Predicting Stock Price Movements Based on Different Categories of News Articles

Yauheniya Shynkevich[1], T.M. McGinnity[1,2], Sonya Coleman[1], Ammar Belatreche[1]
[1]Intelligent Systems Research Centre, Ulster University, Derry, UK
[2]School of Science and Technology, Nottingham Trent University, Nottingham, UK
shynkevich-y@email.ulster.ac.uk, {tm.mcginnity, sa.coleman, a.belatreche}@ulster.ac.uk

Abstract—**Publications of financial news articles impact the decisions made by investors and, therefore, change the market state. It makes them an important source of data for financial predictions. Forecasting models based on information derived from news have been recently developed and researched. However, the advantages of combining different categories of news articles have not been investigated. This research paper studies how the results of financial forecasting can be improved when news articles with different levels of relevance to the target stock are used simultaneously. Integration of information extracted from five categories of news articles partitioned by sectors and industries is performed using the multiple kernel learning technique for predicting price movements. News articles are divided into the five categories of relevance to a targeted stock, its sub industry, industry, group industry and sector while separate kernels are employed to analyze each one. The experimental results show that the simultaneous usage of five news categories improves the prediction performance in comparison with methods based on a lower number of news categories.**

I. Introduction

Stock price movements are driven by publications of financial news. Investors' decisions are made based on the available information considering how a new piece of data will influence the market. News articles incorporate information about a firm's fundamentals, the activities in which a firm is involved and the expectations of other market participants about future price changes [1], [2]. An enormous amount of textual information is provided by real-time trading applications with a tremendously increased broadcasting speed [3]. For over a decade researchers have been working on automated frameworks that analyze large amounts of financial news articles, extract relevant information and use it for financial forecasting [4]. As it has been shown in previous research works [5], the fluctuations of stock prices and the publications of related news articles are strongly related. Existing data mining methods were employed and expanded to study how news items affect the stock price [6], [7]. According to the reviewed literature, a predefined criterion is generally employed by researchers to select news articles relevant to an analyzed stock from a large collection of documents. After relevant articles are selected, the researchers tend to treat every article as having potentially the same impact on stock prices. Thus far, no previous research has been found regarding simultaneous analysis of different categories of news articles divided according to their levels of relevance to the target stock.

This research paper studies how the simultaneous use of financial news articles with different levels of relevance to the target stock can give an advantage in financial news-based forecasting. To achieve this goal, the Global Industry Classification Standard (GICS) is employed for dividing stocks by industries and sectors and for assigning their corresponding news articles to five categories: stock-specific (SS), sub-industry-specific (SIS), industry-specific (IS), group-industry-specific (GIS) and sector-specific (SeS) news items. Every article from a large database is examined based on its relevance to the target stock and included into corresponding news categories. The SS subset of news consists of the articles directly relevant only to the target stock. The SIS, IS, GIS and SeS subsets of data are formed from articles relevant to the stock's sub industry, industry, group industry and sector respectively.

The Multiple Kernel Learning (MKL) technique is often employed for integrating different types of data [8], [9], [10], [11]. It uses several kernels for learning from separate subsets of data. The MKL approach applied in this study utilizes from two to fifteen separate kernels each assigned to either a SS, SIS, IS, GIS or SeS subset of news articles. The experiments show that an attempt to divide articles into different categories, analyze them separately and then combine the resulting predictions into a single decision demonstrates a promisingly improved performance when compared with approaches based on a single subset of news.

The remainder of the paper is organized as follows. Section II provides a review of the related literature and explains the methodology utilized. Section III specifies raw data, outlines data pre-processing and the machine learning approaches used and describes the performance metrics used for analysis. Section IV discusses experimental results. Section V contains conclusions and incentives for future work.

II. Literature Review

A. Representation of News Articles

Trading systems provide a huge amount of textual data to capital market traders. Official announcements, analysts' recommendations, financial journals, discussion boards and news feeds from news wire services are examples of information available to investors [6].

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.107

Wüthrich et al. [12] made the first attempt to use textual information for stock market prediction. A dictionary of terms obtained from a domain expert was utilized for assigning weightings to features and generating probabilistic rules to predict daily price changes of five stock indices. A trading strategy based on the system predictions demonstrated that positive returns can be gained using financial news. Lavrenko et al. [13] developed the Analyst system that included language models, used the price time series and classified the incoming news. The authors showed that profits can be produced using the designed system. Gidofalvi and Elkan [14] created a system for predicting short term price movements. News articles were aligned, scored using linear regression in relation to the NASDAQ index and then assigned with an "up", "down" or "unchanged" label. The author concluded that the stock behavior is strongly correlated with the content of a news article from 20 minutes prior to 20 minutes after the publication. Chan [15] examined monthly returns using headlines about particular companies and found that the publications of bad news cause a strong negative market drift. Kloptchenko et al. [16] focused on official company reports and confirmed their ability to indicate the company's future performance. For example, a change in the written style of a report can indicate a significant change in company productivity. The relationships between companies' average returns and their media coverage were examined by Fang and Peress in [17]. The authors concluded that stocks with high coverage significantly underperformed compared with stocks not featured in the media. Garcia [18] investigated the relationship between sentiments of articles in the New York Times and stock returns and concluded that news content helps to forecast stock returns and investor sentiments have an outstanding effect during recessions. Tetlock [19] examined the interactions between the content of daily articles published by the Wall Street Journal and the stocks. The findings showed that highly pessimistic news cause a decrease in market prices and notably increase trading volume. In [20], Schumaker and Chen investigated the benefits of grouping financial news by similar industries and sectors. The proposed research is similar to [20] in that it utilises articles with different levels of relevance but is different in their usage. Schumaker and Chen examined only one category of news articles at a time and compared the performance of the predictive system based on articles relevant to either the stock itself, its sub industry, industry, group industry or sector, or the whole universal dataset of news. In our proposed predictive system articles from several categories are used simultaneously. To date, no existing research work has involved the division of financial news items into different categories, processing them separately and then integrating multiple predictions made based on these independent news categories into a single prediction decision.

The textual data pre-processing is an essential part of text mining. With respect to financial forecasting based on news, the target of the pre-processing is to extract important information from a given set of news articles that signals a change in a price, and to represent it in a machine learning friendly format. Mittermayer [21] suggested to divide the pre-processing procedure into three preparatory steps: feature extraction, feature selection and feature representation. This terminology was applied in later works [1].

The feature extraction step involves the generation of a list of features that sufficiently describe the documents [21]. Schumaker et al. [6] compared several textual analysis techniques applied to financial articles, including the Bag-of-Words, Noun Phrasing, Named Entities and Proper Nouns approaches. The authors claimed that Proper Nouns showed better results than the others. Hagenau et al. [1] explored the performance of the Bag-of-Words, Noun Phrases, N-Grams (a sequence of N words) [22] and word combinations techniques and found that the word combinations approach significantly outperformed the others. In this study, the Bag-of-Words approach is used for feature extraction, where symbols such as pronouns, articles and prepositions as well as numbers, punctuation marks and stop words are removed from the data. Then word stemming techniques are usually applied to every word. Semantically empty terms are eliminated and the remaining words are utilized to represent the article. This method is often preferred in research studies for its intuitive meaning and simplicity.

When expressive features are selected from the list of all extracted features, those containing the least information are neglected [21]. In some research works, a dictionary containing a list of terms selected by domain experts is utilized [12]. Others employ statistical information about the articles, e.g. term frequency - inverse document frequency (TF*IDF) [9], [21], [23], [24]. Recently, the use of external market feedback was proposed in a number of research works. Wang, Liu and Dou [10] used the Chi-square test to select features for predicting volatility. Hagenau et al. [1] examined the efficiency of the Bi-normal separation and Chi-square test methods for evaluating the explanatory ability of a word. The external market feedback was employed in both methods and showed promising results. In this study, the Chi-square test is selected for feature selection.

In feature representation, the whole set of documents is represented in a machine learning friendly format [1]. For example, a feature vector of *n* elements is formed from each document, where *n* is the number of selected features [21]. The fact that a feature appears in a document is usually regarded as an important factor. In [25], the membership value was computed for each term and binary representation was used for assigning weights in the developed trading system. In other research works, real values are assigned to the weights. Luss and d'Aspremont [23] employed TF*IDF calculations for feature weights when predicting abnormal returns. In [10], the direction of changes in volatility were forecast with TF*IDF values used as weights. Because of their popularity in the reviewed literature, the TF*IDF values are used in this study for computing feature weights for each data point.

After all steps of the pre-processing are completed, the processed articles need to be aligned with the price data and subsequently labelled. The news articles are generally classified into two (positive or negative) or three (positive, negative or neutral) classes. Each class corresponds to the effect of a published news article on an asset price. Rachlin et al. [25] specified five classes to highlight the degree of influence of a news item. In this study, news articles were classified into having positive or negative effect.

B. Predicting from News Articles

Different artificial intelligence approaches are employed for learning from financial textual data for predicting the market reaction; examples include Artificial Neural Networks (ANN) [26]; Naïve Bayes [14] and Support Vector Machines (SVM) [1], [5], [24]. In [2], the impact of financial news articles on Chinese stock markets was investigated where Support Vector Regression (SVR) was used to show that releases of online financial news have negative impacts on the market. The performances of Naïve Bayes, k-Nearest Neighbour (kNN), ANN and SVM classifiers were compared in [3] where an approach for supporting risk and investment management was designed based on the text analysis and machine learning. Taking into account both classification results and computational efficiency, the paper recommends to use the SVM classifier for learning. The Naïve Bayes and SVM techniques were employed by Antweiler and Frank [27] for classifying the messages extracted from Yahoo Finance and Raging Bull websites into bearish, bullish or neutral. SVM slightly outperformed Naïve Bayes in terms of out-of-sample accuracy. Hagenau et al. [1] utilized SVM for classifying the effect of a message on the market price into positive or negative. The authors state that a pilot comparison study showed that SVM performs better than Naïve Bayes and ANN. Considering the previous findings, the SVM method is recognized as the most promising machine learning technique for text classification [1]. SVM employs the principle of structural risk minimization that minimizes the upper limit of the expected risks and construct a robust model to avoid the over-fitting problem. Many machine learning approaches use the empirical risk minimization principle to minimize the error of training that may lead to over-fitting.

Additionally, the ensemble methods are actively used for forecasting financial markets. The results obtained from base learners can be comparatively enhanced using these methods. An ensemble learning algorithm is a computational intelligence approach that integrates a set of base learners into a single model [28]. Recently, the MKL technique was used in financial forecasting to combine several kernels for learning from different features extracted from financial news articles and price data [8], [9], [10], [11]. In [9], information from market prices and financial news articles was integrated using the MKL method. The results state that the MKL approach outperforms models based on simple feature combinations and a single information source. In [10], the MKL model with RBF kernels was suggested for predicting volatility movements. MKL showed higher performance than single kernel methods. The news articles used in both papers [9] and [10] are written in traditional Chinese, hence the developed models were not tested on English news articles. In [8], a stock price prediction system that uses time series price data, numerical dynamics of news and comments and semantic analysis of their content was presented. The model extracts features from these sources of data and forms separate subsets of features, which are then integrated and analyzed by the MKL method. However, the reviewed literature showed no evidence that MKL has previously been employed to analyze different news data categories for financial forecasting.

In the current study, MKL was utilized as the main machine learning approach and SVM and kNN were used for comparison purposes. For this purpose, an implementation of the MKL, SVM and kNN algorithms proposed by Sonnenburg et al. [29] in the SHOGUN toolbox was employed. This toolbox was previously used in several experimental studies [8], [10]. It does concurrent estimation of the parameters and optimal weights by repeating the training procedure used for a simple SVM.

III. Research Design

This section gives details about the design of the news-based predictive system. It explains how news categories were specified, describes raw textual data, and discusses the data pre-processing, machine learning approaches and performance metrics used for evaluation.

A. Industry Classification

GICS aims to support asset management and investment research. It was designed by Standard & Poor's (S&P), a financial services company, and Morgan Stanley Capital International, an independent provider of global indices, products and services. According to the GICS structure, companies are allocated to four categories: sub industry, industry, group industry and sector. Schumaker and Chen [20] utilized GICS to explore the advantages of using financial news articles grouped by similar sectors and industries. In the current study, the GICS classification was used to allocate news articles between five news categories. The categories correspond to the target stock and to all stocks from the target stock's sub industry, industry, group industry and sector. In this study, 16 stocks from the S&P 500 stock market index that belong to the Health Care sector were selected for analysis. Only those stocks that have more than 400 articles published about them during the period of study were chosen.

B. News Articles Data

Financial news articles published about the stocks of interest during a five year period from September 1, 2009, to September 1, 2014, were downloaded from the LexisNexis database. LexisNexis contains news releases from major newspapers, and was used in previous research works [17]. Each article in this database is supplemented with a list of relevant companies and corresponding relevance scores. A relevance score specifies a degree of relevance of an article to a company listed among relevant companies and is expressed as a percentage. It measures the quality of the match between a company and an article, which is based on the keyword frequency, its weight and location within the document. Three providers of news items that showed sufficient press coverage of stocks constituting the S&P 500 market index were selected: McClatchy-Tribune Business News, PR Newswire and Business Wire. The Health Care sector includes 53 stocks from the S&P 500 index. News articles relevant to each of these stocks published during the period of study were downloaded from the LexisNexis database. The total number of the collected articles is 51,435. For every article, the following information was stored: month, day, year, heading, body, a list of relevant companies, a list of relevant tickers and a list of corresponding relevance scores. For every article, a corresponding date of publication is constructed from the day, month and year. The heading and body were combined into a pool of words used as the textual data for information extraction. In order to define the article's degree of relevance to the companies, lists of tickers and corresponding relevance scores were examined. An automated procedure for forming a set of articles relevant to the target company was employed. It checked every article for its relevance to the target company. The article was selected only if the target company's ticker was among relevant tickers to the article and a corresponding relevance score was greater than or equal to 85%. When forming a set of articles related to the sub industry, industry, group industry or sector, every article is examined to determine whether at least one ticker of a company from the sub industry, industry, group industry or sector of interest, respectively, is present among the article's tickers, and then whether its corresponding relevance score is higher than or equal to 85%. The article was selected for further analysis only if these conditions were satisfied. As a result, five subsets of news items corresponding to SS, SIS, IS, GIS and SeS were formed.

After the formation of the subset of articles was completed, articles published on the same date were checked for originality. All unique articles published on the same date were concatenated. This step was necessary to eliminate repetition because some articles were downloaded several times for different stocks or news sources. This procedure was carried out separately through every textual data subsets. In the forecasting system, predictions were made only for dates following the dates when at least one article was published about the target stock. A number of data points was defined at that stage for every stock. A single data instance corresponds to a date when an article, specific to a target stock, was published. When forming other data subsets, articles relevant to other news categories were collected only for those dates.

C. *Historical prices data*

Historical prices were used for selecting the most expressive features from news articles and for labelling the data instances. Time series of prices were downloaded from Yahoo! Finance, a publicly available website [30]. The features were selected based on the market feedback defined as a stock price move on the next trading day following the day of publication. The price move was defined as the difference between the open and close stock prices on the next trading day. A two class classification problem was considered in this research work. Labels 'Up' or 'Down' that corresponded to an increase or decrease in the target stock price, respectively, were assigned to each data instance. Daily data were used in the analysis due to the lack of intraday data. However, it is worth noting that daily price observations were utilized in a number of previous research works on financial forecasting from textual data [12], [27], that demonstrated that the market reacts slowly to new pieces of information and this reaction can be captured and explored using daily data. Table I gives details on the stocks used, the total number of data points and the fractions of each class.

D. *Textual data pre-processing*

When designing a news-based predictive system, the textual data pre-processing is particularly important. According to the Bag-of-Words approach that was used for feature extraction, every article was prepared as follows. At the beginning, numbers, websites' addresses, emails, hyperlinks, punctuation, and symbols other than letters were filtered out. Next, capital letters were converted to lowercase, words consisting of one or two characters and stop words were removed. Then stems were extracted from each word using Porter's stemming algorithm [31]. Finally, a list of unique features extracted from the dataset was formed and features occurred in two or less articles were eliminated.

TABLE I. DESCRIPTION OF ANALYZED STOCKS AND THEIR DATASETS

Sector	Group Industry	Industry	Sub Industry	Ticker	Company Name	# data points	'Up' labelled data points, %	'Down' labelled data points, %
Health Care	Health Care Equipment & Services	Health Care Equipment & Supplies	Health Care Equipment	MDT	Medtronic plc	715	53.93%	46.07%
				A	Agilent Technologies Inc	691	55.81%	44.19%
				ABT	Abbott Laboratories	569	49.30%	50.70%
				BSX	Boston Scientific Corporation	542	57.78%	42.22%
				JNJ	Johnson & Johnson	508	49.61%	50.39%
				BAX	Baxter International Inc	463	45.22%	54.78%
				PKI	PerkinElmer Inc	451	52.68%	47.32%
				COV	Covidien plc	440	51.82%	48.18%
				STJ	St. Jude Medical Inc	416	47.12%	52.88%
		Health Care Providers & Services	Health Care Distributors	BMY	Bristol-Myers Squibb Company	647	54.66%	45.34%
				MCK	McKesson Corporation	520	53.85%	46.15%
			Managed Health Care	AET	Aetna Inc	844	52.61%	47.39%
				CI	Cigna Corp	812	57.14%	42.86%
				UNH	UnitedHealth Group Inc	598	51.68%	48.32%
				HUM	Humana Inc	486	56.20%	43.80%
				WLP	WellPoint, Inc	480	55.00%	45.00%

The scores were computed for each unique feature using the Chi-square method based on the market feedback as a sum of normalized deviations [1]:

$$\chi^2 = \sum_{j=1}^{4} \left(O_{ij} - E_{ij}\right)^2 \Big/ E_{ij}, \qquad (1)$$

where i corresponds to the feature order, O_{ij} and E_{ij} are its observed and expected frequencies within the set of news, and j indicates four possible events: the feature occurred within positive news, $j=1$; it occurred within negative news, $j=2$; it did not occur within positive news, $j=3$; it did not occur within negative news, $j=4$. Thus, the observed frequency of the feature occurring in positive news is defined as a fraction of positive news where this feature appears. The observed frequencies of the feature occurring in negative news and not occurring in positive or negative news are defined in a similar way. If a feature does not imply any positive or negative meaning, it should occur uniformly among positive and negative news. Thereby, the expected frequency of the feature occurring in positive or negative news is the general frequency of the feature occurring in all documents. Analogously, the expected frequency of the feature not occurring in positive or negative news is the general frequency of the feature not occurring within the whole set of news. Therefore, when a feature occurs uniformly among positive and negative news articles, it has a zero Chi-square value. If a feature appears in positive articles significantly more often than in negative articles or vice versa, its Chi-square value is higher than zero. Once the Chi-square scores are calculated for every feature, a list of unique features is sorted in descending order of the scores and 500 features having the highest Chi-square values are chosen as an input. The use of 500 features is considered to be a sufficient representation of news articles. In [1], 567 features were selected for the Bag-of-Words approach.

As a final preliminary step, the dataset of articles is converted to a format appropriate for machine learning. Here, each news article is transformed into a feature vector of 500 elements [21] where TF*IDF values are computed to represent each feature. When a feature does not occur in an article, its TF*IDF value equals zero. So, the result is a sparse matrix of size 500*[number of data points] where each value is equal to the corresponding TF*IDF value. It is worth noting that the SS, SIS, IS, GIS and SeS subsets of news are processed separately through the above described procedure, and different lists of unique features and therefore different feature matrices are constructed for each subset. After textual pre-processing is completed, labels 'Up' or 'Down' are assigned to data points. Each data point contains a label and 500 feature values for each of five subsets.

E. Machine learning techniques

In MKL, the resulting kernel is a linear combination of several sub-kernels:

$$K_{comb}\left(x,y\right) = \sum_{j=1}^{K} \beta_j K_j\left(x,y\right), \quad \beta_j \geq 0, \ \sum_{j=1}^{K} \beta_j = 1. \qquad (2)$$

where $K_{comb}(x,y)$ is the combined kernel, K is a number of kernels, and β_j are weights learnt for each sub-kernel $K_j(x,y)$. MKL allows the assignment of a separate kernel to each category of news articles. In this study the MKL technique with different combinations of linear, polynomial and Gaussian kernels was utilized. Five news categories, SS, SIS, IS, GIS and SeS, were used where separate kernels were utilized for learning from different categories. In order to determine a combination of news categories that leads to the highest prediction performance, a number of combinations were considered. Firstly, each subset of news articles was used independently to predict price movements. In this case, only one kernel of either linear, polynomial or Gaussian type was required, hence SVM was utilized. The kNN approach was also employed to learn from each subset for comparison. Secondly, combinations of subsets were used for prediction starting from a combination of the SS and SIS subsets with different types of kernels. After that the subsets of broader news categories, IS, GIS and SeS, were added consecutively. When a certain kernel type was utilized, a separate kernel of this type was employed to learn from each subset of news. The most complex combination included subsets for all five news categories with three different types of kernels assigned to each subset. Weights learnt for every kernel indicate its usefulness and contribution towards making the final decision.

This procedure below was followed separately for every considered stock. The dataset was divided into training, validation and testing datasets in a chronological order. First 50% of the data points were used for training the system. The subsequent 25% of the data points were used for validation, a phase required to tune model parameters. The C parameter, defined as a penalty rate for misclassification used within MKL and SVM, is required to be tuned. Furthermore, when Gaussian and polynomial kernels were utilized, the Gaussian kernel width and the degree of the polynomial were also tuned during the validation. A grid search was employed to identify good parameter combinations where values of C and gamma were selected from exponentially growing sequences $C=\{2\cdot 10^{-3}, 2\cdot 10^{-2}, \ldots, 2\cdot 10^{7}\}$ and $\gamma=\{2\cdot 10^{-15}, 2\cdot 10^{-13}, \ldots, 2\cdot 10^{-1}\}$ respectively as suggested in [32]. For kNN, an optimal number of neighbors was found during the validation using the grid search. The range of values was selected based on an empirical rule of thumb proposed in [33] where the choice of k is set equal to the square root of the number of training instances. In this study, the number of training data points varies from 208 to 422. A slightly broader range of $k=\{5,6,\ldots,30\}$ was utilized to select the optimal number of nearest neighbors. The remaining 25% of the data instances were utilized to test the developed predictive system. The accuracy was used to measure the model performance with different settings of parameters during the validation phase.

F. Performance Measures

For each of 16 stocks, the predictive performance of the machine learning techniques was measured using the prediction accuracy and return per trade. Prediction accuracy is a measure commonly used in pattern recognition to characterize the classification performance of a machine learning technique [30]. Return per trade is used for evaluating the performance of a trading system. It is important to determine the price direction, which is described by the accuracy. However, correctly identifying large price movements is more valuable than identifying small movements. Mistakes in predicting movements with almost zero return have little lasting effect on the overall performance of the trading system. To study the developed predictive system from a trading point of view, it was evaluated as a trading system using simulated trading. Each time the system predicted an increase in the price on the following trading day (an 'Up' movement) based on news articles published on the current day, it was regarded as a buy signal and an amount of money X was invested in the considered stock on the following day at the opening price. At the end of the trading day the stock was sold. The return per trade was computed as:

$$R_t = (C_t - O_t)/O_t \tag{3}$$

where O_t and C_t are the open and close price on the trading day following the publication day of news articles, R_t is the return from a trade. When a 'Down' price movement was predicted, it was considered as a sell signal. In this case, an amount X of the stock was short sold at the open price on the next trading day and bought back at the end of that day, and the return per trade was computed as:

$$R_t = (O_t - C_t)/O_t \tag{4}$$

IV. Experimental Results

This section focuses on the experimental results produced by the developed news-based prediction system. Both accuracy and return presented in the tables below were averaged over 16 analyzed stocks.

A. The SVM and kNN approaches

This subsection analyzes the performance across all levels of the GISC classification. Table II describes the prediction results obtained using the SVM with different types of kernels and kNN machine learning techniques applied to either SS, SIS, IS, GIS or SeS datasets. This study is conducted in a similar way to [20] where the system was trained on the different GICS classification levels but the universal set of news articles was not taken into consideration in this study. The highest accuracy and return obtained for every subset were highlighted using bold font in the table. SVM outperforms kNN for all stocks and datasets in terms of both accuracy and return. Comparing the use of different kernels, SVM based on a polynomial kernel generally performs slightly better than that based on Gaussian and linear kernels in terms of both accuracy and return, however all three types of kernels display comparatively good results. It is important to highlight that the accuracy of predictions increases when a broader range of news articles is considered. The highest metrics obtained among all data subsets are underlined in Table II and correspond to the group industry data. This indicates that more important information can be extracted from the news articles relevant to the whole group industry rather than from only news articles specific to the stock and its industry. However, considering a much broader range of the articles relevant to the whole sector slightly decreases the overall performance. In [20], the measures peaked for the sector-based model and then steadily decreased when the level of generalization was reduced. This behavior is similar to the behavior observed in the current study but the level of GISC classification that produced the highest results is different. The difference is likely to be caused by the way the experiments were conducted, for instance, the usage of different of datasets.

B. The MKL approach

Table III represents the experimental results obtained using the MKL approach using different sets of data and different combinations of kernels. The highest accuracy and return obtained for each data subset are highlighted in bold. The first section of Table III presents the results where only SS and SIS data were included in the analysis. For treating both data subsets equally, different types of kernels were taken in pairs so that each subset is analyzed by the same set of kernels. The following combinations of kernels were considered: two Gaussian, two linear, two polynomial, a combination of two Gaussian and two linear, a combination of two Gaussian and two polynomial, a combination of two linear and two polynomial, and finally a combination of two Gaussian, two linear and two polynomial kernels, respectively. The highest prediction accuracy (74.76%) was achieved when all types of kernels were employed for learning. This value is higher than the highest accuracies obtained using SVM for the SS and SIS subsets, 68.86% and 73.34% respectively. A number of kernel combinations returned the same maximum return of 0.39%. This results are consistent with [34] and confirm that the MKL method that simultaneously analyzes the SS and SIS subsets produces higher results than SVM and kNN that considers only one type of data at a time.

The second column of Table III displays results of the simultaneous analysis of the SS, SIS and IS data. The combinations of kernels were structured as in the previous section, but three kernels of each type were taken instead of two. The highest accuracy (77.36%) and return per trade (0.45%) were achieved using polynomial kernels. The same accuracy and return were achieved for a combination of polynomial and linear kernels, but zero weights were assigned to linear kernels for all 16 stocks, which indicated that linear kernels did not contribute to the overall prediction decision. As discussed in [34], the selection of the parameter *C* is the most likely reason why linear kernels receive zero weights when combined with Gaussian and/or polynomial kernels. The optimal value of this parameter for Gaussian and polynomial kernels typically lies in a range (1:200), whereas this value for linear kernels typically lies in a range (2000:2000000). The difference in optimal values is likely to cause zero weights to be learnt for linear kernels when it is combined with Gaussian and polynomial kernels. Both accuracy and return achieved using MKL based on three subsets are higher than those of MKL based on two subsets, and than SVM and kNN based on any single subset. These results show that the inclusion of the industry related news improves the prediction performance.

In the third column of Table III, when four categories of news articles are used in MKL, the highest achieved accuracy (79.14%) and return (0.48%) were again obtained using polynomial kernels. As in the previous section, the highest results were also obtained for a combination of linear and polynomial kernels, but zero weights were assigned to the linear kernels. It indicates that adding group industry news improves the performance of the MKL approach.

TABLE II. EXPERIMENTAL RESULTS OBTAINED FOR THE SVM AND KNN APPROACHES

Machine Learning Technique	Data subset (total period of study is from September 1, 2009, to September 1, 2014)									
	Stock-specific		Sub-industry-specific		Industry-specific		Group-industry-specific		Sector-specific	
	Accuracy	Return	Accuracy	Return	Accuracy	Return	Accuracy	Return	Accuracy	Return
SVM, Gaussian	**68.86%**	**0.39%**	72.83%	0.25%	74.36%	0.39%	**76.45%**	0.39%	74.72%	0.39%
SVM, Linear	67.61%	0.23%	72.57%	0.33%	74.12%	0.39%	76.04%	**0.43%**	74.79%	0.38%
SVM, Polynomial	67.31%	0.30%	**73.34%**	**0.37%**	**74.82%**	**0.40%**	75.81%	**0.43%**	**75.07%**	**0.40%**
kNN	54.97%	0.09%	57.97%	0.14%	57.65%	0.14%	57.28%	0.15%	57.07%	0.12%

TABLE III. EXPERIMENTAL RESULTS FOR THE MKL APPROACH

Data subsets (total period of study is from September 1, 2009, to September 1, 2014)											
SS and SIS data			SS, SIS and IS data			SS, SIS, IS and GIS data			SS, SIS, IS, GIS and SeS data		
Kernels	Accuracy	Return	Kernels	Accuracy	Return	Kernels	Accuracy	Return	Kernels	Accuracy	Return
2 Gaussian	64.74%	0.23%	3 Gaussian	63.02%	0.22%	4 Gaussian	66.54%	0.29%	5 Gaussian	62.78%	0.25%
2 linear	70.58%	**0.39%**	3 linear	72.29%	0.36%	4 linear	72.90%	0.37%	5 linear	71.53%	0.35%
2 polynomial	74.34%	**0.39%**	3 polynomial	**77.36%**	**0.45%**	4 polynomial	**79.14%**	**0.48%**	5 polynomial	80.24%	0.49%
2 Gaussian & 2 linear	65.75%	0.25%	3 Gaussian & 3 linear	65.23%	0.25%	4 Gaussian &4 linear	67.95%	0.30%	5 Gaussian & 5 linear	64.89%	0.28%
2 Gaussian & 2 polynomial	74.59%	**0.39%**	3 Gaussian & 3 polynomial	76.27%	0.41%	4 Gaussian & 4 polynomial	78.50%	0.44%	5 Gaussian & 5 polynomial	**82.40%**	**0.53%**
2 linear & 2 polynomial	74.31%	**0.39%**	3 linear & 3 polynomial	**77.36%**	**0.45%**	4 linear & 4 polynomial	**79.14%**	**0.48%**	5 linear & 5 polynomial	80.30%	0.49%
2 Gaussian, 2 linear & 2 polynomial	**74.76%**	**0.39%**	3 Gaussian, 3 linear & 3 polynomial	76.27%	0.41%	4 Gaussian, 4 linear & 4 polynomial	78.50%	0.44%	5 Gaussian, 5 linear & 5 polynomial	**82.40%**	**0.53%**

TABLE IV. WEIGHTS ASSIGNED TO DIFFERENT KERNELS AND DATA SUBSETS WHEN THE HIGHEST PERFORMANCE IS ACHIEVED

Data subset	Kernel Type		Total
	Gaussian	Polynomial	
Stock-specific	14.62%	11.96%	26.58%
Sub-industry-specific	6.12%	9.70%	15.82%
Industry-specific	6.54%	8.43%	14.98%
Group-industry-specific	4.27%	12.33%	16.60%
Sector-specific	3.58%	22.45%	26.03%
Total	35.13%	64.87%	100.00%

Finally, all five categories of news articles were utilized for price movements' prediction. The last column of Table III presents the obtained results. The highest prediction accuracy (82.40%) and return (0.53%) were obtained using a combination of Gaussian and polynomial kernels, and a combination of all three types of kernels with zero weights for the linear kernels. The simultaneous analysis of five categories of news articles provides the highest results among all considered combinations which highlights the importance of inclusion of different types of new articles specifying separate sub-learners explicitly for each category of data. Table IV represents the weights assigned to the Gaussian and polynomial kernels when the highest accuracy and return were achieved. Polynomial kernels received higher weights than Gaussian. Stock-specific and sector-specific news obtained high weights of approximately 26% in total whereas weights for other categories of news equal approximately 15%. The weights indicate that every category of news articles contributed when the final prediction decision was made while stock-specific and sector-specific news provide more useful information for prediction. The results in Tables III and IV highlight the importance of distinguishing between different categories of news, treating them separately when pre-processing the information, and integrating the extracted and learnt information at the later stage instead of combining everything in the early stage and expecting the machine learning method to learn from mixed data.

V. CONCLUSION AND FUTURE WORK

This research study explores whether the simultaneous usage of different financial news categories can provide an advantage in financial prediction system based on news. Five categories of news articles were considered: news relevant to a target stock and news relevant to its sub industry, industry, group industry and sector. Each category of news articles was pre-processed independently and five different subsets of data were constructed. The MKL approach was used for learning from different news categories; independent kernels were employed to learn from each subset. A number of different types of kernels and kernel combinations were used. The findings have shown that the highest prediction accuracy and return per trade were achieved for MKL when all five categories of news were utilized with two separate kernels of the polynomial and Gaussian types used for each news category. The highest kernel weights were assigned to the polynomial kernels indicating that this kernel type contributes the most to the final decision. The SVM and kNN methods based on a single category of news, either SS, SIS, IS, GIS or SeS, demonstrated worse performance than MKL. These results indicate that dividing news items into different categories based on their relevance to the target stock and using separate kernels for learning from these categories allows the system to learn and utilize more information about the future price behavior which gives an advantage for more accurate predictions. The achieved results are promising and can be enhanced further. The introduction of additional data sources such as historical prices can potentially improve the performance. Other possible directions of future work involve applying the proposed approach to more volatile sectors and stock indices such as Nasdaq 100 and testing the algorithm's performance using the in-crisis and out of crisis data.

References

[1] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems*, vol. 55, no. 3, pp. 685–697, 2013.

[2] X. Zhao, J. Yang, L. Zhao, and Q. Li, "The impact of news on stock market: Quantifying the content of internet-based financial news," in *Proc. of the 11th Intl DSI & 16th APDSI Joint meeting*, 2011, pp. 12–16.

[3] S. S. Groth and J. Muntermann, "An intraday market risk management approach based on textual analysis," *Decision Support Systems*, vol. 50, no. 4, pp. 680–691, 2011.

[4] G. Mitra and L. Mitra, *The handbook of news analytics in finance*. Wiley-Finance, 2011.

[5] G. Fung, J. Yu, and H. Lu, "The predicting power of textual information on financial markets," *IEEE Intelligent Informatics Bulletin*, vol. 5, no. 1, 2005.

[6] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, 2009.

[7] M. Mittermayer and G. Knolmayer, "Text mining systems for market response to news: A survey," vol. 41, no. 184. University of Bern, 2006.

[8] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction," in *Proc. of the IEEE 9th Intl Conference on Dependable, Autonomic and Secure Computing*, 2011, pp. 800–807.

[9] X. Li, C. Wang, J. Dong, and F. Wang, "Improving stock market prediction by integrating both market news and stock prices," *Database and Expert Systems Applications, Lecture Notes in Computer Science*, vol. 6861, pp. 279–293, 2011.

[10] F. Wang, L. Liu, and C. Dou, "Stock Market Volatility Prediction: A Service-Oriented Multi-kernel Learning Approach," in *Proc. of IEEE 9th Intl Conference on Services Computing*, 2012, vol. d, pp. 49–56.

[11] C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2177–2186, 2011.

[12] B. Wiithrich, D. Permunetilleke, and S. Leung, "Daily prediction of major stock indices from textual www data," in *Proc. of the 4th Intl Conference on Knowledge Discovery and Data Mining*, 1998.

[13] V. Lavrenko, M. Schmill, and D. Lawrie, "Mining of concurrent text and time series," in *Proc. of the 6h ACM Intl Conference on Knowledge Discovery and Data Mining*, 2000.

[14] G. Gidófalvi and C. Elkan, "Using news articles to predict stock price movements," *Department of Computer Science and Engineering, University of California*. 2001.

[15] W. S. Chan, "Stock price reaction to news and no-news: drift and reversal after headlines," *Journal of Financial Economics*, vol. 70, no. 2, pp. 223–260, 2003.

[16] A. Kloptchenko, T. Eklund, B. Back, J. Karlsson, H. Vanharanta, and A. Visa, "Combining data and text mining techniques for analysing financial reports," *Intl Journal of Intelligent Systems in Accounting and Finance Management*, vol. 12, no. 1, pp. 29 – 41, 2004.

[17] L. Fang and J. Peress, "Media Coverage and the Cross-section of Stock Returns," *The Journal of Finance*, vol. LXIV, no. 5, pp. 2023–2052, 2009.

[18] D. Garcia, "Sentiment during recessions," *The Journal of Finance*, vol. 68, no. 3, pp. 1267–1300, 2013.

[19] P. C. Tetlock, "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[20] R. Schumaker and H. Chen, "A quantitative stock prediction system based on financial news," *Information Processing & Management*, vol. 45, no. 5, pp. 571–583, Sep. 2009.

[21] M. Mittermayer, "Forecasting intraday stock price trends with text mining techniques," in *Proc. of the 37th Annual Hawaii Intl Conference on System Sciences*, 2004, pp. 1–10.

[22] M. Butler and V. Kešelj, "Financial Forecasting using Character N-Gram Analysis and Readability Scores of Annual Reports," in *Proc. of the 22nd Canadian Conference on Artificial Intelligence: Advances in Artificial Intelligence*, 2009, pp. 39–51.

[23] R. Luss and A. D'Aspremont, "Predicting abnormal returns from news using text classification," *Quantitative Finance*, vol. 15, no. 6, pp. 999–1012, 2015.

[24] Y. Zhai, A. Hsu, and S. Halgamuge, "Combining news and technical indicators in daily stock price trends prediction," *Advances in Neural Networks*, vol. 4493, pp. 1087–1096, 2007.

[25] G. Rachlin, M. Last, D. Alberg, and A. Kandel, "Admiral: A data mining based financial trading system," in *Proc. of the IEEE Symposium on Computational Intelligence and Data Mining*, 2007, no. Cidm, pp. 720–725.

[26] S. Simon and A. Raoot, "Accuracy Driven Artificial Neural Networks in Stock Market Prediction," *International Journal on Soft Computing*, vol. 3, no. 2, pp. 35–44, 2012.

[27] W. Antweiler and M. Frank, "Is all that talk just noise? The information content of internet stock message boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.

[28] C. Cheng, W. Xu, and J. Wang, "A Comparison of Ensemble Methods in Financial Market Prediction," in *Proc. of the 5th Intl Joint Conference on Computational Sciences and Optimization*, 2012, pp. 755–759.

[29] S. Sonnenburg and G. Rätsch, "The SHOGUN machine learning toolbox," *Journal of Machine Learning Research*, vol. 11, pp. 1799–1802, 2010.

[30] Y. Shynkevich, T. M. McGinnity, S. Coleman, Y. Li, and A. Belatreche, "Forecasting stock price directional movements using technical indicators: investigating window size effects on one-step-ahead forecasting," in *Proc. of the IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, 2014, pp. 341–348.

[31] M. Porter, *An algorithm for suffix stripping*. 1980, pp. 313–316.

[32] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," *Department of Computer Science, National Taiwan University*, 2010.

[33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd Editio. Wiley-Interscience, 2000.

[34] Y. Shynkevich, T. M. McGinnity, S. Coleman, and A. Belatreche, "Stock Price Prediction based on Stock-Specific and Sub-Industry-Specific News Articles," in *Proc. of the IEEE Intl Joint Conference on Neural Networks*, 2015.