IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2018, accepted November 8, 2018, date of publication November 16, 2018,
date of current version December 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2881689

Stock Prediction via Sentimental Transfer Learning

XIAODONG LI[1], HAORAN XIE[2], RAYMOND Y. K. LAU[3], (Senior Member, IEEE), TAK-LAM WONG[4], AND FU-LEE WANG[5]

[1]College of Computer and Information, Hohai University, Nanjing 210098, China
[2]Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
[3]Department of Information Systems, City University of Hong Kong, Hong Kong
[4]Department of Computing Studies and Information, Douglas College, New Westminster, BC V3M 5Z5, Canada
[5]School of Science and Technology, The Open University of Hong Kong, Hong Kong

Corresponding author: Xiaodong Li (xiaodong.c.li@outlook.com)

The work of X. Li was supported in part by the National Key R&D Program of China under Grant 2018YFC0407901, in part by the National Natural Science Foundation of China under Grant 61602149, and in part by the Fundamental Research Funds for the Central Universities under Grant 2016B01714. The work of H. Xie was supported in part by the Internal Research of The Education University of Hong Kong under Grant RG 92/2017-2018R and in part by the Research Grants Council of Hong Kong Special Administrative Region, China, under Grant UGC/FDS11/E03/16. The work of R. Y. K. Lau was supported in part by the Research Grants Council of the Hong Kong under Projects CityU 11502115 and CityU 11525716 and in part by NSFC Basic Research Program under Project 71671155.

ABSTRACT Stock prediction is always an attractive problem. With the expansion of information sources, news-driven stock prediction based on sentiments of social media, such as sentiment polarities in financial news, becomes more and more popular. However, the distributions of news articles among different stocks are skewed, which makes stocks with few news have few training samples for their prediction models, and thus leads to low prediction accuracy in the stock predictions. To address this problem, we propose sentimental transfer learning, which transfers sentimental information learned from news-rich stocks (source) to the news-poor ones (target), and prediction performances of the later ones are, therefore, improved. In this approach, the financial news articles of both the source and target stocks are first mapped into the same feature space that is constructed by sentiment dimensions. Second, we develop three different transfer principles in order to explore different transfer scenarios: 1) the source and target stocks' historical price time series are highly correlated; 2) the source and target stocks are in the same sector and the former is the most news-rich one in the sector; and 3) the source stock has the highest prediction performance in validation data set. Third, a majority voting mechanism is designed based on the principles. The voting mechanism is to select the most proper source stock from the candidate stocks that are generated by different principles. Stock predictions are finally made based on the prediction models trained on the selected stocks. Experiments are conducted based on the data of Hong Kong Stock Exchange stocks from 2003 to 2008. The empirical results show that sentiment transfer learning can improve the prediction performance of the target stocks, and the performances are better and more stable with the source stocks selected by the voting mechanism.

INDEX TERMS Sentiment analysis, stock prediction, transfer learning.

I. INTRODUCTION

Stock prediction is always attractive to both computer science researchers and finance practitioners. As the development of social media, information about companies and stocks are no longer limited to numerical financial market data. Market behaviors are more and more influenced by new information sources, such as no-structured news articles with sentiment polarities, and these new information sources have been frequently exploited by stock prediction models. There have been many existing studies that analyze textual financial news articles using a machine learning framework [1]–[4]. One of the most critical problems that remains to be solved is how to improve prediction performances for the stocks that have few financial news. As shown in Figure 1, in HSI market index,[1] the distribution of financial news articles is highly imbalanced, and therefore prediction models that are trained on stocks with fewer articles have much worse prediction performances than the ones trained on news-rich stocks. How to solve the problem becomes an interesting research topic.

[1]Hang Seng Index

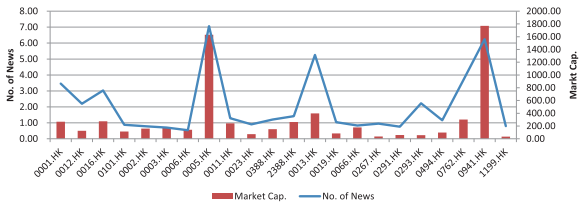


FIGURE 1. Average number of news articles among stocks of HSI (2003-2008).

In this paper, we propose sentimental transfer learning to improve the prediction performances of the news-poor stocks. Sentimental transfer learning in stock prediction is a specific extension of transfer learning, which, by different transfer principles, transfers the various information mined from the news-rich stocks (source) to the news-poor ones (target) in a sentimental feature space. In sentimental transfer learning, the financial news articles from both news-rich and news-poor stocks are firstly mapped into the same feature space that is constructed by sentiment dictionaries. Secondly, following intuitions in pair trading, we develop three different transfer principles to guide the transfer process, which are 1) the source and target stocks' historical price time series are highly correlated, 2) the source and target stocks are in the same sector and the former is the most news-rich one in the sector, and 3) the source stock has the highest prediction performance in validation data set. Thirdly, a voting mechanism is designed based on those three principles, which ranks candidate stocks generated by each principle, and the top-ranked source stock further improves the target stock prediction accuracy. Finally, instances of both the source and target stocks are used together to train the prediction model within the sentiment feature space, and the model is plugged into a stock prediction framework proposed in [5] to make predictions.

Experiments are designed and conducted based on stock prices and news articles of Hong Kong Stock Exchange (year 2003-2008). Models using three different transfer principles are trained and evaluated by standard metrics respectively, and the experimental results show that the sentimental transfer learning approach can improve the stock prediction accuracy in most of the testing cases. In addition, the sentimental transfer learning with voting mechanism on different transfer principles produce more stable prediction performances than single-principle based sentimental transfer learning.

The rest of this paper is organized as follows. Section II reviews the work on news impact analysis and the transfer learning. Section III presents the proposed sentimental transfer learning. Section IV firstly introduces the experimental design, secondly reports the experimental results and lastly gives some discussions. Section V draws the conclusions of this paper.

II. RELATED WORK

In this section, we briefly review previous works that are related to 1) financial news sentiment analysis based stock prediction, 2) sentiment dictionaries, and 3) transfer learning, in order to provide a background for our proposed research.

A. NEWS SENTIMENT ANALYSIS

In order to make accurate stock prediction, many previous studies in the finance and computer science domain frame the problem as a stock price return prediction task, and focus on the impact of news sentiment on the stock price returns [6]–[8]. These analysis exploit the affective aspects of the words that are used in the news articles. Niederhoffer [9] analyze *New York Times* and classify 20 years of headlines into 19 predefined semantic categories from extreme-bad to extreme-good. He also analyzes how markets react to the news of different categories and finds that the markets have a tendency to overreact to the bad news. Davis *et al.* [10] analyze the effects of optimistic or pessimistic language used in news on firms' future performance. They have two conclusions: 1) there is a bias between the readers' expectation and the writers' intension, and 2) readers react strongly to both the content and the affective side of the reports which violate their expectations. Wang *et al.* [11] adopts mutual information-based sentimental analysis methodology and propose a prediction model with extreme learning machine to enhance the prediction accuracy as well as prediction speed. Tetlock [12] extracts and quantifies the optimism and pessimism of *Wall Street Journal* reports, and observes that trading volume tends to increase after pessimism reports and high pessimism scored reports tend to be followed by a down trend and a reversion of market prices. In their later work [13], Tetlock *et al.* use Harvard IV-4 psychological dictionary to analyze the fraction of negative words in *Dow Jones News Service* and *Wall Street Journal* stories about S&P 500 firms from 1980 through 2004, where only positive and negative dimensions are exploited.

B. SENTIMENT DICTIONARIES

In the sentiment analysis and applications, the sentiment dictionary is the cornerstone and widely employed. The construction approaches of the sentiment dictionary can be categorized into two groups:

- *Semi-Automatic: Seed* words are firstly selected manually for the dictionary construction, based on which, the dictionary is then automatically expanded by user defined rules.
- *Manual:* The dictionary is purely constructed by linguistic experts. This kind of dictionary is usually smaller in size than the one constructed semi-automatically, but more accurate.

There have been many typical research results on this subject. Hatzivassiloglou and McKeown [14] select several seed adjective words and classify the rest into positive and negative groups based on two rules: 1) adjectives that are separated by "and" have the same polarity and 2) adjectives that are separated by "but" have opposite polarity. Wiebe [15] also classifies adjectives by polarity. His rules are based on adjectives' tone and orientation clusters. Kim and Hovy [16]

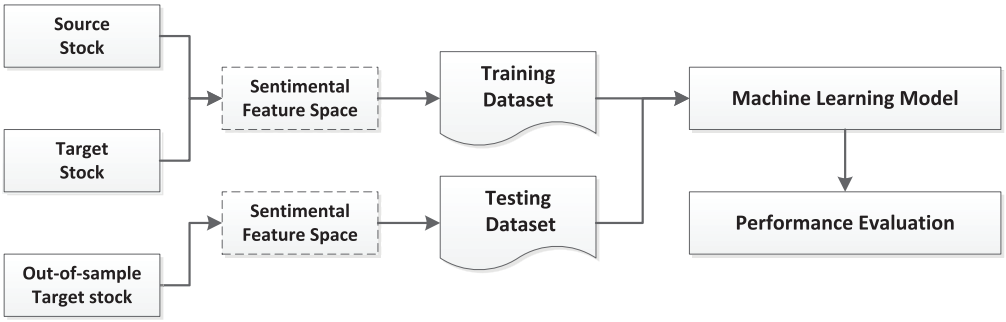


FIGURE 2. The workflow of sentimental transfer learning.

use WordNet to expand the selected seed words. Their generation rules are: 1) synonyms have the same polarity and 2) antonyms have the opposite polarity. In addition, the strength of a word polarity is measured: 1) the number of the word's synonyms that are in the WordNet quantifies the strength and 2) strength is neutral when the word's strength is below a predefined threshold. Nasukawa and Yi [17] weighs more on local sentiment than global sentiment. Their hypotheses consider that human evaluators would agree more on local sentiment than the global one. Similarly, the local sentiments are also analyzed by Godbole *et al.* [18]. General Inquirer[2] is a computer-assisted approach for content analysis of textual data. The Harvard IV-4 dictionary within the General Inquirer Augmented Spreadsheet contains more than 10,000 words and 182 sentiment dimensions. Loughran and McDonald [19] provide a manually made financial sentiment dictionary[3] which contains 6 sentiment dimensions. These two dictionaries are adopted in our research to construct to sentiment feature space.

C. TRANSFER LEARNING

The purpose of the transfer learning is to transfer knowledge learned from a task in a source domain to a task in a target domain. In the survey paper [20], transfer learning approaches can be categorized into four groups, among which instance transfer technique is a widely adopted idea which reuses part of the instances in the source domain and helps the learning task in the target domain. One important issue of the instance transfer is the feature space construction (Moreno *et al.* [21] and Cao *et al.* [22]). Word vector space in bag-of-words approach is usually employed in text mining and news impact analysis [1], [3], [5]. However, while encountering cross-stock news impact analysis, the approach may not work well due to large variances in word feature space. In contrast, we make use of the sentiment space and map both the source stock news articles and target stock articles into the same sentiment feature space. In this way, the variances in feature space are reduced and the instance transfer learning is expected to work well. Another important issue is transfer learning adaptiveness or transfer principle which decides what instances are proper to be transferred (Zhang *et al.* [23] and Zhao *et al.* [24]). In this paper, based on the knowledge of both finance domain and computer science domain, we develop three transfer principles considering different perspectives of source stock selection, and further use a voting mechanism to enhance the transfer learning performances.

[2]http://www.wjh.harvard.edu/~inquirer/Home.html
[3]http://www3.nd.edu/~mcdonald/Word_Lists.html

III. SENTIMENTAL TRANSFER LEARNING

The workflow of the sentimental transfer learning is shown in Figure 2. Each step in the workflow is illustrated in the following subsections.

A. SENTIMENT FEATURE SPACE

All the news articles are basically texts. Following the preprocessing method in text mining, each article can be represented by a word frequency vector X, where each element x_i in X indicates the number of occurrences of a specific word w_i (stop words are removed) in the article. In the first step of the workflow, Xs are further mapped into a sentimental feature space $\mathcal{S}$ which is constructed by a certain sentiment dictionary D, by

$$S = X \odot D, \tag{1}$$

where $\odot$ is a dictionary-search operation which searches w_i in D and finds the corresponding sentiment vector s_i. As the number of sentimental dimension is fixed in the dictionary, each word can be represented by a sentiment feature vector of the same length. Thus, each news article can be represented by a sentiment feature vector by summing up all words' vectors. S for X is then calculated by,

$$S = \sum_{i|w_i \in D} x_i \cdot s_i, \tag{2}$$

where s_i is the sentiment vector of w_i, and x_i is the word frequency of w_i. Only the words that are contained in the sentiment dictionary ($w_i \in D$) are considered.

B. TRANSFER LEARNING AND PRINCIPLES

In the second step of the news-driven stock prediction, given a domain $\mathcal{D} = \{\mathcal{S}, P(X)\}$, where $\mathcal{S}$ is the sentimental feature space and $P(X)$ is the marginal probability distribution of news articles, the prediction task consists of two components: 1) a label space $\mathcal{Y}$ and 2) a news impact prediction function $f(\cdot) = P(y|x)$, and the task is denoted by

$$\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}. \tag{3}$$

A source stock is in a source domain $\mathcal{D}_\alpha$ and a target stock in a target domain $\mathcal{D}_\beta$. All the news articles of both the source stock and target stock are collected, and the source stock in $\mathcal{D}_\alpha$ is preprocessed in step one and represented by data set

$$D_\alpha = \{(s_{\alpha_1}, y_{\alpha_1}), (s_{\alpha_2}, y_{\alpha_2}), \ldots, (s_{\alpha_{|D_\alpha|}}, y_{\alpha_{|D_\alpha|}})\}, \tag{4}$$

where $s_{\alpha_i} \in \mathcal{S}_\alpha$, $y_{\alpha_i} \in \mathcal{Y}_\alpha$, and $|D_\alpha|$ is the size of the data set. Correspondingly, target domain data set is

$$D_\beta = \{(s_{\beta_1}, y_{\beta_1}), (s_{\beta_2}, y_{\beta_2}), \ldots, (s_{\beta_{|D_\beta|}}, y_{\beta_{|D_\beta|}})\}, \tag{5}$$

where $s_{\beta_i} \in \mathcal{S}_\beta$, $y_{\beta_i} \in \mathcal{Y}_\beta$, and $0 \leq |D_\beta| \ll |D_\alpha|$. Given the source stock domain $\mathcal{D}_\alpha$, an impact prediction task $\mathcal{T}_\alpha$, a target stock domain $\mathcal{D}_\beta$ and an impact prediction task $\mathcal{T}_\beta$, transfer learning is to help the target task improve the prediction performance of $f_\beta(\cdot)$. In this paper, all the news articles of both the source stock and the target stock within trading hours are collected. Sentimental transfer learning employs sentiment dictionaries to construct the sentiment space for both S_α and S_β, and follows instance transfer approach which 1) divides $\mathcal{S}_\beta$ into a training set S_β^{in} and an out-of-sample set S_β^{out}, and 2) transfers the instances in S_α to $\mathcal{T}_\beta$ which combines S_α with S_β^{in} and train a new model to help improve the performance of $f_\beta(\cdot)$.

One critical question during the instance transfer is how to select the source stock, in another word, the adaptiveness of the transfer. Since the source stock in $\mathcal{T}_\alpha$ is fundamentally a different company comparing with the target stock in $\mathcal{T}_\beta$, the selection of the source stocks needs to follow either finance domain knowledge or statistical properties of the stocks. In this paper, we follow finance heuristics and develop three different transfer principles to examine the sentiment transfer learning, which are 1) correlation-based principle, 2) number-of-news-based principle and 3) validation-performance-based principle, where different principles focus on different perspectives:

- The first transfer principle is based on price correlations, where instances are selected from the source stock that has the highest historical price time series correlation with the target stock.
- The second transfer principle is based on the number of news, where instances are selected from the source stock that has the largest number of news articles within the same sector as the target stock.
- The third transfer principle is based on the validation performance of each stock without any transfer learning, where instances are selected from the source stock that has the best validation performance within the same sector as the target stock.

C. STOCK PREDICTION

In the third step of stock prediction, a classification machine learning model is trained based on the combined training data set, where the classification model during the modeling can be considered as an *abstract model* which can be instantiated by any classification machine learning model (e.g., SVMs in the Experiments in Section IV-D),

$$D_{tr} = S_\alpha \cup S_\beta^{in}, \tag{6}$$

and tested by a testing data set

$$D_{tt} = S_\beta^{out}. \tag{7}$$

All the preprocessed instances are fed into the machine learning model, and each individual sentiment dimension is taken as one input feature. As there are three different transfer principles and each of them selects a candidate source stock, we further employ a majority voting mechanism to enhance the source stock selection, where the stock voted by more transfer principles is considered to be the only candidate source stock and is expected to have more robust performances in the stock prediction evaluations.

IV. EXPERIMENTS AND DISCUSSIONS

A. DATA SETS

The stocks we investigate are listed in Hong Kong Stock Exchange. In this work, we focus on the Hang Seng Index (HSI) constituents which are the stocks with bigger capital comparing with the rest stocks in the market.[4] There are 4 sectors in HSI, i.e., Commerce, Finance, Properties and Utilities, which have totally 50 stocks in 2013. Since not all the 50 stocks are the HSI constituents through the years in the data set, we remove the stocks that were added after the starting date (2003-01-01) and keep the rest according to the change history of the constituents. After preprocessing, 22 stocks are left in the universe.

A news archive from FINET[5] is employed. The news archive contains both company-specific and market related news from Jan. 2003 to Mar. 2008. Each piece of news is tagged with a time stamp showing the time the news is released, which helps classify news by dates. Stock symbols of companies that are mentioned in the news are listed at the end of the article, which helps establish the mapping from the news articles to stocks and vice versa.

Following the approach in [3], we use a constant split method for the data set, and the whole data set is split into three parts for different purposes: 1) from Jan. 2003 to Dec. 2005 is the training data set; 2) from Jan. 2006 to Dec. 2006 is the validation data set; and 3) from Jan. 2007 to Mar. 2008 is the independent testing data set.

[4] http://www.hsi.com.hk/HSI-Net/HSI-Net

[5] News data is available at http://www.finet.hk/mainsite/index.htm

B. SENTIMENT DICTIONARIES FOR SENTIMENTAL FEATURES

As reviewed in Section II-B, the manually constructed dictionaries are usually more accurate than the semi-automatically constructed ones because of the careful selection by linguistic experts. To make the experiments more solid, we use two manually constructed sentiment dictionaries and one semi-automatically constructed sentiment dictionary, i.e.,

1) Loughran-McDonald financial sentiment dictionary (LMD). LMD is constructed by Loughran and McDonald [19], which contains more than 3911 words and 6 sentiment dimensions. We use LMD version 2012.
2) Harvard IV-4 sentiment dictionary (HVD).[6] HVD contains more than 10,000 words, and 182 sentiment dimensions are categorized into 15 groups.
3) SenticNet sentiment dictionary (SN). SN is built by Cambria *et al.* [25], [26], which exploits *sentic computing* for concept-level sentiment analysis. In the 3.0 version, SN has 5 sentiment dimensions.

C. METRICS

We obtain daily quotes (i.e., Open, High, Low, and Close prices) from Yahoo! Finance[7] for the stocks selected. Each news article within the day is labeled by

$$y = \begin{cases} +1 & \text{if } r \geq \theta \\ 0 & \text{if } -\theta < r < \theta \\ -1 & \text{if } r \leq -\theta, \end{cases} \tag{8}$$

where θ is a threshold, and r is simple return which is calculated based on Open and Close prices,

$$r = \frac{Close - Open}{Open}. \tag{9}$$

Accuracy (*acc*) is selected to evaluate the classification performance, which is defined as,

$$acc = \frac{hits}{all}, \tag{10}$$

and

$$hits = t_{++} + t_{00} + t_{--}, \tag{11}$$

$$all = t_{++} + t_{00} + t_{--} + f_{0+} + f_{-+} + f_{+0} + f_{-0} + f_{+-} + f_{0-}, \tag{12}$$

TABLE 1. The definition of terms in *acc*.

	predict +	predict 0	predict −
true +	t_{++}	f_{+0}	f_{+-}
true 0	f_{0+}	t_{00}	f_{0-}
true −	f_{-+}	f_{-0}	t_{--}

where $t_{++}, t_{00}, t_{--}, f_{0+}, f_{-+}, f_{+0}, f_{-0}, f_{+-}$ and f_{0-} are defined in Table 1.

[6] http://www.wjh.harvard.edu/~inquirer/homecat.htm
[7] Price data is available at http://finance.yahoo.com/

The threshold θ that discretizes r largely influences the *acc*. Suppose we arbitrarily choose a θ and follow the hypothesis in finance domain that the distribution of stock price returns is Gaussian with *fat tail*, the true label -1, 0 and $+1$ have following probabilities,

$$\mathrm{P}_{-1} = \int_{-\infty}^{-\theta} \mathrm{pdf}_{Gaussian}(x)\mathrm{d}x, \tag{13}$$

$$\mathrm{P}_{0} = \int_{-\theta}^{\theta} \mathrm{pdf}_{Gaussian}(x)\mathrm{d}x, \tag{14}$$

$$\mathrm{P}_{+1} = \int_{\theta}^{+\infty} \mathrm{pdf}_{Gaussian}(x)\mathrm{d}x. \tag{15}$$

Given the label distribution without extra learning, people can conduct a random draw based on the distribution and make predictions. If the random draw prediction is eventually the same as the true label, *hits* in *acc* will increase. To formulate, *acc* can be calculated by

$$acc = \mathrm{P}_{-1}^2 + \mathrm{P}_{0}^2 + \mathrm{P}_{+1}^2. \tag{16}$$

Since θ and $-\theta$ are symmetric around 0,

$$\mathrm{P}_{-1} = \mathrm{P}_{+1}, \tag{17}$$

$$\mathrm{P}_{0} = 1 - 2\mathrm{P}_{+1}, \tag{18}$$

substitute Equation ((17)) and ((18)) into ((16)), and denote P_{+1} as P, Equation ((16)) becomes

$$acc = 6\mathrm{P}^2 - 4\mathrm{P} + 1. \tag{19}$$

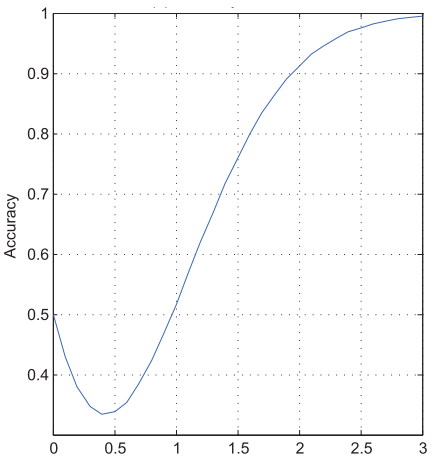


FIGURE 3. Accuracy along with θ.

TABLE 2. Target stocks.

	LMD	HVD	SN
Commerce	0941.HK	0762.HK	0762.HK
Finance	2388.HK	0388.HK	0388.HK
Properties	0012.HK	0016.HK	0012.HK
Utilities	0003.HK	0003.HK	0003.HK

If we plot *acc* along with the change of θ in Figure 3, we find that the chart has a up-side-down *Sine* function shape,

TABLE 3. Commerce correlation.

	0013.HK	0019.HK	0066.HK	0267.HK	0291.HK	0293.HK	0494.HK	0762.HK	0941.HK	1199.HK
0013.HK		0.4514	0.4087	0.3883	0.3885	0.4105	0.3280	0.4787	0.5887	0.3596
0019.HK			0.4022	0.3600	0.2962	0.4421	0.3524	0.3779	0.3972	0.2848
0066.HK				0.4034	0.3304	0.3064	0.3154	0.3844	0.4067	0.3120
0267.HK					0.4233	0.3861	0.3120	0.4290	0.4232	0.3795
0291.HK						0.2896	0.2709	0.3858	0.4160	0.4255
0293.HK							0.3054	0.3356	0.3920	0.2729
0494.HK								0.3064	0.3569	0.2854
0762.HK									**0.6522**^{1,2,3}	0.3729
0941.HK										0.3689
1199.HK										

TABLE 4. Finance correlation.

	0005.HK	0011.HK	0023.HK	0388.HK	2388.HK
0005.HK		0.4879	0.3298	-0.0234	**0.3820**^{1}
0011.HK			0.3326	**0.0002**^{2,3}	0.3220
0023.HK				-0.0195	0.2785
0388.HK					-0.0242
2388.HK					

TABLE 5. Properties correlation.

	0001.HK	0012.HK	0016.HK	0101.HK
0001.HK		0.6322	**0.7386**^{2}	0.4078
0012.HK			**0.6931**^{1,3}	0.4213
0016.HK				0.4076
0101.HK				

TABLE 6. Utilities correlation.

	0002.HK	0003.HK	0006.HK
0002.HK		0.4012	0.4135
0003.HK			**0.4043**^{1,2,3}
0006.HK			

with the minimum value at P $= \frac{1}{3}$, which is equivalent to flipping a fair coin with three outcomes for the prediction. On the other side, *acc* increases along with θ on the right hand side of the minimum point, which means when the value of θ is great, most of the instances are labeled with 0 and the probability of correct guess becomes high. In another word, the accuracy can be increased only by manipulating the labeling method without any change of the learning model. Based on the explanations, θ cannot be too large as it will bring too much *acc* increment. On the other hand, θ cannot be too small, since θ should be greater than the basic market transaction cost which is 0.003 (30 bps) in Hong Kong market. To balance both the constraints, we choose $\theta = 0.003$ (30 bps).

D. BASELINES

We incorporate the approaches proposed by Li *et al.* [3] as the baselines in the experiment, where news articles of each stocks are firstly mapped into a sentiment feature space by using a sentiment dictionary, secondly stock prediction models are trained based on the instances newly constructed, and finally the trained models are evaluated. We replicate their methods, and for each sector and each sentiment dictionary we find the target stock that has relative low prediction performance and few news articles. All the target stocks are listed in Table 2.

Support vector machines (SVMs), which is one of the most popular model in text mining, is adopted in the experiment. During the training and validation phases, the parameters of SVMs are tuned, and the best combination of the parameters is kept for the independent

TABLE 7. Average number of news articles.

Commerce		Finance	
0013.HK	5.2577	0005.HK	**7.0639**
0019.HK	1.0396	0011.HK	1.2871
0066.HK	0.8350	0023.HK	0.9079
0267.HK	0.9546	0388.HK	1.2033
0291.HK	0.7621	2388.HK	1.4175
0293.HK	2.2276		
0494.HK	1.1650		
0762.HK	3.6944		
0941.HK	**6.2494**		
1199.HK	0.7903		
Properties		**Utilities**	
0001.HK	**3.4604**	0002.HK	**0.7864**
0012.HK	2.1976	0003.HK	0.6995
0016.HK	3.0326	0006.HK	0.5339
0101.HK	0.8760		

TABLE 8. Validation results without STL.

Commerce				Finance			
Symbol	LMD	HVD	SN	Symbol	LMD	HVD	SN
0013.HK	0.4245	0.4327	0.4245	0005.HK	0.7560	**0.7480**	0.7420
0019.HK	**0.5474**	0.4421	0.4421	0011.HK	**0.8000**	0.7400	**0.8200**
0066.HK	0.3684	0.4035	0.4561	0023.HK	0.3878	0.3776	0.4184
0267.HK	0.4324	**0.5000**	0.4595	0388.HK	0.3821	0.3089	0.3902
0291.HK	0.4194	0.4677	**0.5000**	2388.HK	0.3333	0.3953	0.4031
0293.HK	0.4218	0.3810	0.4218				
0494.HK	0.4043	0.4043	0.4787				
0762.HK	0.3721	0.3395	0.3302				
0941.HK	0.3482	0.3968	0.3522				
1199.HK	0.4444	0.4568	0.4568				
Properties				**Utilities**			
Symbol	LMD	HVD	SN	Symbol	LMD	HVD	SN
0001.HK	0.4229	0.4317	0.3789	0002.HK	**0.4600**	**0.4700**	0.4000
0012.HK	0.3567	0.4076	0.3694	0003.HK	0.4333	0.3889	0.4000
0016.HK	0.4670	0.4061	0.4162	0006.HK	0.4462	0.4000	**0.4462**
0101.HK	**0.4949**	**0.4747**	**0.4848**				

TABLE 9. Accuracy results in validation data set.

LMD	#corr		#news		#perf		Baseline	
Commerce	0762.HK	0.3442	0013.HK	0.3740	0019.HK	**0.4006**	0941.HK	0.3482
Finance	0005.HK	**0.6332**	0005.HK	**0.6332**	0011.HK	**0.5590**	2388.HK	0.3333
Properties	0016.HK	**0.4068**	0001.HK	**0.3724**	0101.HK	**0.3672**	0012.HK	0.3567
Utilities	0006.HK	0.3613	0002.HK	**0.4579**	0002.HK	**0.4579**	0003.HK	0.4333
HVD	#corr		#news		#perf		Baseline	
Commerce	0941.HK	**0.3571**	0941.HK	**0.3571**	0267.HK	**0.3529**	0762.HK	0.3395
Finance	0011.HK	**0.5516**	0005.HK	**0.6220**	0005.HK	**0.6220**	0388.HK	0.3089
Properties	0001.HK	**0.4292**	0001.HK	**0.4292**	0101.HK	0.4054	0016.HK	0.4061
Utilities	0006.HK	0.3613	0002.HK	**0.4263**	0002.HK	**0.4263**	0003.HK	0.3889
SN	#corr		#news		#perf		Baseline	
Commerce	0941.HK	**0.3398**	0941.HK	**0.3398**	0291.HK	**0.3574**	0762.HK	0.3302
Finance	0011.HK	**0.4529**	0005.HK	**0.5845**	0011.HK	**0.4529**	0388.HK	0.3902
Properties	0016.HK	**0.4153**	0001.HK	**0.3724**	0101.HK	0.3633	0012.HK	0.3694
Utilities	0006.HK	**0.4258**	0002.HK	**0.4158**	0006.HK	**0.4258**	0003.HK	0.4000

TABLE 10. Accuracy results in independent testing data set.

LMD	#corr		#news		#perf		Baseline	
Commerce	0762.HK	0.3413	0013.HK	0.3518	0019.HK	**0.3820**	0941.HK	0.3583
Finance	0005.HK	**0.5684**	0005.HK	**0.5684**	0011.HK	**0.4780**	2388.HK	0.3257
Properties	0016.HK	**0.3918**	0001.HK	**0.3631**	0101.HK	0.3592	0012.HK	0.3611
Utilities	0006.HK	0.3260	0002.HK	**0.4275**	0002.HK	**0.4275**	0003.HK	0.4198
HVD	#corr		#news		#perf		Baseline	
Commerce	0941.HK	**0.3556**	0941.HK	**0.3556**	0267.HK	**0.3662**	0762.HK	0.3409
Finance	0011.HK	**0.4536**	0005.HK	**0.5112**	0005.HK	**0.5112**	0388.HK	0.3611
Properties	0001.HK	**0.3824**	0001.HK	**0.3824**	0101.HK	**0.3693**	0016.HK	0.3462
Utilities	0006.HK	0.3304	0002.HK	**0.4275**	0002.HK	**0.4275**	0003.HK	0.3457
SN	#corr		#news		#perf		Baseline	
Commerce	0941.HK	**0.3606**	0941.HK	**0.3606**	0291.HK	**0.3720**	0762.HK	0.3523
Finance	0011.HK	**0.3918**	0005.HK	**0.4870**	0011.HK	**0.3918**	0388.HK	0.3750
Properties	0016.HK	**0.4021**	0001.HK	0.3689	0101.HK	0.3721	0012.HK	0.3958
Utilities	0006.HK	**0.3965**	0002.HK	**0.4084**	0006.HK	**0.3965**	0003.HK	0.3765

testing phase. To be specific, we use RBF kernel SVMs. The parameters γ and C are tuned through a grid search method, which searches in a two-dimension space, $\gamma \in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$ and $C \in \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$, totally $9 \times 10 = 90$ combinations.

E. TRANSFER PRINCIPLES

In the experiments, we examine the sentiment transfer learning via developing three different transfer principles, which are 1) correlation based principle (#corr), 2) No. of news based principle (#news) and 3) validation performance based principle (#perf). Different principles select and transfer instances of stocks from different perspectives,

1) *#corr:* In Table 3, Table 4, Table 5 and Table 6, correlation statistics of each sector are listed. Numbers in bold font indicate that the source stock has the highest historical price time series correlation with the target stock (using different sentiment dictionaries give different target stocks, and the sentiment dictionaries are indexed by 1: LMD, 2: HVD and 3: SN) in the same sector.
2) *#news:* In Table 7, the average number of news for each stock is listed, where numbers in bold font indicate the largest number of news in the sector.
3) *#perf:* The performances figures are listed in Table 8. Numbers in bold font and underlined indicate the best and the poorest performers in the sector, respectively.

F. EXPERIMENT RESULTS

Experiment results based on three transfer principles are presented in Table 9 and Table 10. Each table is divided into three trunks corresponding to three sentiment dictionaries. In each trunk, there are four columns corresponding to three transfer principles and one baseline, where the baselines are the approaches without any sentiment transfer learning.

For each transfer principle, there are 12 pairs of performance comparisons between the one with sentiment transfer learning and the baseline. The numbers in bold font indicate that with sentiment transfer learning, the prediction performance outperforms the baseline. As illustrated in the Table 9, #corr achieves 9 better performers, #news achieves 11 better performers and #perf achieves 10 better performers among the 12*3 comparisons in the validation data set, and in Table 10, #corr achieves 9 better performers, #news achieves 11 better performers and #perf achieves 10 better performers among the 12*3 comparisons in the independent testing data set. These results show that with sentiment transfer learning, the prediction performance of the target stock can be improved.

On the other hand, it is also shown in the results that models with sentiment transfer learning do not win in all the comparisons based on single transfer principle. To address this problem, we further develop a more robust method to select the source stocks: instead of using the transfer principles separately, we adopt a second source stock selection layer by using a majority voting scheme, which is to select the source stock that is chosen by at least two transfer principles. As shown in Table 9 and Table 10, the source stocks that are chosen by the scheme are underlined, and their prediction performances are consistently better than the baselines in all the comparisons.

V. CONCLUSION

In this paper, we studied the problem that how to improve the performances of news-driven stock prediction models that are trained based on stocks with few financial news. We propose sentimental transfer learning which transfers knowledge learned from news-rich stocks (source) to the news-poor ones (target). During the transfer, three different transfer principles are developed and incorporated to select the source stocks: 1) the source and target stocks' historical price time series are highly correlated; 2) the source and target stocks are in the same sector and the former is the most news-rich one in the sector; 3) the source stock has the highest prediction performance in validation data set. A majority voting mechanism that combines the three principles to enhance the robustness of the source stock selection is further employed. Financial news articles of both the source and target stocks are mapped into the same feature space that are constructed by sentiment dimensions. Stock predictions are made based on the prediction models trained within the feature space. From the experiments conducted on real stock market data, it is shown that prediction models with sentiment transfer learning outperform the baselines that are purely based on models without any transfer learning. Comparing with the performances that are based on single principle stock selection, the results based on the principle majority voting mechanism show that performances are consistently better in both the validation and independent testing comparisons.

ACKNOWLEDGMENT

This paper was presented at the BigComp 2017 [1] (extend from 2 pages to 8 pages). Some contents from the conference version are re-used in this journal article. The new contents of this article are more than 30% according to the regulation of the published journal, which can be summarized in the following aspects: 1) development of different transfer principles; 2) a majority voting mechanism based on transfer principles; and 3) experiments from more perspectives are conducted to test the transfer principles.

REFERENCES

[1] X. Li, H. Xie, T.-L. Wong, and F. L. Wang, "Market impact analysis via sentimental transfer learning," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2017, pp. 451–452.

[2] X. Li, X. Huang, X. Deng, and S. Zhu, "Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information," *Neurocomputing*, vol. 142, pp. 228–238, Oct. 2014, doi: 10.1016/j.neucom.2014.04.043.

[3] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment analysis," *Knowl.-Based Syst.*, vol. 69, pp. 14–23, Oct. 2014, doi: 10.1016/j.knosys.2014.04.022.

[4] X. Li, J. Cao, and Z. Pan, "Market impact analysis via deep learned architectures," in *Neural Computing and Applications*. London, U.K.: Springer, 2018, doi: 10.1007/s00521-018-3415-3.

[5] X. Li, H. Xie, Y. Song, S. Zhu, Q. Li, and F. L. Wang, "Does summarization help stock prediction? A news impact analysis," *IEEE Intell. Syst.*, vol. 30, no. 3, pp. 26–34, May 2015, doi: 10.1109/MIS.2015.1.

[6] W. Liu, H. Yan, and J. Xiao, "Extracting multiple news attributes based on visual features," *J. Intell. Inf. Syst.*, vol. 38, no. 2, pp. 465–486, Apr. 2012, doi: 10.1007/s10844-011-0163-6.

[7] H. Wu, L. Kuang, F. Wang, Q. Rao, M. Gong, and Y. Li, "A multiobjective box-covering algorithm for fractal modularity on complex networks," *Appl. Soft Comput.*, vol. 61, pp. 294–313, Dec. 2017, doi: 10.1016/j.asoc.2017.07.034.

[8] F. Wang, H. Zhang, K. Li, Z. Lin, J. Yang, and X.-L. Shen, "A hybrid particle swarm optimization algorithm using adaptive learning strategy," *Inf. Sci.*, vols. 436–437, pp. 162–177, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025518300380

[9] V. Niederhoffer, "The analysis of world events and stock prices," *J. Bus.*, vol. 44, no. 2, pp. 193–219, 1971.

[10] A. K. Davis, J. M. Piger, and L. M. Sedor, "Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases," in *Proc. Federal Reserve Bank St. Louis Work. Paper*, 2006, pp. 1–12.

[11] F. Wang, Y. Zhang, Q. Rao, K. Li, and H. Zhang, "Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction," *Soft Comput.*, vol. 21, no. 12, pp. 3193–3205, Jun. 2017, doi: 10.1007/s00500-015-2003-z.

[12] P. C. Tetlock, "Giving content to investor sentiment: The role of media in the stock market," *J. Finance*, vol. 62, no. 3, pp. 1139–1168, 2007.

[13] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals," *J. Finance*, vol. 63, no. 3, pp. 1437–1467, 2008.

[14] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proc. 35th Annu. Meeting Assoc. Comput. Linguistics 8th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 1997, pp. 174–181.

[15] J. Wiebe, "Learning subjective adjectives from corpora," in *Proc. AAAI/IAAI*, 2000, pp. 735–740.

[16] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. 20th Int. Conf. Comput. Linguistics*, 2004, Art. no. 1367.

[17] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proc. 2nd Int. Conf. Knowl. Capture*, 2003, pp. 70–77.

[18] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," in *Proc. Int. Conf. Weblogs Social Media*, vol. 2, 2007, pp. 1–4.

[19] T. Loughran and B. McDonald, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *J. Finance*, vol. 66, no. 1, pp. 35–65, 2011.

[20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[21] O. Moreno, B. Shapira, L. Rokach, and G. Shani, "TALMUD: Transfer learning for multiple domains," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, New York, NY, USA, 2012, pp. 425–434, doi: 10.1145/2396761.2396817.

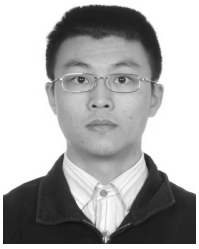
[22] B. Cao, N. N. Liu, and Q. Yang, "Transfer learning for collective link prediction in multiple heterogenous domains," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 159–166.

[23] Y. Zhang, B. Cao, and D.-Y. Yeung, "Multi-domain collaborative filtering," in *Proc. 26th Conf. Uncertainty Artif. Intell.*, 2010, pp. 725–732.

[24] L. Zhao, S. J. Pan, E. W. Xiang, E. Zhong, Z. Lu, and Q. Yang, "Active transfer learning for cross-system recommendation," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 1205–1211. [Online]. Available: http://dl.acm.org/citation.cfm?id=2891460.2891628

[25] E. Cambria, R. Speer, C. Havasi, and A. Hussain, "Senticnet: A publicly available semantic resource for opinion mining," in *Proc. Artif. Intell.*, 2010, pp. 14–18.

[26] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis," in *Proc. FLAIRS Conf.*, 2012, pp. 202–207.



XIAODONG LI received the Ph.D. degree in computer science from the City University of Hong Kong. He is currently an Associate Professor with the College of Computer and Information, Hohai University. He has served as a Principal Investigator of a Sub-Project in the National Key R&D Program of China, and a Project in the National Natural Science Foundation of China. His research interests include artificial intelligence, machine learning, information retrieval, and data mining, especially big data applications to algorithmic trading in finance and water resource management in hydrology.

HAORAN XIE received the Ph.D. degree in computer science from the City University of Hong Kong. He is currently an Assistant Professor with The Education University of Hong Kong. He has published about 120 refereed international journals and conference papers. His research interest includes artificial intelligence, educational technology, and big data. He has served as a Committee Member/Co-Chair of SeCoP, IWUM, SETE, ADDI, IEEE U-Media, WISE, and ICWL. He is also served as a Guest Editors in the *International Journal of Machine Learning and Cybernetics*, *Neurocomputing*, the *International Journal of Distance Education Technologies*, and *Web Intelligent Journal*.

RAYMOND Y. K. LAU (M'00–SM'08) is currently an Associate Professor with the Department of Information Systems, City University of Hong Kong. He has authored over 140 refereed international journals and conference papers. His research work has been published in renowned journals, such as *Management Information Systems (MIS) Quarterly*, *INFORMS Journal on Computing*, *ACM Transactions on Information Systems*, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE INTERNET COMPUTING, the *Journal of MIS*, and *Decision Support Systems*. His research interests include big data stream analytics, social media analytics, information retrieval, and agent-mediated e-commerce. He is a Senior Member of the ACM.

TAK-LAM WONG received the Ph.D. degree in systems engineering and engineering management from The Chinese University of Hong Kong. He is currently a Faculty Member with the Douglas College, Canada. He has published over 100 publications, including PAMI, TKDE, TOIS, and TOIT. His research interests include Web mining, data mining, information extraction, machine learning, and knowledge management. He also has served as a committee member in more than 20 international conferences.

FU-LEE WANG received the Ph.D. degree in systems engineering and engineering management from The Chinese University of Hong Kong. He was the Vice President of the Caritas Institute of Higher Education and a Faculty Member with the City University of Hong Kong. He is currently a Professor and the Dean of the School of Science and Technology, The Open University of Hong Kong. He has over 200 publications and has received 16 grants with a total of more than $20 million Hong Kong dollars. His research interests include e-business, e-learning, financial engineering, and information retrieval.