Sentiment Analysis on English Financial News[1]

Guzel Rinatovna Eremeeva[2], Ekaterina Vladimirovna Martynova[3], Amina Aidarovna Khakimova[3], & Liliya Ernstovna Ilikova[3]

Abstract

This article is devoted to comparison of various approaches towards sentiment analysis (polarity classification) of the English financial news. The purpose of the work is to check, weather higher quality English texts would have higher classification quality and the links between quality of polarity classification and the architectures of the chosen systems. The approaches, presented on SemEval-2017 competition in the appropriate category of tasks, are taken as the basis. A specially collected dataset is used as the source of English financial news. In the end, a conclusion is made about the relationship between the quality of the source of English financial news and the quality of the news data sentiment analysis, made by the compared systems.

Keywords: Linguistics; Text; Sentiment Analysis; Polarity Detection.

1. Introduction

Sentiment analysis is widely used to analyze people opinions and moods in various English texts. Sentiment analysis of financial domain is different from other domains. The sentiments of English texts are important source that forms market dynamics globally (Goonatilake & Herath, 2007; Van de Kauter et al., 2015).

In the literature there are different ways to formalize the model of opinions. Different terminology is also used. In English, this area of study is commonly referred to as opinion mining and sentiment analysis. Despite the fact that the key is only one of the characteristics of the opinion, it is the problem of classification of the key is the most frequently studied in our days. This can be explained by several reasons:

1. The definition of the author and the topic is much more difficult than the classification of key, so it makes sense to first solve a simpler problem, and then switch to the rest (Van de Kauter et al., 2015).

---

[1] Please cite this paper as follows:

Eremeeva, G. R., Martynova, E.V., Khakimova, A. A., Ilikova, L. E. (2019). Sentiment analysis on English financial news. *Journal of Research in Applied Linguistics, 10*(SP), 574-582.

[2] Kazan Federal University; *guzeliksanova@rambler.ru*

[3] Kazan Federal University

2. In many cases, we only need to determine the key, because other characteristics are already known to us. For example, if we collect opinions from blogs, usually the authors of the opinions are the authors of the posts, i.e. we do not need to determine the author. Also, we often already know the topic: for example, if we search for the keyword "Windows 8" in Twitter, then we only need to determine the tone of the found tweets. Of course, this does not work in all cases, but only in most of them. But these assumptions can greatly simplify the already difficult task (Cortis et al., 2017).

Sentiment analysis finds its practical application in different fields:

- sociology — collect data from social. networks (such as religious beliefs)
- political science — collect data from blogs about political views of the population
- marketing — analyze Twitter to find out which laptop model is most in demand
- medicine and psychology — determine depression in social users. Networks'.

Sentiment analysis is usually defined as one of the tasks of computational linguistics, i.e. it is assumed that we can find and classify sentiment using natural language processing tools (such as taggers, parsers, etc.) (Kar et al., 2017). Making a large generalization, we can divide the existing approaches into the following categories:

1. Rule-based approaches
2. Dictionary-based approaches
3. Machine learning with a teacher
4. Machine learning without a teacher

The first type of system consists of a set of rules, applying which the system makes a conclusion about the tonality of the text. For example, for the sentence "I love juice", you can apply the following rule:

If the predicate ("love") is included in the positive set of verbs ("love", "adore", "like", "approve" ...) and there is no negation in the sentence, then classify the key as "positive".

Many commercial systems use this approach, despite the fact that it requires a lot of costs, because for the good operation of the system you need to make a large number of rules. Often the rules are tied to a specific domain (for example, "restaurant theme") and when you change the domain ("camera review") you need to re-make the rules (Pivovarova et al., 2017). Nevertheless, this approach is the most accurate if there is a good rule base, but it is not interesting for research.

Dictionary-based approaches use so-called affective lexicons (tonal dictionaries) to analyze text. In a simple form, a tonal dictionary is a list of words with the meaning of the key for each word (Soboleva, 2015).

To analyze the text, you can use the following algorithm: first, assign each word in the text to its key value from the dictionary (if it is present in the dictionary), and then calculate the general tone of the entire text. There are different ways to calculate the General tonality. The simplest of them is the arithmetic mean of all values. More difficult — to train the classifier (eg. neural network.)

Machine learning with a teacher is the most common method used in research. Its essence is to train the machine classifier on a collection of pre-labeled texts, and then use the resulting model to analyze new documents. It is about this method that I will tell you further.

Machine learning without a teacher is probably the most interesting and at the same time the least accurate method of sentiment analysis. One example of this method is automatic document clustering.

Stylistic sentiment (John & Vechtomova, 2017) of the text helps the author to express:

- attitude to reality,
- values, ideals, beliefs,
- mental attitude,
- self-assessment, assessment of the subject of speech and the addressee on an emotional (not rational) basis.

sentiment has a powerful impact on the psycho-emotional sphere of the addressee. Emotional and expressive influence is transmitted

- language means (morphology, vocabulary (including tropeic), word formation, phonics),
- text models (rhetorical and compositional techniques: figures of speech, architectonics) (Moore & Rayson, 2017).

If the thematic and logical chains of the text are linear, the sentiment (as well as modality) is located in the text by fields of different emotional and volitional saturation. This means that in different parts of the text the author concentrates linguistic, compositional means, emotionally enhancing the significance of the fact, evaluation, will.

The type of textual sentiment depends on the type of publication, the editorial policy of the mass media. In modern journalism, you can find three main types of stylistic sentiment corresponding to the "mood" (profile) of publications (Cabanskiet et al., 2017):

1) opposition (negative tone, often resulting in verbal aggression),

2) approving (positive tone, basic speech tactics – compliment),

3) emphasized objective (neutral stylistic tone).

The "opposition" texts are characterized by an aggressive tone, when the publicist, instead of thoroughly and objectively analyzing the arguments of the other side, seeks to seize the initiative and discredit his "opponent" by any means. The main speech tactic in this type of text is labeling (Saleiro et al., 2017). The absence of logical arguments is veiled by psychoemotional images that paint a negative picture of modern reality. The author's position in such texts is totalitarian, i.e. there is no dialogue with the opponent, but the suppression of someone else's opinion (Moore & Rayson, 2017).

In the texts of pro-government publications, the communicative task of which is to assert in the mass consciousness the ideology adopted by the ruling elite, the stylistic tone, on the contrary, receives an approving coloring, social and personal ideologems are surrounded by a "sweet" context: approval and exhortation (persuasion, Council). In texts of this type, the addressee uses tricks, in particular, the replacement of logical arguments with positively colored emotional definitions, which contributes to the promotion of a given idea and its assimilation through emotions, not reason.

The tone of the emphatic objective: the author seeks to show its veracity and impartiality (Soboleva, 2017). The texts do not contain explicit signals of evaluation, but when carefully read, the impact occurs at the deep level of the text (the number of positive and negative facts and details, the compositional strengthening of the given meanings with the help of variable repetitions, etc.) (Ghosh et al., 2015). Here the influencing force on consciousness of the addressee is implicit, but not less effective as does not cause rejection and objection.

To gain more insight about sentiments in English financial texts, SemEval, International Workshop on Semantic Evaluation, organized a special task, “Fine-Grained Sentiment Analysis on Financial Microblogs and News”, for the competition of 2017.

Our purpose was to investigate the results of the competition and to perform our own evaluation of selected competitors on the Corpus of Business News, a source of English Financial Texts that is larger and more informative than the texts, provided on the competition. The Corpus of Business News was introduced and provided by the competitor of SemEval-2017, HCS team from the University of Helsinki (Cortis K. et al., 2017).

Several competitors provide their systems in open-source format, so it is possible to reproduce and re-evaluate the works. In this article, we limited our task to comparing only two systems: the system (Kar et al., 2017), by RiTUAL-UH team, and the system (Pivovarova et al., 2017), provided by the authors of the Corpus of Business News, HCS team (Soboleva, 2015).

2. Methodology

The papers by the competitors provide detailed explanations of their systems. The main architectures of the models are shown below:

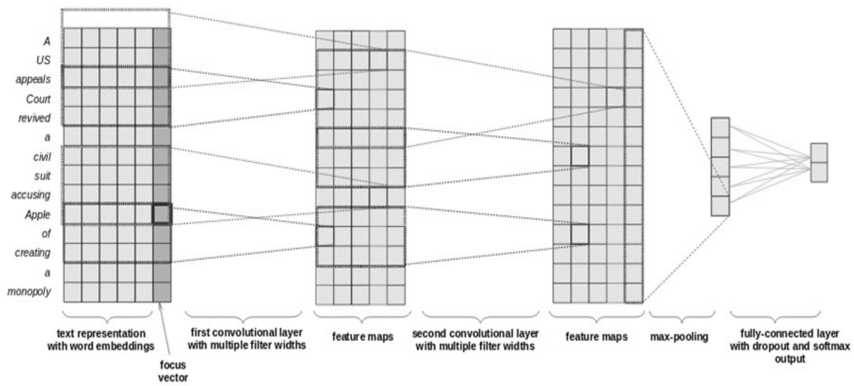

Fig. 1. Main Architecture of CNN Model by HCS Team

HCS system relies on CNN model. RiTUAL-UH system has the following structure, that combines CNNs and Bi GRUs:

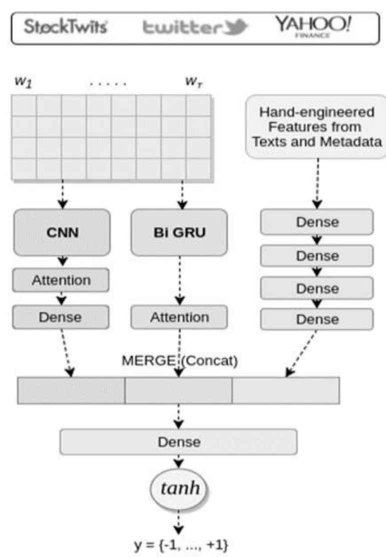

Fig. 2. Main Architecture of the System by RiTUAL-UH Team

Cosine similarity (John & Vechtomova, 2017), the same metric, as used on SemEval-2017, was chosen here for model evaluation. Its formula, where G is the vector of gold standard scores and P is the vector of scores predicted by the system, is presented below:

$$score = weight \times cosine(G, P),$$

$$weight = \frac{|P|}{|G|}, cosine(G,P) = \frac{\sum_{i=1}^{n} G_i \times P_i}{\sqrt{\sum_{i=1}^{n} G_i^2} \times \sqrt{\sum_{i=1}^{n} P_i^2}}$$

2.1. Technical Details and Comments: HCS

The system was tested in the same configuration as initially given, as it is already trained on other data samples from the Corpus of Business News with good configuration options (Moore & Rayson, 2017). But it might be worthy to experiment further with configurations in ‘polarity/cnn-classifiers’ and to retrain the model.

The implementation of HCS system, adapted for a new data source, could be found here: https://bitbucket.org/sentimentals/semeval_2017_solution_analysis/

2.2. Technical Details and Comments: RiTUAL-UH

This system uses caching heavily. Lots of additional features should be generated and saved as pkl-s before launching the model’s training (Cabanski et al., 2017).

Binary bigrams and trigrams generation results in 1-3Gb file each, later this requires more than 16 Gb RAM to train (that’s why finally we didn’t use them). Almost all features were restored (via edited and adapted code from `prepare_data` folder) and were saved to pkl-s, and the rest of features were not used on the competition by the claims of the authors.

The system is trained on 1693 samples from the Corpus of Business News, and tested on the remaining ones from 5000 samples.

While HCS system threats the task as classification problem and returns results as probabilities, RiTUAL-UH system uses regression model.

The system is quite heavy for experimenting, initially it was adapted for working on cluster.

Due to the heaviness of the system, further experiments with this system are not very realistic without rich hardware resources (Saleiro et al., 2017).

The implementation of RiTUAL-UH system, adapted for a new data source, could be found here: https://bitbucket.org/sentimentals/semeval-2017-task-5-ritualuh

3. Results

We investigated the results of the competition and evaluated HCS and RiTUAL-UH systems on the Corpus of Business News:

HCS and RiTUAL-UH Systems on the Corpus of Business News

Name	Cosine similarity on test set	Test set size	Cosine similarity from the competition results
HCS (fulltest)	0.85085689	19926	0.68
HCS (on the same data as RiTUAL-UH)	0.84095291	3307	0.68
RiTUAL-UH	*0.8630541*	*3307*	*0.744*

4. Discussion

As we can see, both systems benefit strongly from using larger English texts from the Corpus of Business News. Also, the difference in the polarity detection quality is lower on the bigger texts. Although RiTUAL-UH system requires more resources and includes heavy feature generation process, its results are better, but still close to HCS results (and much closer, than in the competition). HCS achieved much higher results than on the competition data, which may show, that HCS system is better in extracting information directly from the data, while in RiTUAL-UH system additional features provide the higher quality.

The current work was limited to comparison of two competitors, however reevaluating the following systems may be interesting as future work:

- UW-FinSent at SemEval-2017 Task 5: Sentiment Analysis on Financial News Headlines using Training Dataset Augmentation (Soboleva, 2015) (14 place, 2nd track), link to the implementation: https://github.com/v1n337/semeval2017-task5/blob/master/
- Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines (John & Vechtomova, 2017) (4 place, 2nd track), link to the implementation: https://github.com/apmoore1/semeval
- HHU at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Data using Machine Learning Methods (Moore & Rayson, 2017) (4 place, 1st track), link to the implementation: https://github.com/tocab/SemEval2017Task5
- FEUP at SemEval-2017 Task 5: Predicting Sentiment Polarity and Intensity with Financial Word Embeddings (Cabanski et al., 2017) (9 place, 1st track), link to the implementation: https://github.com/saleiro/SemEval2017-Task5

5. Summary

We compared two systems, that were created for solving Task-5 from SemEval-2017, adapting them for new data source with richer English texts, and observed results with higher quality in general and less difference in quality.

6. Conclusion

Creating a system of sentiment analysis is a challenging task, but quite feasible, if the data is available for training and pre-specified domain (the subject). When using machine learning, it is important to test different parameters to find the ones that work best on the test data. In particular, you need to test different classification algorithms (NB, SVM), feature set (unigrams, bigrams, character N-grams), feature weighing function. There are still a lot of ways to improve sentiment classification, such as using tonal dictionaries, additional linguistic features (for example, parts of speech), and general ways to improve machine learning (boosting, bagging, etc.).

Acknowledgments

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

We would like to express deep gratitude to Professor Roman Yangarber for his advice and to University of Helsinki for help and for providing computational cluster.

References

Cabanski, T., Romberg, J., & Conrad, S. (2017). HHU at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Data using Machine Learning Methods. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, C. 832-836.

Cortis, K. et al. (2017). Semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017*), 519-535.

Ghosh, A., et al. (2015). Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015),* 470-478.

Goonatilake, R., & Herath, S. (2007). The volatility of the stock market and news. *International Research Journal of Finance and Economics, 3*(11), 53-65.

John, V., & Vechtomova, O. (2017). Sentiment analysis on financial news headlines using training dataset augmentation. Association for Computational Linguistics.

Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), (pp. 869-873). arXiv:1707.09448.

Kar, S., Maharjan, S., & Solorio, T. (2017). RiTUAL-UH at SemEval-2017 task 5: Sentiment analysis on financial data using neural networks. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 877-882.

Moore, A., & Rayson, P. (2017). Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines. *Proceedings of the 11th International Workshop on Semantic Evaluation*. Arxiv Preprint Arxiv:1705.00571.

Pivovarova, L. et al. (2017). HCS at SemEval-2017 task 5: Polarity detection in business news using convolutional neural networks. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017),* 842-846.

Saleiro, P. et al. (2017). Feup at SemEval-2017 task 5: Predicting sentiment polarity and intensity with financial word embeddings. *Proceedings of the 11th International Workshop on Semantic Evaluation.* Arxiv Preprint Arxiv:1704.05091.

Soboleva, N. P. (2015). Expanded metaphor and double actualization of phraseological units in advertising texts. *Journal of Language and Literature, 6*(1), 282-286.

Soboleva, P. N. (2017). Linguoculturological aspects of the contextual use of phraseological units in advertising slogans. Vestnik Tomskogogosudarstvennogo universiteta. *Filologiya, 45*, 139-149. 10.17223/19986645/45/10.

Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications, 42*(11), 4999-5010.