A Multilayer Perceptron based Ensemble Technique for Fine-grained Financial Sentiment Analysis

A and **B** and **C** and **D** and **E**
`{}@iitp.ac.in`

Abstract

In this paper, we propose a novel method for combining deep learning and classical feature based models using a Multi-Layer Perceptron (MLP) network for financial sentiment analysis. We develop various deep learning models based on Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). These are trained on top of pre-trained, autoencoder-based, financial word embeddings and lexicon features. An ensemble is constructed by combining these deep learning models and a classical supervised model based on Support Vector Regression (SVR). We evaluate our proposed technique on a benchmark dataset of SemEval 2017. The propose model shows impressive results on two datasets, i.e. microblogs and news headlines datasets. Comparisons show that our proposed model performs better than the existing state-of-the-art systems for the above two datasets by 2.0 and 4.1 cosine points, respectively.

1 Introduction

Microblog messages and news headlines are freely available on Internet in vast amount. Dynamic nature of these texts can be utilized effectively to analyze the shift in the stock prices of any company (Goonatilake and Herath, 2007). By keeping a track of microblog messages and news headlines for financial domain one can observe the trend in stock prices, which in turn, allows an individual to predict the future stock prices. An increase in positive opinions towards a particular company would indicate that the company is doing well and this would be reflected in the increase in company stock prices and vice-versa. Benefits of such analysis are two-fold: (i). an individual can take informed decision before buying/selling his/her share; and (ii). an organization can utilize this information to forecast its economic situation.

Sentiment prediction is a core component of an end-to-end stock market forecasting business model. Thus, an efficient sentiment analysis system is required for real-time analysis of financial text originating from the web. Sentiment analysis in financial domain offers more challenges (as compared to product reviews domains etc.) due to the presence of various financial and technical terms along with numerous statistics. Coarse-level sentiment analysis in financial texts usually ignores critical information towards a particular company, therefore making it unreliable for the stock prediction. In fine-grained sentiment analysis, we can emphasize on a given company without losing any critical information. For example, in the following tweet sentiment towards $APPL (APPLE Inc.) is *positive* while *negative* towards $FB (Facebook Inc.).

'$APPL going strong; $FB not so.'

In literature, many methods for sentiment analysis from financial news have been described. O'Hare et al. (2009) used word-based approach on financial blogs to train a sentiment classifier for automatically determining the sentiment towards companies and their stocks. Authors in (Schumaker and Chen, 2009) use the bag-of-words and named entities for predicting stock market. They successfully showed that the stock market behavior is based on the opinions. A fine-grained sentiment annotation scheme was incorporated by (de Kauter et al., 2015) for predicting the explicit and implicit sentiment in the financial text. An application of multiple regression model was developed by (Oliveira et al., 2013).

In this paper, we propose a novel Multi-Layer Perceptron (MLP) based ensemble technique for fine-grained sentiment analysis. It combines the outputs of four systems, one is feature-driven supervised model and the rest three are deep learning based.

We further propose to develop an enhanced word representation by learning through a stacked denoising autoencoder network (Vincent et al., 2010) using word embeddings of Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) models.

For evaluation purpose we use datasets of SemEval-2017 'Fine-Grained Sentiment Analysis on Financial Microblogs and News' shared task (Keith Cortis and Davis, 2017). The dataset comprises of financial short texts for two domains i.e. microblog messages and news headlines. Comparisons with the state-of-the-art models show that our system produces better performance.

The main contributions of the current work are summarized as follows: a) we effectively combine competing systems to work as a team via MLP based ensemble learning; b) develop an enhanced word representation by leveraging the syntactic and semantic richness of the two distributed word representation through a stacked denoising autoencoder; and c) build a state-of-the-art model for sentiment analysis in financial domain.

2 Proposed Methodology

We propose a Multi-Layer Perceptron based ensemble approach to leverage the goodness of various supervised systems. We develop three deep neural network architecture based models, *viz.* Convolution Neural Network (CNN) (Kim, 2014), Long Shrort Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) network (Cho et al., 2014)). The other model is based on Support Vector Regression (SVR) (Smola and Schölkopf, 2004) based feature-driven system.

The classical feature based system utilizes a diverse set of features (c.f. Section 2.D). On the other hand we train a CNN, a LSTM and a GRU network on top of distributed word representations. We utilize Word2Vec and GloVe word representation techniques to learn our financial word embeddings. Since Word2Vec and GloVe models capture syntactic and semantic relations among words using different techniques

(Word2Vec: given a context, a word is predicted or vice-versa; GloVe: count-based model utilizing word co-occurrence matrix), some applications adapt well to Word2Vec while others perform well on GloVe model. We, therefore, attempt to leverage the richness of both the models by using a stacked denoising autoencoder.Finally, we combine predictions of these models using the MLP network in order to obtain the final sentiment scores. An overview of the proposed method is depicted in Figure 1.
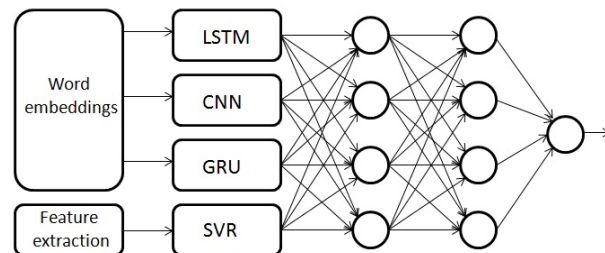

Figure 1: MLP based ensemble architecture.

A. Convolution Neural Network (CNN): Literature suggests that CNN architecture had been successfully applied for sentiment analysis at various level (Kim, 2014; Akhtar et al., 2016; Singhal and Bhattacharyya, 2016). Most of these works involve classification tasks, however, we adopt CNN architecture for solving the regression problem. Our proposed system employs a convolution layer followed by a max pool layer, 2 fully connected layers and an output layer. We use 100 different filters while sliding over 2, 3 and 4 words at a time. We employ all these filters in parallel.

B. Long Short Term Memory Network (LSTM): LSTMs (Hochreiter and Schmidhuber, 1997) are a special kind of recurrent neural network (RNN) capable of learning long-term dependencies by effectively handling the vanishing or exploding gradient problem. We use two LSTM layers on top of each other having 100 neurons in each. This was followed by 2 fully connected layers and an output layer.

C. Gated Recurrent Unit (GRU): GRUs (Cho et al., 2014) are also a special kind of RNN which can efficiently learn long-term dependencies. A key difference of GRU with LSTM is that, GRU's recurrent state is completely exposed at each time step in contrast to LSTM's recurrent state which

controls its recurrent state. Thus, comparably GRUs have lesser parameters to learn and training is computationally efficient. We use two GRU layers on top of each other having 100 neurons in each. This was followed by 2 fully connected layers and an output layer.

D. Feature based Model (SVR): We extract and implement following set of features to learn a Support Vector Regression (SVR) (Smola and Schölkopf, 2004) for predicting the sentiment score in the continuous range of -1 to +1.

- Word Tf-Idf: Term frequency-inverse document frequency (tf-idf) is a numerical statistic that is intended to reflect how important a word is to a document in a corpus. We consider tf-idf weighted counts of continuous sequences of n-grams (n=2,3,4,5) at a time.

- Lexicon Features: Sentiment lexicons are widely utilized resources in the field of sentiment analysis. Its application and effectiveness in sentiment prediction task had been widely studied. We employ two lexicons i.e. Bing Liu opinion lexicon (Ding et al., 2008) and MPQA (Wilson et al., 2005) subjectivity lexicon for news headlines domain. First we compile a comprehensive list of positive and negative words form these lexicons and then extract the following lexicon driven features.

Agreement Score : This score indicates the polarity of the sentence i.e. whether the sentence takes a polar or neutral stance. If the agreement score is 1 then it implies that the instance is of having either high positive or negative sentiment whereas, a 0 agreement score indicates a mixed or disharmony in the positive and negative sentiment implying the sentence is not polar (Rao and Srivastava, 2012).

$$A = 1 - \sqrt{1 - \left| \frac{T_{pos} - T_{neg}}{T_{pos} + T_{neg}} \right|}$$

Class score : Each text is assigned a class score indicating an overall sentiment value. The class score is -1, 0 or +1 depending on whether T_{pos} is less than, equal to or greater than T_{neg}. This helps in detecting the correct class of the sentence.

We also use four Twitter specific sentiment lexicons. These are NRC (Hashtag Context, Hashtag Sentiment, Sentiment140, Sentiment140 Context)

lexicons (Kiritchenko et al., 2014; Mohammad et al., 2013) which associate a positive or negative score to a token. Following features are extracted for each of these: i) *positive*, *negative* and *net* count. ii) maximum of *positive* and *negative* scores. iii) sum of *positive*, *negative* and *net* scores.

- Vader Sentiment: Vader sentiment (Gilbert, 2014) score is a rule-based method that generates a compound sentiment score for each sentence between -1 (extreme negative) and +1 (extreme positive). It also produces ratio of positive, negative and neutral tokens in the sentence. We obtain score and ratio of each instance in the datasets and use as feature for training.

Network parameters for CNN, LSTM & GRU: In the fully connected layers we use 50 and 10 neurons , respectively for the two hidden layers. We use *Relu* activations (Glorot et al., 2011) for intermediate layers and *tanh* activation in the final layer. We employ 20% *Dropout* (Srivastava et al., 2014) in the fully connected layers as a measure of regularization and *Adam* optimizer (Kingma and Ba, 2014) for optimization.

E. MultiLayer Perceptron (MLP) based Ensemble: Ensemble of models improves the prediction accuracy by combining the outputs of all the individual models. It exploits the strengths of all the participating models. Some of these exiting ensemble techniques cover a wide variety of traditional approaches such as bagging, boosting, majority voting, weighted voting (Xiao et al., 2013; Remya and Ramya, 2014; Ekbal and Saha, 2011) etc. In recent times Particle Swarm Optimization based ensemble technique (Akhtar et al., 2017) has been proposed for aspect based sentiment analysis. However, our current work differs significantly *w.r.t.* the methodology that we adapt as well as the problem domain that we focus on.

In this work we propose a new ensemble technique based on MLP which learns on top of the predictions of candidate models. We use a small MLP network consisting of 2 hidden layers (4 neurons in each) and an output layer. We use *Relu* activations for hidden layers and *tanh* activation in the final layer. We employ 25% dropout in the intermediate layers and use Adam optimizer during backpropagation. The output of this network serves as the final prediction value.

We separately train and tune all the four models (Section 2.A - 2.D) for both the domains. Evaluation shows that results of these individual models are encouraging, and an effective combination of these through the proposed ensemble further increases the performance.

Word Embeddings: Distributed representation models such as Word2Vec and Glove are generally very effective in a multitude of natural language processing tasks. Quality of any word embedding directly depends upon two entities: a) in-domain corpus and b) size of the corpus. Pre-trained word embeddings of Word2Vec (PWE-W2V) and GloVe (PWE-GLV) serve general purpose rather than focusing on a specific domain. Since we are addressing the problem in financial text we train and use our own embedding for financial domain corpus (FWE-W2V & FWE-GLV). We collected 126,000 financial news articles from Google News having a total of 92 million tokens. Although the corpus size is not as large as pre-trained word embedding corpus, it works reasonably well (c.f Table 1).

We observe that Word2Vec and GloVe word embeddings are quite competitive. In some cases GloVe has the advantage while in others Word2Vec performs better. Therefore, we derive a new hybrid word embedding model using a stacked denoising autoencoders (DAWE). A denoising autoencoder (Vincent et al., 2010) is a neural network which is trained to reconstruct a clean repaired input from a noisy version of the input. Following (AP et al., 2014), we combine pretrained Word2Vec and GloVe representation of a word and fed it to the network in order to capture the richness of both representations. The input to the network is a combined 600 dimensional vector (300 W2V + 300 GLV) with statistically added *salt-and-pepper* noise.

In total we employ five different word embedding models for both the domains. Dimension of all these word embeddings are set to 300. While training our deep learning models, we keep the word embeddings dynamic so that they can be fine-tuned during the process.

3 Experiments, Results and Analysis

Dataset: We evaluate our proposed approach on the benchmark datasets of SemEval-2017 shared task 5 on 'fine-grained sentiment analysis on financial microblogs and news' (Keith Cortis and Davis, 2017). The two datasets consist of financial texts from Microblogs (Twitter and StockTwits) and News, respectively. There are 1,700 and 1,142 instances of microblog messages and news headlines in the training data. The test dataset comprises of 800 microblog messages and 491 news headlines.

Experiments: We use Python based libraries Keras and Scikit-learn for the implementation. Following the shared task guideline we use cosine similarity as the metric for evaluation. Cosine score represents the degree of agreement between predicted and actual values.

Table 1 shows evaluation of our various models. In microblog dataset we obtain the best cosine similarities of 0.724, 0.727, 0.721 and 0.765 for CNN, LSTM, GRU and feature based systems, respectively. Similarly, for news datasets we obtain the best cosine similarities of 0.722, 0.720, 0.721 and 0.760. It can be observed that results for all the models are numerically comparable, however, on a qualitative side they are quite contrasting in nature. Figure 2 shows the contrasting nature of different models for microblog messages. The predicted output of different models (i.e. CNN, LSTM, GRU and SVR) are often complimentary in nature. In some case, one model predicts correctly (or, closer to the gold output), while other models make incorrect predictions and the *vice-versa*. We also observe the same phenomena for news headline. Motivated by this contrasting behavior we choose to combine the predictions of these models.
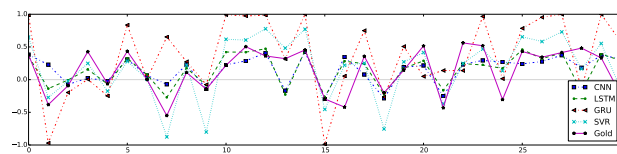

Figure 2: Contrasting nature of different models *w.r.t.* to the gold standard values; Sample size:30.

Consequently, we construct an ensemble by taking the best performing deep learning (CNN, LSTM and GRU each) and classical feature based (SVR) models using a MLP network. The proposed ensemble yields enhanced cosine scores of 0.797 and 0.786 for the microblog messages and news headline, respectively.

For comparison we choose two state-of-the-art systems (ECNU (Lan et al., 2017) and Fortia-FBK (Mansar et al., 2017)) which were the best per-

forming systems at SemEval-2017 shared task 5. ECNU reported to have obtained cosine similarity of 0.777 in microblog as compared to 0.797 cosine similarity of our proposed system, whereas, for news headlines Fortia-FBK reported cosine similarity of 0.745. ECNU employed several regressors on top of optimized feature set obtained through *hill climbing* algorithm. For the final prediction, authors averaged the predictions of different regressors. Fortia-FBK trained a CNN with the assistance of sentiment lexicons for predicting the sentiment score. It should be noted that both the systems (ECNU and Fortia-FBK) utilize different approaches to achieve the stated cosine similarities on two different domains. The proposed approach of ECNU does not perform well for the news headlines, and Fortia-FBK reported results only for the news headlines. Our proposed system performs better compared to these existing systems for both the domains.This shows that our system is more robust and generic in nature in predicting the sentiment scores. We also perform statistical significance test on the obtained results and observe that the performance gain is significant with p-value = 0.00747. Table 2 depicts the comparative results on the test datasets.

	Models	Microblog	News
Convolutional neural network (CNN)			
C1	PWE-W2V CNN	0.705	**0.722**
C2	PWE-GLV CNN	0.721	0.697
C3	FWE-W2V CNN	0.710	0.705
C4	FWE-GLV CNN	**0.724**	0.714
C5	DAWE CNN	0.697	0.698
Long short term memory (LSTM)			
L1	PWE-W2V LSTM	0.700	0.704
L2	PWE-GLV LSTM	0.715	0.683
L3	FWE-W2V LSTM	**0.727**	0.680
L4	FWE-GLV LSTM	0.717	0.691
L5	DAWE LSTM	0.722	**0.720**
Gated Recurrent Unit (GRU)			
G1	PWE-W2V GRU	0.689	**0.721**
G2	PWE-GLV GRU	0.713	0.705
G3	FWE-W2V GRU	0.715	0.687
G4	FWE-GLV GRU	0.713	0.703
G5	DAWE GRU	**0.721**	0.712
Feature - SVR			
F1	Tf-idf + Lexicon + Vader	0.752	0.749
F2	Tf-idf + Lexicon + Vader + PWE-W2V	0.740	0.731
F3	Tf-idf + Lexicon + Vader + PWE-GLV	0.758	0.745
F4	Tf-idf + Lexicon + Vader + FWE-W2V	0.709	0.702
F5	Tf-idf + Lexicon + Vader + FWE-GLV	0.732	0.725
F6	Tf-idf + Lexicon + Vader + DAWE	**0.765**	**0.760**
Ensemble			
E1	C4 + L3 + G5 + F6 (MLP)	**0.797**	0.765
E2	C1 + L5 + G1 + F6 (MLP)	0.779	**0.786**

Table 1: Cosine similarity score of various models on test dataset.

Systems	Microblogs	News
ECNU (Lan et al., 2017)	0.777	0.710
Fortia-FBK (Mansar et al., 2017)	-	0.745
Proposed System	**0.797**	**0.786**

Table 2: Comparison with the state-of-the-art systems.

3.1 Error Analysis

We also perform qualitative error analysis on the obtained results and observe that the proposed system faces problems in the following scenarios:
• Presence of implicit negation makes it a non-trivial task for the proposed system to predict the sentiment and intensity correctly. For the example given below, the overall negative sentiment is altered because of the presence of the word '*breaks*'.
Example : *Tesco breaks its downward slide by cutting sales decline in half*
Predicted: -0.694 **Actual:** 0.172

• The proposed system often confuses when the input text contains a question (?) mark.
Example : *Is $FB a BUY? Topeka Capital Markets thinks so*
Predicted: 0.363 **Actual:** -0.373.

4 Conclusion

In this paper, we have presented an ensemble network of deep learning and classical feature driven models. Evaluation on the benchmark datasets show that it performs remarkably well to identify bullish and bearish sentiments associated with companies in financial texts. We have implemented a variety of linguistic and semantic features for our analysis of the noisy text in Tweets and news headlines. Our proposed approach achieves state-of-the-art performance with the increments of 2.0 and 4.1 points over the existing systems for the tasks of sentiment prediction of financial microblog and news data. In future, we would like to extend our work by creating an end to end stock prediction system where the system would predict the future stock prices based on the sentiment score and stock value of the company.

References

Md Shad Akhtar, Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Feature selection and ensemble construction: A two-step method for aspect

based sentiment analysis. *Knowledge-Based Systems* 125:116 – 135.

Md Shad Akhtar, Ayush Kumar, Asif Ekbal, and Pushpak Bhattacharyya. 2016. A Hybrid Deep Learning Architecture for Sentiment Analysis. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers, December 11-16, 2016*. Osaka, Japan, pages 482–493.

Sarath Chandar AP, Stanislas Lauly, Hugo Larochelle, Mitesh Khapra, Balaraman Ravindran, Vikas C Raykar, and Amrita Saha. 2014. An Autoencoder Approach to Learning Bilingual Word Representations. In *Advances in Neural Information Processing Systems*. pages 1853–1861.

KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *CoRR* abs/1409.1259. http://arxiv.org/abs/1409.1259.

Marjan Van de Kauter, Diane Breesch, and Vronique Hoste. 2015. Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications* 42(11):4999 – 5010.

Xiaowen Ding, Bing Liu, and Philip S Yu. 2008. A Holistic Lexicon-Based Approach to Opinion Mining. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, pages 231–240.

Asif Ekbal and Sriparna Saha. 2011. Weighted Vote-Based Classifier Ensemble for Named Entity Recognition: A Genetic Algorithm-Based Approach. *ACM Transactions on Asian Language Information Processing* 10:9:1–9:37.

CJ Hutto Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep Sparse Rectifier Neural Networks. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*. Journal of Machine Learning Research - Workshop and Conference Proceedings, volume 15, pages 315–323.

Rohitha Goonatilake and Susantha Herath. 2007. The Volatility of the Stock Market and News. *International Research Journal of Finance and Economics* 3(11):53–65.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural computation* 9(8):1735–1780.

Andre Freitas Tobias Daudert Manuela Huerlimann Manel Zarrouk Keith Cortis and Brian Davis. 2017. SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News. In *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*. ACL, Vancouver, Canada, pages 519–535.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. pages 1746–1751.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980. http://dblp.uni-trier.de/db/journals/corr/corr1412.html.

Svetlana Kiritchenko, Xiaodan Zhu, and Saif M. Mohammad. 2014. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research (JAIR)* 50:723–762.

Man Lan, Mengxiao Jiang, and Yuanbin Wu. 2017. ECNU at SemEval-2017 Task 5: An Ensemble of Regression Algorithms with Effective Features for Fine-grained Sentiment Analysis in Financial Domain. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL, Vancouver, Canada, pages 888–893.

Youness Mansar, Lorenzo Gatti, Sira Ferradans, Marco Guerini, and Jacopo Staiano. 2017. Fortia-FBK at SemEval-2017 Task 5: Bullish or Bearish? Inferring Sentiment towards Brands from Financial News Headlines. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL, Vancouver, Canada, pages 817–822.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. Lake Tahoe, NV, USA, pages 3111–3119.

Saif Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, Georgia, USA, pages 321–327.

Neil O'Hare, Michael Davy, Adam Bermingham, Paul Ferguson, Páraic Sheridan, Cathal Gurrin, and Alan F. Smeaton. 2009. Topic-dependent Sentiment Analysis of Financial Blogs. In *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*. ACM, New York, NY, USA, TSA '09, pages 9–16.

Nuno Oliveira, Paulo Cortez, and Nelson Areal. 2013. On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. In

EPIA. Springer, volume 8154 of *Lecture Notes in Computer Science*, pages 355–365.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-2014)*. ACL, Doha, Qatar, pages 1532–1543.

Tushar Rao and Saket Srivastava. 2012. Analyzing stock market movements using twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*. IEEE Computer Society, pages 119–123.

K. R. Remya and J. S. Ramya. 2014. Using weighted majority voting classifier combination for relation classification in biomedical texts. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. pages 1205–1209.

Robert P. Schumaker and Hsinchun Chen. 2009. Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System. *ACM Transactions on Information Systems* 27(2).

Prerana Singhal and Pushpak Bhattacharyya. 2016. Borrow a Little from your Rich Cousin: Using Embeddings and Polarities of English Words for Multilingual Sentiment Classification. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016): Technical Papers, December 11-16, 2016*. Osaka, Japan, pages 3053–3062.

Alex J. Smola and Bernhard Schölkopf. 2004. A Tutorial on Support Vector Regression. *Statistics and Computing* 14(3):199–222.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15:1929–1958.

Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* 11(Dec):3371–3408.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 347–354.

Tong Xiao, Jingbo Zhu, and Tongran Liu. 2013. Bagging and Boosting Statistical Machine Translation Systems. *Artif. Intell.* 195:496–527.