NEUROSENT-PDI at SemEval-2018 Task 1: Leveraging a Multi-Domain Sentiment Model for Inferring Polarity in Micro-blog Text

Mauro Dragoni
Fondazione Bruno Kessler
Via Sommarive 18
Povo, Trento, Italy
`dragoni@fbk.eu`

Abstract

This paper describes the NeuroSent system that participated in SemEval 2018 Task 1. Our system takes a supervised approach that builds on neural networks and word embeddings. Word embeddings were built by starting from a repository of user generated reviews. Thus, they are specific for sentiment analysis tasks. Then, tweets are converted in the corresponding vector representation and given as input to the neural network with the aim of learning the different semantics contained in each emotion taken into account by the SemEval task. The output layer has been adapted based on the characteristics of each subtask. Preliminary results obtained on the provided training set are encouraging for pursuing the investigation into this direction.

1 Introduction

Sentiment Analysis is a natural language processing (NLP) task (Dragoni et al., 2015) which aims at classifying documents according to the opinion expressed about a given subject (Federici and Dragoni, 2016a,b). Many works available in the literature address the sentiment analysis problem without distinguishing domain specific information of documents when sentiment models are built. The necessity of investigating this problem from a multi-domain perspective is led by the different influence that a term might have in different contexts. The idea of adapting terms polarity to different domains emerged only in the last decade (Blitzer et al., 2007; Dragoni and Petrucci, 2017). Multi-domain sentiment analysis approaches discussed in the literature focus on building models for transferring information between pairs of domains (Dragoni, 2015; Petrucci and Dragoni, 2015). While on the one hand such approaches allow to propagate specific domain information to others, their drawback is the neces-

sity of building new transfer models every time a new domain has to be analyzed. Thus, such approaches do not have a great generalization capability of analyzing texts, because transfer models are limited to the N domains used for building the models.

The NeuroSent tool applied in SemEval 2018 Task 1 (Mohammad et al., 2018) leverages on the following pillars: (i) the use of word embeddings for representing each word contained in raw sentences; (ii) the word embeddings are generated from an opinion-based corpus instead of a general purpose one (like news or Wikipedia); (iii) the design of a deep learning technique exploiting the generated word embeddings for training the sentiment model; and (iv) the use of multiple output layers for combining domain overlap scores with domain-specific polarity predictions.

The last point enables the exploitation of linguistic overlaps between domains, which can be considered one of the pivotal assets of our approach. This way, the overall polarity of a document is computed by aggregating, for each domain, the domain-specific polarity value multiplied by a belonging degree representing the overlap between the embedded representation of the whole document and the domain itself. Within the SemEval 2018 Task 1 challenge, we consider with the term *domain* one of the emotions that have been considered into the provided datasets.

2 Related Work

Sentiment analysis from the multi-task and multi-domain perspective is a research field which started to be explored only in the last decade. According to the nomenclature widely used in the literature (see (Blitzer et al., 2007; Dragoni and Petrucci, 2017)), we call *domain* a set of documents about similar topics, e.g. a set of reviews

about similar products like mobile phones, books, movies, etc.. The massive availability of multi-domain corpora in which similar opinions are expressed about different topics opened the scenario for new challenges. Researchers tried to train models capable to acquire knowledge from a specific domain and then to exploit such a knowledge for working on documents belonging to different ones. This strategy was called domain adaptation. The use of domain adaptation techniques demonstrated that opinion classification is highly sensitive to the domain from which the training data is extracted. The reason is that when using the same words, and even the same language constructs, we may obtain different opinions, depending on the domain. The classic scenario occurs when the same word has positive connotations in one domain and negative connotations in another one, as we showed within the examples presented in Section 1.

Several approaches related to multi-domain sentiment analysis have been proposed. Roughly speaking, all of these approaches rely on one of the following ideas: (i) the transfer of learned classifiers across different domains (Blitzer et al., 2007; Pan et al., 2010; Bollegala et al., August 2013; Xia et al., May-June 2013), and (ii) the use of propagation of labels through graph structures (Ponomareva and Thelwall, 2013; Tsai et al., March 2013; Dragoni et al., April 2015; Dragoni, 2015, 2017; Petrucci and Dragoni, 2017, 2016, 2015; Dragoni et al., 2014; Dragoni and Petrucci, 2018).

While on the one hand such approaches demonstrated their effectiveness in working in a multi-domain environment, on the other hand they suffered by the limitation of being influenced by the linguistic overlap between domains. Indeed, such an overlap leads learning algorithms to infer similar polarity values to domains that are similar from the linguistic perspective.

The adoption of evolutionary algorithms within the sentiment analysis research field is quite recent. First studies focused on the use of evolutionary solutions for modeling financial indicators by starting from investors sentiments (Yamada and Ueda, 2005; Chen and Chang, 2005; Huang et al., 2012; Yang et al., 2017; Simoes et al., 2017). Here, the evolutionary component was used for learning the trend of financial indicators with respect to the sentiment information extracted from opinions provided by the investors. With respect

to these papers, we propose an approach adopting evolutionary computation to a more fine-grained level where the evolution component affects also the polarities of opinion concepts.

Studies considering the use of evolutionary algorithms for optimizing the polarity values of opinion concepts have been proposed only recently (Ferreira et al., 2015; Onan et al., 2016, 2017). However, these works focused on learning candidate refinements of opinion concepts polarity without considering the context dimension associated with them. A variant of this problem is the use of polarity adaptation strategy in the field of social media and microblogs (Alahmadi and Zeng, 2015; Wang et al., 2014; Keshavarz and Abadeh, 2017; Hu et al., 2016; Fu et al., 2016; Gong et al., 2016).

With respect to state of the art, this work represents the first exploration of evolutionary algorithms for multi-domain sentiment analysis with the aim of learning multiple dictionaries of opinion concepts. Moreover, we differ from the literature by do not considering the propagation of polarity information across domain (i.e., we keep them completely separated) in order to avoid transfer learning drawbacks.

3 System Implementation

NeuroSent has been entirely developed in Java with the support of the Deeplearning4j library [1] and it is composed by following two main phases:

- Generation of Word vectors (Section 3.1): raw text, appropriately tokenized using the Stanford CoreNLP Toolkit, is provided as input to a 2-layers neural network implementing the skip-gram approach with the aim of generating word vectors.

- Learning of Sentiment Model (Section 3.2): word vectors are used for training a recurrent neural network with an output layer customized based on the addressed subtask. The customizations have been explained in Section 4.

In the following subsections, we describe in more detail each phase by providing also the settings used for managing our data.

3.1 Generation of Word Vectors

The generation of the word vectors has been performed by applying the skip-gram algorithm on

---

[1]https://deeplearning4j.org/

the raw natural language text extracted from the smaller version of the SNAP dataset (McAuley and Leskovec, 2013). The rationale behind the choice of this dataset focuses on three reasons:

- the dataset contains only opinion-based documents. This way, we are able to build word embeddings describing only opinion-based contexts.

- the dataset is multi-domain. Information contained into the generated word embeddings comes from specific domains, thus it is possible to evaluate how the proposed approach is general by testing the performance of the created model on test sets containing documents coming from the domains used for building the model or from other domains.

- the dataset is smaller with respect to other corpora used in the literature for building other word embeddings that are currently freely available, like the Google News ones. [2] Indeed, as introduced in Section 1, one of our goal is to demonstrate how we can leverage the use of dedicated resources for generating word embeddings, instead of corpora's size, for improving the effectiveness of classification systems.

The aspect of considering only opinion-based information for generating word embeddings is one of the peculiarity of our system. While embeddings currently available are created from big corpora of general purpose texts (like news archives or Wikipedia pages), ours are generated by using a smaller corpus containing documents strongly related to the problem that the model will be thought for. On the one hand, this aspect may be considered a limitation of the proposed solution due to the requirement of training a new model in case of problem change. However, on the other hand, the usage of dedicated resources would lead to the construction of more effective models.

Word embeddings have been generated by the Word2Vec implementation integrated into the Deeplearning4j library. The algorithm has been set up with the following parameters: the size of the vector to 64, the size of the window used as input of the skip-gram algorithm to 5, and the minimum word frequency was set to 1. The reason for

---
[2]https://github.com/mmihaltz/word2vec-GoogleNews-vectors

which we kept the minimum word frequency set to 1 is to avoid the loss of rare but important words that can occur in domain specific documents.

3.2 Learning of The Sentiment Model

The sentiment model is built by starting from the word embeddings generated during the previous phase.

The first step consists in converting each textual sentence contained within the dataset into the corresponding numerical matrix $\mathbf{S}$ where we have in each row the word vector representing a single word of the sentence, and in each column an embedding feature. Given a sentence s, we extract all tokens t_i, with $i \in [0, n]$, and we replace each t_i with the corresponding embedding $\mathbf{w}$. During the conversion of each word in its corresponding embedding, if such embedding is not found, the word is discarded. At the end of this step, each sentence contained in the training set is converted in a matrix $\mathbf{S} = [\mathbf{w}^{\langle 1 \rangle}, \dots, \mathbf{w}^{\langle n \rangle}]$.

Before giving all matrices as input to the neural network, we need to include both padding and masking vectors in order to train our model correctly. Padding and masking allows us to support different training situations depending on the number of the input vectors and on the number of predictions that the network has to provide at each time step. In our scenario, we work in a many-to-one situation where our neural network has to provide one prediction (sentence polarity and domain overlap) as result of the analysis of many input vectors (word embeddings).

Padding vectors are required because we have to deal with the different length of sentences. Indeed, the neural network needs to know the number of time steps that the input layer has to import. This problem is solved by including, if necessary, into each matrix $\mathbf{S}_k$, with $k \in [0, z]$ and z the number of sentences contained in the training set, null word vectors that are used for filling empty word's slots. These null vectors are accompanied by a further vector telling to the neural network if data contained in a specific positions has to be considered as an informative embedding or not.

A final note concerns the back propagation of the error. Training recurrent neural networks can be quite computationally demanding in cases when each training instance is composed by many time steps. A possible optimization is the use of truncated back propagation through time (BPTT)

that was developed for reducing the computational complexity of each parameter update in a recurrent neural network. On the one hand, this strategy allows to reduce the time needed for training our model. However, on the other hand, there is the risk of not flowing backward the gradients for the full unrolled network. This prevents the full update of all network parameters. For this reason, even if we work with recurrent neural networks, we decided to do not implement a BPTT approach but to use the default backpropagation implemented into the DL4J library.

Concerning information about network structure, the input layer was composed by 64 neurons (i.e. embedding vector size), the hidden RNN layer was composed by 128 nodes, and the output layers with a different number of nodes based on the addressed subtask. The network has been trained by using the Stochastic Gradient Descent with 1000 epochs and a learning rate of 0.002.

4 The Tasks

The SemEval 2018 Task 1 is composed by a set of five subtasks aiming to attract systems able to automatically determine the intensity of emotions and the intensity of sentiment of tweets' authors. Then, organizers included also a multi-label emotion classification task for tweets. For each task, there were provide separate training and test datasets for four languages: English, Arabic, and Spanish. The proposed system implements a strategy only for the English language. Below, we provide a summary of the five subtasks including how we configured the output layer of our neural network.

Subtask #1: EI-reg Given a tweet and an emotion E, the system has to determine the intensity of E that best represents the mental state of the tweet's author by providing a real-valued score between 0 and 1. Here, four emotions are considered: anger, fear, joy, and sadness. Separated datasets have been provided for training the system. The output layer of our neural network is composed by a single neuron implementing the SIGMOID activation function.

Subtask #2: EI-oc Given a tweet and an emotion E, the system has to classify the tweet into one of four ordinal classes of intensity of E that best represents the mental state of the tweet's author. Also here, four emotions are considered:

anger, fear, joy, and sadness. Separated datasets have been provided for training the system. The output layer of our neural network is composed by four neurons and the SOFTMAX strategy has been implemented for selecting the most candidate emotion intensity class.

Subtask #3: V-reg Given a tweet, the system has to determine the valence of a sentiment that best represents the mental state of tweet's author by providing a real-valued score between 0 and 1. The output layer of our neural network is composed by a single neuron implementing the SIGMOID activation function.

Subtask #4: V-oc Given a tweet, the system has to classify it into one of seven ordinal classes (from -3 to 3) corresponding to various levels of positive and negative sentiment intensity. The output layer of our neural network is composed by seven neurons and the SOFTMAX strategy has been implemented for selecting the most candidate emotion intensity class.

Subtask #5: E-c Given a tweet, the system has to classify it as a *neutral*, or *no emotion* or as one, or more, of eleven given emotions that best represent the mental state of the tweet's author. The eleven emotions are: anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust. The output layer of our neural network is composed by eleven neurons implementing the SIGMOID activation function. This way, each emotion has been managed separately.

The NeuroSent system has been applied to all five subtasks. In Section 5, we report the preliminary results obtained by NeuroSent on the training set compared with a set of baselines.

5 In-Vitro Evaluation

The NeuroSent approach have been preliminarily evaluated by adopting the Dranziera protocol (Dragoni et al., 2016).

The validation procedure leverages on a five-fold cross evaluation setting in order to validate the robustness of the proposed solution. The approach has been compared with four baselines: Support Vector Machine (SVM) (Chang and Lin, 2011), Naive Bayes (NB) and Maximum Entropy (ME) (McCallum, 2002), and Convolutional Neural Network (Chaturvedi et al., 2016).

In Table 1, we provide for subtasks two, four, and five the average Pearson correlation obtained

Approach	Task #1.1	Task #1.2	Task #1.3	Task #1.4	Task #1.5
Support Vector Machine	0.3189	0.4890	0.3698	0.5145	0.3498
Naive-Bayes	0.2944	0.4956	0.3544	0.5387	0.4167
Maximum Entropy	0.2765	0.5073	0.3025	0.5777	0.4178
CNN Architecture	0.2433	0.6037	0.2466	0.5895	0.5487
NeuroSent	**0.2187**	**0.6687**	**0.2079**	**0.6241**	**0.5814**

Table 1: Results obtained on the training set by NeuroSent and by the four baselines.

on the five folds in which the training set has been split. While, for subtasks one and three, we provide the average mean square error.

The obtained results demonstrated the suitability of NeuroSent with respect to the adopted baselines. We may also observed how solutions based on neural networks obtained a significant improvement with respect to the others for the Tasks #1.2 and #1.4.

Then, for Tasks #1.2, #1.4, and #1.5, we performed a detailed error analysis concerning the performance of NeuroSent. In general, we observed how our strategy tends to provide false negative predictions. An in depth analysis of some incorrect predictions highlighted that the embedded representations of some positive opinion words are very close to the space region of negative opinion words. Even if we may state that the confidence about positive predictions is very high, this scenario leads to have a predominant negative classification for borderline instances.

On the one hand, a possible action for improving the effectiveness our strategy is to increase the granularity of the embeddings (i.e. augmenting the size of the embedding vectors) in order to increase the distance between the positive and negative polarities space regions. On the other hand, by increasing the size of embedding vectors, the computational time for building, or updating, the model and for evaluating a single instance increases as well. Part of the future work, will be the analysis of more efficient neural network architectures able to manage augmented embedding vectors without negatively affecting the efficiency of the platform.

6 Conclusion

In this paper, we described the NeuroSent system presented at SemEval 2018 Task 1. Our system makes use of artificial neural networks to classify tweets by polarity or for detecting emotion levels. The results obtained on the training set

demonstrated that the adopted solution is promising and worthy of investigation. Therefore, future work will focus on improving the system by exploring the integration of sentiment knowledge bases (Dragoni et al., 2015) in order to move toward a more cognitive approach.

References

Dimah Hussain Alahmadi and Xiao-Jun Zeng. 2015. Twitter-based recommender system to address cold-start: A genetic algorithm based trust modelling and probabilistic sentiment analysis. In *27th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2015, Vietri sul Mare, Italy, November 9-11, 2015*, pages 1045–1052. IEEE Computer Society.

John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic*. The Association for Computational Linguistics.

Danushka Bollegala, David J. Weir, and John A. Carroll. August 2013. Cross-domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Trans. Knowl. Data Eng.*, 25(8):1719–1731.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM TIST*, 2(3):27:1–27:27.

Iti Chaturvedi, Erik Cambria, and David Vilares. 2016. Lyapunov filtering of objectivity for spanish sentiment model. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 4474–4481. IEEE.

An-Pin Chen and Yung-Hua Chang. 2005. Using extended classifier system to forecast s&p futures based on contrary sentiment indicators. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2005, 2-4 September 2005, Edinburgh, UK*, pages 2084–2090. IEEE.

Mauro Dragoni. 2015. Shellfbk: An information retrieval-based system for multi-domain sentiment

analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval '2015, pages 502–509, Denver, Colorado. Association for Computational Linguistics.

Mauro Dragoni. 2017. Extracting linguistic features from opinion data streams for multi-domain sentiment analysis. In *Proceedings of the 3rd International Workshop at ESWC on Emotions, Modality, Sentiment Analysis and the Semantic Web co-located with 14th ESWC 2017, Portroz, Slovenia, May 28, 2017.*, volume 1874 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Mauro Dragoni and Giulio Petrucci. 2017. A neural word embeddings approach for multi-domain sentiment analysis. *IEEE Trans. Affective Computing*, 8(4):457–470.

Mauro Dragoni and Giulio Petrucci. 2018. A fuzzy-based strategy for multi-domain sentiment analysis. *Int. J. Approx. Reasoning*, 93:59–73.

Mauro Dragoni, Andrea Tettamanzi, and Célia da Costa Pereira. 2016. DRANZIERA: an evaluation protocol for multi-domain opinion mining. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23-28, 2016.* European Language Resources Association (ELRA).

Mauro Dragoni, Andrea G. B. Tettamanzi, and Célia da Costa Pereira. 2014. A fuzzy system for concept-level sentiment analysis. In *Semantic Web Evaluation Challenge - SemWebEval 2014 at ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers*, volume 475 of *Communications in Computer and Information Science*, pages 21–27. Springer.

Mauro Dragoni, Andrea G. B. Tettamanzi, and Célia da Costa Pereira. 2015. Propagating and aggregating fuzzy polarities for concept-level sentiment analysis. *Cognitive Computation*, 7(2):186–197.

Mauro Dragoni, Andrea Giovanni Battista Tettamanzi, and Célia da Costa Pereira. April 2015. Propagating and aggregating fuzzy polarities for concept-level sentiment analysis. *Cognitive Computation*, 7(2):186–197.

Marco Federici and Mauro Dragoni. 2016a. A knowledge-based approach for aspect-based opinion mining. In *Semantic Web Challenges - Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*, volume 641 of *Communications in Computer and Information Science*, pages 141–152. Springer.

Marco Federici and Mauro Dragoni. 2016b. Towards unsupervised approaches for aspects extraction. In *Joint Proceedings of the 2th Workshop on Emotions, Modality, Sentiment Analysis and the Semantic Web and the 1st International Workshop on Extraction and Processing of Rich Semantics from*

Medical Texts co-located with ESWC 2016, Heraklion, Greece, May 29, 2016., volume 1613 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Lohann Ferreira, Mariza Dosciatti, Júlio C. Nievola, and Emerson Cabrera Paraiso. 2015. Using a genetic algorithm approach to study the impact of imbalanced corpora in sentiment analysis. In *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2015, Hollywood, Florida. May 18-20, 2015.*, pages 163–168. AAAI Press.

Peng Fu, Zheng Lin, Hailun Lin, Fengcheng Yuan, Weiping Wang, and Dan Meng. 2016. Quantifying the effect of sentiment on topic evolution in chinese microblog. In *Web Technologies and Applications - 18th Asia-Pacific Web Conference, APWeb 2016, Suzhou, China, September 23-25, 2016. Proceedings, Part I*, volume 9931 of *Lecture Notes in Computer Science*, pages 531–542. Springer.

Lin Gong, Mohammad Al Boni, and Hongning Wang. 2016. Modeling social norms evolution for personalized sentiment classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.

Yan Hu, Xiaofei Xu, and Li Li. 2016. Analyzing topic-sentiment and topic evolution over time from social media. In *Knowledge Science, Engineering and Management - 9th International Conference, KSEM 2016, Passau, Germany, October 5-7, 2016, Proceedings*, volume 9983 of *Lecture Notes in Computer Science*, pages 97–109.

Chien-Feng Huang, Tsung-Nan Hsieh, Bao Rong Chang, and Chih-Hsiang Chang. 2012. A comparative study of regression and evolution-based stock selection models for investor sentiment. In *2012 Third International Conference on Innovations in Bio-Inspired Computing and Applications, Kaohsiung City, Taiwan, September 26-28, 2012*, pages 73–78. IEEE.

Hamidreza Keshavarz and Mohammad Saniee Abadeh. 2017. ALGA: adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs. *Knowl.-Based Syst.*, 122:1–16.

Julian J. McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12-16, 2013*, pages 165–172. ACM.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. `http://mallet.cs.umass.edu`.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018.

Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.

Aytug Onan, Serdar Korukoglu, and Hasan Bulut. 2016. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Syst. Appl.*, 62:1–16.

Aytug Onan, Serdar Korukoglu, and Hasan Bulut. 2017. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Inf. Process. Manage.*, 53(4):814–833.

Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pages 751–760. ACM.

Giulio Petrucci and Mauro Dragoni. 2015. An information retrieval-based system for multi-domain sentiment analysis. In *Semantic Web Evaluation Challenges - Second SemWebEval Challenge at ESWC 2015, Portorož, Slovenia, May 31 - June 4, 2015, Revised Selected Papers*, volume 548 of *Communications in Computer and Information Science*, pages 234–243. Springer.

Giulio Petrucci and Mauro Dragoni. 2016. The IRMUDOSA system at ESWC-2016 challenge on semantic sentiment analysis. In *Semantic Web Challenges - Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*, volume 641 of *Communications in Computer and Information Science*, pages 126–140. Springer.

Giulio Petrucci and Mauro Dragoni. 2017. The IRMUDOSA system at ESWC-2017 challenge on semantic sentiment analysis. In *Semantic Web Challenges - 4th SemWebEval Challenge at ESWC 2017, Portoroz, Slovenia, May 28 - June 1, 2017, Revised Selected Papers*, volume 769 of *Communications in Computer and Information Science*, pages 148–165. Springer.

Natalia Ponomareva and Mike Thelwall. 2013. Semi-supervised vs. cross-domain graphs for sentiment analysis. In *Recent Advances in Natural Language Processing, RANLP 2013, 9-11 September, 2013, Hissar, Bulgaria*, pages 571–578. RANLP 2013 Organising Committee/ACL.

Carlos Simoes, Rui Ferreira Neves, and Nuno Horta. 2017. Using sentiment from twitter optimized by genetic algorithms to predict the stock market. In *2017 IEEE Congress on Evolutionary Computation, CEC 2017, Donostia, San Sebastián, Spain, June 5-8, 2017*, pages 1303–1310. IEEE.

Angela Charng-Rurng Tsai, Chi-En Wu, Richard Tzong-Han Tsai, and Jane Yung jen Hsu. March 2013. Building a concept-level sentiment dictionary based on commonsense knowledge. *IEEE Int. Systems*, 28(2):22–30.

Zhitao Wang, Zhiwen Yu, Zhu Wang, and Bin Guo. 2014. Investigating sentiment impact on information propagation and its evolution in microblog. In *2014 International Conference on Behavioral, Economic, and Socio-Cultural Computing, BESC 2014, Shanghai, China, October 30 - November 1, 2014*, pages 33–39. IEEE.

Rui Xia, Chengqing Zong, Xuelei Hu, and Erik Cambria. May-June 2013. Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IEEE Int. Systems*, 28(3):10–18.

Takashi Yamada and Kazuhiro Ueda. 2005. Explanation of binarized time series using genetic learning model of investor sentiment. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2005, 2-4 September 2005, Edinburgh, UK*, pages 2437–2444. IEEE.

Steve Y. Yang, Sheung Yin Kevin Mo, Anqi Liu, and Andrei Kirilenko. 2017. Genetic programming optimization for a sentiment feedback strength based trading strategy. *Neurocomputing*, 264:29–41.