Decision Support Systems xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning

Yauheniya Shynkevich [a,*], T.M. McGinnity [a,b], Sonya A. Coleman [a], Ammar Belatreche [a]

[a] Intelligent Systems Research Centre, Ulster University, BT48 7JL Derry, UK
[b] School of Science and Technology, Nottingham Trent University, Nottingham, UK

ARTICLE INFO

Article history:
Received 2 September 2015
Received in revised form 15 February 2016
Accepted 1 March 2016
Available online xxxx

Keywords:
Stock price prediction
Financial news
Text mining
Multiple kernel learning
Decision support systems

ABSTRACT

The market state changes when a new piece of information arrives. It affects decisions made by investors and is considered to be an important data source that can be used for financial forecasting. Recently information derived from news articles has become a part of financial predictive systems. The usage of news articles and their forecasting potential have been extensively researched. However, so far no attempts have been made to utilise different categories of news articles simultaneously. This paper studies how the concurrent, and appropriately weighted, usage of news articles, having different degrees of relevance to the target stock, can improve the performance of financial forecasting and support the decision-making process of investors and traders. Stock price movements are predicted using the multiple kernel learning technique which integrates information extracted from multiple news categories while separate kernels are utilised to analyse each category. News articles are partitioned according to their relevance to the target stock, its sub-industry, industry, group industry and sector. The experiments are run on stocks from the Health Care sector and show that increasing the number of relevant news categories used as data sources for financial forecasting improves the performance of the predictive system in comparison with approaches based on a lower number of categories.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Investors make investment decisions based on the information available to market participants. News articles bring new information to the market. They contain news about a company, the activities in which it is involved, its fundamentals and what is expected by market participants about its future price changes [1,2]: stock prices are driven by these publications. With the development of the internet, finance-related websites and applications constantly provide a large amount of textual data containing new information. A system capable of efficiently utilising this new data to predict future changes in prices is required to support the decision making of investors and traders. Researchers have been studying the influence of news articles and developed several automated frameworks that consider large amounts of financial news. These frameworks extract relevant information and employ it to forecast prices and their changes [3]. As has been shown in previous research [4], there is a strong relationship between stock prices fluctuations and publications of relevant news. The effect that news items have on stock prices has been studied using existing data mining techniques [5–7]. According to the related literature, researchers usually employ a predefined criterion for selecting news articles from a large collection of textual information. Generally, only news articles highly relevant to an analysed stock are selected. After that, equal importance is given to all articles so that every article is treated as impacting the stock price to the same extent. So far no previous studies employ articles that are divided into different news categories and analysed simultaneously yet differently based on their relevance to the analysed stock, which is the focus of this paper.

This paper investigates whether financial news articles that have different degrees of relevance to the target stock can provide an advantage in financial news-based forecasting when used simultaneously and appropriately. Toward this end, the considered stocks are assigned to the corresponding sub-industries, industries, group industries and sectors according to the Global Industry Classification Standard (GICS) as in [8]. Then news published about these stocks is allocated to different news categories. We consider five news categories; these are stock-specific (SS), sub-industry-specific (SIS), industry-specific (IS), group-industry-specific (GIS) and sector-specific (SeS) news items. The experiments are performed on stocks from the S&P 500 index belonging the Health Care sector. News categories are formed from a large database downloaded from the LexisNexis database. News items are allocated to the corresponding categories based on their relevance to the target stock. The SS subset of data includes articles that are only relevant to the target stock. News articles, that are relevant to at least one stock from a list of stocks belonging to the target stock's sub-industry, are

Abbreviations: SS, stock-specific; SIS, sub-industry-specific; IS, industry-specific; GIS, group-industry-specific; SeS, sector-specific.

* Corresponding author.
E-mail address: shynkevich-y@email.ulster.ac.uk (Y. Shynkevich).

http://dx.doi.org/10.1016/j.dss.2016.03.001
0167-9236/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: Y. Shynkevich, et al., Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.03.001

assigned to the SIS subset of news. Similarly, news articles, relevant to all stocks within the relevant industry, group industry and sector to which the target stock belongs, form the IS, GIS and SeS subsets respectively. A detailed explanation of how the news is allocated to different categories is given in Section 3.2.

Integration of different data types is often performed by the Multiple Kernel Learning (MKL) method [9–12]. Several kernels are used for learning different data subsets. MKL is applied in this study and it utilises from two to fifteen kernels assigned to either SS, SIS, IS, GIS or SeS subset of articles. The results show that an attempt to allocate news articles into different categories, pre-process them separately, learn from them and integrate their predictions into a single prediction decision improve the prediction performance in comparison with approaches based on a single news subset.

The remainder of the paper is organised as follows. Section 2 gives an overview of the relevant literature. Section 3 discusses the raw dataset, data pre-processing techniques, machine learning approaches and performance metrics utilised for analysis. Section 4 describes the experimental results. Section 5 concludes the research work and outlines directions for future work.

2. Related work

An extensive review of the research articles published about financial predictions using text mining is presented in [5]. All systems employing text mining for financial prediction have some of the components illustrated in Fig. 1. Textual data obtained from online sources and market price data are used as an input to the predictive system, and values predicting the market are outputted from it.

2.1. Early works

Wüthrich et al. [13] were the first to try to use textual information for financial forecasting. The authors used knowledge of a domain expert to obtain a dictionary of terms that were later used to assign feature weightings and generate probabilistic rules. Daily price changes were predicted for five stock indices and a trading strategy was formed based on the predictions. The resulting returns were positive and confirmed that profit can be gained with the use of financial news. Lavrenko et al. [14] proposed the Analyst system that employed language models, utilised time series of prices and classified news articles. The authors showed that the designed system is capable of producing profit. Gidofalvi and Elkan [15] developed a system that predicted short term price movements using news articles. Articles were scored using linear regression to the NASDAQ index and assigned with a “down”, “unchanged” or “up” label. The authors stated that the behaviour of stock prices is strongly correlated with the information in news articles starting from 20 min prior to 20 min after its publication. Headlines of news published about companies were examined in [16]. The authors claimed that bad news enforced a strong negative market drift. In [17], official company reports were considered and their ability to indicate future performance of a firm was shown. For instance, a change in written style of documents may indicate a significant change in firm's productivity.

2.2. Key related research

Approaches to the financial forecasting that exist in the literature mainly differ in three general aspects: the dataset, the textual pre-processing methods and the machine learning algorithm. Correspondingly, Table 1 reviews the key related research relevant to the work presented in this paper and provides details about the choices of datasets, textual pre-processing and machine learning techniques made in those papers.

Schumaker and Chen [8] tried to group financial news by similar sectors and industries and studied the predictability of related stock prices based on the news. The authors showed that the ability to predict stock prices varies for different news groups. Schumaker and Chen used only one news group at a time and examined the forecasting performance achieved using articles from the whole dataset of news or relevant to either a stock, its sub-industry, industry, group industry or sector. The research proposed in this paper adopts an idea to partition articles by sectors and industries from [8] to create subsets of news articles divided according to their relevance to the target stock. However, these subsets are used simultaneously in order to benefit from news published about the target stock and other stocks across the target stock's industry and sector. The proposed predictive system employs the concurrent use of news articles from all categories. To the best of our knowledge, no existing research has focussed on the simultaneous use of financial news items from different industrial categories and subcategories. Therefore, this paper investigates the importance of including news articles having different stock relevance levels to forecast stock price changes.

Hagenau, Liebmann and Neumann [18] designed a stock price prediction system that uses text mining to automatically read corporate announcements and financial news articles and employs market reaction

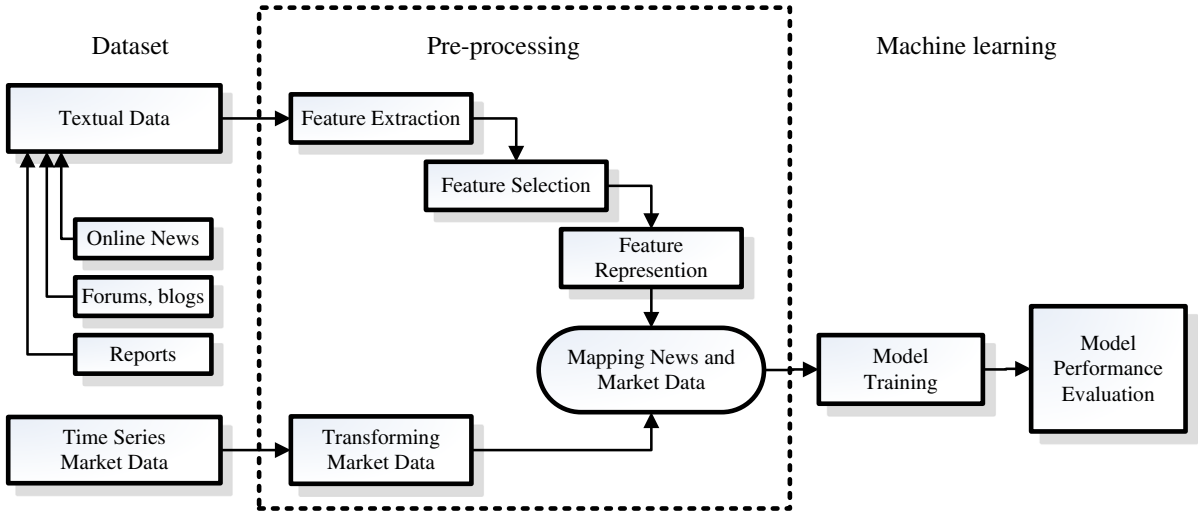

Fig. 1. Typical components of the news-based financial forecasting system.

Please cite this article as: Y. Shynkevich, et al., Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.03.001

Download English Version:

https://daneshyari.com/en/article/6948462

Download Persian Version:

https://daneshyari.com/article/6948462

Daneshyari.com