

A high-angle photograph of a massive crowd of runners participating in a marathon on a suspension bridge. The runners are densely packed, filling the entire width of the bridge deck. The bridge's steel structure, including cables and support beams, is visible on either side of the runners. The scene is captured from a high vantage point, looking down the length of the bridge.

The Congestion Count

a tool for trading breakouts

Following the Leaders

using the COT report
to trade the S&P 500

Intermarket Analysis

a review of John Murphy's
latest book

ProStick vs Candlestick

comparing trading strategies
using closing and modal prices

FUNDAMENTAL DATA TO SATISFY THE CHARTIST

by Khurshid Ahmad together with Tugba Taskaya Temizel, David Cheng, Pensiri Manomaisupat, Saif Ahmad, Lee Gillam, Haitham Trablousi and Matthew Casey

Can market sentiment be visualized in the same way as share prices or any other financial asset? If so, technical analysts could use such charts for fundamental confirmation of their technical views. A team of researchers at the University of Surrey present a prototype system which they have developed for an EU-sponsored project - GIDA (Generic Information-based Decision Assistant) - the aims of which include the development of a method for automatically generating trading signals from financial texts.

The financial trader receives market information in at least two forms: numbers and text. While various types of chart can be used to display numerical data in a form that is easy to digest and analyse, the financial trader is expected to read literally hundreds of stories. Even if this only means reading the headlines or skim-reading, this is still a monumental task. Surveys of financial traders show that rather than using informative, complex techniques, simple visualisation techniques were always preferred (Saltz and Steinbach, 1997). Clearly, a system that can simultaneously deal with numeric and textual data would be of benefit.

It is the prototype of such a system being developed for an EU sponsored project (GIDA) that we describe briefly. Despite the fact that textual information is not as discretely time indexed, this prototype demonstrates how news reports can be organised and analysed in order to assess and quantify the general mood of the trading market. The assessment is about the sentiment, which is conveyed in an individual news item: does the news item express a positive sentiment or

The prototype:

- ◆ can manage a collection of news stories and historical time-series;
- ◆ can analyse the news stories for market sentiment;
- ◆ can correlate financial time-series with a time-series of keywords describing the positive and negative mood of the market.

a negative sentiment? Such expression is deeply embedded in language. Nevertheless, one can argue that the occurrence of words related to progress, profits, rising share prices may boost traders' confidence and the opposite, traumatic accidents, losses, falling share prices, will dent the traders' confidence.

Visualising market sentiment

Reports about the financial markets sometimes use metaphorical terms to indicate how the traders feel about the market: terms like

'bear' and 'bull' markets are well-known. News reports contains other expressions like: FTSE set to fall, Royal and Sun hit by asbestos suit, Energy tax hurts business.

Each of the underlined verbs fall, hit, hurts conveys a negative sentiment. On the other hand, verbs potentially like rise, jump, and climb may convey a positive sentiment. It is true that a word in natural language may have a range of meanings: the token rose may mean the flower rose; it can be a name, as in Mr/Ms Rose in addition to meaning 'increase in the value of a share or currency'. However, in financial texts tokens like rose, fall have a restricted meaning [14]. This is particularly true when these words are used in the following contexts:

{Shares, currencies} rose/fell {by} X percent
{Shares, currency} rose/fell {to} X percent

One problem-solving technique (heuristic) would be to use the frequency of these two verbs only if they occur in the context described above. The use of this heuristic allows us to quantify the market sentiment in

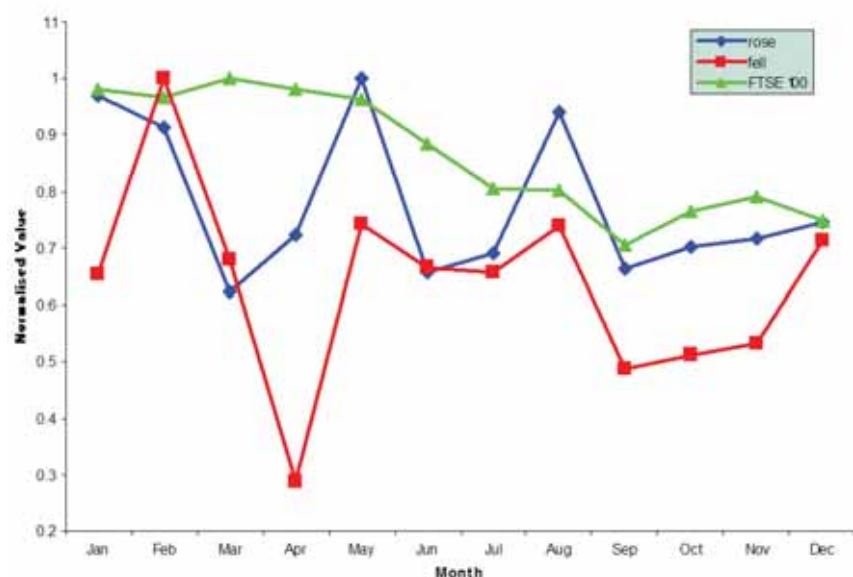


Figure 1.

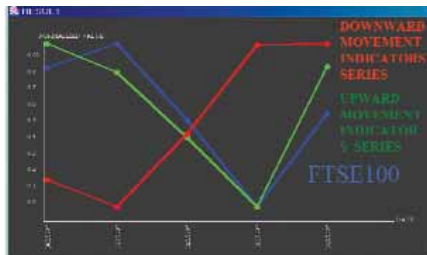


Figure 2. The visualisation system on the web shows weekly news. The system displays content of individual news items together with the keywords associated with the content.

as much as a computer system can ever evaluate any sentiment. We have analysed 2 million words of text comprising financial news supplied by Reuters for the year 2000. These texts were analysed by our text analysis system, System Quirk (which performs text analysis for our prototype) and we extracted the frequency of the verbs rose and fell. These frequencies were then plotted together with the FTSE 100 index. Figure 1 shows how the sentiment time-series correlates with FTSE.

We have performed the analysis over a larger collection of financial news from Reuters over a 3-year period (2000-2002), comprising over 10 million words, and found that the heuristics mentioned above hold as well. In addition, many other sentiment verbs and prepositions (up, down, adrift) have been identified and used to compute the market sentiment.

Typically, a news agency like Reuters or Bloomberg will supply over 2000 stories per day comprising between 300-800 words. Reuters news stories give, amongst other information, industry sector and country information. Queries against a database of news stories can be handled provided the stories have been pre-indexed on to the date of publication, keywords characterising the content of the news, and, in our case, key verbs and prepositions characterising the

sentiments, with positive and negative, expressed in and by each of news stories. Once a query has been successfully matched, it is important that the news story be displayed in full together with the frequency of the keyword, key verbs and prepositions. A visualisation system for viewing a collection of text, queried according to the various attributes of its constituent texts, has been developed in Java as a key component to our

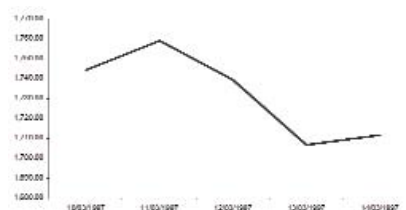


Figure 3. The visualisation system can work with Reuters news categorisation.

financial trading prototype (SATISFI: "Sentiment and Time-series: Financial Analysis System"). The system is written in Java and is accessible on the world-wide web (see Figure 2). Our visualisation system can work with Reuters country and industry sector categories together with Reuters supplied keywords (see Figure 3).

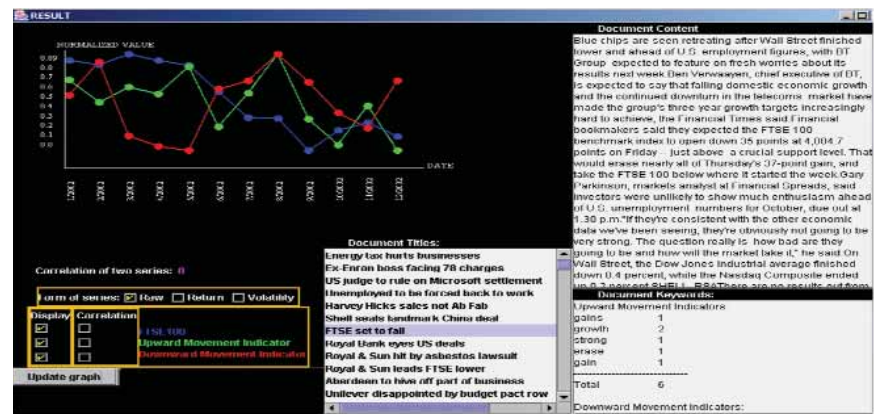


Figure 4. SATISFI prototype

The SATISFI system

SATISFI was developed to show the correlation between a time-series of an instrument and market sentiment. The time-series of an instrument is seldom used in its raw form, rather derived forms are used: return and volatility. The index return is the logarithmic difference between two consecutive values and volatility is a measure of how fast or slowly the instrument moves up or down in terms of its value. Figure 4 shows how the prototype system works in terms of its four components: each helping to visualise either the market sentiment or an instrument's time variation. The value of the correlation helps the user to make a buy/sell decision.

SATISFI has four major components that have been integrated as shown in Figure 4.

I. Time-series Display: SATISFI can display three time-series at a time. These time-series comprise of FTSE-100 close index values, upward movement indicators and downward movement indicators. As discussed above, upward and downward movement indicators are the quantification of the market sentiment expressed in financial news. Over 70 terms each have been identified for conveying 'good' and 'bad' news. For example upward movement indicators would contain terms like 'up, rise, growth' etc. while down-

ward movement indicators would contain terms like 'down, fall' etc. The movement indicator time-series are synthesized by counting these movement indicator terms within the financial news published for a particular day. Each time-series is normalised for proper display purposes. SATISFI is capable of displaying the above time-series in three forms:

- a. Raw form denotes the original time-series.
- b. Return form refers to the logarithmic difference between two consecutive values.
- c. Volatility (historical volatility) is the relative rate at which the time-series moves up or down.

II. Time-series Correlation: SATISFI provides the user the facility of cross correlating two series in any form (raw, return, volatility). Any series can be shifted forward or backward and cross correlation recalculated to determine whether the market is followed by the news or vice versa.

III. Document Display: This comprises of two parts:

- a. Document Titles: Clicking a dot (date) on any of the time-series, displays the corresponding date's news titles.
- b. Document Content: The content of any document title can be viewed by clicking that news title.

IV. Document Analysis: Whenever a document title is selected from the news list, the extracted sentiment keywords along with the frequencies are displayed in "Document Keywords" area. Positive sentiment keyword analysis details appear under the title of "Upward Movement Indicators" and negative sentiment keyword analysis details appear under the title of "Downward Movement Indicators".

Case Study

In order to investigate the effectiveness of the SATISFI system a real-life case study was identified for us by a capital management and research consultancy, JRC Berlin, (who are also the lead partner in the EU spon-

sored GIDA project). According to JRC market experts, there is a well-recorded turning point, from a low position on a day before to a high on the day followed by a low on the

day after, in the European Stock Markets. The two major stock exchanges are the London Stock Exchange (with the FTSE100 index of 100 leading UK companies) and its

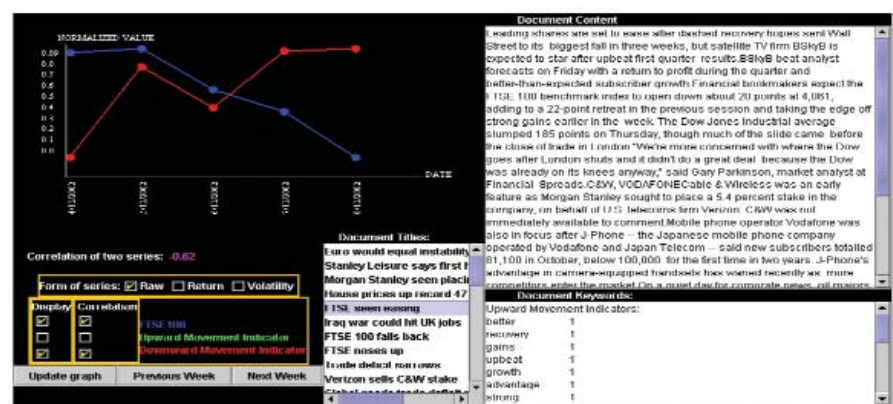


Figure 5. The chart displays DAX100 index between 10 and 14 March 2002.

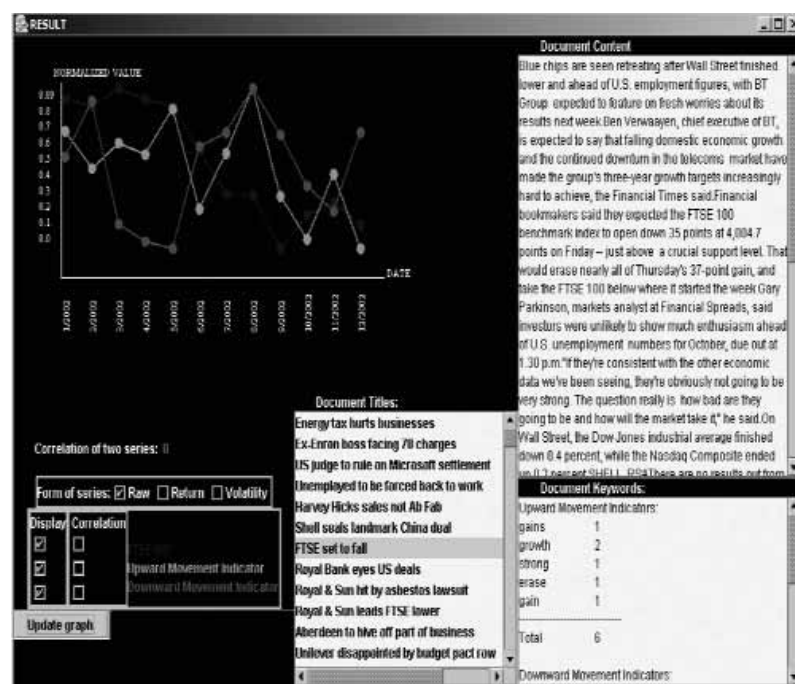


Figure 6. The SATISFI interface depicts three time-series together: FTSE index (blue), upward movement indicators series (green) and downward movement indicators series.

"SURVEYS OF FINANCIAL TRADERS SHOW THAT RATHER THAN USING INFORMATIVE COMPLEX TECHNIQUES, SIMPLE VISUALISATION TECHNIQUES WERE ALWAYS PREFERRED."

then German counterpart (with the DAX 100 GDX). The high point of FTSE100 (and DAX 100) in the recorded turning point case study is that of the 11th March 1997 and two low points are on the 10th and 12th March 1997: a week of changing indices (see Figure 5).

In order to see whether there was a corresponding change in the sentiment related to the UK markets to the FTSE100's behaviour, we selected a week's financial news from Reuters (10th to 14th March 1997) and extracted the potential sentiment expressing words according to our heuristic. There were 15,387 news items produced by Reuters during that week. Figure 6 shows the correlation between the FTSE100 and the upward movement indicator: the Pearson correlation moment is 0.88; visually the change in FTSE100 and the upward movement indicator appears to be almost identical. We have also plotted the downward movement indicator, which is anti correlated with both the FTSE100 and the upward movement indicator (correlation moments of -0.85 and -0.57 respectively).

Two major feedbacks were received from JRC. One feedback was to show the positive news when 'positive sentiment time-series' was clicked and to show negative news when 'negative sentiment time-series' was clicked. The other feedback was to show the major news that caused the turning points.

According to JRC: Rather than displaying news headlines, visualisation of the sentiment of the news is much more beneficial in terms of providing a general overview of the news with respect to stock market in a quick way.

Afterword

SATISFI shows how it is possible to visualise data and information contents in two different modalities. The system is currently under evaluation by JRC Berlin and by the reviewers of the GIDA project appointed by the EU. The initial results are interesting. The end-users have suggested that instead of using SATISFI as a prediction system, it should be used in investigating past behaviour of the market. For instance, using the visualisation system to find out which of the turning points in the value of an instrument (from down to up or up to down) were caused by a specific news item. SATISFI provides a simple visualisation technique for identifying and displaying a complex feature of the news it receives. Furthermore, its ability to receive and display a numerical time-series of financial data, and its ability to correlate the series with an index of market sentiment should be its notable feature.

Khurshid Ahmad is Professor of Artificial Intelligence at the University of Surrey and Principal Investigator of the GIDA Project. Taskaya Temizel and Cheng were Research Assistants of Prof Ahmad on the Project and worked on visualization and document manager respectively. Manomaisupat, S.Ahmad, and Trabloussi are doctoral students of Prof Ahmad and researched on text categorization, time series analysis and local grammar respectively. Gillam is a Research Assistant of Prof Ahmad and has worked as an information architect on a number of EU projects. Casey was a Project Coordinator and now Lecturer in Neural Computing. All the authors are based at the Department of Computing, University of Surrey.

References:

- J.Saltz, J.Steinbach. "Wayfinding in Layered Information Worlds", CODATA Euro-American Workshop Visualization of Information and Data: Where We Are and Where Do We Go From Here?, 24-25 June 1997, pp. 6-7.
- K. Ahmad, D. Cheng, T. Taskaya, S. Ahmad, L. Gillam, P. Manomaisupat, H. Trabloussi, A. Hippisley. "The mood of the (financial) markets: In a corpus of words and of pictures", Proceedings of Corpus Linguistic 2003. Lancaster University, UK. <http://www.computing.surrey.ac.uk/iemmg/publications.html>