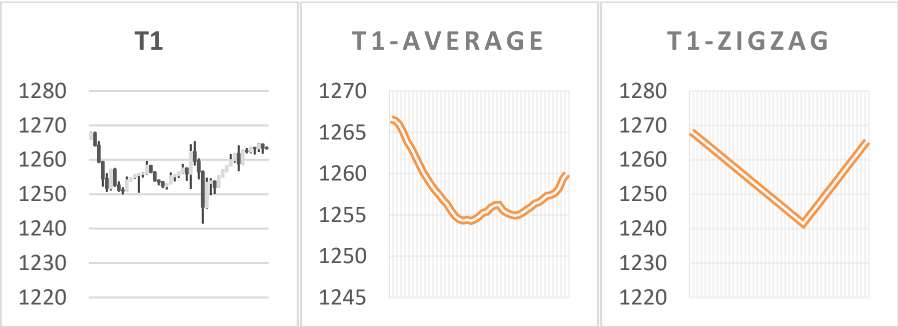

Figure 16: Representation of (a) an original time series with (b) averaging and (c) zigzag models

The Euclidean difference between original time series T, and Zigzag and moving average models are shown in Figure 17. Mean absolute error for the Zigzag model is *6.3317* and for the averaging model it is *3.854*. This is a general case for these noise elimination methods. Averaging method follows the data more closely, and Zigzag does not. For a trading scenario the more important issue is to model data that is closer at the trend reversal points. For *T1* time frames < *0, 28, 43* > are important points to model. For these points Average-T1 difference is < *1.2566, 9.272, 3.3229* > and Zigzag-T1 difference is < *0.35, 4.82, 2.0873* >. Since Zigzag is focused on representing extremum points in a time series more closely, it is sensible to use Zigzag to model trends and trend reversals.

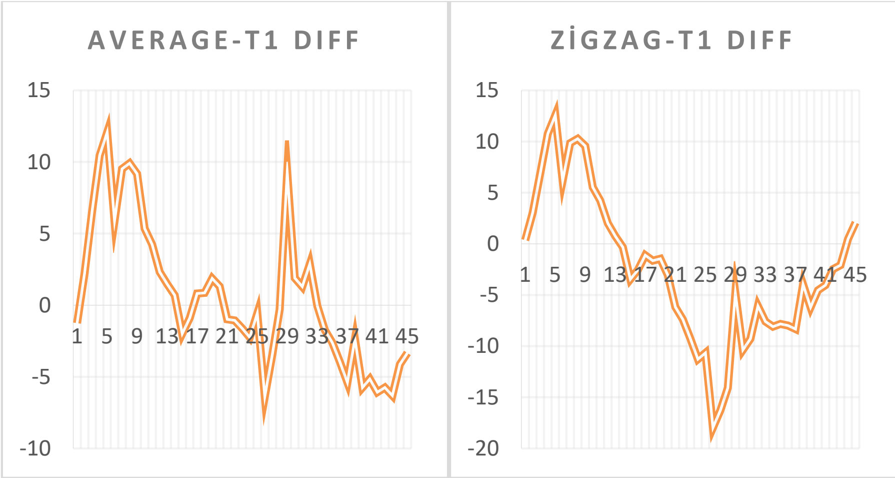

Figure 17: Euclidean difference of T1 with (a) moving average model and (b) zigzag model

5.3 Modified Zigzag Indicator with Thickness

Zigzag indicator is used to highlight the significant highs and lows of the instrument's historic values and eliminates the noise in the data, however it does not provide a method to indicate the compliance of the time series data with Zigzag lines. We add a *thickness* variable to the Zigzag indicator to represent the compliance of the time series data with the Zigzag$_{thick}$ lines.

Given time series $T = < t_1, t_2, \dots, t_n > (t_i \in N^+)$ and thickness k Zigzag Z$_{thick}$ satisfies the following:

1. $Z_{thick} =< z_1, z_2, \dots, z_n > \subset T$
2. $\forall z_i \; z_i \neq z_{i+1}$
3. $\forall \, Zigzag \; point \; z_i \in Z_{thick} = < z_1, .., z_n > \rightarrow ((z_i > z_{i-} \wedge z_i > z_{i+}) \oplus (z_i < z_{i-} \wedge z_i < z_{i+}))$
4. $\forall t_i \in T, \forall z_i \in Z_{thick} \rightarrow z_i - k \leq t_i \leq z_i + k$

Using *thickness* as an indication of motif compliance requires the consideration of two different variables; first one being the value of the thickness variable itself and the second one being the amount of *coverage*, which represents the amount of bars that are required to be contained in the thick Zigzag line.

Given a multivariate time series $MT = << low,\ high >_1, < low\ , high >_2, \ldots, < low, high >_n>$ where $o_m \in R$, $Z_{thick} =< z_1, z_2, \ldots, z_n > \subset T$ with thickness k, and coverage *C*, *C*% of the elements in *MT* satisfies the following:

1. $\forall t_i\ \in MT, \forall\ Zigzag\ point\ z_i\ \in Z_{thick} \rightarrow\ z_i - k \leq t_i \leq z_i + k$

There could be scenarios where requiring full coverage which would require a very high thickness is less optimal than a partial coverage which would require a lower thickness. In each scenario, a high bar percentage value and a low thickness value would indicate a higher motif compliance. Similarly a low bar percentage and high thickness would indicate a low motif compliance. The Zigzag indicator with thickness effectively creates an error function to describe a *covering motif* [5] which can be used to select *thinnest* motifs available. Two sample motifs with low and high compliances to the thick Zigzag lines are provided in Figure 18 (a) and (b) respectively.

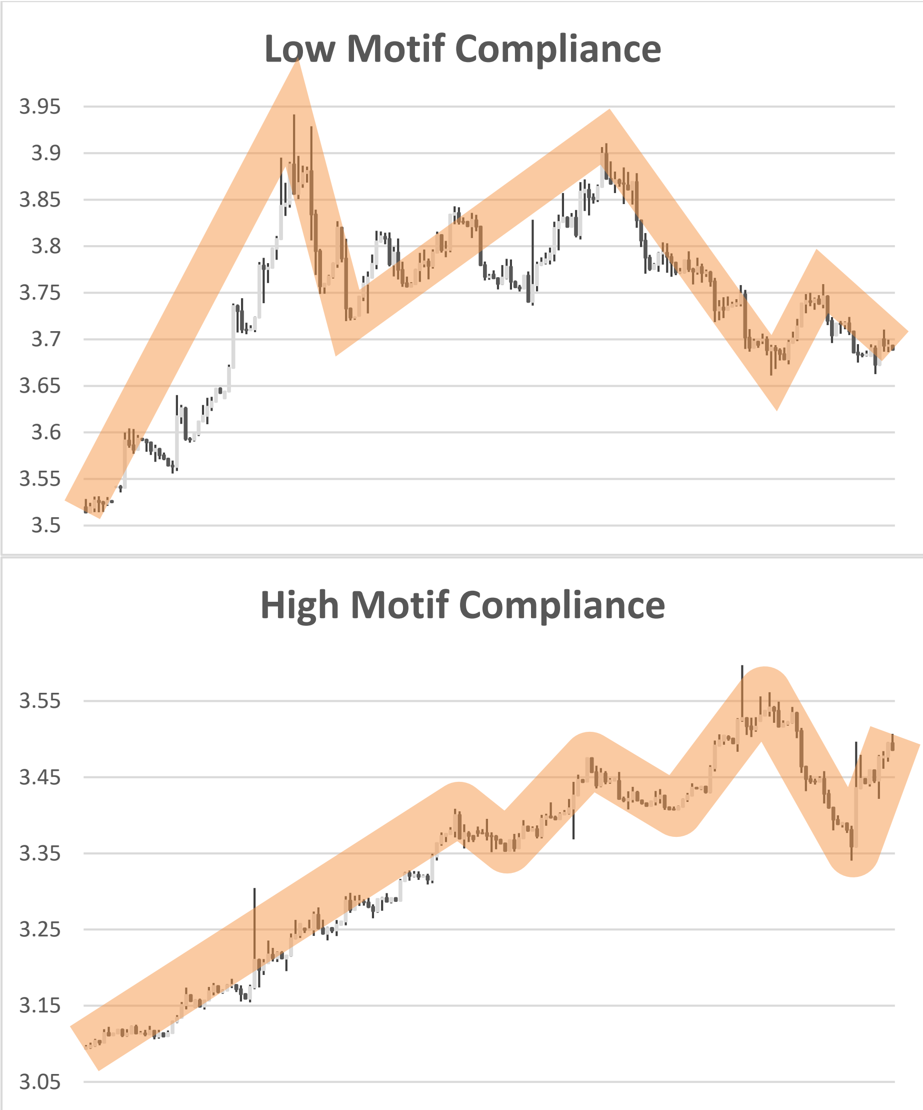

Figure 18: A thick Zigzag line with (a) low motif compliance and (b) high motif compliance

5.4 Detection of the Signal Point in a Motif

A financial time series in the Forex environment is a multivariate time series consisting of observations made at equal intervals. Each observation is a vector $O = < open, close, low, high >$. Even though most distance measures and mining algorithms are invariant to the start time and sampling interval of the time series [28, 29], the case for the Forex market is different [30]. Therefore the detection of a signal point is crucial for the success of forecasting. Detection of the signal point in a motif problem is similar to change point detection in time series. A survey of change point detection methods can be found in [31]. Our approach includes time series transformations to the original time series and is supervised by its nature.

For a given financial time series T Zigzag indicator creates a list of values $V = <zz_1, zz_2, \dots, zz_n>$ such that $\left(<zz_1, zz_3, \dots, zz_{2x+1}> \in ZZ_{low} \wedge <zz_2, zz_4, \dots, zz_{2x}> \in ZZ_{high}\right) \oplus \left(<zz_1, zz_3, \dots, zz_{2x+1}> \in ZZ_{high} \wedge <zz_2, zz_4, \dots, zz_{2x}> \in zz_{low}\right)$ where $ZZ_{low} = \{zz_{low}: zz_{low} \in T\} \Leftrightarrow \forall\, o_x \in \{o_{i-depth} .. o_{i+depth}\} : o_i \le o_x \quad \vee \quad ZZ_{high} = \{zz_{high}: zz_{high} \in T\} \Leftrightarrow \forall\, o_x \in \{o_{i-depth} .. o_{i+depth}\} : o_i \ge o_x$. In other words, the values generated by the Zigzag indicator are Zigzag high and Zigzag low values in alternating order. Also, the Zigzag values are members of the original time series, however for a specific interval a Zigzag can be valued at any of the four possible values of the observation vector.

Zigzag creates high and low values for specific time intervals, however for the remaining intervals no values are generated. Ordering all of the values generated by Zigzag in a set $T_V = \{(t_i, v_i) | t_i < t_{i+1}, i = 1..n-1\}$ would result in a sparse time series. A sparse time series has many zero-valued observations as opposed to non-zero observations. The ratio between the length of the time series and the number of non-zero observations is defined as the sparsity factor *s* of the time series. In our application, the sparsity factor *s* is dependent on Zigzag depth *d*. A higher Zigzag depth would result in more zero-valued observations. For the encoding of the Zigzag time series T_V we use a length encoding method [32]. In the length encoded series T_{Ve} we replace the contiguous *k* zeroes in T_V with a (*k*). For instance for the sparse Zigzag time series $T_V = < 1.2569, *, *, *, *, 1.2437, *, *, *, 1.2701 >$, the length encoded series is $T_{Ve} = < 1.2569, (4), 1.2437, (3), 1.2701 >$. In this example, 1.2569 and 1.2701 are high Zigzag points and 1.2437 is the low Zigzag point. All three points are Zigzag points, whereas the remaining 7 points are *insignificant data* points.

Our algorithm classifies the motifs in terms of the number of *significant* Zigzag points. A class *C* of Zigzag contains time series

For a motif $T_V =< v_1, v_2, \ldots, v_n >$, the n[th] point in the original time series is a signal point if the motif belongs to a cluster *C* where *Reward/RiskRatio(C)* > *e*, and *e* is a predefined error. The signal price for a motif, p_{zz} at point *n* in $T_V =< v_1, v_2, \ldots, v_n >$ is equal to the closing price $close_i$ of the n[th] point in the original time series $T =< o_1, o_2, \ldots, o_n >$ where $o_i =< open_i, close_i, low_i, high_i >$.

The clustering and classification of motifs in different time frames require time series transformations such as amplitude shifting, uniform amplification and uniform time scaling as discussed in Section III.

5.5 Motif Reward/Risk Ratio

Trades in the financial markets come with rewards and risk. The ratio of these two numbers is called the reward/risk ratio [33] and is an indicator of the profitability of the trade position.

Definition 30 - Reward/risk ratio is the ratio of possible profits to possible losses. Reward is the movement of price in the expected direction, which would result in profits and the risk is the movement of price in the unexpected direction, which would result in losses.

Every detected motif and signal associated with the motif should not result in an open trade position. The reward associated with the motif should be high enough to bear the risk. In our system the reward/risk ratio of motif clusters are continuously monitored. Once a trade signal is received from a motif, the trade position is opened if and only if the motif ends up in one of the most rewarding clusters.

Reward/Risk ratio is essentially used to classify a cluster of motifs using a feedback loop where future data in the time series effects the predefined class label of the cluster.

Definition 31 – *High until low* is the difference between the decision point of a short motif and the highest point in the event horizon of the motif that is recorded before the lowest point in the event horizon of the motif. It represents the risks associated with a short motif in points.

Definition 32– *Low until high* is the difference between the decision point of a long motif and the lowest point in the event horizon of the motif that is recorded before the highest point in the event horizon of the motif. It represents the risks associated with a long motif in points.

ZZMOP predicts future movement after the decisive point of the motif. Despite a motif being a short motif a “high until low” value (i.e. the risk) is taken into account to allow the future decrease in price to be achieved. For a short motif the “low” value forecasted represents the reward. A sample short motif having a reward/risk ratio of 3:1 means that when the increase after the decisive point is proportional to 10 pips, the decrease after the maximum drawdown point is proportional to 40 pips which would place the absolute difference between decisive point and maximum profit point proportional to 30 pips. In this scenario possible losses at maximum drawdown point is valued at 10 pips where the possible rewards at maximum profit point is 30 pips. Hence the reward/risk ratio is 3:1.

5.6 Motif Directional Bias

Prices in the financial markets fluctuate in both directions in a specific timeframe but with different magnitudes. One can benefit from a motif in both directions, as long as take profit and stop loss locations are determined properly. Since the event horizon of a motif contains movements in both directions, a reward/risk ratio can be computed for both directions too.

For a motif to be used for a sell position, the reward/risk ratio of the motif in the sell direction should be higher than the reward/risk ration in the buy direction. For a motif to carry a meaningful information with respect to the direction of trade, the reward/risk ratios between sell and buy positions should have a significant difference. Otherwise the motif can be used in both directions and hence does not point to a certain type of trade.

Definition 33 – Given a motif *M*, *directional bias* of *M* can be calculated as the absolute difference of short and long risk reward ratios: $e_M = |\ Reward\ Risk\ Ratio_{short} - Reward\ Risk\ Ratio_{long}|$.

For a given motif, the reward/risk ratio can be computed once the prices materialize in the event horizon of the motif. We provide sample price action in the event horizon of different motifs in Figure 19. Figure 19 (a) contains the price action of a buy motif with high directional bias, (b) a sell motif with high directional bias, (c) a motif with low directional bias.

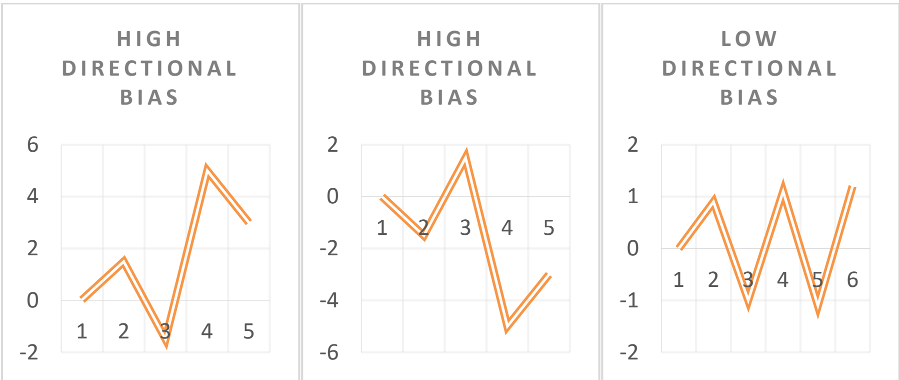

Figure 19: Price changes in the event horizon of a (a) buy motif with high directional bias (b) sell motif with high directional bias (c) motif with low directional bias

5.7 Trade Parameter Detection

Before entering a trade with a discovered motif, our algorithm determines the lot size for the order, the stop loss and take profit locations. This is done using the previously recorded motifs in the cluster. For each type of order a separate set of values are forecasted to determine these values.

For sell orders, a drop in the price is expected in the future. The expected amount of drop is supposed to be the take profit location. However, there might be a slight increase in the price before a drop occurs. Our system should be able to overlook this much of a loss at any time to reach the forecasted drop and take profit. Another possibility is that the increase in the price continues and price does not drop in the event horizon; in which case the system should stop the losses. Hence the stop loss location should account for these two possible price movements. The stop loss location for a sell order is called “high until low” and the take profit location is called “low”.

For buy orders, a hike in the price is expected in the future, which will be the take profit location. A decrease in price prior to the hike should be allowed to reach the hike forecasted. A higher decrease should be evaluated as a false signal and losses shall be suffered. The stop loss location for a sell order is called “low until high” and the take profit location is called “high”. Our system trains with previous data in a motif’s cluster to determine the appropriate take profit and stop losses to maximize gains for each type of motif.

For both types of orders, the expected ultimate value in the event horizon of the motif reflects the reliability of the signal. If the expected ultimate value is very high for a motif, the likelihood of a profit for the motif is higher. In that case the lot size of the order can be increased accordingly to maximize profits. The same logic applies to sell orders.

CHAPTER 6

EXPERIMENTS AND RESULTS

In this section we provide details about the dataset we use, parameters we experiment with and the results we have obtained with the three different approaches we have carried out throughout this work.

6.1 TI-MFT Results

In our TI-MFT algorithm we are making use of the RSI and CCI technical indicators. We also determine optimal trade parameters for a given trade interval. Both the indicators and trades have certain parameters that adapt them to the interval at hand. We discuss the characteristics of our dataset and how we adapt parameters for the aforementioned components to our data.

6.1.1 Characteristics of Our Dataset

Two types of historical data are collected by our system. First is price data with 1 Minute intervals which summarizes the opening, closing, low and high values for the given interval. 1 Minute interval data values are used for training purposes. Second data is the real time price data which contains all the price changes that have happened in the currencies. Real time price data is used for testing and simulation purposes. Both data is collected for currency pairs including USD, EUR, GBP, and CHF. Summary statistics of the data can be found in Table 4. Total data points are the number of 1 Minute bars, minimum and maximum values represent the highest and lowest values recorded, average value is the average of 1 Minute closes, increases and decreases in value show a 1 Minute increase or a 1 Minute decrease in the closing price of the currency pair. For training 1 Minute values are used.

Table 4: 1 minute exchange rate (training data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value	Decreases in Value
EURCHF	1821600	1.488	1.007	1.253	836640	984960
EURGBP	1823040	0.913	0.776	0.839	914400	908640
EURUSD	1823040	1.493	1.187	1.334	933120	889920
GBPCHF	1821600	1.711	1.146	1.493	884160	937440
GBPUSD	1823040	1.718	1.423	1.589	914400	908640
USDCHF	1823040	1.172	0.706	0.941	891360	931680

Since our system makes use of real life trading parameters such as stop loss and take profit, for testing and model simulation, real time price data are used. Statistics for these values can be found in Table 5.

Table 5: Real time exchange rate (testing data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value (%)	Decreases in Value (%)
EURCHF	59948502	1.488	1.007	1.248	45.88	54.12
EURGBP	67221512	0.913	0.776	0.826	50.12	49.88
EURUSD	62915513	1.493	1.187	1.316	51.12	48.88
GBPCHF	76187806	1.711	1.146	1.498	47.99	52.01
GBPUSD	65708389	1.718	1.423	1.591	50.22	49.78
USDCHF	59461873	1.172	0.706	0.939	48.84	51.16

6.1.2 RSI Related Parameters

RSI can be used to detect the strength of the instrument that is being traded. While a higher than typical RSI value might suggest an increase in the price, a contiguously higher RSI might suggest a correction in the instrument thus a decrease in the price.

In our work we do not use static values of a single RSI indicator, rather we use crossovers of different RSI indicators to trigger buying and selling positions. Therefore we have a fast changing and a slow changing RSI indicator. The idea is that the fast and the slow indicators reflect changes at different time scales, so a divergence in their direction might be indicative of tensions accumulating in the price. The parameters searched throughout our work are outlined in Table 6.

Table 6: RSI related parameters

Parameter Name	Values Experimented	Value Used	Explanation
RSI Fast Periods to Measure	5, 10, 14, 21, 30, 50, 100, 200	Dynamic Selection	The fast moving RSI considers a lower number of periods and is hence more prone to oscillation.
RSI Slow Periods to Measure	10, 14, 21, 30, 50, 100, 200, 500	Dynamic Selection	The slow moving RSI will be the average that gets crossed above and below by the fast moving RSI.
Price to Measure	Open, High, Low, Close, Median	Close	RSI can be applied to any type of price. Our experiments show that applying the RSI to the closing price of the instrument delivers the best results.

6.1.3 CCI Related Parameters

CCI warns the traders of extreme market conditions and emergence of possible new trends. Like the RSI indicator CCI is a short term indicator. We do not statically use a CCI indicator to trade on certain extreme values, rather we use two different CCI values and consider their crossovers to enter and exit trades.

Therefore like the RSI indicator we have a fast changing and a slow changing CCI indicator. The parameters searched throughout our work are outlined in Table 7.

Table 7: CCI related parameters

Parameter Name	Values Experimented	Value Used	Explanation
CCI Fast Periods to Measure	5, 10, 14, 21, 30, 50, 100, 200	Dynamic Selection	The fast moving CCI considers a lower number of periods and is hence more prone to oscillation.
CCI Slow Periods to Measure	10, 14, 21, 30, 50, 100, 200, 500	Dynamic Selection	The slow moving CCI will be the average that gets crossed above and below by the fast moving CCI.
Lambert Variable	0.005, 0.010, 0.015, 0.05, 0.1	0.010	Lambert variable effects the oscillation amount of the CCI indicator. A higher Lambert variable results in lower oscillation.

6.1.4 Trading Related Parameters

A realistic trading system requires certain parameters such as trading volume, stop loss or take profit. In our system we have empirically experimented with several values for these parameters that are common in the industry and selected the best performing ones. To determine the best performing parameters for our system, all possible combinations of the parameters have been exhaustively searched resulting in 18.432.000 experiments per each forecasting period. These parameters are outlined in Table 8.

Table 8: Trading system parameters

Parameter Name	Values Experimented	Selected Value	Explanation
Trading Volume (Lot)	0.01, 0.1, 0,25, 0,5, 1	0.1	Forex providers generally enable 0.1 or 1 lot trades. For a small account mini-lots (0.1 lot) or micro-lots (0.01 lot) are better suited. In our experiments using higher lot sizes resulted in defaulting accounts. Lower lot sizes resulted in less profits.
Stop Loss (Pips)	2, 5, 10, 20, 50, 100	Dynamic Selection	When a currency pair moves in the opposite direction of an open trade, stop loss amount determines how much loss is accepted before closing position. In our experiments using fixed stop losses proved less profitable, therefore the stop loss is dynamically selected based on reverse price-action estimate.
Take Profit (Pips)	2, 5, 10, 20, 50, 100	Dynamic Selection	A trade position cannot be kept open forever even though it is in profit, since markets fluctuate. Take profit determines how much profit a single open trade can make. In our experiments using fixed take profits proved less profitable, therefore the take profit is dynamically selected based on price-action estimate.

6.1.5 Genetic Algorithm Parameters

The genetic evolution process requires certain parameters which directly affect the performance. We have empirically experimented with several values and selected the best performing ones in our previous work in [46]. The parameters that we have used are outlined in Table 9.

Table 9: Genetic algorithm parameters

Parameter Name	Values Experimented	Selected Value	Explanation
Maximum Number of Generations	10, 20, 50, 100, 500	100	This is the maximum number of generations that are allowed to evolve.
Mutation Probability	0.01, 0.10, 0.20, 0.50	0.10	The probability of a mutation on a gene.
Elitism Rate	0.05, 0.10, 0.20, 0.50	0.10	The rate of genes that are to be preserved for the next generation due to their superior fitness.
Population Size	50, 100, 200	50	Size of the population is the number of genes that are present in each evolution cycle.
Active Genetic Operators	N/A	Crossover, Gaussian Mutation	Genetic operators to be applied to current genes to create the next offspring.
Selection Method	N/A	Tournament Selector	Selection mechanism for the current genes from the current population to next cycle.

6.1.6 Performance of Our System

Different trading strategies and input compositions were used with the system and different results in terms of performance were obtained. Obtained results are presented in Table 10.

Table 10: Experimental results

Validity/ Training Interval	Weekly Validity/ Monthly Training		Daily Validity/ Weekly Training		2 Hour Validity/ Daily Training	
Currency	Exhaustive	Genetic	Exhaustive	Genetic	Exhaustive	Genetic
EURCHF	56513	56082	58812	57999	57996	55983
EURGBP	62525	57928	61415	59303	64183	64366
EURUSD	72244	72349	70888	70821	73515	69816
GBPCHF	51265	49921	53669	53680	54627	52814
GBPUSD	48952	49016	46819	47007	49002	50479
USDCHF	60057	60172	61473	61375	60813	60001
Total Balance	351556	345468	353076	350185	360136	353459

Exhaustively searching the ~18 million possible parameter combinations takes significant amount of time. In our experiments we have used an early stopping criteria for underperforming models; an experiment is aborted once 10% of the account balance is lost. Even though many of the experiments are terminated early due to this criteria, experiments take too long to complete when real time data is used for simulation. To allow exhaustive parameter selection to complete in time, parameter selection intervals has been limited as in Table 11. For longer intervals (i.e. 2 days parameter selection for 2 hour parameter validity) exhaustive method cannot be completed within the required time.

Table 11: Simulation runtimes for different models

	Exhaustive	Genetic	Exhaustive	Genetic	Exhaustive	Genetic
Parameter Validity	Weekly	Weekly	Daily	Daily	2 Hours	2 Hours
Parameter Selection Interval	Monthly	Monthly	Weekly	Weekly	Daily	Daily
Number of Trials	18432000	5000	18432000	5000	18432000	5000
Simulation Time (Minutes)	7852	136	1216	21	114	2
Available Simulation Time (Minutes)	10080	10080	1440	1440	120	120

In the genetic algorithm assisted parameter optimization model, a population of 50 parameters are evolved through 100 generations, which accounts to roughly 5000 parameter combination trials. This cuts down the running time significantly. Our current model uses two technical indicators and hence does not require too many parameter combinations. A more advanced system with a higher number of technical indicator parameters would benefit ever more from genetic algorithm assisted parameter optimization. The runtimes of experimental simulations are provided in Table 11.

Table 11 shows that an exhaustive search for all the parameters can be barely completed in the available simulation time. Further parameter refining for additional indicators or trade parameters cannot be done in the available time. Genetically optimized trading models have comparable performance to exhaustively optimized models as outlined in Table 10. Thus we use “2 Hour Validity/Daily Training with Genetic Parameter Optimization” as our reference model for our discussion.

The system cannot generate successive buy signals or successive sell signals due to the nature of signals – they require crossovers. Algorithmic performance of the system is measured with accuracy and f-measure and is comparable with similar systems [46] as shown in Table 12. Trading performance of the system is measured with acquired account balances and is given in Table 13.

Table 12: RSI and CCI based medium frequency trading algorithm performance

Traded Currencies	Precision	Recall	Accuracy	F-Measure
EURCHF	0,5372	0,5744	0,5612	0,5562
EURGBP	0,6208	0,6113	0,6105	0,6174
EURUSD	0,625	0,6445	0,6344	0,6399
GBPCHF	0,4835	0,5123	0,5273	0,4994
GBPUSD	0,5228	0,4748	0,5129	0,4966
USDCHF	0,5632	0,5701	0,5864	0,5662

As Table 13 outlines, our approach beats a revenue model based on interest rates for every currency. The highest profits are recorded for the EUR/USD pair where the one year end experiment results in a combined 39.62% profits with an account balance of 69816 USD. The lowest account balance of 50479 USD is recorded for the GBP/USD pair with a combined 2.39% profits, which is still above the interest rate for both of these currencies. The highest profits in a single month was for the EUR/CHF pair with 9.3% in September and the lowest profit in a single month was for the EUR/CHF pair with -9.4% in March. Average monthly profit for the system was 1.67% taking all currencies into consideration. Figure 20 shows the evolution of balance and exchange rate for the currencies during the trading period.

Table 13: RSI and CCI based medium frequency trading performance

Currency		EUR/ CHF	EUR/ GBP	EUR/ USD	GBP/ CHF	GBP/ USD	USD/ CHF
01.01.2015	Start	50000	50000	50000	50000	50000	50000
	End	51440	54710	54322	48164	48015	54525
	Profits	0,029	0,094	0,086	-0,037	-0,04	0,091
02.01.2015	Start	51440	54710	54322	48164	48015	54525
	End	54485	53325	55513	49842	50889	58818
	Profits	0,059	-0,025	0,022	0,035	0,06	0,079
03.01.2015	Start	54485	53325	55513	49842	50889	58818
	End	59193	50938	58159	48586	50969	62214
	Profits	0,086	-0,045	0,048	-0,025	0,002	0,058
04.01.2015	Start	59193	50938	58159	48586	50969	62214
	End	53605	54333	56298	48001	49909	66080
	Profits	-0,094	0,067	-0,032	-0,012	-0,021	0,062
05.01.2015	Start	53605	54333	56298	48001	49909	66080
	End	56521	54528	61300	49654	49012	68005
	Profits	0,054	0,004	0,089	0,034	-0,018	0,029
06.01.2015	Start	56521	54528	61300	49654	49012	68005
	End	56385	53743	59826	51554	52549	72484
	Profits	-0,002	-0,014	-0,024	0,038	0,072	0,066
07.01.2015	Start	56385	53743	59826	51554	52549	72484
	End	53318	56452	64181	51933	48471	66047
	Profits	-0,054	0,05	0,073	0,007	-0,078	-0,089
08.01.2015	Start	53318	56452	64181	51933	48471	66047
	End	51744	54736	70034	50003	53047	62560
	Profits	-0,03	-0,03	0,091	-0,037	0,094	-0,053
09.01.2015	Start	51744	54736	70034	50003	53047	62560
	End	56615	59597	67849	51174	54415	64962
	Profits	0,094	0,089	-0,031	0,023	0,026	0,038
10.01.2015	Start	56615	59597	67849	51174	54415	64962
	End	60318	62029	69915	47858	51034	62216
	Profits	0,065	0,041	0,03	-0,065	-0,062	-0,042
11.01.2015	Start	60318	62029	69915	47858	51034	62216
	End	56504	62625	72203	51265	48952	60099
	Profits	-0,063	0,01	0,033	0,071	-0,041	-0,034
12.01.2015	Start	56504	62625	72203	51265	48952	60099
	End	55983	64366	69816	52814	50479	60001
	Profits	-0,009	0,028	-0,033	0,03	0,031	-0,002

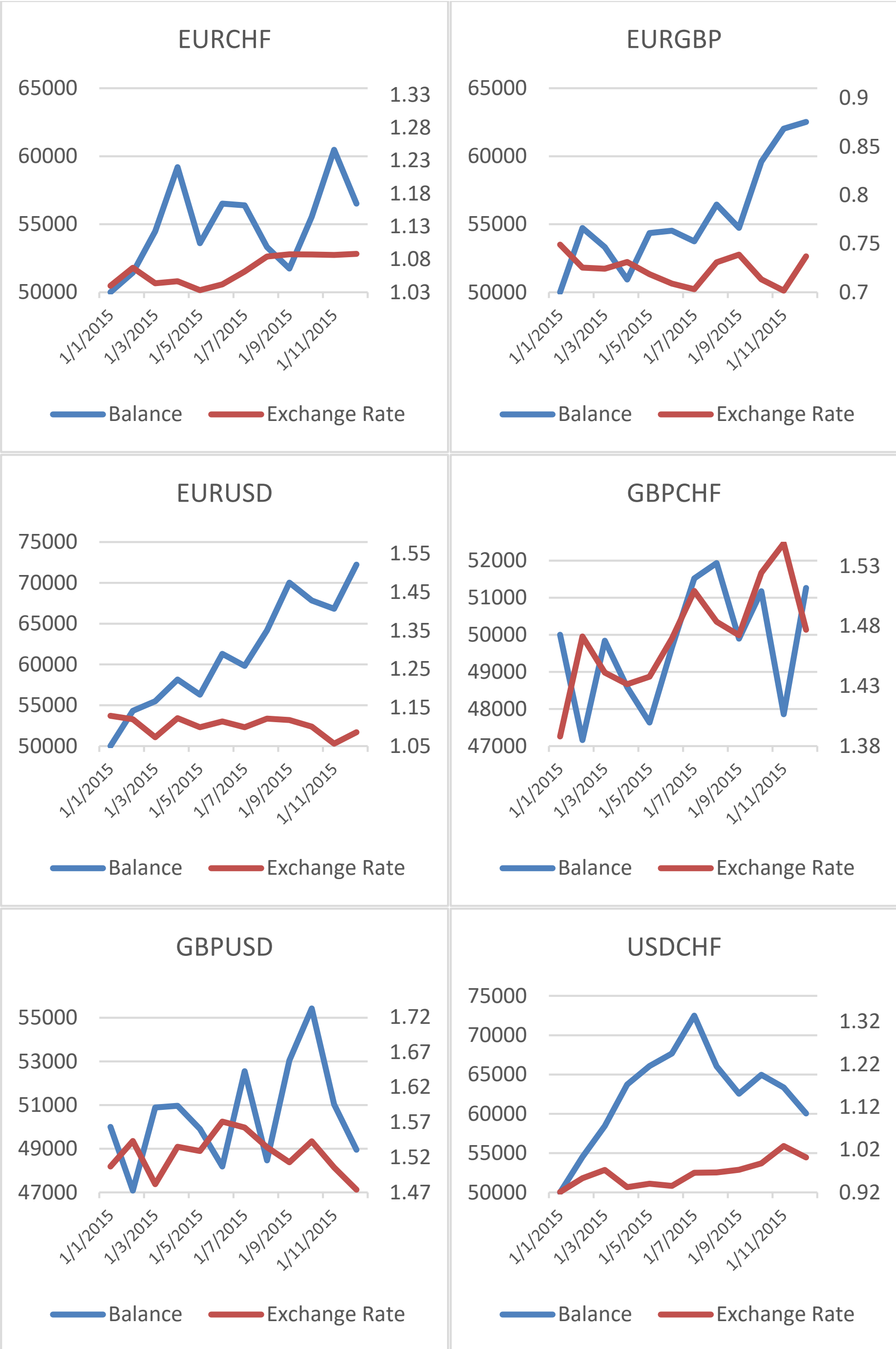

Figure 20: Evolution of balance and exchange rate for (a) EURCHF (b) EURGBP (c) EURUSD (d) GBPCHF (e) GBPUSD (f) USDCHF currencies between 01.01.2015 and 12.01.2015

A summary of trading statistics regarding the trades made with the TI-MFT algorithm is provided in Table 14.

Table 14: TI-MFT algorithm trade statistics

Currency	EURCHF	EURGBP	EURUSD	GBPCHF	GBPUSD	USDCHF
Open Positions	6032	5408	5569	5513	6297	6195
Profit Positions	3385	3302	3533	2907	3230	3633
Loss Positions	2647	2106	2036	2606	3067	2562
Average Profits	7,2	6,4	9,5	6,8	7,9	6,4
Average Loss	-6,9	-3,2	-6,8	-6,5	-8,2	-5,2
Total Profit Pips	5983	14366	19816	2814	479	10001
Maximum Account Draw-Down %	13,38	9,62	7,14	7,63	10,9	14,29

6.2 SBT-DAP Results

Our first approach is using trend deterministic technical indicator signals with strength bias. In this section results of the SBT-DAP algorithm which implements this approach are provided.

6.2.1 Characteristics of Our Dataset

The transactions and data in the Forex market is determined in real-time by an electronic network of Forex brokers, liquidity providers, banks and other financial institutions. Historical data is available through several different channels, and the data used in our work is obtained from TrueFX [21], which is also used in related work. Two types of historical data are collected by our system. First is daily price data which summarizes the opening, closing, low and high values for the day. Daily values are used for training purposes. Second data is the real time price data which contains all the price changes that have happened in the currencies. Real time price data is used for testing and simulation purposes. Both data is collected for currency pairs including USD, EUR, GBP, and CHF. Summary statistics of the data can be found in Table 15. Total data points are the number of trade days, minimum and maximum values represent the highest and lowest values recorded, average value is the average of daily closes, increases and decreases in value show a daily increase or a daily decrease in the closing price of the currency pair. For training daily values are used.

Table 15: Daily exchange rate (training data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value	Decreases in Value
EURCHF	1265	1.488	1.007	1.253	581	684
EURGBP	1266	0.913	0.776	0.839	635	631
EURUSD	1266	1.493	1.187	1.334	648	618
GBPCHF	1265	1.711	1.146	1.493	614	651
GBPUSD	1266	1.718	1.423	1.589	635	631
USDCHF	1266	1.172	0.706	0.941	619	647

Since our system makes use of real life trading parameters such as stop loss and take profit, for testing and model simulation, real time price data are used. Statistics for these values can be found in Table 16.

Table 16: Realtime exchange rate (testing data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value	Decreases in Value
EURCHF	59948502	1.488	1.007	1.248	45.88	54.12
EURGBP	67221512	0.913	0.776	0.826	50.12	49.88
EURUSD	62915513	1.493	1.187	1.316	51.12	48.88
GBPCHF	76187806	1.711	1.146	1.498	47.99	52.01
GBPUSD	65708389	1.718	1.423	1.591	50.22	49.78
USDCHF	59461873	1.172	0.706	0.939	48.84	51.16

The set of parameters stored for each financial instrument are listed below in Table 17.

Table 17: Symbol parameters

Parameter Name	Parameter Definition
Symbol Name	The name of the financial instrument (i.e. EUR/USD, GBP/CHF)
Open	The opening value of the financial instrument for the given trade day.
Close	The closing value of the financial instrument for the given trade day.
Low	The lowest value the financial instrument is traded for in the given trade day.
High	The highest value the financial instrument is traded for in the given day.
Volume	The total volume of trade performed on that financial instrument.

After the raw price data is collected for each relevant financial instrument, the system computes certain technical indicators and stores these values in the database for further use. The technical indicators that are automatically computed for each financial instrument are listed below in Table 18. There are hundreds of technical indicators available to traders –a brief list can be found in [34]- and many more are being developed continuously.

Table 18: Technical indicators[1]

Indicator Name	Definition	Formula
Moving Average Convergence Divergence (MACD)	MACD is a trend following momentum indicator and it shows the relationship of two moving averages of prices.	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1})$
Relative Strength Indicator (RSI)	RSI compares magnitude of gains and losses to determine overbought and oversold conditions.	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} UP_{t-i}/n)/(\sum_{i=0}^{n-1} DW_{t-i}/n)}$
Simple Moving Average (SMA)	SMA is the average of closing prices of a financial instrument in a given time period.	$\frac{c_t + c_{t-1} + .. + c_{t-9}}{n}$
Weighted Moving Average (WMA)	WMA is a weighted average of closing prices of a financial instrument in a given time period.	$\frac{(10)c_t + (9)c_{t-1} + .. + c_{t-9}}{n + (n-1) + .. + 1}$
Commodity Channel Index (CCI)	CCI is an oscillator indicator which determines if an instrument is overbought or oversold.	$\frac{M_t - SM(n)_t}{0.015 D_t}$
Stochastic K% (SK%)	SK% indicator compares closing price to the price range over a given time period.	$\frac{c_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100$

New technical indicators are created based on their success in previous market data so using new indicators in historical market data creates a forward bias. Therefore in our work we are using legacy technical indicators which have been available to traders in our training/testing interval of 2010-2015. The indicators we have selected have been widely used by other researches in related work [10,35,36,37,38].

[1] C_t is the closing price, L_t is the low price and H_t the high price of the instrument at time t; $DIFF_t = WMA(12)_t – WMA(26)_t$, LL_t is the lowest low and HH_t is the highest high in the last t days. M_t is the median price ($H_t+L_t+C_t/3$). $SM(n)_t$ is the simple average of median price for n days. UP and DW show the count of upward and downward price changes, respectively.

Table 19: Trend deterministic technical indicator signals

Signal Name	Definition	Formula	Time Period (Days)
MACD	MACD indicator follows the financial instrument’s strength.	$MACD(n)_{t-1} > MACD(n)_t \rightarrow -1$ $MACD(n)_{t-1} < MACD(n)_t \rightarrow 1$	5, 15
RSI	RSI determines when a financial instrument is overbought and oversold.	$RSI(n)_t > 70 \rightarrow -1$ $RSI(n)_t < 30 \rightarrow 1$	5, 15
SMA	SMA provides a smoothing for the price data.	$MAVG(n)_t > c_t \rightarrow -1$ $MAVG(n)_t < c_t \rightarrow 1$	5, 10, 20, 50, 100, 200
WMA	WMA provides a smoothing for the price data that gives more importance to the more recent price action.	$WMAVG(n)_t > c_t \rightarrow -1$ $WMAVG(n)_t < c_t \rightarrow 1$	5, 10, 20, 50, 100, 200
WMA SMA Crossovers	WMA value increases and decreases in a higher pace than a SMA. Crossovers between WMA and SMA signals uptrend or downtrend.	$WMAVG(n)_{t-1} < MAVG(n)_{t-1} \wedge WMAVG(n)_t > MAVG(n)_t \rightarrow 1$ $WMAVG(n)_{t-1} > MAVG(n)_{t-1} \wedge WMAVG(n)_t < MAVG(n)_t \rightarrow -1$	5, 10, 20, 50, 100, 200
CCI	CCI determines overbought and oversold conditions.	$CCI(n)_t > 200 \rightarrow -1$ $CCI(n)_t < -200 \rightarrow 1$	5, 15
SK%	SK% oscillator shows the direction of price movement.	$SK\%(n)_{t-1} > SK\%(n)_t \rightarrow -1$ $SK\%(n)_{t-1} < SK\%(n)_t \rightarrow 1$	5,15

Traders in the exchange market routinely use technical indicators to derive signals that may point to a trade opportunity. Based on experience, analysts have attempted to associate situations of possible stress in the price-determining process with specific conditions on the values of technical indicators. Upon meeting such conditions, a technical indicator signal can be generated. In its formal form a technical indicator signal is a formula which usually outputs a ternary value (i.e. -1, 0, 1) representing a buy, sell or no action decision. The formula is based on the technical indicator values for the given instrument and the selected time period. For each of the above technical indicators, a trend deterministic technical indicator signal is generated in our system. Table 19 outlines the rules for trend determination for each of the technical indicators. The time periods for the indicators (i.e. 5, 10, 20, 50, 100 and 200) were selected based on previous experience of traders and academic research [9,10,27,35,36,38,39,40].

6.2.2 Genetic Algorithm Parameters

The parameters used for the genetic algorithm are the ones that were previously used in our TI-MFT algorithm and outlined in Table 9.

Since execution time of genetic algorithms depend on fitness calculation and each fitness calculation needs training and testing of the SVM, parameters directly affecting the length of the run (i.e. population size, maximum allowed evolutions) are selected to be small numbers out of necessity.

6.2.3 SVM Parameters

An SVM requires certain parameters which directly affect its performance. We have empirically experimented with several values for these parameters that have been mentioned in related research and selected the best performing ones. To determine the best performing parameters for our SVM, all possible combinations of the parameters have been exhaustively searched resulting in 1200 experiments per each training day. The parameters that we have used are outlined in Table 20.

Table 20: SBT-DAP SVM parameters

Parameter Name	Values Experimented	Selected Value	Explanation
Kernel Type	Polynomial, Radial Basis	Polynomial	Polynomial function kernels achieved superior performance in related work [17, 27]. Thus we use the polynomial function kernel in our system.
Function Degree (d)	1, 2, 3, 4	2	Degree of polynomial function characterizes how well the model fits to the data. A very high degree could result in overfitting.
Regularization Parameter (c)	0.5, 1, 5, 10, 100	1	Regularization parameter controls margin and misclassification error.
Training Interval (Days)	10, 20, 40, 80, 160, 320	80	A model in our system is trained with the given number of days before making decisions in validation or actual data sets.
Validation Interval (Days)	10, 20, 40, 80, 160	20	A trained model in our system is validated with the given number of days.

6.2.4 System Parameters

A realistic trading system requires certain parameters such as trading volume, stop loss or take profit. In our system we have empirically experimented with several values for these parameters that are common in the industry and selected the best performing ones. To determine the best performing parameters for our system, all possible combinations of the parameters have been exhaustively searched resulting in 32768 experiments per each training day. These parameters are outlined in Table 21.

6.2.5 Performance of Our System

Different trading strategies and input compositions were used with the system and different results in terms of performance were obtained. Obtained results are presented below. For exchange ratios, buy and sell decisions are materialized based on the closing value of the decision date, since the market is open for 24 hours. The system can generate successive buy signals or successive sell signals, which would be meaningless in a real life scenario.

Performance of the system is measured with accuracy and f-measure criteria as is the case in related work [10, 27, 39]. Computation of these are made with precision and recall which are calculated from True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) decision counts. The equations regarding these criteria are provided in the Appendix.

In the first experiment the usefulness of dynamically adapting parameters via genetic algorithm is tested. In all currency pairs, the learning model is fed with all available input data (i.e. raw data and trend deterministic technical indicator data). The results are shown in Table 22. It can be seen that the system achieves different accuracy values in different pairs. The most accurately predicted pair is EURGBP and a 59.63% accuracy is recorded.

Table 21: SBT-DAP trading system parameters

Parameter Name	Values Experimented	Selected Value	Explanation
Account Balance (USD)	N/A	1000	Most experiments use 1000 or 10000 USD initial account balance. Choice of balance effects trading volume.
Trading Volume (Lot)	0.1, 1	0.1	Forex providers generally enable 0.1 or 1 lot trades. For a small account mini-lots (0.1 lot) are better suited.
Trading Leverage	10, 20, 50, 100	100	Trading leverage determines the amount of impact a change in the currency pair makes in the open trade. A 100 leverage means a 1% change in the pair results in 100% change in the account balance.
Stop Loss (Pips)	10, 20, 50, 100	10	When a currency pair moves in the opposite direction of an open trade, stop loss amount determines how much loss is accepted before closing position.
Trailing Stop Loss (Pips)	10, 20, 50, 100	10	An open trade might result in earnings first and losses later. For those cases a trailing stop loss allows the trader to stop the losses from a profit. This variable determines how much loss from profit is are accepted.
Take Profit (Pips)	10, 20, 50, 100	100	A trade position cannot be kept open forever even though it is in profit, since markets fluctuate. Take profit determines how much profit a single open trade can make.
Strength Category Granularity (SCG)	1, 2, 5, 10	10	Amount of strength categories the SVM will train for to distinguish a weak buy/sell signal from a strong buy/sell signal. A value of 1 would result in binary buy/sell signals.
Strength Categorization Bucket Size (SCBS)	5, 10, 50, 100	5	The amount of pips per strength category between the current trading day and next hill top or bottom.
Hill Size for Strength Tagging Training Data (Days)	5, 10, 50, 100	10	Our algorithm will tag the training data based on hill tops and hill bottoms. This number determines how high or low the peaks and bottoms will be.

Table 22: Single currency pair trading with all available parameters

Pair Name	Precision	Recall	Accuracy	F-Measure
EURCHF	0,5663	0,5289	0,5688	0,5470
EURGBP	0,5543	0,6090	0,5963	0,5804
EURUSD	0,5293	0,5532	0,5399	0,5410
GBPCHF	0,5189	0,5440	0,5580	0,5312
GBPUSD	0,5071	0,5486	0,5434	0,5270
USDCHF	0,5396	0,5759	0,5799	0,5572

When the parameters are dynamically adapted to the trading period using genetic algorithm, and all the input data is not fed to the learning model, the trading system records observable improvements as shown in Table 23.

Table 23: Single currency pair trading with dynamically adapting parameters

Pair Name	Precision	Recall	Accuracy	F-Measure
EURCHF	0,6678	0,5852	0,6297	0,6238
EURGBP	0,5543	0,6219	0,6068	0,5862
EURUSD	0,5293	0,5883	0,5962	0,5572
GBPCHF	0,6458	0,6125	0,6320	0,6287
GBPUSD	0,6504	0,6364	0,6382	0,6433
USDCHF	0,6527	0,6273	0,6400	0,6397

Prediction accuracy in all currency pairs increase in an average of 5.94% where the largest increase is achieved in GBPUSD pair with 9.48% when parameters are dynamically adapted.

Starting to use multiple currencies and allowing strength biased trading with dynamically adapting parameters, results are improved significantly as shown in Table 24. The average accuracy in 3 currency trading systems is 69.59% while the accuracy of the 4 currency trading system is 78.78% both of which are higher than single currency pair trading systems.

Table 244: Strength biased trading with dynamically adapting parameters

Traded Currencies	Precision	Recall	Accuracy	F-Measure
EUR-GBP-CHF	0,7745	0,6716	0,7223	0,7194
EUR-GBP-USD	0,7008	0,6964	0,6962	0,6986
EUR-USD-CHF	0,6404	0,6770	0,6593	0,6582
GBP-USD-CHF	0,7068	0,6908	0,7059	0,6987
EUR-GBP-USD-CHF	0,8040	0,7673	0,7878	0,7852

A graphical comparison of the above results are presented in Figure 21. AAP stands for All Available Parameters and DAP stands for Dynamically Adapting Parameters. SCPT models use Single Currency Pair Trading and SBT models use Strength Biased Trading. Figure shows that DAP and SBT improves accuracy. The best results are achieved in the four currency pool with SBT and DAP.

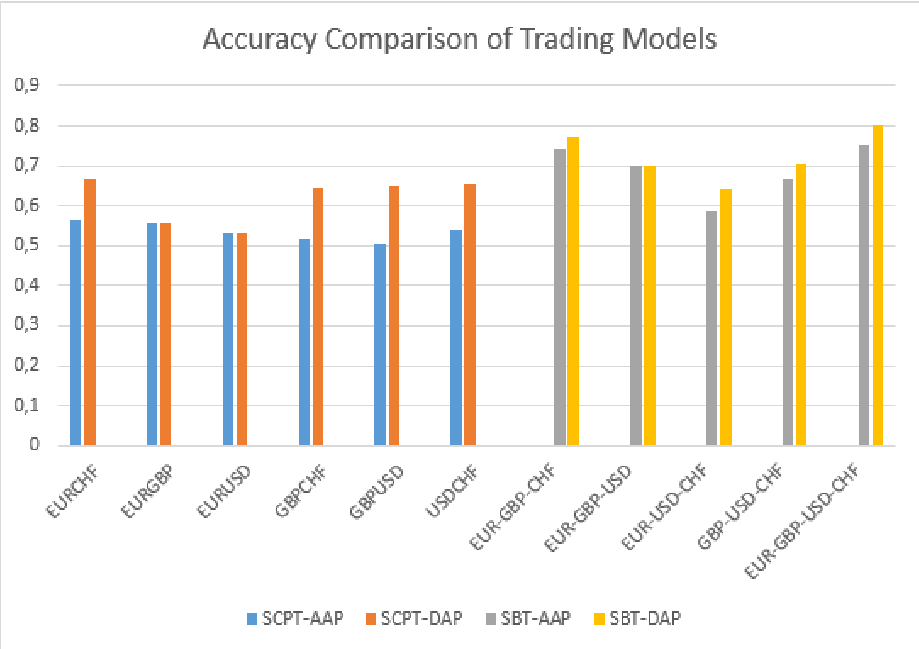

Figure 21: Accuracy comparison of different trading models

All the above results include trend deterministic data being used in combination with raw data. Due to space limitations we cannot include the detailed results obtained with raw data and trend deterministic data alone. However using raw data and trend deterministic data alone results in 5% and 7% less accuracy on the average respectively.

To embody the currency strength and strength bias concept, Figure 22 illustrates the strength of the GBPUSD currency throughout trading days between 10/05/2011 and 10/11/2011. The blue line (i.e. the dashed line) indicates the price fluctuations of the currencies, the values are denoted in the primary y-axis (i.e. left axis). The orange line (i.e. dotted line) denoting actual strength is the strength of the currency determined from actual future data and would be the optimal strength to predict. The gray line (i.e. continuous line) is the strength of the currency determined by our model. The values of these series are denoted in the secondary y-axis (i.e. right axis). In our charts sometimes only the gray line is seen, this is due to an exact match (i.e. both direction and magnitude-wise) between the predicted and actual strength. For our trading system exactly matching the predicted strength to actual strength is not necessary. However matching the direction of the strength values is crucial (i.e. if the actual strength is negative, a negative predicted strength improves the directional symmetry).

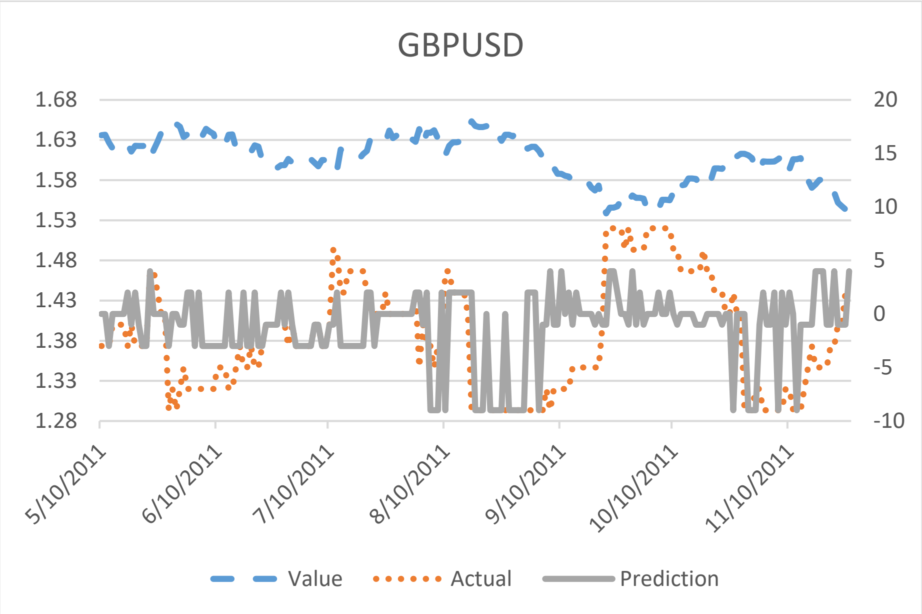

Figure 22: Strength of GBPUSD between 10/05/2011 and 10/11/2011

At 10/08/2011 the actual and predicted strengths for all the currencies available in our system are depicted below in Table 25.

Table 25: Actual and predicted currency pair strength at 10/08/2011

Currency Pair	Actual Strength	Predicted Strength
EURCHF	8	1
EURGBP	3	1
EURUSD	1	3
GBPCHF	8	0
GBPUSD	4	2
USDCHF	8	6

The strengths of the individual currencies are then calculated via equation (1) and results are outlined below in Table 26.

Table 26: Individual currency strengths at 10/08/2011

Currency	Strength
CHF	-24
EUR	12
GBP	9
USD	3

Using equations (2), (3) and (4) and Table 14, we can conclude that qmin is CHF and qmax is EUR since a qmax/qmin pair exists, our system will go long on EURCHF pair. The pair is valued at 1,02637 at the day. In the next day -11/08/2011- EURCHF will be valued at 1,08394. Which is a ~575 pip increase. Since our system employs a 100 pip take profit our earnings will be stopped at 100 pips. With a leverage of 1:100 that would result in 100% profit for the given day.

6.2.6 Comparison of Performance

Comparing our work to related work is difficult since we are using a currency strength biased trading approach with a pool of four major currencies. The state of the art models do not try to increase their directional symmetry by selecting a strong currency against a weak one, rather they predict in single exchange rated. Therefore we compare our single currency trading approach with the related work. For the comparison we have implemented Kamruzzaman and Sarker's [35] algorithm SCG-ANN, Stella and Villa's [41] algorithm CTBNC, Shen, Chao and Zhao's [42] algorithm DBN-CRBM, Moosa and Burns' approach [24] TVP and Anastasakis and Mort's algorithm [43] AC_NNGMDH.

Table 27: Performance comparison of single currency pair trading systems

		SCPT AAP	SCPT DAP	AC NNGM	TVP	DBN CRBM	CTBNC	SCG ANN
EUR/ CHF	Precision	0,5663	0,6678	0,4750	0,6162	0,7005	0,6936	0,5250
	Recall	0,5289	0,5852	0,4326	0,5416	0,5724	0,5858	0,4758
	Accuracy	0,5688	0,6297	0,4727	0,5842	0,6221	0,6340	0,5162
	F-Measure	0,5470	0,6238	0,4528	0,5765	0,6300	0,6351	0,4992
EUR/ GBP	Precision	0,5543	0,5543	0,5858	0,5039	0,6000	0,6110	0,6205
	Recall	0,6090	0,6219	0,5345	0,5527	0,6175	0,5924	0,6147
	Accuracy	0,5963	0,6068	0,5363	0,5466	0,6130	0,5940	0,6145
	F-Measure	0,5804	0,5862	0,5590	0,5272	0,6086	0,6016	0,6176
EUR/ USD	Precision	0,5293	0,5293	0,4861	0,5509	0,5818	0,5648	0,5370
	Recall	0,5532	0,5883	0,5016	0,5622	0,5891	0,5631	0,5790
	Accuracy	0,5399	0,5962	0,4897	0,5506	0,5782	0,5529	0,5632
	F-Measure	0,5410	0,5572	0,4937	0,5565	0,5854	0,5639	0,5572
GBP/ CHF	Precision	0,5189	0,6458	0,5342	0,5147	0,6287	0,6547	0,5635
	Recall	0,5440	0,6125	0,5307	0,4847	0,5848	0,5982	0,5710
	Accuracy	0,5580	0,6320	0,5447	0,4988	0,6032	0,6190	0,5826
	F-Measure	0,5312	0,6287	0,5325	0,4992	0,6060	0,6252	0,5672
GBP/ USD	Precision	0,5071	0,6504	0,5449	0,4992	0,6331	0,6409	0,6252
	Recall	0,5486	0,6364	0,5635	0,5121	0,6271	0,6214	0,6455
	Accuracy	0,5434	0,6382	0,5600	0,5103	0,6272	0,6240	0,6398
	F-Measure	0,5270	0,6433	0,5540	0,5056	0,6301	0,6310	0,6352
USD/ CHF	Precision	0,5396	0,6527	0,5557	0,5412	0,5557	0,6446	0,4830
	Recall	0,5759	0,6273	0,5111	0,5210	0,5705	0,6036	0,5137
	Accuracy	0,5799	0,6400	0,5229	0,5324	0,5782	0,6193	0,5237
	F-Measure	0,5572	0,6397	0,5325	0,5309	0,5630	0,6234	0,4979

All algorithms use the same data set obtained from TrueFX. Same spreads and commissions defined in our system are applied. For each algorithm paper's origin country time-zone is used. The results are presented in Table 27.

In their original work Kamruzzaman and Sarker make [35] weekly forecasts for AUD against five major forex currencies. The highest directional accuracy recorded in the given work by SCG-ANN algorithm is 0.7714 for the AUD/GBP exchange rate. The weekly data used is from years 1991 to 2002. There are 65 testing data points which result in the given performance. In our experiments this model's best performances were recorded for GBP pairs (i.e. GBPCHF, GBPUSD and EURGBP). The highest accuracy was achieved at GBPUSD pair and is 0.6398

Stella and Villa [41] have used a continuous time Bayesian network classifier for predicting intraday values of foreign exchange rates. The predictions have been made in EUR/USD, GBP/USD and EUR/CHF exchange rates. The work uses three different data sets (i.e. TrueFX, Dukascopy and GainCapital) and different directional accuracies have been recorded in different data sets. In our experiments highest recorded performance of the CTBNC algorithm is 0.6340 for EUR/CHF.

Shen, Chao and Zhao's work [42] achieve their highest accuracies in GBP/USD exchange rate. The achieved accuracy is 0.6362. Forecasts are performed weekly and there are only 52 testing data points. In our experiments the best performance recorded is again on the GBP/USD with an accuracy of 0.6272 for the DBN-CRBM algorithm.

Mossa and Burns' work [26] use three different intervals –monthly, quarterly and every six months- to predict the exchange rates of CAD, GBP, JPY and USD pairs. The highest directional accuracy is once again achieved for the GBP/USD exchange rate and is 0.72. This is better than our top GBP/USD forecast of 0.6382, however this performance achieved with 12 data points and predictions are made in six month intervals as opposed to our daily forecast approach. When forecasts are made for the same currency quarterly the accuracy falls to 0.56 and for monthly forecasts the accuracy is 0.48. In our experiments the best accuracy achieved by TVP is at EUR/CHF pair with an accuracy of 0.5842.

Anastasakis and Mort [43] forecasts daily values of GBP, USD, DM and JPY pairs. The data uses 1362 data points, which is similar to the size of our data set. Authors report GBP/DEM performance in their work. This currency is comparable to our work since historical EUR/GBP prices are fixed based on GBP/DEM pair [44]. The directional accuracy values for six months of testing data are reported and the highest accuracy recorded in a given month is 0.6818 while the lowest accuracy recorded is 0.3182. The average performance on the given exchange rate is 0.5382. In our experiments NNGMDH achieved its highest accuracy 0.5600 in the GBP/USD pair.

The comparison of performances of our system and related work is presented in Figure 23. SCPT-DAP outperforms SCPT-AAP at each currency pair. With five algorithms and six currency pairs present SCPT-DAP outperforms 25 out of 30 performance figures. SCG-ANN outperforms SCPT-DAP in two of six available pairs: EUR/GBP and GBP/USD pairs with margins of 0.0077 and 0.0016 respectively. CTBNC outperforms SCPT-DAP in one of six available pairs: EUR/CHF with a margin of 0.0043. DBN-CRBM outperforms SCPT-DAP in one of six available pairs: EUR/GBP with a margin of 0.0062. In all the remaining instances SCPT-DAP performs better than the competition.

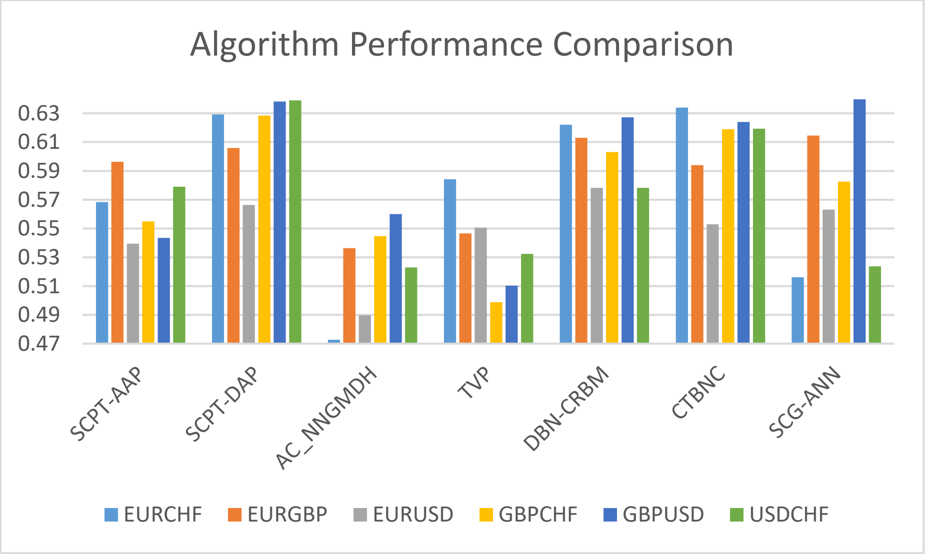

Figure 23: Single currency pair algorithm performance comparison

Figure 24 shows that when Strength Biased Trading approach is used, SBT-DAP outperforms all the remaining algorithms significantly. Whilst the highest directional symmetries recorded in Single Currency Pair algorithms belongs to SCPT-DAP with 0.6400 in USD/CHF and SCG-ANN with 0.6398 in GBP/USD, SBT-DAP's lowest directional symmetry is 0.6593 for EUR-USD-CHF pool and highest directional symmetry is 0.7878 for the EUR-GBP-USD-CHF pool.

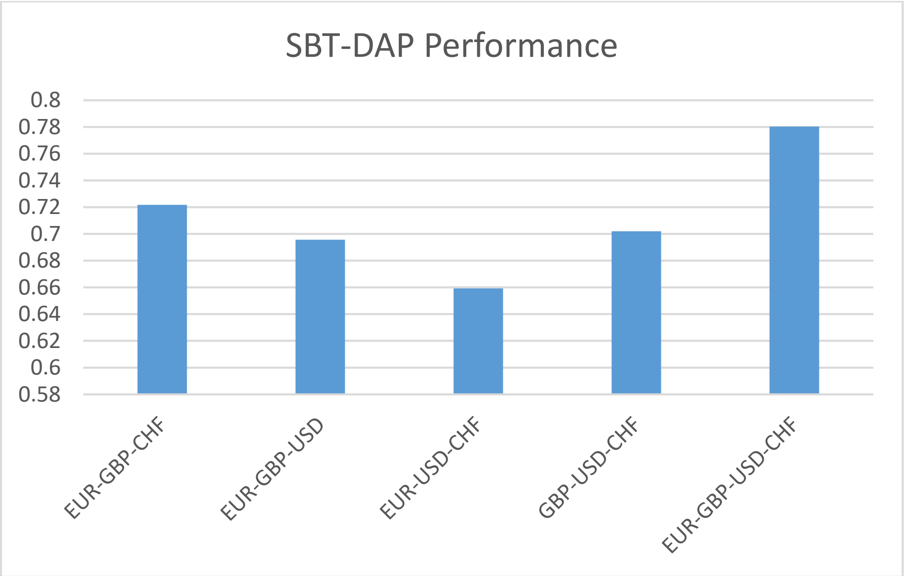

Figure 24: Strength biased currency trading performance

A summary of trading statistics regarding the trades made with the SBT-DAP algorithm is provided in Table 28.

Table 28: SBT-DAP algorithm trade statistics

Currencies	EUR-GBP-USD-CHF
Open Positions	3497
Profit Positions	2755
Loss Positions	742
Average Profits	88,9
Average Loss	-127,6
Total Profit Pips	150216
Maximum Account Draw-Down %	13,38

6.3 ZZMOP Results

In our ZZMOP algorithm we are making use of the Zigzag technical indicator, Expectation Maximization clustering algorithm and SVMs. All three of these components have certain parameters that adapt them to the problem at hand. We discuss the characteristics of our dataset and how we adapt parameters for the aforementioned components to our data in this section.

6.3.1 Characteristics of Our Dataset

Two types of historical data are collected by our system. First is price data with 15 Minute intervals which summarizes the opening, closing, low and high values for the given interval. 15 Minute interval data values are used for training purposes. Second data is the real time price data which contains all the price changes that have happened in the currencies. Real time price data is used for testing and simulation purposes. Both data is collected for currency pairs including USD, EUR, GBP, and CHF. Summary statistics of the data can be found in Table 29. Total data points are the number of 15 Minute bars, minimum and maximum values represent the highest and lowest values recorded, average value is the average of 15 Minute closes, increases and decreases in value show a 15 Minute increase or a 15 Minute decrease in the closing price of the currency pair. For training 15 Minute values are used.

Table 29: 15 minute exchange rate (training data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value	Decreases in Value
EURCHF	121440	1.488	1.007	1.253	55776	65664
EURGBP	121536	0.913	0.776	0.839	60960	60576
EURUSD	121536	1.493	1.187	1.334	62208	59328
GBPCHF	121440	1.711	1.146	1.493	58944	62496
GBPUSD	121536	1.718	1.423	1.589	60960	60576
USDCHF	121536	1.172	0.706	0.941	59424	62112

Since our system makes use of real life trading parameters such as stop loss and take profit, for testing and model simulation, real time price data are used. Statistics for these values can be found in Table 30.

Table 30: Real time exchange rate (testing data) statistics

Instrument	Total Data Points	Max. Value	Min. Value	Avg. Value	Increases in Value (%)	Decreases in Value (%)
EURCHF	59948502	1.488	1.007	1.248	45.88	54.12
EURGBP	67221512	0.913	0.776	0.826	50.12	49.88
EURUSD	62915513	1.493	1.187	1.316	51.12	48.88
GBPCHF	76187806	1.711	1.146	1.498	47.99	52.01
GBPUSD	65708389	1.718	1.423	1.591	50.22	49.78
USDCHF	59461873	1.172	0.706	0.939	48.84	51.16

The set of parameters stored for each financial instrument are the same with our previous model, and is listed in Table 17.

6.3.2 Zigzag Related Parameters

Zigzag can detect any number of points given a large enough history window. Our system can then use the Zigzag discovered price points to discover motifs. The size of the window, the minimum required depth of Zigzag points and number of Zigzag points to use are Zigzag related parameters of our algorithm. These are detailed in Table 31.

Table 31: Zigzag related parameters

Parameter Name	Values Experimented	Value Used	Explanation
Window Size in Bars	50, 100, 150, 200	150	Our algorithm will run the Zigzag indicator in a historic window of price bars. This number determines how large that window will be.
Zigzag Depth	8, 12, 16, 20	16	Zigzag will look for tops and bottoms for a specific period of bars. This number determines the required minimum depth for a top or bottom to be formed.
Zigzag Pattern Length	5, 6, 7, 8, 9, 10	7	Zigzag will look for this many Zigzag points to form a motif. A very short motif would be too common to be meaningful, and a very long motif would be too scarce.
Zigzag Thickness	5, 10, 30, 50	10	Zigzag thickness is the amount of pips Zigzag line will extend in both directions to cover prices occurring in the instruments historic data.
Zigzag Coverage	0.5, 0.6, 0.7, 0.8, 0.9, 1	0.8	Zigzag coverage is the percent of price action that has to be covered by the thick Zigzag line.

The number of points to take into account when clustering and classifying possible motifs is particularly important for the success of the system. We are detailing the selection of 7 Zigzag points for our motifs based on the experiments with different motif lengths to determine the optimal length of a motif in Table 32.

Table 32: Performance of different motif lengths

Motif Length	Average Short Reward/ Risk Ratio	Average Long Reward/ Risk Ratio	Top Short Reward/ Risk Ratio	Top Long Reward/ Risk Ratio	Average Reward/ Risk Ratio Difference
5	2,0880	2,0511	3,2293	2,6919	0,7630
6	2,0969	2,1037	3,0239	2,5817	0,6615
7	2,2250	2,1827	3,3267	3,0474	0,6763
8	2,2036	2,1296	3,0884	2,9385	0,5712
9	2,2180	2,1887	3,2179	2,8030	0,5849
11	2,2791	2,1648	2,9968	2,7557	0,4844
13	2,2960	2,2408	3,1513	2,9476	0,6079

The different motif lengths resulted in motif clusters with different characteristics and although top efficiencies were achieved by motifs with length 7, there were no clear winner since some other motif lengths resulted in higher average reward/risk ratio and reward/risk ratio differences. To make a selection, we have awarded each first, second and third place three, two and one points respectively. Based on this criteria motifs with length 7 were selected since they have acquired the highest score.

The thickness and coverage properties of the modified Zigzag indicator contribute significantly to the directional bias of the discovered motifs and hence the success of our algorithm. Figure 25 depicts the (a) variation of average cluster directional bias and (b) directional bias of the cluster with highest directional bias with respect to Zigzag thickness with a fixed Thick Zigzag coverage value of 0.8. Figure 26 depicts the variation aforementioned variables with respect to Thick Zigzag coverage with a fixed Zigzag thickness of 10.

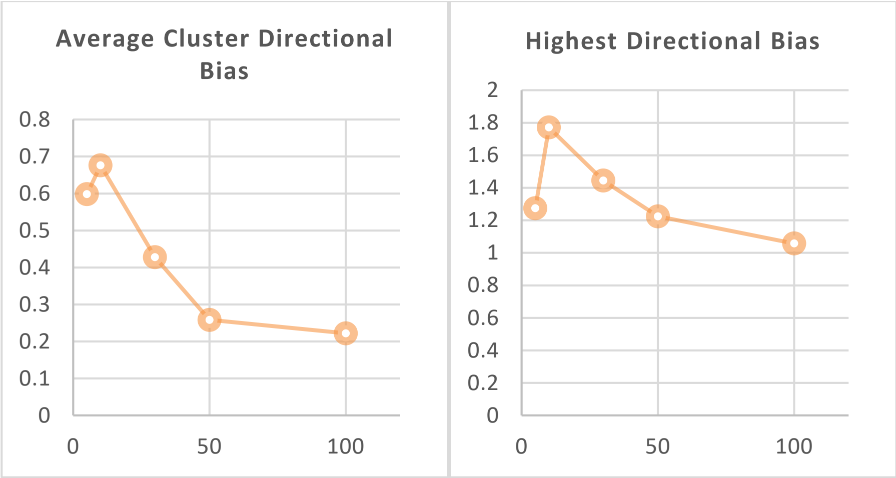

Figure 25: Variation of (a) average cluster directional bias with respect to Thick Zigzag thickness, (b) highest cluster directional bias with respect to Zigzag thickness with fixed Thick Zigzag coverage of 0.8

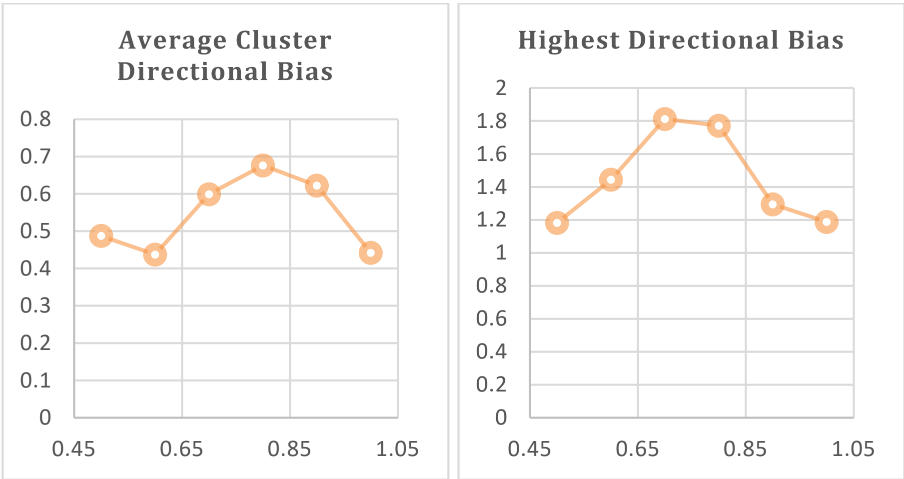

Figure 26: Variation of (a) average cluster directional bias with respect to Thick Zigzag coverage, (b) Highest cluster directional bias with respect to Thick Zigzag coverage with fixed Zigzag thickness of 10

6.3.3 Clustering Related Parameters

The Zigzag motifs detected by our system are clustered via the Expectation Maximization (EM) clusterer [45]. The EM algorithm operates iteratively to find the maximum likelihood estimates of parameters in a statistical model. The EM model depends on unobserved latent variables. EM iterations alternate between performing the expectation step and the maximization step. The expectation step creates the function for expectation of the log-likelihood with the current estimate for the parameters. The maximization step computes the parameters maximizing the expected log-likelihood found on the expectation step. The estimates are used iteratively to determine the distribution of the latent variables in the next expectation step. The EM clusterer uses certain parameters to operate, which we have exhaustively searched with 432 trials to find the optimal. The used parameters are explained in Table 33.

Table 33: EM parameters

Parameter Name	Values Experimented	Value Used	Explanation
Iterations	100, 500, 1000	500	The maximum number of iterations to perform. A low number would result in a solution that is not converged. A high number would result in unnecessary computation.
Folds	10, 50, 100	10	Number of folds to use when cross-validating to find the best number of clusters.
Minimum Log Likelihood Improvement	0.00001, 0.0001, 0.001, 0.01	0.0001	The minimum improvement in cross-validated log likelihood used to consider increasing the number of clusters when cross-validating to find the best number of clusters.
Minimum Standard Deviation	0.00001, 0.0001, 0.001, 0.01	0.00001	The minimum value for standard deviation when calculating normal density.
Simple K-Means Runs	10, 50, 100	50	Number of runs of k-means to perform.

With the parameters specified in Table 31 and 33, EM clusterer creates 23 clusters for the interval 2010-2015 with a 15 minute operational timeframe. Number of clusters are determined by EM’s built-in cross validation, and are not pre-specified. Among the soft-assigned cluster numbers, the cluster number with highest probability is assumed as the final cluster number for any individual pattern. Clusters with top three highest short and long position efficiencies are highlighted below. The average statistics regarding the clusters are specified below in Table 34 and displayed in Figure 27 and 28.

Table 34: Clusters created by EM for 2010-2015 interval

Cluster Number	Pattern Count	Average Pattern Length	Short Position Reward/ Risk Ratio	Long Position Reward/ Risk Ratio	Reward/ Risk Ratio Difference
0	126	142,09	2,8512	1,7218	1,1294
1	534	103,01	2,3607	1,8544	0,5063
2	166	139,53	1,9180	2,4896	0,5716
3	267	132,80	1,8988	2,1156	0,2168
4	310	133,48	2,0464	2,2881	0,2416
5	109	154,15	1,6734	2,5769	0,9035
6	196	130,73	2,2928	2,0434	0,2494
7	195	122,04	2,8373	1,6496	1,1877
8	478	108,51	2,5654	1,8966	0,6687
9	105	151,11	2,9010	1,7912	1,1098
10	374	116,91	2,0763	1,9058	0,1704
11	216	140,06	3,2293	1,4582	1,7711
12	151	176,78	2,1294	1,8646	0,2647
13	216	133,43	1,6528	2,9476	1,2947
14	341	128,65	2,1239	1,9483	0,1756
15	459	102,88	2,0389	2,6718	0,6328
16	152	146,03	2,4670	1,5849	0,8821
17	518	121,86	1,7670	2,7252	0,9581
18	392	118,46	1,7768	2,9415	1,1646
19	254	129,70	2,3306	1,8516	0,4789
20	532	107,69	1,8644	2,4398	0,5754
21	290	140,25	2,1870	1,9863	0,2006
22	224	126,80	2,0254	2,2271	0,2016

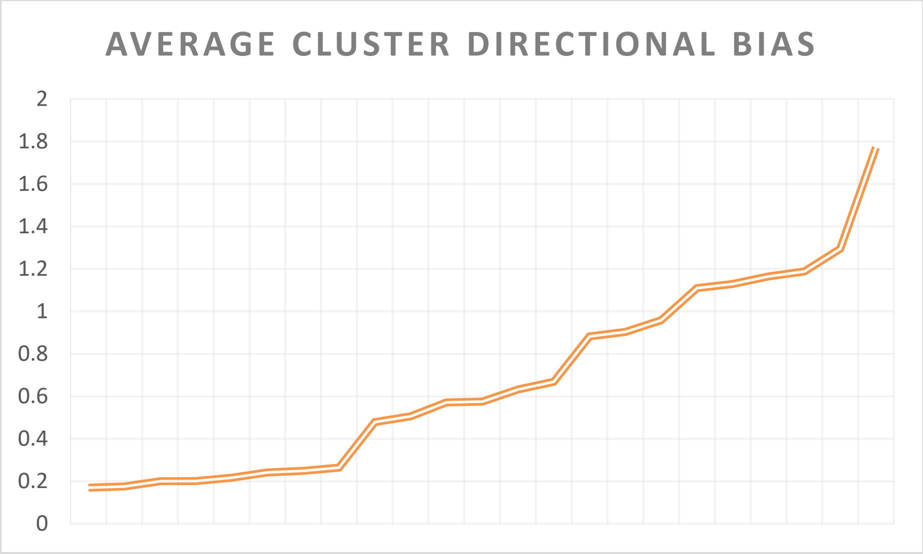

Figure 27: Average directional bias of motifs in different clusters

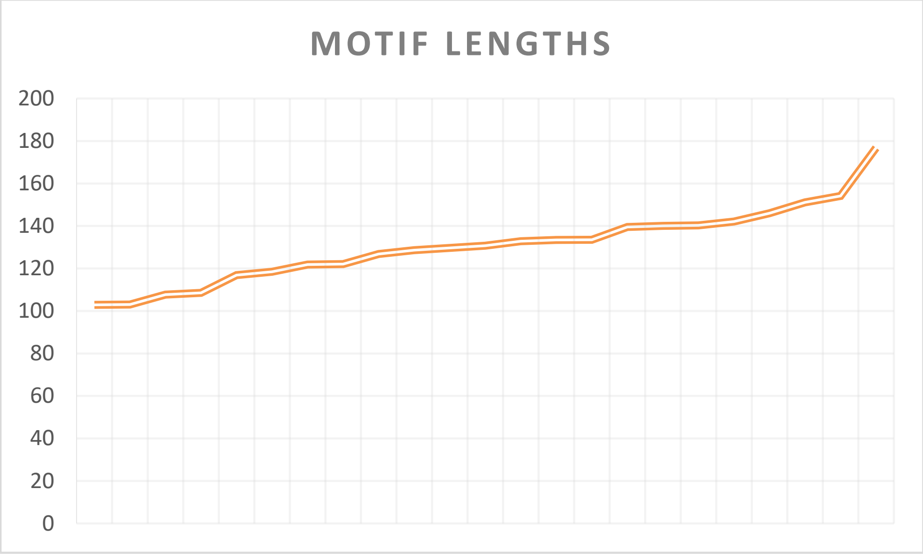

Figure 28: Distribution of average lengths of motifs in different clusters

6.3.4 SVM Related Parameters

An SVM requires certain parameters which directly affect its performance. We have empirically experimented with several values for these parameters that have been mentioned in related research and selected the best performing ones. To determine the best performing parameters for our SVM, all possible combinations of the parameters have been exhaustively searched resulting in 540 experiments per each training interval. The parameters that we have used are outlined in Table 35.

6.3.5 Trading Related Parameters

A realistic trading system requires certain parameters such as trading volume, stop loss or take profit. In our system we have empirically experimented with several values for these parameters that are common in the industry and selected the best performing ones. To determine the best performing parameters for our system, all possible combinations of the parameters have been exhaustively searched resulting in 30720 experiments per each forecasting period of 15 minutes for an interval of 5 years. These parameters are outlined in Table 36.

Table 35: SVM parameters

Parameter Name	Values Experimented	Value Used	Explanation
Kernel Type	Polynomial, Radial Basis	Polynomial	Polynomial function kernels achieved superior performance in related work [17, 27]. Thus we use the polynomial function kernel in our system.
Function Degree (d)	3, 4, 5	4	Degree of polynomial function characterizes how well the model fits to the data. A very high degree could result in overfitting.
Regularization Parameter (nu)	0.00001, 0.0001, 0.01, 0.1, 0.25, 0.5	0.0001	Regularization parameter controls margin and misclassification error.
Attribute Normalization	Yes / No	Yes	Input attributes are normalized or left with assigned bucket values across the training set.
Training Interval (Years)	1, 2, 3, 4, 5	5	A model in our system is trained with the given number of years before making decisions in validation or actual data sets.

Table 36: Trading system parameters

Parameter Name	Values Experimented	Selected Value	Explanation
Trading Volume (Lot)	0.01, 0.1, 0,25, 0,5, 1	Dynamic Selection	Forex providers generally enable 0.1 or 1 lot trades. For a small account mini-lots (0.1 lot) or micro-lots (0.01 lot) are better suited. In our experiments using fixed lot sizes proved less profitable, therefore the lot size is dynamically selected based on trade profitability analysis.
Stop Loss (Pips)	10, 20, 50, 100	Dynamic Selection	When a currency pair moves in the opposite direction of an open trade, stop loss amount determines how much loss is accepted before closing position. In our experiments using fixed stop losses proved less profitable, therefore the stop loss is dynamically selected based on reverse price-action estimate.
Take Profit (Pips)	10, 20, 50, 100	Dynamic Selection	A trade position cannot be kept open forever even though it is in profit, since markets fluctuate. Take profit determines how much profit a single open trade can make. In our experiments using fixed take profits proved less profitable, therefore the take profit is dynamically selected based on price-action estimate.

6.3.6 Sample Chart Patterns and Motifs Detected by ZZMOP

As specified in Algorithm 4 and 5 for each currency, at the completion of each bar, the Zigzag indicator is applied to the historic bars of the currency to create Zigzag points. Zigzag points are discontinuous and are alternating between highs and lows, as is the case with legacy chart patterns. In this section we will show two legacy chart patterns and two new motifs detected by our system.

Legacy Chart Patterns

Two famous chart patterns are Head and Shoulders (H&S) and Inverse Head and Shoulders (IH&S) chart patterns. For H&S and IH&S chart patterns to be formed, at least 7 points are required on a chart. Sample H&S and IH&S chart patterns found by ZZMOP are shown in Figure 29.

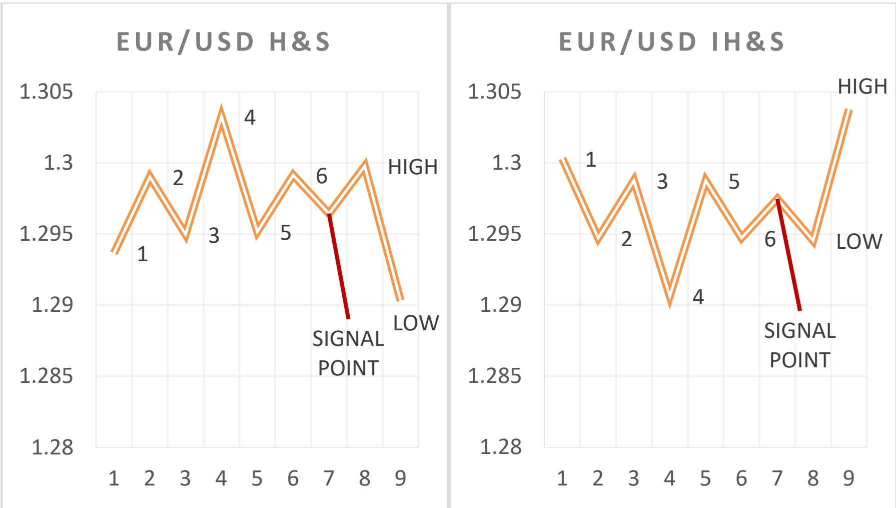

Figure 29: A sample (a) H&S pattern (b) IH&S pattern discovered by ZZMOP

The points defining the chart patterns are numbered 1-7 from left to right. The sample chart patterns are less than perfect, for instance 1 and 3 are not equal to each other, and they are also not equal to 5 and 7 which is by definition the case for H&S and IH&S chart patterns. The local tops 2 and 6 are also not equal to each other. However it is very hard to find a perfect chart pattern in the real world. In the presence of the defining characteristics of the H&S, a chart pattern should be accepted as an H&S chart pattern –same holds true for IH&S. These characteristics are the presence of three hills where the first and third (i.e. the shoulders) hills are smaller than the second hill (i.e. the head). Therefore a preliminary elimination in our algorithm looks for at least 7 Zigzag points in the history window of a bar which have these characteristics of an H&S chart pattern.

Novel Motifs

Our system does not only detect previously established chart patterns, it also discovers new motifs and groups them into clusters with similar future price action characteristics. Two sample motifs that are detected by our algorithm are presented in Figure 30. The 7th points in the motifs are the decision points and marked with "Decision" label. The future price action for the motifs show an average of the future behavior of motif instances.

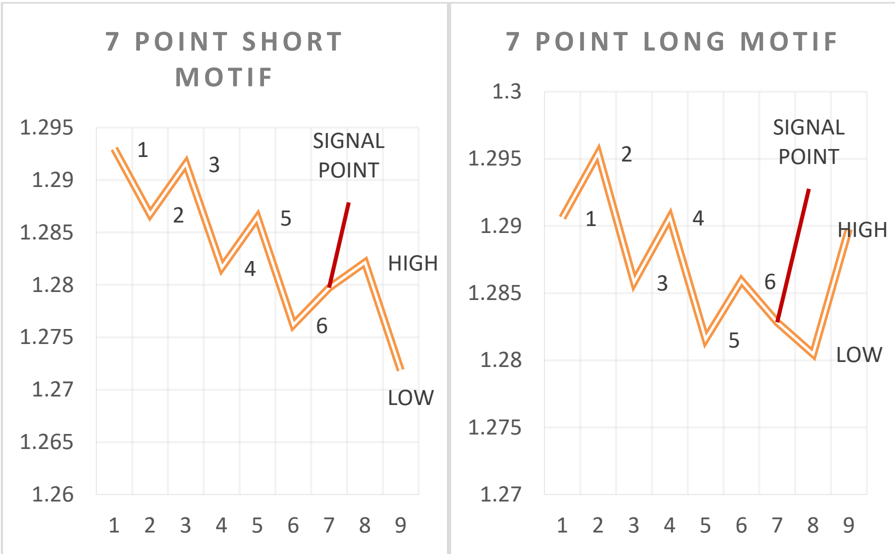

Figure 30: Sample novel (a) short pattern (b) long pattern discovered by ZZMOP

It can be observed when averaged, the behavior of the future prices result in an amount of inverse price. In cases where the inverse price movement is equal to the desired price movement, the motifs are inefficient due to commissions. For these motifs, the time series are T_{short} = *<1.29299, 1.28672, 1.29156, 1.28164, 1.28644, 1.27619, 1.27972, 1.28215, 1.27186>* and T_{long} = *<1.29061, 1.29542, 1.28582, 1.29064, 1.28158, 1.28593, 1.28283, 1.28047, 1.28979>*. The reward/risk ratio of these motifs are R_{short}= *3.2345* and R_{long} = *2,9491*.

6.3.7 Performance of Our System

Different trading strategies and input compositions were used with the system and different results in terms of performance were obtained. Obtained results are presented in Table 37. For exchange ratios, buy and sell decisions are materialized based on the closing value of the decision date, since the market is open for 24 hours. The system can generate successive buy signals or successive sell signals, which could be still meaningful due to their different signal efficiencies and take-profit and stop-loss locations.

Table 37: Zigzag based pattern mining algorithm performance

Traded Currencies	Precision	Recall	Accuracy	F-Measure
EURCHF	0,6953	0,6121	0,6577	0,6510
EURGBP	0,6283	0,6363	0,6334	0,6323
EURUSD	0,6358	0,6721	0,6548	0,6534
GBPCHF	0,6221	0,6044	0,6189	0,6131
GBPUSD	0,6881	0,6702	0,6737	0,6790
USDCHF	0,6445	0,6425	0,6508	0,6435

6.3.8 Comparison of Performance

The comparison of performances between related work and our SCPT-DAP system has been presented in Table 27. In Table 38 we present performance values obtained with ZZMOP and compare it to our previous work SCPT-DAP. With six algorithms and six currency pairs present ZZMOP performs as the best model in five of the currency pairs, and takes the third place in the GBP/CHF pair. SCPT-DAP and CTBNC outperform ZZMOP in GBP/CHF. In all the remaining instances ZZMOP performs better than the competition. In Table 39 an overview of directional accuracies is also provided

Table 38: Performance comparison of currency pair trading systems

		ZZMOP	SCPT DAP
EURCHF	Precision	0,6953	0,6678
	Recall	0,6121	0,5852
	Accuracy	0,6577	0,6297
	F-Measure	0,6510	0,6238
EURGBP	Precision	0,6283	0,5543
	Recall	0,6363	0,6219
	Accuracy	0,6334	0,6068
	F-Measure	0,6323	0,5862
EURUSD	Precision	0,6358	0,5293
	Recall	0,6721	0,5883
	Accuracy	0,6548	0,5962
	F-Measure	0,6534	0,5572
GBPCHF	Precision	0,6221	0,6458
	Recall	0,6044	0,6125
	Accuracy	0,6189	0,6320
	F-Measure	0,6131	0,6287
GBPUSD	Precision	0,6881	0,6504
	Recall	0,6702	0,6364
	Accuracy	0,6737	0,6382
	F-Measure	0,6790	0,6433
USDCHF	Precision	0,6445	0,6527
	Recall	0,6425	0,6273
	Accuracy	0,6508	0,6400
	F-Measure	0,6435	0,6397

Table 39: Performance summary of currency pair trading systems

Currency Pair	ZZMOP	SCPT DAP	TVP	DBN CRBM	CTBNC	SCG ANN
EURCHF	0,6577	0,6297	0,5842	0,6221	0,6340	0,5162
EURGBP	0,6334	0,6068	0,5466	0,6130	0,5940	0,6145
EURUSD	0,6548	0,5962	0,5506	0,5782	0,5529	0,5632
GBPCHF	0,6189	0,6320	0,4988	0,6032	0,6190	0,5826
GBPUSD	0,6737	0,6382	0,5103	0,6272	0,6240	0,6398
USDCHF	0,6508	0,6400	0,5324	0,5782	0,6193	0,5237
Average	0,6482	0,6238	0,5372	0,6037	0,6072	0,5733

A summary of trading statistics regarding the trades made with the ZZMOP algorithm is provided in Table 40.

Table 40: ZZMOP algorithm trade statistics

Currency	EURCHF	EURGBP	EURUSD	GBPCHF	GBPUSD	USDCHF
Open Positions	3011	3216	3497	2995	3284	3113
Profit Positions	1737	2037	2290	1854	2212	2026
Loss Positions	1274	1179	1207	1141	1072	1087
Average Profits	34,5	41,9	56,2	61	36,3	28,7
Average Loss	31,3	49,8	47,4	47,9	35,7	27,5
Total Profit Pips	99816	144039	185862	167793	118578	88048
Maximum Account Draw-Down %	25,52	19,93	34,46	32,77	24,61	19,83

CHAPTER 7

CONCLUSION

In this paper we present three different approaches to forecasting exchange rates. Our first approach is based on genetic algorithms and support vector machines for building a foreign exchange rate prediction model. At the first stage raw price data and trend deterministic technical indicators are used for input variable pool. Then the inputs are dynamically adapted for each trading interval to better represent the fluctuations in a given pair. Lastly prediction is done with a strength bias which further increases directional symmetry in exchange rate prediction.

Our first approach shows that basic price data and trend deterministic technical indicator signals can be used in conjunction with learning models such as support vector machines to forecast price changes in financial markets such as the Forex market. The success of the system heavily depends on the selection of inputs, learning model and decision support mechanisms. The proposed strength biased trading strategy proves useful in terms of directional symmetry and profits. Since the strength bias property allows the system to select the strongest currency against the weakest currency, directional symmetry is higher than related work.

Our second approach uses a modified version of the Zigzag technical indicator, expectation maximization and support vector machines for predicting short term trends in financial time series found in the foreign exchange market. At the first stage raw price data is processed with Zigzag to create Zigzag points. Thickness and coverage properties are introduced to Zigzag indicator to determine the compliance of prices to motifs. Then the points are clustered each trading interval to better represent the future movement types. The upcoming motifs are also clustered to find similarities with previous motifs. Lastly prediction is done for multiple parameters to determine with trading parameters would result in optimal profits.

Our second approach shows that technical indicator data can be used to discover motifs in conjunction with learning models such as support vector machines and clustering algorithms to forecast price changes in financial markets such as the Forex market. The success of the system depends on the selection of motif discovery algorithm, learning model and decision support mechanisms. The proposed strategy proves useful in terms of directional symmetry and profits.

Our last approach presents a pure technical indicator based trading mechanism for trading short term trends in financial time series found in the foreign exchange market. At the first stage a pair of RSI and CCI indicators are used to create buy and sell signals. The crossovers between the fast and slow moving indicators are used to buy and sell the instruments. At the second stage, trading parameters such as take profit and stop loss are optimized for the given indicator signals. Lastly orders are submitted in a test environment to simulate trades. The system is retrained in a sliding window manner and parameters for the indicators and trades are updated to keep the results optimal.

This approach shows that combination of different technical indicators and parameters can be used to create profitable trade models that forecast price changes in financial markets such as the Forex market. The success of the system depends on the selection of technical indicators, indicator parameters and trade parameters.

Due to the nature of trade in the Forex market, stop loss and take profit locations of the system are more important than when an order is submitted; therefore a Forex trading system should focus heavily on stop loss and take profit parameters. The optimization of the trade parameters are done both exhaustively and genetically. Genetically optimized models perform similar to the exhaustively optimized models. The simulation runtimes of genetic models are a fraction of their peers. Further advancements of the trading algorithm mandates a parameter selection mechanism such as the genetic algorithms due to the required runtime of the exhaustive parameter search.

All the systems proposed in this work are automatic and require no human intervention. This is practical in a real time trading scenario, since the trades need to be instantaneous.

For forecasting a combination of raw price data, technical indicator signals, technical indicator generated motifs are used, but a rule based financial instrument selection mechanism can also be implemented for further profits and directional symmetry.

REFERENCES

[1] http://www.federalreservehistory.org/Events/DetailView/34 as accessed on, April 2017.

[2] http://www.investopedia.com/university/forex-currencies/ as accessed on, April 2017.

[3] A. M. Denton, C. A. Basemann, and D. H. Dorr, “Pattern-based time-series subsequence clustering using radial distribution functions”, Knowl Inf Syst (2009) 18: 1. doi:10.1007/s10115-008-0125-7.

[4] Rakthanmanon, T., Keogh, E.J., Lonardi, S. et al., "MDL-based time series clustering", Knowl Inf Syst (2012) 33: 371. doi:10.1007/s10115-012-0508-7

[5] A. Mueen. Enumeration of time series motifs of all lengths. Data Mining (ICDM), 2013 IEEE 13th International Conference on. IEEE, pp. 547-556, 2013.

[6] P. Esling and C. Agon. Time Series Data Mining. ACM Computing Surveys, 45(1): Article 12, 2012.

[7] Von Neumann, John, and Oskar Morgenstern. Theory of games and economic behavior. Princeton university press, 2007.

[8] http://www.investopedia.com/categories/forex.asp as accessed on April 2017.

[9] Brown, D. P., Robert H. J., “On Technical Analysis”, The Review of Financial Studies, 2, 1989, pp 527-551.

[10] Pang, S., Song, L. and Kasabov, N., “Correlation-aided support vector regression for forex time series prediction”, Neural Computation & Applications, 2011, pp. 1193-1203

[11] Huang, Chenn-Jung, Dian-Xiu Yang, and Yi-Ta Chuang. "Application of wrapper approach and composite classifier to the stock trend prediction." Expert Systems with Applications 34.4 (2008): 2870-2878.

[12] Lee, Ming-Chi. "Using support vector machine with a hybrid feature selection method to the stock trend prediction." Expert Systems with Applications 36.8 (2009): 10896-10904.

[13] Bolton, Arthur Hamilton. The Elliott Wave Principle: A Critical Appraisal. Monetary Re-search Limited, 1976.

[14] Gaucan, Violeta. "How to use Fibonacci retracement to predict forex market." Journal of Knowledge Management, Economics and Information Technology 1.2 (2011).

[15] Wilder, J. Welles. New concepts in technical trading systems. Trend Research, 1978.

[16] Lambert, Donald R. "Commodity channel index: Tool for trading cyclic trends." Technical Analysis of Stocks & Commodities 1 (1983): 47.

[17] Khemchandani, R., Jayadeva, Chandra, S., “Regularized least squares fuzzy support vector regression for financial time series forecasting”, Expert Systems with Applications, 2009, pp. 132-138.

[18] Kampouridis, M., Otero, F. E. B, “Heuristic procedures for improving the predictability of a genetic programming financial forecasting algorithm”, Soft Computing, 2015, pp. 1-16.

[19] Davis, L., “Handbook of Genetic Algorithms”, Van Nostrand Reinhold, 1991, pp. 136-140.

[20] Meffert, K., "JGAP - Java Genetic Algorithms and Genetic Programming Package", http://jgap.sf.net, 2014.

[21] Tick-by-Tick Real-Time and Historical Market Rates, http://www.truefx.com/, 01/14/2016.

[22] Nesreen, A., Atiya, A., Gayar, N. and Shishiny, H., “An Empirical Comparison of Machine Learning Models for Time Series Forecasting”, Econometric Reviews, 2010, pp. 594-621.

[23] Cortes, C., Vapnik, V., "Support-Vector Networks", Machine Learning, 1995, pp. 20.

[24] Moosa, I., Burns, K., “The unbeatable random walk in exchange rate forecasting reality or myth?”, Journal of Macroeconomics, 2014, pp. 69-81.

[25] Cheung, Y., Chinn, M. D., Pascual, A. G., “Empirical exchange rate models of the nineties: are they fit to survive?”, Journal of International Money Finance 24, 2005, pp. 1150-1175.

[26] Leitch, G., Tanner, J. E., “Economic Forecast Evaluation: Profits Versus the Conventional Error Measures”, The American Economic Review, 1991, pp. 580-590.

[27] Patel, J., Shah, S., Thakkar, P., Kotecha, K., “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques”, Expert Systems with Applications, 2015, pp. 259-268.

[28] C. Faloutsos and C. Faloutsos. Fast subsequence matching in time-series databases. ACM SIGMOD Record, 23(2):419–429, 1994.

[29] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and, E. Keogh. Experimental comparison of representation methods and distance measures for time series data. Data Min-ing and Knowledge Discovery, 26(2):275–309, 2013.

[30] J. Danielsson, R. Payne. Real trading patterns and prices in spot foreign exchange markets. Journal of International Money and Finance.Vol 21, Issue 2, pp203-222, 2002.

[31] Aminikhanghahi, S. & Cook, D.J., "A survey of methods for time series change point detection", Knowl Inf Syst (2016). doi:10.1007/s10115-016-0987-z

[32] Run-length encoding. https://en.wikipedia.org/wiki/Run-length_encoding.

[33] Risk/reward ratio. http://www.investopedia.com/terms/r/riskrewardratio.asp.

[34] A List of Technical Indicators, http://www.investopedia.com/active-trading/technical-indicators/, 12/13/2015.

[35] Kamruzzaman, J., Sarker, R., “Forecasting of Currency Exchange Rates Using ANN: A Case Study”, IEEE Int. Conf. Neural Networks & Signal Processing, 2003, pp. 793-797.

[36] Wang, S., Yu, L., Lai, K., Huang, W., “Developing and Assessing an Intelligent Forex Rolling Forecasting and Trading Decision Support System for Online E-Service”, International Journal of Intelligent Systems, 2007, pp. 475-499.

[37] Chen, M. Y., Chen, K., Chiang, H.., Huang, H., Huang, M., “Comparing extended classifier system and genetic programming for financial forecasting: an empirical study”, Soft Computing, 2007, pp. 1173-1183.

[38] Bahrepour, M., Akbarzadeh, M.,Yaghoobi, M., Naghibi, S., “An adaptive ordered fuzzy time series with application to FOREX”, Expert Systems with Applications, 2011, pp. 475-485.

[39] Lu, C., Lee, T. and Chiu, C., “Financial time series forecasting using independent component analysis and support vector regression”, Decision Support Systems, 2009, pp. 115-125.

[40] Yao, J., Tan, C., “A case study on using neural networks to perform technical forecasting of forex”, Neurocomputing, 2000, pp. 79-98.

[41] Villa, S., Stella, F., “A Continuous Time Bayesian Network Classifier for Intraday FX Prediction”, Quantitative Finance, 2012, pp. 1-20.

[42] Shen, F., Chao, J., Zhao, J., “Forecasting exchange rate using deep belief networks and conjugate gradient method”, Neurocomputing, 2015, pp. 243-253.

[43] Anasataskis, L., Mort, N., “Exchange rate forecasting using a combined parametric and nonparametric self-organising modelling approach”, Expert Systems with Applications, 2009, pp. 12001-12011.

[44] Determination of the Euro Conversion Rate, https://www.ecb.europa.eu/press/pr/date/1998/html/pr981231_2.en.html, 13/13/2015.

[45] Dempster, A.P., Laird, N.M., Rubin, D.B, "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society, 1977, pp. 1–38.

[46] Ozorhan, O. M., Toroslu, I.H., Sehitoglu, O. T., “A Strength-Biased Prediction Model for Forecasting Exchange Rates Using Support Vector Machines and Genetic Algorithms”, Soft Computing, 2016, pp. 1-19.

APPENDIX A

ALGORITHMIC SUCCESS RATE COMPUTATION

A.1 Formulae Regarding Computation of Precision, Recall, Accuracy and F-measure

$$Precision_{positive} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$Precision_{negative} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}$$

$$Recall_{positive} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$Recall_{negative} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}$$

$$\mathrm{F} - \mathrm{measure} = \frac{2 \text{ x Precision x Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

VITA

Current Address:

Central Bank of the Republic of Turkey,

Department of Information Technologies,

06100, Ulus – Ankara, Turkey

Permanent Address:

Beştepe – Ankara, Turkey

onur.ozorhan@gmail.com

RESEARCH INTERESTS

- Time-series Forecasting
- Information Extraction
- Machine Learning
- Data Mining

EDUCATION

Ph.D. Computer Engineering, 2011 - 2017

Middle East Technical University, Ankara, Turkey

Graduate School of Natural and Applied Sciences,

Department of Computer Engineering

Advisor: Prof. Dr. İsmail Hakkı Toroslu

M.S. Computer Engineering, 2006 – 2010

Middle East Technical University, Ankara, Turkey

Graduate School of Natural and Applied Sciences, Department of Computer Engineering

Thesis Supervisor: Prof. Dr. Nihan Kesim Çiçekli

Thesis: A Service Oriented Peer to Peer Web Service Discovery Mechanism with Categorization

B.S. Computer Engineering, 2001 – 2005

Bilkent University, Ankara, Turkey

Faculty of Engineering,

Department of Computer Science

Honor Graduate

PUBLICATIONS

- M. O. Özorhan, N. Çiçekli. "Web Service Discovery: A Service Oriented, Peer-to-Peer Approach with Categorization", *Computer and Information Sciences II/Springer*, September 2011.
- M. O. Özorhan, İ.H. Toroslu, O. T. Şehitoğlu. "A strength-biased prediction model for forecasting exchange rates using support vector machines and genetic algorithms", *Soft Computing*, June 2016.