# Classifying News Stories to Estimate the Direction of a Stock Market Index

Brett Drury
LIAAD-INESC
Porto, Portugal
brett.drury@gmail.com

Luis Torgo
Fac. Sciences
LIAAD/INESC
Porto, Portugal
ltorgo@inescporto.pt

J. J. Almeida
University of Minho
Braga, Portugal
jj@di.uminho.pt

*Abstract*—**News can contain information which may provide an indication of the future direction of a share or stock market index. The possibility of predicting future stock market prices has attracted an increasing numbers of industry practitioners and academic researchers to this area of investigation. Popular approaches have relied upon either: models constructed from manually selected training or manually constructed dictionaries. A potential flaw of manually selecting data is that the effectiveness of the trained model is dependent upon the ability of the human annotator. An alternative approach is to align news stories with trends in a specific market. A negative story is inferred if it co-occurs with a market losing value where as positive story is associated with a rise. This approach may have its flaws because news stories may co-occur with market movements by chance and consequently may inhibit the construction of a robust classifier with data gathered with this method.**

**This paper presents a strategy which combines a: rule classifier, alignment strategy and self-training to induce a robust model for classifying news stories. The proposed method is compared with several competing methodologies and is evaluated with: estimated F-Measure and estimated trading returns. In addition the paper provides an evaluation of classifying a news story with it's: headline, description or story text. The results demonstrate a clear advantage for the proposed methodology when evaluated by estimated F-Measure. The proposed strategy also produces the highest trading returns. In addition the paper clearly demonstrates that a news story's headline provides the greatest assistance for classification. The models induced from headlines gained the highest estimated F-Measure and trading returns for each strategy with the exception of the alignment method which performed uniformly poorly.**

***Keywords: constrained learning, finance, news***

## I. INTRODUCTION

The inference of the direction of the price of a stock or share based upon the information contained in news stories has become an increasingly popular area of research for both the academic researcher and the commercial practitioner. News may contain timely information which can assist in the prediction of the prospects of economic actors. Information contained in news typically is concerned with the present or the future which is contrary to numeric data which typically describes the past. The potential of news information has generated a number of approaches which have had varying degrees of success. The published methods for predicting market returns based upon news information can be categorized as one of the following strategies: manually created rules, models learnt from manually selected data and manually constructed dictionaries. The methods have one central flaw that they rely upon a human annotator and therefore effectiveness of the strategies may depend upon the human's ability. This paper describes a strategy which uses a combination of: manual constructed rules with automatically constructed dictionaries, alignment of stories with sharp market movements and self-training to construct a robust model to classify news stories. The rest of the paper will describe the following: related work, data acquisition, news story classification and evaluation.

### *A. Related Work*

The research literature has described a number of prototypes for predicting stock market's reaction to news. The first recorded approach was by the trader Victor Niederhoffer in the early 1970's. The stories from the day's newspapers were organised into 19 separate categories with a sliding polarity scale (positive to negative) [1]. Trends were inferred from the aggregation of the polarity information. This manual approach would have been slow. The advent of machine readable news has allowed a number of systems to automatically classify news stories, eliminating the lag of Niederhoffer's approach. J.D Thomas' Ph.D.[2] dissertation was arguably a descendent of Niederhoffer's strategy. Thomas created hand-crafted rules which assigned a news story to one of 39 categories, although it was noted by Mittermayer[3] that there was no published trading strategy. Wuthrich[4] attempted to classify stories which were published outside of market hours (over-nights). The classification strategy relied upon a manually constructed dictionary which contained 423 features. Mittermayer[3] reported that the dictionary was unpublished. A further system which relied upon a manually constructed dictionary was NewsCats[5] which categorized press releases into three categories: (1. buy, 2. sell or 3. no-recommendation).

An alternative to manually constructed rules and dictionaries is the alignment of news stories to market movements [6]. The alignment methodology infers that a news story is negative if it is published in the same time frame as a negative trend and a news story is positive if it co-occurs with a positive trend. There may be some flaws with this methodology: 1. news stories may be published out of time with market trends, 2. a news story may co-occur with a trend by chance, 3. a news story may contrary information to a market trend, for example a positive story co-occurring with a negative trend, 4. a market trend may be illusionary because the market can move without news information. In addition Lavrenko[6] limited the methodology to stocks in single companies and selected stories where the company name was in the headline. This approach excludes indirect news such as macro-economic information. The Lavrenko[6] alignment methodology may not be suitable to market indexes because indexes can move on non-company information.

### B. Data Acquisition

A large number of news stories (>300,000) were collected from Really Simple Syndication (RSS) feeds during the period from October 2008 until June 2010. The crawler ran at the same time each day. The following information from the RSS feed was stored in a database (RDBMS): headline, description, published date and story text. Inter-day stock price data was collected from Yahoo Finance.

## II. News Story Classification

The aim of this paper is to evaluate the effectiveness of the proposed method to classify newspaper stories. The evaluation process is a comparison of the proposed method and several competing strategies. This section will describe each method in detail.

### A. Model Constructed from Rule Selected Data

The rule classifier was designed to identify and score event and sentiment phrases. The economic literature suggests that markets can react to events [7] and sentiment information [8]. The rule classifier modelled event or sentiment phrases as a triple: Economic Actor (company, organization, market etc), Verb/Adjective, Object (profits, unemployment, etc). The rules were described in Jape [9] which are regular expressions for manipulating annotations in GATE[10]. A full description of the rule classifier is provided by Drury and Almeida [10]. The authors claim that the rule classifier returned: recall of 0.71 and a precision 0.94 for sentiment phrases, and a recall of 0.83 and a precision 0.84 for event phrases when evaluated against a gold standard.

The experiments for this paper required the rule classifier to select stories by scoring the headline of a story. The story is assumed to be negative if the headline was scored with a value of less than 0 where as the story is positive if the headline is scored with a positive value. If the rule classifier abdicates, i.e. it fails to assign a score to a story headline then the story is not assigned a category. The rule categorized data is then used as training data for a classifier. The rule classifier ensures that the story headline is either a predefined: 1. negative or positive event or 2. negative or positive sentiment phrase.

### B. Alignment of Market Data

This strategy selected data by labelling news stories by their co-occurrence with a single market movement. This strategy relied upon single movements in the FTSE-100 market index because it was assumed that if the market moves sharply then this movement will be reflected in the published news stories. A positive day was assumed to be when the market moved by more than 1.7% and a negative day when FTSE-100 lost more than 2.11%. There was a lower threshold for the positive alignment because the time period crawled contained significantly more "negative" days than "positive". The data returned by this strategy is then used to induce a model from a classifier.

### C. Hybrid of Rules and Alignment

This strategy is the first part of the proposed strategy. This strategy attempts to mitigate the flaws of a rule classifier and alignment with a simple voting strategy. A rule classifier applies a label to each news story in the same manner described earlier. The returned data from the rule classifier is then aligned with sharp market movements. Stories which were labelled equally by both strategies were used as training data. Stories which were assigned contradictory labels by the two strategies were ignored. The selected documents were used to train a classifier. The strategy ensured that stories which were contrary to a trend were not included in the training set.

### D. Proposed Algorithm

An initial document training set was returned by the previously described "hybrid of rules and alignment" strategy. Three models were induced from the training set. The models were induced from headline, description and story text information. The three models and the previously described rule classifier selected high confidence candidates from documents not in the initial training set. Each classifier had a veto, therefore if one classifier dissented, then the candidate story was not selected. The selected stories were then added to the training set, new models were induced from the complete training set and a new iteration started. The process was continued until a specific stopping condition is met or no new candidates are selected. This process is commonly referred to as self-training. The algorithm is fully described in Fig 1.

## III. Evaluation

As stated earlier the evaluation methodology was two-fold: 1. an estimated F-Measure and 2. trading returns. This section will describe the estimated F-Measure. The F-Measure was estimated with 2 x 5 cross validation with Lingpipe's Language Model implementation. The F-Measure was estimated for models generated from: headline, description and story text information. The results are presented in Fig: 2.

The estimated F-Measure indicates that models constructed from headlines are the most robust classifiers for each strategy with the exception of alignment whose models performed uniformly poorly. The weakest classifiers were constructed from story text with the exception of alignment data. The proposed strategy returns the highest estimated F-Measure for the: headline classifier, description classifier and story text classifier. Its decline in estimated F-Measure from headline classifier to story text classifier was the shallowest of all of the strategies with the exception of the alignment method. The proposed method gained both precision and recall through the self-training process, for example the initial training set returned an estimated recall and precision of 0.66 for the headline classifier. This was increased to a precision of 0.86 and a recall of 0.83. The proposed method demonstrates a clear advantage over the competing strategies when evaluated by F-measure.

**Input**: UL: A list of unlabelled stories
**Input**: LD: A list of labelled stories
**Input**: Const: A predefined constant for classifier confidence
$rc \leftarrow newRuleClassifier()$
$hc \leftarrow TrainWithHeadlines(hc, LD)$
$dc \leftarrow TrainWithDescriptions(dc, LD)$
$sc \leftarrow TrainWithStoryText(sc, LD)$
$ul \leftarrow ()$
$ld \leftarrow ()$
**forall the** $story \in UL$ **do**
    $ruleC \leftarrow rc.classify(story.headline)$
    $headC \leftarrow hc.classify(story.headline)$
    $descC \leftarrow hc.classify(story.description)$
    $textC \leftarrow hc.classify(story.text)$
    ** Check classification confidence **
    **if** $headC.conf < Const$ $or$ $descC.conf < Const$ $or$ $textC.conf < Const$ **then**
        $next$
    ** Check classification agreement **
    **if** $ruleC = headC \wedge headC = descC \wedge descC = textC$ **then**
        $ld \leftarrow ld.add(story, ruleC)$
    **else**
        $ul \leftarrow ul.add(story)$
    ** Termination, no further candidates **
    **if** $ld.Size() = 0$ **then**
        return $LD$
    $LD \leftarrow mergeDocuments(LD, ld)$
    return $(SelfTrain(ul, LD, Const))$

Fig 1: Proposed Algorithm

| Strategy | Headline | Text | Description |
|---|---|---|---|
| Alignment | 0.57 +/- 0.01 | 0.57 +/- 0.01 | 0.57 +/- 0.00 |
| Hybrid | 0.66 +/- 0.04 | 0.57 +/- 0.06 | 0.58 +/- 0.04 |
| Rules | 0.77+/- 0.01 | 0.60+/- 0.01 | 0.65 +/- 0.01 |
| Proposed | 0.84+/- 0.01 | 0.71+/- 0.01 | 0.77+/- 0.01 |

Fig 2: Estimated F-Measure for competing strategies

## IV. TRADING EVALUATION

This section will describe a secondary evaluation. The evaluation will estimate the percentage returns gained by trading on the FTSE-100 index based upon recommendations from classifiers trained by the competing strategies. It may be possible that the estimated F-Measure for the selected training data may not be an accurate indicator of its effectiveness in estimating market movements. There is a caveat to this form of evaluation. The classifiers will have access to all the news stories published on a given day and consequently the evaluation will measure the effectiveness of the classifiers ability to classify a trading day as negative or positive. A strategy which returned a positive return may not be guaranteed to gain a return in a trading environment. In a trading environment a trader will only have access to news published up to the time of trade where as in this evaluation the strategies will have access to the complete day's news when the market opens.

### A. *Experimental Setup*

The data from 2008 and 2009 was reserved for training, whilst the evaluation data were news stories published in 2010. The evaluation determined that a day was either: negative or positive by counting the number of stories of each category on a given day. If there was at least a difference of 5 between the categories then the recommendation would be the dominant category. If there was no clear difference then the day would be determined to be neutral. The "trading strategy" would be: for a "positive day" to buy at the opening price and sell at the end of the day and 2. "short" on a negative day. Shorting implies selling at the opening price and buying at the closing price. On a neutral day no trading action would be undertaken.

Models were induced from: headline information, description text and story text. The evaluation task evaluated each classifier and in combination. The combination of classifiers, classify stories with a majority vote, therefore only two classifiers were required to agree. There were two evaluation cycles: 1. count all classified stories and 2. count only classified stories with a high confidence (>90%). The competing methods were evaluated with a simple metric: the percentage "profit" made through trading activities. The results are presented in Figs: 3 and 4. The majority of strategies made losses with the exception of the rule trained classifier and the proposed method. The proposed method returned the highest "trading profits" when classifying stories with only headline information, but made losses when using either: story text or description information. The use of “high confidence” classifications increased returns. The rule trained classifier was the proposed method's nearest competitor because it made consistent positive returns. The rule trained classifier was the only classifier in the first evaluation cycle to make a positive return when classifying news stories by description information, but the classifier made negative returns in the second evaluation cycle where “high confidence” classifications were used for "trading".

| Strategy | Voting | Headline | Description | Text |
|---|---|---|---|---|
| Alignment | -12.2% | -24.5% | -20.6% | -0.1% |
| Hybrid | -10.6% | -10.6% | -10.6% | -10.6% |
| Rules` | 16.5% | 16.8% | 14.8% | -1.2% |
| Proposed | -10.6% | 33.8% | -6.5% | -12.3% |

Fig 3: Evaluation for high confidence classifications

| Strategy | Voting | Headline | Description | Text |
|---|---|---|---|---|
| Alignment | -12.4% | 2.2% | -19.6% | -3.8% |
| Hybrid | -10.6% | -11.9% | -19.6% | -10.6% |
| Rules` | 18.6% | 30.0% | -10.6% | -5.9% |
| Proposed | 6% | 47.2% | 0.5% | -10.6% |

Fig 4: Evaluation for high confidence classifications

### B. *Evaluation of News Story Characteristics*

The results presented in Figs: 2, 3 and 4 demonstrate clearly that models induced from headline information construct are the most robust. The models induced from headlines consistently gained the highest F-Measure for each competing strategy as well highest returns or the lowest losses. The models induced from description text consistently performed worse than headline classifiers, but better than models induced from story text. This is reflected in trading returns and estimated F-Measure. Models induced from story text were the weakest both in F-Measure and trading returns.

## V. Conclusion and Future Work

A reoccurring problem for systems which attempt to classify news stories is the lack of training data. A further problem is the correct identification of stories that have market influence. Manual classification of stories or manual construction of dictionaries may be a long and laborious process which may yield insufficient training data or incomplete dictionaries. Alignment of news stories with market movements may assist in the selection of news stories for training data, however this method has its flaws. News stories may co-occur with market movements by chance or the news stories may be contrary to a market trend or movement.

This paper presents a proposed method for categorizing news stories into positive or negative categories. A rule classifier selects stories which have an event or sentiment phrase in its headline. These selected stories are then aligned with market movements. Stories which have the same label assigned by both strategies may limit the possibility of a story co-occurring by chance or contains contrary information to the trend. This method may increase the chance of identifying events which may influence the market. The proposed method adds further documents with a self-training strategy.

The proposed method has a clear advantage over the competing methods when evaluated by estimated F-Measure. The trading evaluation is less clear. The proposed strategy obtains the single highest return, which was the model induced from headlines. The proposed methodology performs less well with: majority voting, description and story text. The rule trained classifier return the highest return (or lowest loss) for majority voting and description, however the alignment strategy returned the best returns for the story text classifier. It is reasonable to suggest that the two evaluation tasks indicate that the proposed method has a demonstrable superiority to the competing methods and provides a basis for further work.

### A. *Future Work*

As stated earlier the trading evaluation task is not a realistic simulation of a trading environment because the competing techniques have access to all of the news published on a single trading day. Traders have access to news published up to the time of their trade. An immediate aim of the authors will be to evaluate the competing techniques with news published when the market is closed (over-nights). The successful prediction of the direction of the opening price will be a more realistic simulation. A secondary aim will be to assign a relevance measure to each news story. The assumption that each classified story has an equal effect is arguably a naive approach because certain stories will be more relevant and therefore have a stronger effect on the market index. For example, UK unemployment figures will have a stronger effect on the FTSE-100 than the NASDAQ (an index located in the U.S.A) because unemployment in the UK is more relevant to the FTSE-100. A third priority will be to utilize news volume. The effect of a story to influence a market may increase with its frequency of repetition in separate news sources. The current work provides an evaluation baseline for news story classification.

## References


[1] Taleb, Nassim Nicholas and Lane, Allen., The Black Swan (The impact of the highly improbable). Random House, 2008.

[2] Thomas, James D. News and Trading Rules. s.l. : CiteSeer, 2003.

[3] Mittermayer, M A and Knolmaye, G F. Text Mining Systems for Market Response to News: A Survey. University Of Bern, 2006.

[4] Wuthrich, B, et al., Daily prediction of major stock indices from textual www data. International conference on Knowledge Discovery and Data Mining, 1998.

[5] Mittermayer, Marc-Andre and Knolmayer, Gerhard F., NewsCATS: A News Categorization and Trading System. ICDM, 2006.

[6] Lavrenko, Victor, et al., Language Models for Financial News Recommendation. ACM Press, 2000.

[7] Bondt, De, Werner, F M and Thaler, Richard., Does the Stock Market Overreact? Journal of Finance, 1985.

[8] Devitt, Ann and Ahmad, Khurshid., Sentiment Polarity Identification in Financial News: A Cohesion-based Approach. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007.

[9] Cunningham, et al., GATE: A framework and graphical development environment for robust NLP tools and applications. Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics, 2002.

[10] Drury, Brett and Almeida, J J., Identification of Fine Grained Feature Based Event and Sentiment Phrases from Business News Stories. ACM, 2011.