Currency Exchange Rate Forecasting from News Headlines

Desh Peramunetilleke Raymond K. Wong
School of Computer Science & Engineering
University of New South Wales
Sydney, NSW 2052, Australia
wong@cse.unsw.edu.au

Abstract
We investigate how money market news headlines can be used to forecast intraday currency exchange rate movements. The innovation of the approach is that, unlike analysis based on quantifiable information, the forecasts are produced from text describing the current status of world financial markets, as well as political and general economic news. In contrast to numeric time series data textual data contains not only the effect (e.g., the dollar rises against the Deutschmark) but also the possible causes of the event (e.g., because of a weak German bond market). Hence improved predictions are expected from this richer input. The output is a categorical forecast about currency exchange rates: the dollar moves up, remains steady or goes down within the next one, two or three hours respectively. On a publicly available commercial data set the system produces results which are significantly better than random prediction. The presented techniques can also be applied to predict five outcomes (dollar moves strong up, up, steady, down, or strong down respectively) or only two outcomes (up or down) for any possible time period such as one day ahead or ten hours ahead. The contribution of this research is the smart modeling of the prediction problem enabling the use of content rich text for forecasting purposes.

Keywords: Data mining, foreign exchange, prediction

1. Introduction

The foreign exchange market has changed dramatically over the past twenty five years. The amounts traded are now huge with over a trillion US dollars in transactions executed each day in the foreign exchange market alone. In this increasingly challenging and competitive market, investors and traders need tools to select and analyze the right data from the vast amounts of data available to them to help them make good decisions. This paper specifically describes an approach to forecast short-term movements in the foreign exchange (FX) markets from real-time news headlines and quoted exchange rates based on hybrid data mining techniques.

The basic idea is to automate human thinking and reasoning. Traders, speculators and private individuals anticipate the direction of financial market movements before making an investment decision. To reach a decision, any investor will carefully read the most recent economic and financial news, study reports written by market analysts and market strategists, and carefully weight opinions expressed in various financial journals and news sources. This gives a picture of the current situation. Then knowing how markets behaved in the past in different situations, people will implicitly match the current situation with those situations in the past which are most similar to the current one. The expectation is then that the market now will behave as it did in the past when circumstances were similar. Our approach is automating this process. The news headlines which are taken as input summarize in condensed manner the most important and earth shaking news. News headlines use a restricted vocabulary, containing only relevant information (no sports news for instance) and are written by professionals following strict writing guidelines. This makes these news headlines perfect candidates for automated analysis. Furthermore, exactly these news headlines are received real-time in all the trading rooms around the world. Hence the traders who are actually moving the markets base their expectations precisely on those news headlines. The current situation is then expressed in terms of counts of these keyword records. The current situation is matched with previous situations and their correlation is determined. This allows to conclude what will happen in the future based on what happened in the past in similar situations. This research elaborates and validates this prediction approach.

We show how textual input can be used to forecast intraday currency exchange rate movements. In contrast to numeric time series data textual data contains not only the effect (e.g., the dollar rises against the Deutschmark) but also the possible causes of the event (e.g., because of a weak German bond market). The output is a categorical forecast about currency exchange rates: the dollar moves up, remains steady or goes down within the next one, two or three hours. The presented techniques could also be applied to predict five outcomes (dollar moves strong up, up, steady, down, or strong down respectively) or only two outcomes (up or down).

Much promising research to predict FX movements has already been done. It is well known that purchasing power parity [27] and trade balance [11] are two fundamental factors influencing the long-term movements of exchange rates. For short-term FX prediction, however, the forecasting methods used so far, be they technical analysis [25], statistics or neural nets [12,17], base their predictions on quantifiable information [2,5,6,9,10,13,14,23,24]. As input they usually take huge amounts of quoted exchange rates between various currencies. The innovation of our approach is that we make use of non-numeric and hard to quantify data derived from textual information. In contrast to time series data [32] containing the effect only (e.g., the dollar rises against the Deutschmark) textual information also contains the possible causes of the event (e.g., because of a weak German bond market) [7]. Hence improved predictions are expected from this more powerful input. Goodhart initially attempted to quantify textual news by looking at full news pages of Reuters [8]. But he did not take our approach of looking at potentially market moving word pairs, records and quadruples. The study [36] describes how to manually processing news to enhance the knowledge base of foreign exchange trade support systems.

The rest of the paper is structured as follows. Section 2 contains the technical description of our FX forecasting techniques. One of the major issues investigated is how to preprocess data so as to make them amenable to classification techniques. Section 3 describes the experiments conducted using the data set HFDF93 which can be purchased on-line (via `www.olsen.ch`), the results achieved and a discussion of the findings. Section 4 summarizes this research.

2. Forecasting Techniques

As mentioned before, this section describes the technical details of the suggested forecasting approach.

2.1 Overview

In a typical short term trading environment, FX traders are mainly interested in three mutually exclusive events or outcomes. These three events are to find out whether the change of bid rate in the future between a particular currency and the US dollar will be up, steady or down. Our system predicts which of these three mutually exclusive events will come true. Suppose the exchange rate is moving *x* percent during an interval such s an hour. We predict either of up, steady or down defined as follows: up $\Leftrightarrow$ $x \geq 0.023\%$, steady $\Leftrightarrow$ $-0.023\% < x < 0.023\%$, and down $\Leftrightarrow$ $x \leq -0.023\%$. The percentage change 0.023% of the bid exchange rate is chosen such that each of the outcomes up, down and steady occur about equally likely in the training and testing period.

The major input are news headlines:

```
1993-09-24 08:59:10 "NO MONETARY, FISCAL STEPS IN JAPAN PM'S PLAN - MOF"
1993-09-24 09:00:46 "GERMAN CALL MONEY NEARS 7.0 PCT AFTER REPO"
1993-09-24 09:01:00 "BOJ SEEN KEEPING KEY CALL RATE STEADY ON THURSDAY"
1993-09-24 09:01:06 "AVERAGE RATE FALLS TO 9.28 PERCENT AT ITALY REPO"
1993-09-24 09:02:24 "INSIGHT -ITALIAN BOND FUTURE HEADING FOR 114.00"
1993-09-24 09:04:18 "DUTCH MONEY MARKET RATES LITTLE CHANGED"
```

Each news headline is associated with a time stamp showing the day, hour and minute it was received through a news service such as Reuters. Although it varies on average, about forty news headlines are received every hour. The input data and its flow over time is illustrated in figure 1.

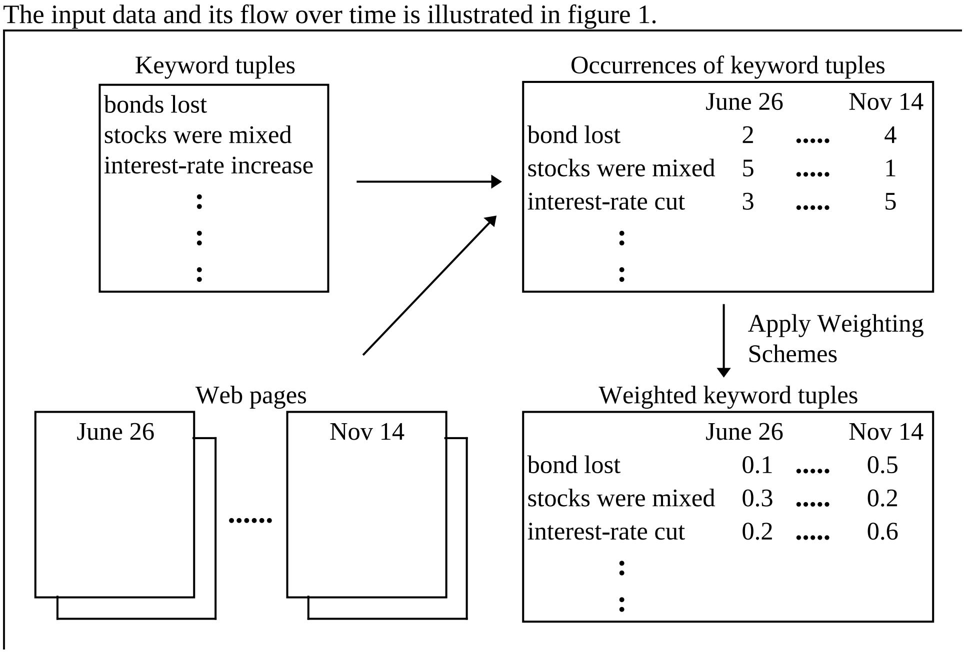

Figure 1: from this information the dollar movement in the period 10-11pm is forecast.

The other source of input is a set of keyword records. These keyword records are provided once by a domain expert and are not changed thereafter. We use over four hundred records consisting of a sequence of two to five words:

```
US, inflation, weak
Bund, strong
Germany, lower, interest, rate
BUBA cut
pound, lower
US, dollar, up
US, dollar, down
Bund gain
Swiss bond steady
```

There is no limitation on the number of keyword records nor on the number of words constituting a record.
The actual currency movements are filtered out from time series of quoted exchange rates.

```
1993-09-24 08:59:32 1.6535
1993-09-24 09:00:00 1.6535
1993-09-24 09:00:20 1.6537
...
1993-09-24 10:00:04 1.6528
1993-09-24 10:00:10 1.6520
```

On 24 Sept 1994, the dollar went down versus the Deutschmark in the period 9 to 10 am, as it depreciated by 0.4% ((1.6528-1.6535)/1.6535).

Given the data described, the prediction is done as follows:

1. The number of occurrences of the keyword records in the news of each time period is counted, see figure 2. The counting of keyword records is case insensitive, stemming algorithms [28] are applied and the system considers not only exact matches. For example, if we have a keyword record "US inflation weak", and a headline contains a phrase "US inflation is expected to weaken", the system counts this as a match.
2. The occurrences of the keywords are then transformed into weights (a real number between zero and one). This way, each keyword gets a weight for each time period, see figure 2. The computation of the weights from their occurrences is described in section 2.3.

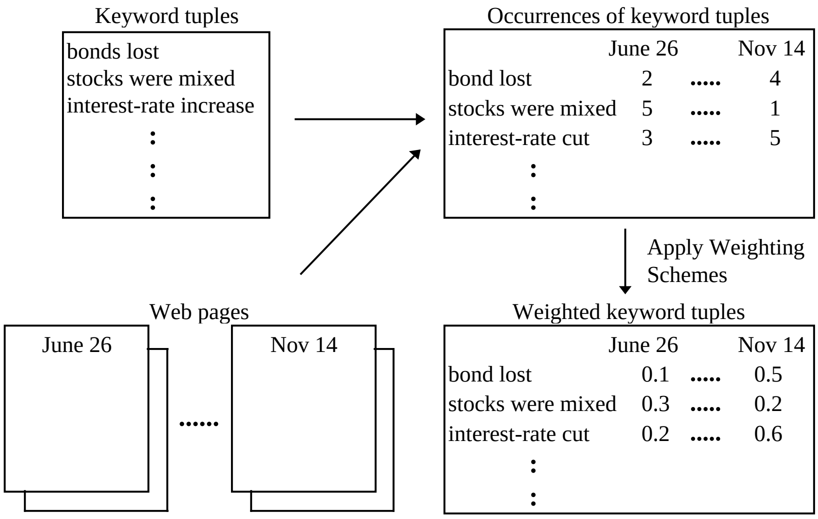

Figure 2: weights are generated from keyword record occurrences.

3. From the weights and the closing values of the training data (the last 60 time periods for which the outcome is known), classification rules are generated [34], see figure 3. The rule generation algorithm is provided in section 2.4.

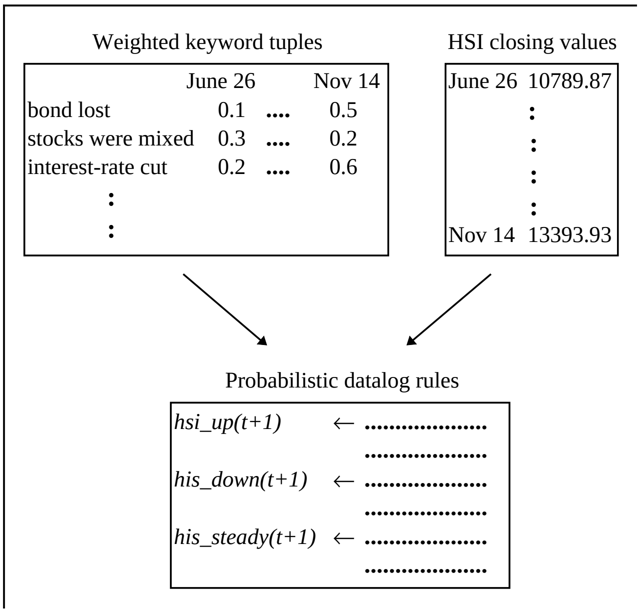

Figure 3: rules are generated from weighted keywords and closing values.

4. The rules are applied to the news of the two most recent periods to yield the prediction. In figure 1, the news received between 8 pm and 9 pm results in keyword record weightings for the time period *t-2*. Peroid *t-1* is from 9 pm to 10 pm. The forecast for period *t*, 10 pm to 11 pm, is computed by evaluating the rules on the weighted keyword records of period *t-1* and *t-2*. Note that only the last two periods, *t-1* and *t-2*, are used to predict the movement in period *t* as this yields the highest prediction accuracy.

Every hour (two and three hours respectively), only the keyword records in the latest news headlines are actually counted. The counts of the previous sixty periods (the training periods) are already known. The outcome (up, steady or down) of the latest training period is determined through reading of the quoted currency exchange rates. Now all three rule sets are regenerated, that is, every hour, the rule generation algorithm is invoked so that the rules reflect the most recent market behavior (markets do not always react the same way to the same piece of news). Finally, the newly generated rules are applied to the latest keyword counts (keyword weights respectively) to yield the prediction for the coming hour.

2.2 Rule Semantics

The classifier expressing the correlation between the keywords and the outcome up for instance is a rule set. Versus conventional rules [4,26], our rules have the advantage that they are able to handle continuous attributes and do not rely on Boolean tests. They have therefore more expressive power [31] by retaining retain the strength of rule classifiers: comprehensible models and relatively fast learning algorithms. For example, suppose that attribute stock_rose has been normalized so that maximum value is 1 and minimum value is 0. A rule like `DOLLAR_UP(T) <- STOCK_ROSE(T-1)` expresses a direct linear relationship between the dollar going up and the weight attached to stock rose. Suppose a second rule, `DOLLAR_UP(T) <- STERLING_ADD(T-1).` The event `DOLLAR_UP` is therefore defined to be `STOCK_ROSE or STERLING_ADD`. The probability of the event `STOCK_ROSE or STERLING_ADD` is computed by *stock+sterling-stock*sterling*, where *stock* denotes the weight derived for keyword `stock rose` as outlined in section 2.3. That is, the rules define the event `DOLLAR_UP` as `STOCK_ROSE or STERLING_ADD` and map this event to a real number. This mapping satisfies the three well known Kolmogoroff axioms [33] and hence the mapping defined by the rules is a probability function in the sense of axiomatic probability theory. The number computed by the rules can therefore be called a probability. Similarly, in information retrieval, weights are computed for individual keywords and mapped to a document relevance number. When this mapping satisfies the Kolmogoroff axioms then it is said to be probabilistic information retrieval and people talk about the probability of a document to be relevant.

The aim of this section is to briefly recall this rule semantics in an informal way. The rule generation algorithm is provided in section 2.4. The following is a sample rule set generated by the system.

```
DOLLAR_UP(T) <--     STOCK_ROSE(T-1), NOT INTEREST_WORRY(T-1),
                     NOT BUND_STRONG(T-2), NOT INTEREST_HIKE(T-2)
DOLLAR_UP(T) <--     STERLING_ADD(T-1), BUND_STRONG(T-2)
DOLLAR_UP(T) <--     YEN_PLUNG(T-1), NOT GOLD_SELL(T-2),
```

```
STOCK_ROSE(T-1)
```

Once these rules are generated from the training data, they are applied to the most recently received news headlines, the news of the last two hours. So the likelihood of the dollar going up depends for instance on the weight computed for `stock rose` in the last hour and on the weight of `bund strong` two hours ago. Suppose the following weights for the last two time periods, say period 60 and 61 in our example:

```
STOCK_ROSE(61)        :     1.0
INTEREST_WORRY(61)    :     0.2
BUND_STRONG(60)       :     0.7
INTEREST_HIKE(60)     :     0.0
STERLING_ADD(61)      :     0.5
YEN_PLUNG(61)         :     0.6
GOLD_SELL(60)         :     0.1
```

Applying the rules on those weights computes the probability of the dollar going up within the next hour. More specifically, the rules compute how likely the dollar moves up from the beginning to the end of period 62, i.e. how likely it moves up from 10 pm to 11 pm:

```
DOLLAR_UP(62)    = 1*(1-0.2)*(1-0.7)*(1-0) + 0.5*0.7 + 0.6*(1-0.1)*1
                     // likelihood that first rule true, or second
                     // rule true, or third rule true
                   - 0
                     // since first and second rule are contradictory
                   - 1*(1-0.2)*(1-0.7)*(1-0)* 0.6*(1-0.1)
                     // likelihood that first and third rule  are
                     // both true; note stock_rose is taken only once
                   - 0.5*0.7*0.6*(1-0.1)*1
                     // likelihood that second and third rule true
                   + 0
                     // three rule bodies together are contradictory
                 = 0.811
```

The same way a probability for dollar steady and down respectively is computed.
If the rules also have attached a confidence expressing the accuracy of the rules, then the rule evaluation is the same except that each term stemming from rule r_i will additionally be multiplied with $conf(r_i)$. For example, suppose that the three rules above have attached confidence 0.9, 0.8 and 0.7 respectively. The evaluation `DOLLAR_UP(62)` yields now

```
1*(1-0.2)*(1-0.7)*(1-0)*0.9 + 0.5*0.7*0.8 + 0.6*(1-0.1)*1*0.7
- 1*(1-0.2)*(1-0.7)*(1-0)* 0.6*(1-0.1)*0.9*0.7
- 0.5*0.7*0.6*(1-0.1)*1*0.8*0.7
= 0.512
```

2.3 Computation of Keyword Record Weights

This section describes how the weights are generated from the money markets news headlines. The computation of weights is illustrated in figure 2. The weight generation makes use of two input sources: the news headlines and the keyword records. For each training period a weight is generated for each keyword record from the news headlines received in this period. For every consecutive time period the weights generated may be different. There is a long history of text retrieval using keyword weighting to rank documents [21,22,28,29]. In contrast to these approaches, however, we consider not single keywords but word pairs, triples etc. Furthermore, our aim is not to find out which documents are most relevant with respect to a query, but rather to discover corelations between keyword records and currency movements. In the following subsections we investigate three different methodologies to compute the relevant weights.

2.3.1 Boolean Method

Suppose the time period for which forecasts are made is one hour. Let from 9 am to 10 am be period t. The next time window refers to the period from 10 am to 11 am and so on. Then the system checks whether in some news headline arriving in period t a keyword record i occurs at least once. If so, the value of $w_i(t)$ is set to one, otherwise $w_i(t)$ is set to zero. $w_i(t)$ is the weight of record i for time window t. The term frequency $TF_i(t)$ is the number of occurrences of keyword record i in a particular time window t.

2.3.2 TF x IDF Method

This method consists of three components, term frequency, discrimination factor and normalization. The term frequency alone is not a good indicator of the record importance with respect to a particular time window. This is due to the fact that if a keyword record appears frequently, the keyword record is not necessarily a characteristic indicator for the strength or weakness of the US dollar. Therefore, a new component is introduced that favors keyword records concentrated in only a few time windows. We use inverse document frequency [28]. In our case, inverse document frequency is defined as follows:

$$IDF_i = \log\left(\frac{N}{DFi}\right)$$

where N is the number of time windows in the training data and DF_i is the number of time windows containing record i at least once. The weight $w_i(t)$ of keyword i is calculated by multiplying the term frequency $TF_i(t)$ with the document discrimination IDF_i. In addition, the weight has to be normalized to obtain a value between zero and one. Therefore, it is divided by the maximum number of times record i occurs in any training time window.

2.3.3 TF x CDF Method

Another potentially useful concept is category frequency *CF* [28]. For each possible category (bid exchange rate of dollar up, down and steady) the *CF* of a keyword record is the number of time windows containing the keyword record in that particular category. Table 1 shows category frequency of keyword records.

$$w_i(t) = TF_i(t) \times IDF_i \times \left(\frac{1}{\max_t\{TF_i(t) \times IDF_i\}}\right)$$

keyword record	\$ up	\$ down	\$ steady
US, inflation, weak	20	2	10
Germany, lower, interest, rate	8	0	7
Bund, strong	1	4	12

Table 1 : Category frequency of keyword records.

The Category Discrimination (*CDF*) is derived from *CF*.

$$CDF_i = \frac{\max\,(CF_{i,up},\, CF_{i,down},\, CF_{i,steady})}{DF_i}$$

where DF_i is the number of time windows containing keyword tuple i at least once. For each record, the sum of its category frequencies is equal to the number of time windows that it appears in the training data. The weight $w_i(t)$ of record i is calculated by multiplying the term frequency $TF_i(t)$ with the category discrimination CDF_i. Finally, $w_i(t)$ is again divided by the maximum number of times record i occurs in any time window. This again assures that $w_i(t)$ is a weight between zero and one.

2.4 Rule Generation

For many data mining and discovery tasks, a rule-based approach has proven useful [1,15,16]. We also take a rule-based approach.

$$w_i(t) = TF_i(t) \times CDF_i \times \left(\frac{1}{\max_t\{TF_i(t) \times CDF_i\}}\right)$$

The algorithm generating the rules relies on the notion of *most general rule* . A most general rule is one which has only one positive literal in its body involving either variable *t-1* or *t-2*. The following are most general rules.

```
DOLLAR_UP(T) <-- STOCK_ROSE(T-1)
DOLLAR_UP(T) <-- BUND_STRONG(T-1)
DOLLAR_UP(T) <-- INTEREST_WORRY(T-1)
DOLLAR_UP(T) <-- STOCK_ROSE(T-2)
…
```

A rule *r* is *specialized* to rule *s*, denoted *r>s*, by appending an additional literal to the body of *r*. Suppose *r* is the rule `DOLLAR_UP(T) <-- STOCK_ROSE(T-1).` The following are specializations of *r*.

```
DOLLAR_UP(T) <-- STOCK_ROSE(T-1), BUND_STRONG(T-1)
DOLLAR_UP(T) <-- STOCK_ROSE(T-1), BUND_STRONG(T-2)
DOLLAR_UP(T) <-- STOCK_ROSE(T-1), INTEREST_WORRY(T-1)
DOLLAR_UP(T) <-- STOCK_ROSE(T-1), NOT INTEREST_WORRY(T-1)
…
```

Suppose the head of rule *r* is `DOLLAR_UP` (the cases `DOLLAR_STEADY` and `DOLLAR_DOWN` are analogous). The confidence of rule *r*, denoted *conf(r),* is defined as follows:

$$conf(r) = \frac{\sum_t eval_{\{r\}}(t) \times up(t)}{\sum_t eval_{\{r\}}(t)}$$

where t is a training example, $up(t)$ is 1 if the actual outcome is *up* and 0 otherwise. The evaluation of the single rule r on example t, denoted by $eval_{\{r\}}(t)$, is explained in section 2.2 (see also [33]).

The rule algorithm generating a rule set R is as follows [34].

$R=\emptyset$
while $|R|\leq maxRules$ **do**
{ $C=\{r|$ r is a most general rule$\}$
  repeat
  { $r'=r$
    $C=\{s|r>s\}\cup\{r\}$
    $r=$ the rule $s\in C$ minimizing $mse(R\cup\{s\})$
  }
  until $(r=r')$
  attach $conf(r)$ to r
  $R=R\cup\{r\}$
}
$R'=R$
$R=$ the rule set $S\subseteq R'$ minimizing $mse(S)$

In the inner loop, the algorithm selects the rule s with minimal mean square error (mse) of the rule set $R\cup\{s\}$.

$$mse(R\cup\{s\}) = \sum_t (up(t) - eval_{R\cup\{s\}}(t))^2$$

The evaluation of example t using the rules R generated so far with their confidence plus the rule s is denoted by $eval_{R\cup\{s\}}(t)$. The summation goes over all training examples t and $up(t)$ is defined as before (assuming the rule set to be built is for `dollar_up`; for rule sets steady and down it is analogous). Note that mean square error is used to measure the quality of a rule. This is an appropriate goodness measure for applications where the classification problem is expected to be relatively difficult (no perfect models possible). Regression analysis, neural net learning based on back propagation and nearest neighbor algorithms are also based on mean square error or square distance considerations. The last statement of the algorithm selects that subset S of the generated rules R' which has least mean square error. This is a common rule set simplification and yields the final result R.

2.5 Final Prediction

Once the rules are generated, they are applied to the most recently collected textual news and analysis results. So the likelihood of the dollar going up in the period starting at 10 pm depends for instance on the weight computed for `STOCK_ROSE`. From those probabilities, i.e. how likely the dollar is going up, down or remains steady respectively, the final decision is taken. For example, the final decision is that the dollar moves up. Though maximum likelihood yields fairly good results for making this final decision, we found an improvement over maximum likelihood [3]. This method also proved superior in other applications [35].
Each of the three rule sets (`DOLLAR_UP, DOLLAR_STEADY, DOLLAR_DOWN`) yields a probability saying how likely the respective event will occur. For each rule set j we compute a threshold v_j such that if the computed likelihood $l_j(t)$ is above the threshold then it is taken as true and false otherwise. The threshold is determined by testing the values v_j = 0, 0.05, 0.1, 0.15 , …, or 1 and selecting that threshold which results in the least error on the training examples. Given the three thresholds v_j and the three likelihoods $l_j(t)$, there are three possible cases.

- Exactly one of three likelihoods is above its threshold, i.e. $l_j(t)\geq v_j$ for one j: class j is the final prediction. This case is illustrated in table 2.
- None of the three likelihoods is above its threshold, i.e. $l_j(t)<v_j$ for all j: we compute

$$d_j(t) = \frac{l_j(t) - v_j}{v_j}$$

  and select that j to be true for which the deviation $d_j(t)$ is maximal.
- All likelihoods are above their threshold, i.e. $l_j(t)\geq v_j$ for all j: as before we select that j with maximal deviation $d_j(t)$.

time *t*			
	probability	threshold	binary decision
up	0.811	0.45	1
steady	0.018	0.25	0
down	0.171	0.20	0

Table 2: forecast for time *t* is up

3. Experiments, Results and Discussion

Experiments were conducted using the HFDF93 data available from Olsen & Associates in Zurich. This data set contains FX rate quotes for USD-DEM, USD-JPY money market news headlines plus 3 months maturity inter bank deposit rates of USD and JPY. Beforehand the rule generation part of the system was tested extensively at the Treasury Department of the Union Bank of Switzerland (UBS) by FX dealers by providing manual weights [30] rather than having them generated automatically as is done by this system. Some of the key issues related to the experimental set up are discussed first.

When the system was tested at UBS with manual input of factors, it produced successful results. Traders entered a standard value between very high (1) and very low (0) for twelve factors namely, government policies, political news, rumours, central bank, employment rate, inflation, bonds, capital flow, stock market, money supply, volatility NY and volatility London [30]. But the purpose of the experiments conducted using the HFDF93 data set was to derive such and other factors automatically.

In every experiment, sixty test periods are considered. We chose Sept 1993 as training and test period as this is the last month for which HFDF93 contains data. More precisely, the testing period for one hour predictions is 22 Sept 1993 13:00 GMT time to 27 Sept 1993 10:00 GMT (due to holiday break this period constitutes sixty trading hours). The start date and time is the same when testing two hour forecasts. The testing period for the three hour predictions finishes at the end of the data set HFDF93 which is 30 Sept; it begins on 21 Sept 1993. The first training periods are always those intervals immediately preceding the one to be predicted. Hence the training period when predicting the movement from 9:00 to 10:00 on 27 Sept is 22 Sept 12:00 to 27 Sept 9:00. The following variables were changed during the experimentation: length of the time period (one, two and three hours, see figure 1), weight generation method (the three methods described in section 2.3), different currencies (exchange rate of DM and Yen against US dollar). All experiments were conducted using the same news headlines and the same keyword record definitions. The prediction accuracy of the test data is shown in tables 3 and 4.

weighting method	1 hour	2 hours	3 hours
Boolean	41	48	40
TF x IDF	42	39.5	42.5
TF x CDF	51	42	53

Table 3: prediction accuracy in % for DEM/USD and various time periods.

weighting method	1 hour	2 hours	3 hours
Boolean	38.5	25	31
TF x IDF	28	37	35.5
TF x CDF	46	39	48

Table 4: prediction accuracy in % for JPY/USD and various time periods.

From tables 3 and 4 is observed that data set 1(DEM/USD) produces better results than data set 2 (JPY/USD). This is mainly due to the fact that the keyword records have a greater influence on the DEM than on the Japanese Yen. We also noted that for intraday forecasting, increasing the number of training data makes no significant difference in the prediction performance as the currencies rally or fall on various short-term economic factors and also on the rapidly changing conditions of stock and bond markets.

The results are compared against a standard statistical tool which extrapolates time series data. The highest success rate achieved by using a statistical package was 37 per cent. Our best weighting method has an accuracy of 51 per cent for the same test and training period. Human traders are said to have an accuracy of up to 50 percent for the same intraday prediction task. However, we did not actually succeed in convincing a trader to measure his personal prediction accuracy. There was simply a consensus among the traders that it is actually hard to achieve 50% accuracy. In another experiment we used a feed forward neural net to predict the next outcome (dollar up, steady or down) based on the previous n such outcomes. Varying n between 2 to 10 the average accuracy achieved is 37.5%, though never fifty per cent was reached.

It is obvious that the TF x CDF method is the best and that it performs significantly better than random guessing. Random guessing gives on average 33% accuracy because - by definition of the three possible outcomes up, steady and down - each outcome is about equally likely. We determine the probability of the outcomes as reported in tables 3 to 8 when our system would do random prediction. In this case, each prediction is independent from the other predictions. So we have a Binomial Distribution with mean $n*p$ and variance $n*p*(1-p)$ where p is the probability of success (0.33) and n is the number of times we predict [18]. If *n* is rather large, Binomial distribution is approximately Normal distribution. In the sequel, we consider only the TD x CDF method. The probability that the prediction accuracy is equal or above 51% for random guessing is less than 0.4% when having 60 trials (see table 3). The probability of achieving at least 43.3% prediction accuracy with random guessing is 95% for 60 trials. Each of the outcomes in the first and third column of table 3 and 4 can therefore be achieved by random guessing only with a probability of less than 5%. The outcome of the second column in tables 3 and 4 can be achieved by random guessing with a likelihood of a little more than 5%. However, when taking all 180 forecasts of tables 3, 4, 5 together, then the likelihood of getting the average prediction accuracy 48.6% ((51%+42%+53%)/3) by random guessing is below 0.0001%. The probability of achieving the average accuracy of 44.3% ((46%+39%+48%)/3) as reported in table 4 by random guessing is still below 0.001%. Hence, our system performs almost certainly better than random guessing. For DM and the best performing weighting method, TF*CDF, the results are presented a little more detailed. The third column in table 5 indicates how many times the system predicts up or down and the dollar was actually steady; or, the system predicts steady and the system was actually up or down. The last column indicates the percentage of totally wrong predictions. That is, the system expects the dollar to go up and it moves down, or vice versa. Table 6 shows the distributions of the actual outcomes and the forecasts.

	accuracy	slightly wrong	wrong
DM, 1 hour period	**52%**	**23%**	**25%**
DM, 2 hour period	**41%**	**30%**	**29%**
DM, 3 hour period	**53%**	**22%**	**25%**

Table 5: forecasting accuracy for DM/US dollar.

	distribution of actual outcome			distribution of the forecast		
	up	steady	down	up	steady	down
DM, 1 h	**35%**	**30%**	**35%**	**33.3%**	**30%**	**36.6%**
DM, 2 h	**30%**	**31.6%**	**38.3%**	**35%**	**33.3%**	**22.6%**
DM, 3 h	**36.6%**	**33.3%**	**30%**	**35%**	**31.6%**	**33.3%**

Table 6: distribution of DM/US dollar forecast.

4. CONCLUSIONS

A new approach to forecast intraday exchange rates using news headlines has been introduced. The major difference from other forecasting techniques such as technical analysis or statistics is that the input to the system is different. We take as input textual information which is hard to process but which is rich in contents. The conventional approach is to take numerical time series data and to analyze those data using various techniques. In contrast to numeric time series data our input data contains not only the effect (e.g., the dollar rises against the Deutschmark) but also the possible causes of the event (e.g., because of a weak German bond market). Hence improved predictions are expected from this richer input.

We gave a comprehensible overview on the timely collaboration of our text analyses, preprocessing and forecasting approach. Different ways of pre-processing the news headlines have been suggested and the rule based prediction engine was explained in detail. Extensive experimentation has revealed that the weighting method TF x CDF performs the best. The results were compared with those of a conventional numeric time series analysis tool and two different neural net approaches. It was found that the techniques introduced in this paper outperform other approaches and that our approach is significantly better than random guessing. This reveals the enormous potential of the system and opens up many paths for future research in this area. It is also planned to predict in future other financial markets such as bond markets. Furthermore, it is believed that there are some parts of the system which can be further improved to provide more accurate forecasts. For example, by combining our techniques with other forecasting methodologies a powerful hybrid forecasting system can be built. Finally, it is conceivable that the keyword records can also be generated automatically from a sample of news headlines.

REFERENCES

[1] R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules, VLDB, pp 487- 499, 1994.
[2] P. Birrer and T. Eggenschwiler, Frameworks in the Financial Engineering Domain - An Experience Report. UBILAB, Union Bank of Switzerland, Bahnhofstrasse 45, CH-8021 Zurich, Switzerland,1996.
[3] V. Cho and B. Wüthrich, Towards Real-Time Discovery from Distributed Information Sources, Pacific Asia Conf on KDD and Data Mining (PAKDD98), get from http://www.cs.ust.hk/~beat/bio.html, 1998.
[4] W. Cohen, Fast Effective Rule Induction, 12th Int Conf on Machine Learning, pp 80-89, 1995.
[5] E. Cox and T. J. Schwartz, Software Review: Around the World with Fuzzy Products, AI Expert, March 1990, p44-8.

[6] T. Dhillon, Managing Intelligent Systems in a Banking and Finance Environment, La Trobe University, Bundoora, Victoria 3083 Australia, 1996..
[7] C. Goodhart, News and the Foreign Exchange Market, Journal of International Securities Markets, Vol. 4, pp333-348, 1989.
[8] C. Goodhart, S.G. Hill, S. G. B. Henry and B. Persaran, News Effects in a High Frequency Model of the Sterling-Dollar Exchange Rate, Journal of Applied Econometrics, Vol. 8, pp 1-13, 1993.
[9] S. Goonatilake and P. Treleaven, Intelligent Systems for Finance and Business, Wiley, 1995.
[10] S. Goonatilake and S. Khebbal, Intelligent Hybrid Systems, Wiley, 1995.
[11] P D Grauwe, H Dewachter, et al; Exchange Rate Theory, Blackwell Publications, Cambridge, 1993.
[12] P. Gregory, Neural Networks: Developing an Effective Strategy, Conference - Neural Networks 1990, Blenheim Online.
[13] D. M. Guilaume, M. M. Dacorogna, R R Dave, U. A. Muller, R B Olsen and O. V. Pictet, From the Bird's Eye to the Microscope - A Survey of New Stylized Facts of the IntraDaily Foreign Exchange Markets, DMG.1994-04-06, Olsen & Associates, Seefeldstrassen 233, 8008 Zurich, Switzerland, 1995.
[14] D. M. Guillaume, O V. Pictet and M. M. Dacoragna, On the IntraDaily Performance of GARCH Processes, Olsen & Associates, Seefeldstrassen 233, 8008 Zurich, Switzerland, 1995.
[15] J. Han, Y. Cai, C. Cercone, and Y. Huang, Data-Driven Discovery of Quantitative Rules in Relational Databases, IEEE TKDE, 5(1):29-40, 1993.
[16] K. Hatonen et al, Knowledge Discovery from Telecommunication Network Alarm Databases, Int Conf on Data Engineering, 1996.
[17] J. Hertz, A Krogh, and R. G. Palmer, Introduction to the Theory of Neural Computation. Wiley, 1990.
[18] W.W. Hines, Probability and Statistics in Engineering and Management Science, Wiley, 1990.
[19] K. Jim, J. Lai, and B. Wüthrich, Rule Discovery: Error Measures and Conditional Rule Probabilities, Pacific Asia Conf on Knowledge Discovery and Data Mining, pp. 82-89, 1997.
[20] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, AAAI Press / The MIT Press, pp625, 1996.
[21] E.M. Keen, Query Term Weighting Schemes for Effective Ranked Ouput Retrieval, 15th Int. Online Information Meeting, pp.135-142, 1991.
[22] D.L. Lee, H. Chuang, and K. Seamons, Document Ranking and the Vector-Space Model, IEEE Computer, Theme Issues on Assessing Measurement, April/May, 14(2):67-75, 1997.
[23] U. A Muller et al., Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components, Internal document, Olsen & Associates, Zurich, Switzerland, 1996.
[24] O. V. Pictet et al., Genetic Algorithms with Collective Sharing for Robust Optimization in Financial Applications, Internal document, Olsen & Associates, Zurich, Switzerland, 1996.
[25] J. Pring, Technical Analysis Explained. McGraw-Hill, 1991.
[26] J. R. Quinlan, Simplifying Decision Trees, Int J Man-Machine Studies, 27:221-234, 1987.
[27] F.L. Rivera-Batiz and L.A. Rivera-Batiz, International Finance and Open Economy Macroeconomics 2 nd. MacMillan, 1994.
[28] G. Salton and M.J. McGill, Introduction to Modern Information Retrieval. New York McGraw-Hill, 1983.
[29] G. Salton and C. Buckley, Term Weighting Approaches in Automatic Text Retrieval, Information Processing and Management, Vol. 24, No. 5, pp513-523, 1988.
[30] K. P Sankaran, A System to Forecast Currency Exchange Rates (MPhil thesis). Department of Computer Science, Hong Kong University of Science and Technology, 1996.
[31] P. Schäuble and B Wüthrich, On the Expressive Power of Query Languages. ACM TOIS, 12(1):67-91, 1994.
[32] D Schmidt, A. K. Dittrich, W Dreyer and R. Marti, Time Series, a Neglected Issue in Temporal Database Research? UBILAB, Union Bank of Switzerland, Universitatsstrasse 84, CH-8033 Zurich, Switzerland, 1995.
[33] B Wüthrich, Probabilistic Knowledge Bases. IEEE TKDE, 7(5) 651-698,1995.
[34] B. Wüthrich, Discovering Probabilistic Decision Rules, International Journal of Intelligent Systems in Accounting, Finance and Management, Vol 6:269-277, 1997.
[35] B. Wüthrich, S. Leung, D. C. Peramunetilleke, W. Lam, V. Cho, and J. Zhang, Daily Stock Market Predictions from World Wide Web Data, Int Conf on Knowledge Discovery and Data Mining (KDD98), pp. 269-274, 1998
[36] T. Yagyu, H. Yuize, M. Yoeneda, and S. Fukami, Foreign Exchange Trade Support Expert System, Proc. IFSA'91, Artificial Intelligence, Brussels, pp. 214-217, 1991.