

IHS-RD-Belarus at SemEval-2016 Task 5: Detecting Sentiment Polarity Using the Heatmap of Sentence

Maryna Chernyshevich

IHS Inc. / IHS Global Belarus

131 Starovilenskaya St

220123, Minsk, Belarus

{Marina.Chernyshevich}@ihs.com

Abstract

This paper describes the system submitted by IHS-RD-Belarus team for the sentiment detection polarity subtask on Aspect Based Sentiment Analysis task at the SemEval 2016 workshop on semantic evaluation. We developed a system based on artificial neural network to detect the sentiment polarity of opinions. Evaluation on the test data set showed that our system achieved the F-score of 0.83 for restaurants domain (rank 4 out of 28 submissions) and F-score of 0.78 for laptops domain (rank 4 out of 21 submissions).

1 Introduction

Social media texts found in user review services have a great data-mining potential, as they offer real-time data that can be useful to monitor public opinion on brands, products, events, etc.

Most of recent approaches to sentiment analysis task are based on bag-of-words features, syntactic dependency features, out-of-domain and domain-specific sentiment lexicons, to train a supervised model that predicts the polarity of each given term or aspect category. This approach is very popular but it relies on heavy pre-processing of the data which involves careful choosing of the right features, empirical thresholds and intuitive analysis of the training set (Brun et al., 2014, Saias, 2015).

In this paper, we present an approach to the opinion polarity detection task based on an artificial neural network and sentiment orientation score of words.

2 Task description

The SemEval-2016 shared task on Aspect based Sentiment Analysis focuses on identifying the Opinion target expressions (OTE), the Aspect categories and the sentiment expressed towards each OTE or Aspect category.

The main focus of this paper is polarity subtasks, such as OTE polarity in restaurant subject domain and Aspect category polarity in laptop subject domain.

In the OTE polarity subtask, the input consists of a review sentence and the set of terms or aspect categories. The expected output is a polarity label (positive, negative or neutral) for each of the associated terms or aspect categories.

For example, the system should determine the polarity of *fajitas* and *salads* in the following sentence: *I hated their fajitas, but their salads were great.*

As for the Aspect category polarity, the task is more complicated. In the following sentence, the system has to determine the polarity of display quality (DISPLAY#QUALITY) and display usability (DISPLAY#USABILITY): *The display has a great resolution but has difficulty always seeing the small print.*

The task organizers provided a dataset of customer reviews with manually annotated opinion targets: 2500 sentences for laptop domain and 2000 sentences for restaurant domain.

Evaluation was to be carried out according to Precision, Recall and F1 metrics.

3 System description

The central idea behind our system is the visualization of sentiment orientation in a word sequence using heatmap, that highlights regions with higher or lower “temperature”. The “temperature” of a word is its sentiment polarity, positive or negative, and the intensity is calculated based on sentiment orientation score of word.

3.1 Sentiment orientation lexicon

Sentiment orientation score indicates the strength of association of words (w) with positive (pr) and negative (nr) reviews. Following Turney and Littman (2003), we calculated SO as the difference using Pointwise Mutual Information (PMI) measures:

$$SO(w) = PMI(w, pr) - PMI(w, nr)$$

SO score is positive when the word or phrase tends to occur mostly in positive reviews and negative when the word occurs more often in negative reviews. The magnitude indicates the degree of association.

We followed the lexicon-generation approach proposed by Kiritchenko et al. (Kiritchenko et al., 2014) and when generating the sentiment lexicons we distinguished the terms appearing in negated and affirmative contexts. Sentiment scores were then calculated separately for these two types of contexts.

We created uni-, bi- and tri-gram lexicons based on Yelp restaurant reviews and Amazon reviews.

3.2 Sentiment orientation score

The final SO score of n-gram in a sentence can be affected by some neighbour terms, such as valence shifters and intensifiers (Kennedy and Inkpen, 2006). We created short wordlists and tuned final SO score of n-gram if words from these lists were found in term context.

3.2.1 Valence shifters

Valence shifters are terms that can change the semantic orientation of another term, for example, combining positively valenced words with a nega-

tion flips the positive orientation to a negative orientation. The most important shifters are negations such as *not*, *never*, *none*, *nobody*, etc. (Polanyi and Zaenen, 2004).

As it was mentioned above we included negations, such as *not* and *never*, as n-gram postfix into lexicons. If the n-gram in negative context is not found in lexicon or its raw frequency in review corpus is less than 5, the final SO score is calculated based on calculation rules showed in table 1.

These negation rules are designed in order to improve the sentiment text analysis and are based on simple assumption: the negation flips positive the valence of a word to a negative with the same strength, but the negation doesn’t flip the valence of a negative word, it rather reduces its strength.

SO of word	calculation rule	Example
positive SO	SO=pos_SO * -1	SO(love) = +2.6 SO(not love) = -2.6
negative SO	SO=neg_SO * 0.5	SO(hate) = -2.2 SO(not hate) = -1.1

Table 1: SO calculation rules in negative context.

3.2.2 Intensifiers

Intensifiers are terms that change the degree of the expressed sentiment. For example, in the sentence “*The waterfly cases are very good.*”, the term *very good* is more positive together than just *good* alone. So to calculate the final SO score we multiply sentiment score of n-gram by 3.

On another side, in the sentence “*The waterfly cases are barely good.*”, the term *barely* makes this statement less positive. We created 4 lists of intensifiers of different intensity of affecting the final SO score of n-gram.

3.2.3 Unreality and conditionality

A sentence in reviews can express not only a real user experience, but also an unreal opinion, for example a wish, “*I should have bought something better.*” or “*The laptop may be better.*”. We collected all surface patterns which can express unreality or conditionality. The SO of all positive n-grams in unreal context is flipped to negative with the same strength. The SO of negative n-grams is not changed.

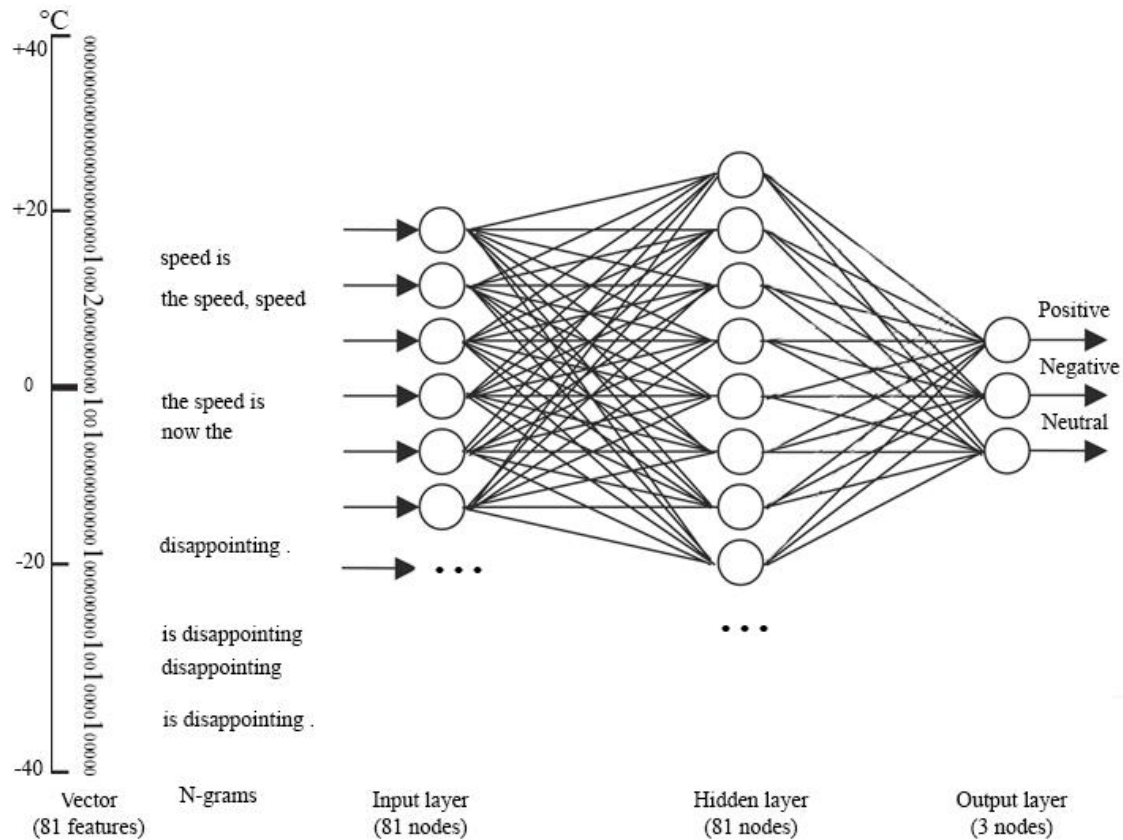


Fig. 1: System architecture

3.3 Neural network architecture

A fully connected multilayer neural network with back-propagation is applied. The network, illustrated in Fig. 1, contains 3 layers:

- input with 81 nodes - one for each feature presenting “temperature” range;
- hidden layer with 80 nodes - each node in the input layer is connected to each node of the hidden layer;
- output layer with 3 nodes - one for each of 3 classes.

As the activation function the sigmoid function is used. We apply dropout to our hidden layers, as described in Srivastava et al. (Srivastava et al., 2014), to prevent network overfitting.

The network architecture is implemented on Keras¹, which is an effective deep learning framework implementation in python.

At the first step, we detect the context of opinion target expressions or category. If the sentence has only one opinion target, the term context will include all words from the beginning to the end of the sentence. If the sentence has many targets, the context will include all words surrounding the term enclosed between two separators. As a separator we consider all punctuation marks, the next opinion target, the end and the beginning of the sentence. In the laptops subtask we had to detect polarity of the category that is not bounded to any word in the sentence. That is why we considered the whole sentence as context for every aspect category.

For training, each term context was represented as a vector of 81 features that resembles the scale of “temperature” from -40 (very negative) to 40 (very positive). Each word according to its SO score was placed on this scale. The value of a feature means the number of words within this range of sentiment orientation.

¹ <http://keras.io/>

Table 2 illustrates a set of n-grams with their final SO scores and “temperature” extracted from the sentence “Now the speed is disappointing”.

	n-gram	SO score	°C
Uni-grams	disappointing	-3.0	-30
	speed	+1.0	+10
Bi-grams	is disappointing	-2.7	-27
	disappointing .	-1.7	-17
	now the	-0.4	-4
	the speed	+1.0	+10
Tri-grams	speed is	+1.4	+14
	is disappointing .	-3.5	-35
	the speed is	-0.1	-1

Table 2: SO scores of n-grams

The number of training epochs is set to 2. More epochs would lead to overfitting because the training set that is relatively small.

4 Results and further work

The 3-fold cross-validation procedure was used to test the system performance. It uses 66% of the data to train and the remaining 33% to assess the accuracy. This is repeated 3 times with a different 33% left. Table 3 represents the results.

	restaurants	laptops
total accuracy	0.82838934	0.8030438
F1 positive	0.882567	0.8479
F1 negative	0.762167	0.796867
F1 neutral	0.0	0.0

Table 3: 3-fold cross-validation results

The results on the official test set are given in Table 4. They are similar to our best results attained during the developments stage.

	restaurants	laptops
total accuracy	0.8393481	0.7790262
F1 positive	0.9078	0.8308
F1 negative	0.753	0.7588
F1 neutral	0.0	0.0

Table 4: Official results

On both sets our system did not recognize neutral (mildly positive or mildly negative sentiment) sentiments, although we used 3 classes classification. The reason is the low number of neutral sentiments in the annotated corpus.

As a future work, we intend to develop more complex artificial neural network architecture and try to use review-long memory about the polarity of previous targets. Also we are going to investi-

gate the influence of words with extremely high or low sentiment orientation score on the SO scores of neighbouring words.

5 Conclusion

We have presented a simple neural network architecture that predicts polarity of word sequence based the heatmap of sentence that highlights regions with higher or lower sentiment intensity. We submitted runs corresponding to the slot3 subtasks, obtaining competitive results. Our submission was ranked 4th out of 28 submissions for restaurants domain and 4th out of 21 submissions for laptops domain.

References

- Alfred. V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling, volume 1*. Prentice-Hall, Englewood Cliffs, NJ.
- Caroline Brun, Diana Nicoleta Popa, Claude Roux. 2014. *XRCE: Hybrid Classification for Aspect-based Sentiment Analysis*. in Proceedings of the eight international workshop on Semantic Evaluation Exercises (SemEval-2014), August 2014, Dublin, Ireland.
- Kennedy, A. and D. Inkpen. 2006. *Sentiment Classification of Movie Reviews using Contextual Valence Shifters*. Computational Intelligence, vol.22(2), pp.110-125, 2006.
- José Saias. 2015. *Sentiue: Target and Aspect based Sentiment Analysis in SemEval-2015 Task 12*. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado.
- Livia Polanyi and Annie Zaenen. *Contextual valence shifters*. In Proceedings of the AAAI Symposium on Exploring Attitude and Affect in Text: Theories and Applications (published as AAAI technical report SS-04-07), 2004.
- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez Zafra, and Gülşen Eryiğit. 2016. *SemEval-2016 Task 5: Aspect Based Sentiment Analysis*. In Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval ’16, San Diego, California, June 2016. Association for Computational Linguistics.
- Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad. 2014. *NRC-Canada-2014: De-*

detecting Aspects and Sentiment in Customer Reviews.
in Proceedings of the eight international workshop on
Semantic Evaluation Exercises (SemEval-2014), Au-
gust 2014, Dublin, Ireland.

Srivastava, N, Hinton, G, Krizhevsky, A, Sutskever, I,
Salakhutdinov, R. *Dropout: A Simple Way to Prevent
Neural Networks from Overfitting.* Journal of Ma-
chine Learning Research 15, 2014