Predicting Market Direction from Direct Speech by Business Leaders

Brett M. Drury[1] and José João Almeida[2]

1 **LIAAD-INESC**
Rua de Ceuta 188,6, Portugal
`Brett.Drury@gmail.com`

2 **University of Minho,**
Braga, Portugal
`jj@di.uminho.pt`

Abstract

Direct quotations from business leaders can communicate to the wider public the latent state of their organization as well as the beliefs of the organization's leaders. Candid quotes from business leaders can have dramatic effects upon the share price of their organization. For example, Gerald Ratner in 1991 stated that his company's products were *crap* and consequently his company (Ratners) lost in excess of 500 million pounds in market value. Information in quotes from business leaders can be used to make an estimation of the organization's immediate future financial prospects and therefore can form part of a trading strategy. This paper describes a contextual classification strategy to label direct quotes from business leaders contained in news stories. The quotes are labelled as either: 1. positive, 2. negative or 3. neutral. A trading strategy aggregates the quote classifications to issue a buy, sell or hold instruction. The quote based trading strategy is evaluated against trading strategies which are based upon whole news story classification on the NASDAQ market index. The evaluation shows a clear advantage for the quote classification strategy over the competing strategies.

1998 ACM Subject Classification I.2.1 Application and Expert Systems

Keywords and phrases Sentiment, Direct Speech, Trading, Business, Markets

Digital Object Identifier 10.4230/OASIcs.SLATE.2012.163

1 Introduction

Direct quotations from business leaders can communicate to the wider public the latent state of their organization as well as the beliefs of its leaders. This information can be used to make an estimation of the organization's immediate future financial prospects. An infamous example of the effect of candid quotes on a company's market value was a speech by Gerald Ratner at the Institute of Directors (IOD) in 1992. He stated: *We also do cut-glass sherry decanters complete with six glasses on a silver-plated tray that your butler can serve you drinks on, all for £4.95. People say, How can you sell this for such a low price?, I say, because it's total crap.* and *We sell a pair of earrings for under £1,which was cheaper than a prawn sandwich from M&S, but probably wouldn't last as last as long.*. The Ratner Group company share price fell rapidly and as a direct result of the speech at the IOD the company lost 500 million pounds in market value. Gerald Ratner was forced to resign and the company changed its name to the Signet Group to disassociate itself from Ratner's speech [16]. This extreme example illustrates the communication of the latent negative attitudes of the company leadership towards its products and the associated market reaction.

1st Symposium on Languages, Applications and Technologies (SLATE'12).
Editors: Alberto Simões, Ricardo Queirós, Daniela da Cruz; pp. 163–172
OASICS OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The identification of *actionable information* in quotes from business leaders is a non-trivial task because quotes can contain obscure linguistic features as well as a rich and varied lexicon [5]. Traditional classification techniques are not suitable for this task [8]. This paper proposes that a *contextual classification technique* [8] for labelling quotes from business leaders can be the basis of a successful stock trading strategy. This *quote based trading strategy* was evaluated on the NASDAQ market index against trading strategies which used whole news classification strategies to issue trading instructions. The evaluation used trading accuracy and points difference to gauge each strategy. The results show a clear advantage for using direct quotes as a basis of trading strategy when compared to the competing strategies. The remainder of the paper will discuss: 1. related work, 2. classification strategies, 3. experiments 4. conclusions.

1.1 Related Work

The use of quotes from business leaders as a basis of a trading strategy seems to be a unique research problem as the literature review failed to identify papers which used quotes for trading. There were a small number of papers which attempted to classify quotes into pre-assigned categories. Balahur et al [1] classified quotes from politicians by using general sentiment dictionaries. The value from the dictionaries for each sentiment word in a quote were summed together and a label applied to the quote. A positive quote has a score of greater than 0, a negative quote would have a score of less than 0 and a neutral quote would have a score of 0.

The research literature was richer when examined for strategies which used general news information to predict the direction of a stock market index. The research literature revealed two dominate strategies for classifying news stories: 1. dictionary based and 2. market alignment. Dictionary approaches use a pre-compiled dictionary which contain n-grams and an associated trading action or classification. The information in the dictionary is used to score information in a group of news stories on a single day. This score is used to issue a: 1. buy, 2. sell or 3. hold trading instruction. An example of a dictionary approach is a system developed by Wuthrich et al [19]. The dictionary which was constructed by Wuthrich contained tuples of words separated with an "AND" Boolean instruction, for example BOND and STRONG. The dictionary was used to analyse news stories published when the financial markets were closed. The number of stories in each category would be counted and sell or buy instruction would be generated for the index. They claim a 21% advantage over and above a trader who would trade by guessing based on an uniform distribution [19]. Another example of a trading system based upon a dictionary based approach was the News Categorization and Trading System *(NewsCATS)* [15]. NewsCATS analysed company press releases, and attempted to predict the company's share price upon information contained in the press release. The dictionary was not published, but the authors state that the dictionary was created by hand. The construction methodology was also not published. The function of the dictionary was to assist NewsCATS to categorise press releases into pre-designated categories. These categories were designed to indicate the influence (positive or negative) of the press release upon the share price. The system's authors claim that they significantly out perform a trader who buys on a random basis after press releases [15].

The market alignment strategies are used to "bootstrap" training data for a classifier. A news story would be labelled: 1. "negative" if it was published at the same time as a fall in a market's value, 2. "positive" if it was published at the same time as a rise in a market's value or 3. "neutral" it was published at the same time as a marginal change in a market's value. The classifier would then classify unlabelled news stories. A trading action would be issued based upon the number of: positive, neutral and negative classification in a given time period. The AEnalyst [13] used market trends to construct language models to recommend news which would effect a share price. Drury et al [9] used market alignment and self-training to construct models to classify news stories into pre-defined categories. This information is used as a basis of a trading strategy.

2 Classification Strategies

The experiments for this compared a trading strategy which used information from quotes from business leaders against trading strategies which used news story classification methods. News story classification strategies classified a either a news story's: 1. headline or 2. story text. There were three competing news story classification techniques: 1. rule selected, 2. market alignment and 3. combination of rule selected and market alignment.

2.1 Quote Classification Strategy

The *quote classification method* for the trading strategy was the one proposed by Drury et al [8]. They proposed a two-step quote classification strategy. The strategy assumed that there was two types of speakers: 1. biased and 2. unbiased.

The *biased group* of speakers were assumed to be *unobjective* when talking about their organization because their job role forced them to be less than truthful [18]. A quote from a member of the biased group can not be taken on "face value", therefore classical sentiment analysis techniques will fail because quotes from this group were overwhelmingly "positive" [8]. Members of this group cannot always behave in this manner because on occasions they are legally compelled to present information about their company in an objective manner, for example profit announcements or warnings. Business leaders who actively lie or mislead about this information commit criminal offences which can lead to long terms of imprisonment [10].

The *unbiased group* consisted of speakers who had job roles where they were compelled to be objective, for example, researchers, traders or analysts. They are compelled to be objective because their job to disseminate objective information to clients who pay fees for this advice. The clients use this information to trade. Biased or untruthful information would impair the client's ability to trade.

The strategy grouped together quotes from members of the biased and unbiased groups, and applied a separate classification strategy to quotes from each group. The classification strategy for quotes from members of the "biased" strategy uses a market alignment strategy to label training data. The labelling procedure labelled quotes into two categories: 1. market moving and 2. non market moving. A quote is labelled as *market moving* if: 1. the share price of the speaker's employer moves significantly or 2. the trading volume of the share of the speaker's employer increases significantly. If the quote did not provoke a price change or trading volume increase it was labelled as "non market moving". The labelling process was manual and therefore time consuming. It was difficult to locate market moving quotes because there were 100 non market moving quotes to 1 market moving quote. A cluster processing was used to increase the number labelled quotes. An initial "seed set" of labelled quotes was clustered with unlabelled quotes. Clusters which contained a single class of labelled quotes had their label propagated to the unlabelled quotes in the cluster. The market moving quotes were labelled into two further categories: 1 negative and 2. positive. The labelling process was a manual process which aligned quotes with share price movements of the speaker's employer. A quote was labelled "positive" if the share price rose whereas a quote would be labelled as "negative" if the share price fell. There were two classifiers induced. One model classified unlabelled quotes into “market moving” and “non market moving”. The second model classified quotes labelled as “non market moving” into “positive” or “negative” categories.

Labelled data for the unbiased group was initially manually selected into: positive and negative categories. The quotes were selected by the affective information in the quote, i.e quotes which contained negative words were labelled as “negative” and quotes which contained positive words were labelled as “positive”. Adjectives, nouns, verbs and adverbs were extracted from the initially labelled data and placed in a dictionary with a label. These terms were expanded with: synonyms and antonyms from Wordnet [11, 14]. The dictionary used linguistic rules to label unlabelled quotes. The linguistic rules labelled a quote “negative” if it contained at least three “negative terms” or positive if it contained at least three “positive terms”. The linguistic rules were based upon Weibe’s high confidence subjectivity classifier [17]. The newly labelled data was used to induce a classifier. Quotes from the unbiased group into “positive” or “negative” categories. A trading signal was issued by summing the quotes from both categories which were labelled as either “positive” or “negative”.

2.2 Rule Selected Data Strategy

The rule selected data was a strategy which relied upon manually constructed linguistic rules to select and label data which was used to train a classifier. The economic literature suggests that events or sentiment reported in the mass-media can provoke a reaction in a market index or share price [4, 12]. The linguistic rules identified and scored event or sentiment phrases in news text. The linguistic rules relied upon dictionaries learnt from the news text. There were dictionaries for: economic actors, verbs, adverbs, adjectives and economic actor properties. Economic actors were either: companies, organizations, market indexes or business leaders. Economic actor properties were nouns associated with economic actors, for example: profits, costs, unemployment, etc. The verb dictionary contained “scored verbs” which linked an economic actor with economic actor properties, for example: “Microsoft (economic actor) profits (economic actor property) *rose* (verb) ”. The sentiment dictionary contained “scored adjectives” which linked an economic actor with economic actor properties.

The linguistic rules modelled event or sentiment phrases as a triple: 1. economic actor, 2. verb and 3. economic actor property for event phrases and 2. 1. economic actor, 2. sentiment and 3. economic actor property for sentiment phrases. Event phrases were assigned a score based on the verb score. The event score was the sum of the verb scores in the event phrase. The verb score could be inverted if an economic actor property in the phrase was labelled as verb modifier. For example, the verb “rise” had a score of “1”, however if “unemployment” was the economic actor property then “rise” was assigned a score of “-1”. Sentiment phrases were scored with the *AVAC algorithm* [2]. The adverbs in the adverbs were used to modify the adjectives scores in the sentiment phrase. The adverbs can: 1. increase sentiment scores, 2. decrease sentiment scores and 3. invert sentiment scores. The rule construction strategy is described in full by Drury and Almeida [6].

The linguistic rules labelled news stories by assigning a score to its headline. A score was calculated for a headline by combining the scores returned by the event and sentiment linguistic rules. A news story was labelled: 1. *negative* if it was assigned a score of less than 0, 2. *positive* if it was assigned a score of greater than 0 and 3. *neutral* if the headline did not contain an entry from the verb or adjective dictionary. The classifier was induced by balancing the number of labelled documents from each category.

2.3 Market Aligned Data Strategy

The market aligned data strategy relied upon rise and falls in the value of a market index to label news stories. The strategy labelled news stories as: 1. "positive" if published on the same day as an increase in the value of a market index, 2. "negative" if published on the same day as a decrease in the value of a market index and "neutral" if published on the same day as a marginal change in the value of a market index. Sharp single day movements were chosen in preference to trends because financial markets can move on non news information and consequently trends can be illusionary [3] whereas single day market movements are more likely to occur is there is an underlying cause. Examples of single movements and their cause are in Table 1.

Table 1 Large Single Day Fluctuations in the FTSE.

Date	FTSE (+/-)	Reason
8th August 2011	-3.39%	Falls in US and Asian Markets
10th/12th Sept 2011	-2.73%	Terrorist Attacks
7th Sept 2008	-1.93%	Financial Crisis

The market alignment strategy evaluated the effect of *market index thresholds* in the labelling process. A market index threshold is the minimum market movement before a news story published on the same day can be labelled: 1. negative, 2. positive or 3. neutral. For example, market threshold of 1% and -1% would infer that stories published on the same day: 1. as an increase in value of a market index more than 1% would be labelled as "positive",2. as a decrease in value of a market index more than 1% would be labelled as "negative" and 3. an increase or decrease in value of a market index between 1% to -1% would be labelled as neutral. The tested market index thresholds were 1% to 9% for "positive" stories and -1% to -9% for "negative" stories.

2.4 Rules with Market Alignment Strategy

This strategy used the rule selected and market alignment strategies to label data. A story would be assigned a label if both strategies agreed on the same label. For example, a story would be labelled "negative" if both the rule selected and the market alignment strategy applied a negative label. The effect of *market index thresholds* was evaluated. The tested market index thresholds were 1% to 9% for "positive" stories and -1% to -9% for "negative" stories.

3 Experiments

The experiments were designed to evaluate the ability of each strategy to estimate the direction of the opening price of the NASDAQ market index when compared to the previous days closing price. This period was chosen because the NASDAQ was closed and the market could not react to the information contained in a quote or news story until the market opens the next day. The competing strategies had access to news stories published the day before until the opening of the market the next day.

The strategiesused a corpus of news stories which were published between October 2008 - April 2011. Headlines, story text, story published date and direct speech (quotations) were extracted from the news stories. The quotations are publicly available in *Minho Quotation*

Resource[1] [5]. The experiments used 300 randomly selected days from which news stories or quotations were used as labelled data. A 100 randomly selected days from from which news stories or quotations were used as testing data. The testing data was drawn from a period of time which was after the latest training day. This selection process was repeated 5 times. The competing strategies had access to the same sets of selected days.

The trading strategies classified a news story or quotation. If a news story or quotation was classified *negative* then it was assigned a score of -1 whereas a *positive* news story or quotation was assigned a score of 1. A neutral story or quotation was assigned a score of 0. A trading action was generated by summing the total news story scores for each day. A *buy* action was issued if the score for a specific day was greater than 0 whereas a *sell* action was issued if the score for a specific day was less than 0. A hold action was issued if the score for a day was 0. The effect of trading thresholds was tested. A trading threshold is the minimum day's score before a trading action could be issued. For example a threshold of 10 ensured that a day's score of 10 or more was required before a buy action was issued, and -10 or less before a sell action was issued. The range tested was from 0 to 90. The experiments also tested the effect of classifier confidence on the results of each strategy. The classifier confidence refers to the minimum confidence required before a classification of a news story or quotation would be accepted by a trading strategy. The tested range was from 0.5 to 1.0. The rules with market alignment and market alignment strategies also evaluated the *market index thresholds* for labelling data. Language models were used as a classifier in all of the experiments.

3.1 Points Difference and Trading Accuracy Evaluation

A measure for evaluation the competing strategies was the mean points difference gained or lost during the trading activities for each trading period. Table 2 holds the average points difference for all tested configurations. A successful strategy produces a positive points difference whereas an unsuccessful strategy produces a negative points difference. The quote strategy was clearly superior when all configurations were considered.

Table 2 Average Trading Profits (All Configurations).

Strategy	Points Difference	Accuracy
Rule Trained (news text)	-0.33±34.41	0.49±0.01
Rule Trained (headlines)	-18.74±54.99	0.47±0.03
Market Alignment (news text)	11.95±83.55	0.44±0.17
Market Alignment (headlines)	11.48±90.32	0.46±0.16
Rules + Market Alignment (news text)	0.57±66.59	0.45±0.16
Rules + Market Alignment (headlines)	-10.89±90.23	0.45±0.15
Quotes	86.98 ±67.08	0.55±0.03

The market based strategies used gains and losses from the NASDAQ to label quotes as either: positive, negative or neutral. The variation of the values used in this labelling strategy had an effect on the mean points difference and trading accuracy. Table3 holds the results for the market based strategies optimal configuration. The quotes strategy returned the highest points differences, but this difference was within the standard deviation of the competing strategies.

[1] Available at: `http://goo.gl/U6quN`

Table 3 Average (Optimal Market Configurations).

Strategy	Positive	Negative	Returns	Accuracy
Align. (text)	4	-3	78.47 ±53.18	0.52 ±0.02
Align. (headlines)	1	-8	75.63 ±68.77	0.55 ±0.05
Rules + Align. (text)	5	-3	51.97 ±49.16	0.53 ±0.02
Rules + Align. (headlines)	4	-3	32.23 ±120.10	0.48 ±0.07
Quotes	NA	NA	86.98 ±67.08	0.54 ±0.03
Rules (text)	NA	NA	-0.33 ±108.24	0.49 ±0.05
Rules (headlines)	NA	NA	-18.74 ±54.96	0.46 ±0.03

The final evaluation was to discover the optimal configurations of decision boundary and classification confidence for the quote and rules strategies, and market values, decision boundary and classification confidence for the market based strategies. Classifier confidence is the minimum confidence value a learner can return for a document before it is used in a trading decision. For example, a trading strategy which uses a classifier confidence of 0.90 will discard all documents classified with a confidence of 0.895 or lower. The decision boundary is the difference between the number of classifications before a trading decision is made. For example, a decision boundary of 90 ensures that the majority class must contain 90 more documents than either of the two minority classes. Table 4 demonstrates the trading accuracies and points differences for the optimal configurations for each strategy. The quote strategy returned the highest points difference and trading accuracy.

Table 4 Trading Returns (Optimal Configurations Confidence / Decision Boundary).

Strategy	Boundary	Confidence	Returns	Accuracy
Align. (text)	0	0.5 - 0.9	110.44 ±52.32	0.53 ±0.01
Align. (headlines)	0	0.5 - 0.9	83.44 ±77.86	0.55 ±0.04
Rules+ Align. (text)	0	0.5	109.86 ±63.24	0.53 ±0.02
Rules+ Align. (headlines)	90	0.5	83.96 ±112.42	0.51 ±0.05
Quotes	0	0.6	233.80 ±60.27	0.58 ±0.04
Rules (headline)	60	0.6	40.76 ±16.20	0.48 ±0.03
Rules (text)	90	1	56.64 ±96.39	0.50 ±0.06

4 Influence of Variables on Quote Strategy Performance

The experimental section described experiments which estimated points differences and trading accuracy for each strategy and for various configurations of the strategy. This section will discuss the configurations of the quote strategy. Visual evidence of the influence of classifier confidence and decision boundary are presented in Figures 1 and 2. The visual evidence suggests that the lower the decision boundary the higher the trading accuracy and points differences. The influence of classifier confidence is unclear. Pearson values were calculated for each of the variables with either points difference or trading accuracy. Pearson values are a value which represents the strength of a correlation between variables. Table5 holds the Pearson scores for each variable. The Person values confirm that there was an identifiable relationship between: 1. decision boundary and trading accuracy and 2. decision boundary and points difference and no identifiable relationship between classifier confidence and trading accuracy or points difference.

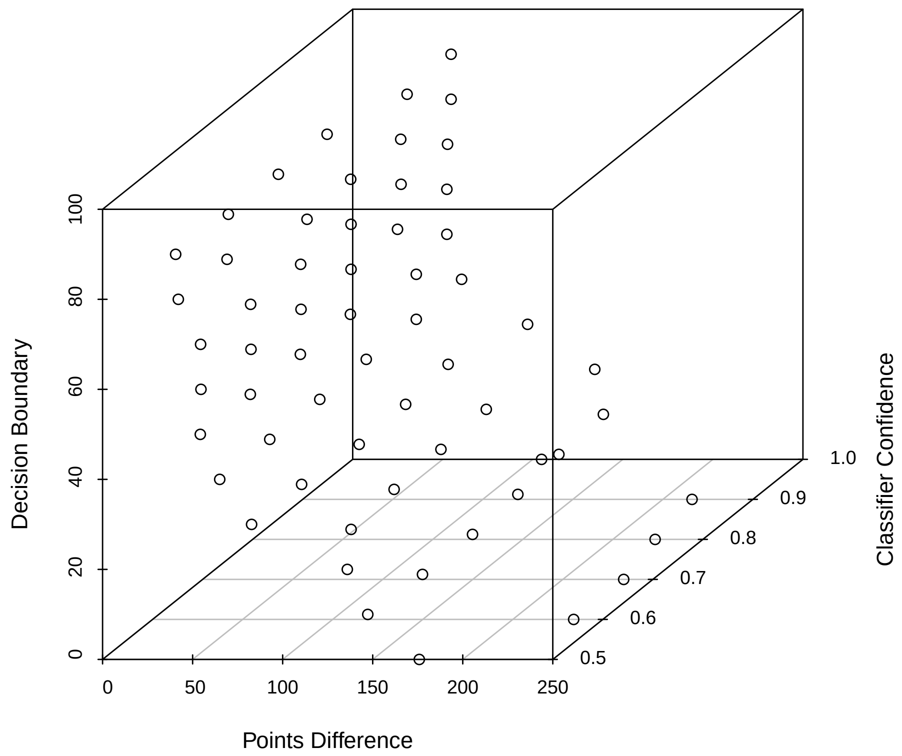

Figure 1 Point difference against classifier confidence and decision boundary.

Table 5 Correlation between variables and trading accuracy and points differences.

Variable 1	Variable 2	Pearson Value
Accuracy	Decision Boundary	-0.31
Trading Profits	Decision Boundary	-0.62
Accuracy	Confidence	0.00
Trading Profits	Confidence	-0.02

The negative correlation between decision boundary and trading accuracy and points difference may indicate that quotations are not equal. A consensus among speakers may not be required because quotes from single *important* speakers may outweigh numerous quotations for less influential people. For example, a single quote from Warren Buffett[2]

[2] Examples of his investment record can be found at `http://goo.gl/AETVm`

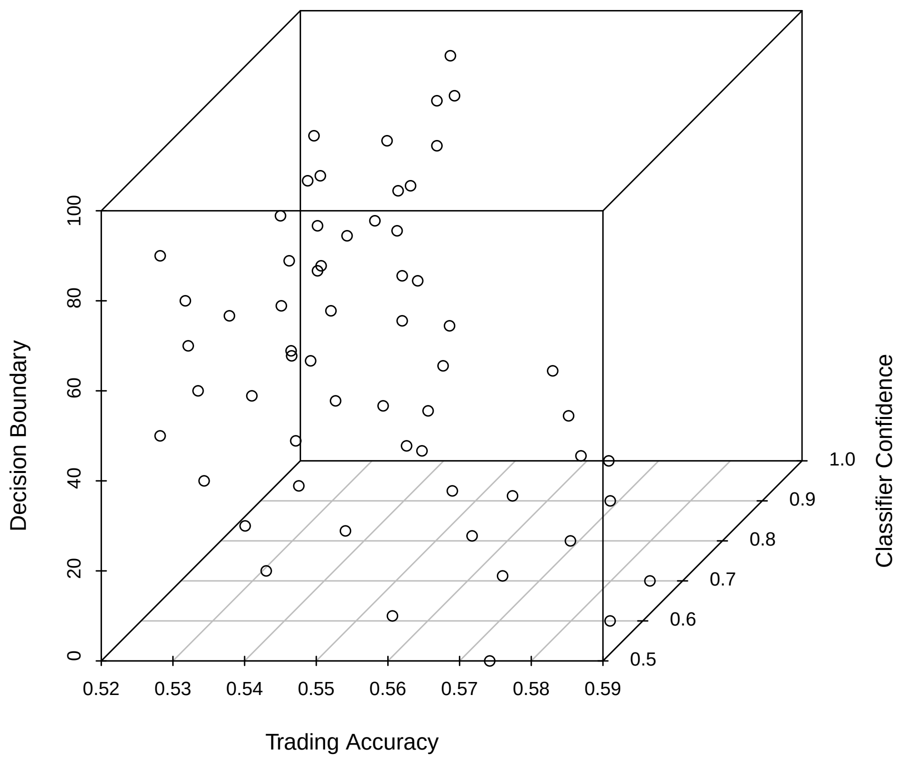

Figure 2 Trading Accuracy against classifier confidence and decision boundary.

can have a strong influence on stock prices, whereas numerous quotes from less well known analysts may have negligible or no effect on stock prices.

5 Conclusion

This paper presented a novel trading strategy which used quotations to estimate the direction of a market index (NASDAQ). It was superior to the competing strategies as it returned the highest points differences in all of the evaluation measures. It returned the highest trading accuracy for two of the three evaluations.

The work presented in this paper did not consider the influence of the speaker and market performance. A strategy which considers the influence of a speaker may achieve better results. It may be possible to identify and score influential speakers with a domain ontology [7]. In addition the results gained by the strategy will have to be tested against the possibility that these results could have been gained by chance. The work presented in this paper demonstrates that trading with *quotations* gains results comparable with traditional rule based and market alignment news classification strategies.

References

1 Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. Sentiment analysis in the news. In Nicoletta Calzolari, editor, *Proceedings of the Seventh conference on International Language Resources and Evaluation*, Valletta, Malta, may 2010.

2 Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato, and V. S. Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

3 Wesley Chan. Stock price reaction to news and no-news. drift and reversal after headlines. *Journal of Financial Economics*, 70(2):223–260, 203.

4 Werner F. M. De Bondt and Thaler R. Does the stock market overreact? *The Journal of Finance*, 40(3):793–805, 1985.

5 Brett Drury and J.J. Almeida. The minho quotation resource. In *LREC*, 2012.

6 Brett Drury and José João Almeida. Identification of fine grained feature based event and sentiment phrases from business news stories. In *WIMS*, page 27, 2011.

7 Brett Drury, Jose Joao Almeida, and M.H. Moreira Morais. Construction and maintenance of a fuzzy temporal ontology from news stories. *International Journal of Metadata, Semantics and Ontologies*, 2012.

8 Brett Drury, Gaël Dias, and Luís Torgo. A contextual classification strategy for polarity analysis of direct quotations from financial news. In *RANLP*, pages 434–440, 2011.

9 Brett Drury, Luís Torgo, and José João Almeida. Classifying news stories with a constrained learning strategy to estimate the direction of a market index. *IJCSA*, 9(1):1–22, 2012.

10 Peter Elkind and Bethany McLean. *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron.* Penguin, 2004.

11 C Fellbaum. *WordNet: An Electronic Lexical Database.* The MIT Press, 1998.

12 Peter Ager Hafez. Construction of market sentiment indices using news sentiment. Technical report, Ravenpack, 2009.

13 Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. Language models for financial news recommendation. In *In Proceedings of the Ninth International Conference on Information and Knowledge Management*, pages 389–396. ACM Press, 2000.

14 Bing Liu. *Web Data Mining: chapter(Opinion Mining).* Springer, 2007.

15 Marc-Andre Mittermayer and Gerhard F. Knolmayer. Newscats: A news categorization and trading system. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 1002–1007. IEEE Computer Society, 2006.

16 Gerald Ratner. *The Rise and Fall... and Rise Again.* Wiley, J, 2007.

17 E. Riloff and J. Weibe. Learning extraction patterns for subjective expressions. In *In Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 105–112, 2003.

18 James Spindler. Why shareholders want their ceos to lie more after dura pharmaceuticals, 2006.

19 Wuthrich, Cho, and Leung. Daily prediction of major stock indices from textual www data. In *4th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, pages 364–368, 1998.