# Stance and Sentiment in Czech

Tomáš Hercig[1,2] and Peter Krejzl[1]

[1] Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic
[2] NTIS—New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Technická 8, 306 14 Plzeň, Czech Republic
{tigi,krejzl}@kiv.zcu.cz
http://nlp.kiv.zcu.cz

**Abstract.** Sentiment analysis is a wide area with great potential and many research directions. One direction is stance detection, which is somewhat similar to sentiment analysis. We supplement stance detection dataset with sentiment annotation and explore the similarities of these tasks. We show that stance detection and sentiment analysis can be mutually beneficial by using gold label for one task as features for the other task. We analysed the presence of target entities for stance detection in the dataset. We outperform the state-of-the-art results for stance detection in Czech and set new state-of-the-art results for the newly created sentiment analysis part of the extended dataset.

## 1 Introduction

During recent years, there have been a lot of research in the area of Natural Language Processing (NLP) related to sentiment analysis [13, 14, 12, 11].

Stance detection can be viewed as a subtask of opinion mining, similar to sentiment analysis. In sentiment analysis, systems determine whether a piece of text is positive, negative, or neutral. Stance detection goes even further and tries to detect whether the author of the text is in favor or against a given target. The main difference to sentiment analysis is that in stance detection, systems are to determine the author's favorability towards a given target and the target may not even be explicitly mentioned in the text. Moreover, the text may express positive opinion about an entity contained in the text, but one can also infer that the author is against the defined target (an entity or a topic). It has been found difficult to infer stance towards a target of interest from tweets that express opinion towards another entity [9].

There are many applications which could benefit from the automatic stance detection, including information retrieval, textual entailment, or text summarization, in particular opinion summarization.

The same stance towards a target may be expressed by positive or negative language. This phenomenon has not yet been thoroughly investigated. The pioneer work in English Tweets [10] annotated stance dataset with additional sentiment labels and show that knowing the sentiment label is beneficial for

stance detection, however they also state that "even though sentiment can play a key role in detecting stance, sentiment alone is not sufficient".

Our goal is to examine how stance and sentiment influence each other in Czech language and either confirm or reject the hypothesis that sentiment labels are beneficial for stance detection.

The rest of this paper is organized as follows. Section 2 presents the related work. The dataset is described in Section 3. The annotation of sentiment is covered in Section 4. Our approach is presented in Section 5. Conducted experiments are described in Section 6. Finally, we conclude in Section 7.

## 2 Related Work

Table 1: Statistics of the SemEval-2016 task "Detecting Stance in Tweets" corpora in terms of the number of tweets and stance labels.

| Target Entity | Total | *In favor* | *Against* | *Neither* |
|---|---|---|---|---|
| Atheism | 733 | 124 (17%) | 464 (63%) | 145 (20%) |
| Climate Change is Concern | 564 | 335 (59%) | 26 (5%) | 203 (36%) |
| Feminist Movement | 949 | 268 (28%) | 511 (54%) | 170 (18%) |
| Hillary Clinton | 934 | 157 (17%) | 533 (57%) | 244 (26%) |
| Legalization of Abortion | 883 | 151 (17%) | 523 (59%) | 209 (24%) |
| All | 4,063 | 1,035 (25%) | 2,057 (51%) | 971 (24%) |

Table 2: Statistics of the Czech corpora in terms of the number of news comments and stance labels.

| Target Entity | Total | *In favor* | *Against* | *Neither* |
|---|---|---|---|---|
| "Miloš Zeman" – Czech president | 2,638 | 691 (26%) | 1,263 (48%) | 684 (26%) |
| "Smoking Ban in Restaurants" – Gold | 1,388 | 272 (20%) | 485 (35%) | 631 (45%) |
| "Smoking Ban in Restaurants" – All | 2,785 | 744 (27%) | 1,280 (46%) | 761 (27%) |

The SemEval-2016 task **Detecting Stance in Tweets**[3] [9] had two subtasks: supervised and weakly supervised stance identification.

The goal of both subtasks was to classify tweets into three classes (*In favor*, *Against*, and *Neither*). The performance was measured by macro-averaged F1-score of two classes (*In favor* and *Against*) denoted $F1_{ma2}$ and by micro-averaged

[3] http://alt.qcri.org/semeval2016/task6/

F1-score for the same two classes denoted $F1_{mi2}$. This evaluation measure does not disregard the *Neither* class, because falsely labelling the *Neither* class as *In favor* or *Against* still affects the scores. We use the same evaluation metrics $F1_{ma2}$, accuracy, and the F1-score of all classes ($F1_{ma3}$).

The supervised task (subtask A) tested stance towards five targets: *Atheism*, *Climate Change is a Real Concern*, *Feminist Movement*, *Hillary Clinton*, and *Legalization of Abortion*. Participants were provided with 2814 labeled training tweets for the five targets.

A detailed distribution of stances for each target is given in Table 1. The distribution is not uniform and there is always a preference towards a certain stance. The distribution reflects the real-world scenario, in which a majority of people tend to take a similar stance [3].

For the weakly supervised task (subtask B), there were no labeled training data but participants could use a large number of tweets related to the single target: *Donald Trump*.

The best results ($F1_{ma2}$ 56.0%, $F1_{mi2}$ 67.8%) for subtask A were achieved by an advanced baseline using SVM classifier with unigrams, bigrams, and trigrams along with character n-grams (2, 3, 4, and 5-gram) as features.

Wei et al. [16] present the best result for subtask B and they ranked close second in subtask A of the SemEval stance detection task. They used a convolutional neural network (CNN) designed according to Kim [5]. They initialized the embedding layer with pre-trained word2vec embeddings. The main difference from Kim's network is the used voting scheme. During each training epoch, several iterations were selected to predict the test set. At the end of each epoch, the majority voting scheme was applied to determine the label for each sentence. This was done over a specified number of epochs and finally the same voting was applied to the results of each epoch. The train and test data were separated according to the stance targets.

Mohammad et al. [10] annotated the SemEval-2016 task **Detecting Stance in Tweets** dataset [9] with sentiment labels and whether the opinion is expressed towards the given stance target. They performed a detailed analysis of the dataset and conducted several experiments. They showed that sentiment label is beneficial for stance detection however it is not sufficient ($F1_{ma2}$ 56.1%, $F1_{mi2}$ 59.6%).

### 2.1 Stance Detection in Czech

The initial research on Czech stance detection has been done in [7]. They collected 1,460 comments from a Czech news server[4] related to two topics – Czech president – *"Miloš Zeman"* (181 *In favor*, 165 *Against*, and 301 *Neither*) and *"Smoking Ban in Restaurants"* (168 *In favor*, 252 *Against*, and 393 *Neither*).

Hercig et al. [3] extended the dataset from [7]. The detailed annotation procedure was described in [4] (in Czech). The whole corpus was annotated by three native speakers. The distribution of stances for each target is given in Table 2.

[4] `www.idnes.cz`

They evaluated Maximum Entropy, SVM and two CNN classifiers. We used the Czech president – *"Miloš Zeman"* dataset[5] to annotate Czech stance detection corpus with sentiment labels. We chose this dataset because of its size and better inter-annotator agreement. The best results for this dataset were achieved by the CNN designed according to Kim [5] and the Maximum Entropy classifier.

## 3 Dataset

The dataset for the target entity "Miloš Zeman" was annotated by one annotator and then 302 comments were also labeled by a second annotator to measure inter-annotator agreement. The dataset for the target entity "Smoking Ban in Restaurants" was independently annotated by two annotators (2,203 comments) and then the majority voting scheme was applied to the gold label selection (third annotator was used to resolve conflicts). The inter-annotator agreement (Cohens $\kappa$) is 0.579 for "Miloš Zeman" and 0.423 for "Smoking Ban in Restaurants".

The inter-annotator agreement for "Smoking Ban in Restaurants" was quite low, thus they selected a subset of the dataset, where the original two annotators assigned the same label as the gold dataset (1,388 comments).

Table 3: Distribution of instances by sentiment and stance in the extended dataset.

| Sentiment/Stance | In Favor | Against | Neither | SUM |
|---|---|---|---|---|
| Positive | 164 (6.2%) | 43 (1.6%) | 20 (0.8%) | 227 (8.6%) |
| Negative | 116 (4.4%) | 614 (23.3%) | 83 (3.1%) | 813 (30.8%) |
| Neutral | 411 (15.6%) | 606 (23.0%) | 581 (22.0%) | 1598 (60.6%) |
| SUM | 691 (26.2%) | 1263 (47.9%) | 684 (25.9%) | 2638 (100%) |

Table 4: Annotator agreement confusion matrix.

| A1/A2 | Positive | Negative | Neutral |
|---|---|---|---|
| Positive | 6 | 0 | 3 |
| Negative | 1 | 49 | 9 |
| Neutral | 12 | 12 | 39 |

[5]This is the only available Czech stance detection dataset we could find. The corpus is available for research purposes at `http://nlp.kiv.zcu.cz/research/sentiment#stance`.

## 4 Annotation

We annotated the Czech president – *"Miloš Zeman"* stance detection dataset with sentiment labels (positive, negative, and neutral). The whole dataset was annotated by one annotator and then a second annotator was used to calculate inter-annotator agreement (Cohens $\kappa$) on 131 comments. The annotators should assign the strongest sentiment to each comment or neutral label when the comment is factual (non-subjective) without anticipating further information (context). The inter-annotator (Cohens $\kappa$) is 0.524% (see the confusion matrix Table 4) and accuracy is 71.8%.

Table 3 shows the distribution of sentiment and stance labels in the extended dataset. While most comments are against the target, the sentiment of most comments is neutral and only a small portion of the dataset is positive. Most of the comments that are in favor of the target are neutral which means that the comments are non-subjective, however the comments against the target are mostly negative and almost none is positive. The comments neither for nor against the target are mostly neutral as expected. For positive sentiment the comment is mostly in favor of target. Negative sentiment most of the time means against the target and neutral sentiment is almost uniformly distributed across stance labels.

We also labeled the comments for the presence of the "Miloš Zeman" entity and the "president" entity. The distribution of entities by stance and sentiment labels is shown in Table 5. The presence of these entities was detected by regular expressions[6].

The extended corpus annotated with sentiment labels and marked for the presence of entities "Miloš Zeman" and "president" is available for research purposes at `http://nlp.kiv.zcu.cz/research/sentiment#stance`.

Table 5: Presence of Entities "Miloš Zeman" and "president".

(a) Presence of Entities by Stance.

| Entity | Miloš Zeman | | President | |
|---|---|---|---|---|
| Present | True | False | True | False |
| In Favor | 364 | 327 | 187 | 504 |
| Against | 688 | 575 | 333 | 930 |
| Neither | 435 | 249 | 212 | 472 |

(b) Presence of Entities by Sentiment.

| Entity | Miloš Zeman | | President | |
|---|---|---|---|---|
| Present | True | False | True | False |
| Positive | 130 | 97 | 69 | 158 |
| Negative | 412 | 401 | 216 | 597 |
| Neutral | 945 | 653 | 447 | 1151 |

[6] `".*\bMZ\b.*|.*eman.*|.*milo(u)?s.*"` and `".*prezident.*|.*president.*"`

## 5 The Approach Overview

For all experiments we use Maximum Entropy classifier from Brainy machine learning library [6]. We evaluate using 20-fold cross-validation to allow comparison with previous work [3].

### 5.1 Preprocessing

We use UDPipe [15] with Czech Universal Dependencies 1.2 models for tokenization, POS tagging, and lemmatization. We further use lower-casing, remove diacritics, and we also replace all characters "y" with the character "i".

### 5.2 Features

This section describes features used in our experiments.

- **Character n-grams (ChN$_n$):** Separate binary feature for each character $n$-gram in the utterance text. We do it separately for different orders $n \in \{5, 7\}$ and remove $n$-grams with frequency $f \leq 2$.
- **First Words (FW):** Bag of first five words with at least 2 occurrences.
- **Last Words (LW):** Bag of last five words with at least 2 occurrences.
- **Emoticons (E):** We used a list of negative emoticons[7] specific to the news commentaries source. The feature captures the presence of an emoticon within the text.
- **Unigram Shape (Sh):** The occurrence of word shape unigram in the text. Word shape assigns words into one of 24 classes[8] similar to the function specified in [1]. We consider unigrams with frequency $f > 2$.
- **Target (TP):** One-hot vector for gold labels of the other task (e.g. sentiment label for stance detection) combined with the presence of the "president" entity (the resulting vector has length 6).
- **Target (TZ):** One-hot vector for gold labels of the other task (e.g. sentiment label for stance detection) combined with the presence of the "Miloš Zeman" entity (the resulting vector has length 6).
- **Text Length (TL):** We map the text length into a one-hot vector with length three and use this vector as binary features for the classifier. The text length belongs to one of three equal-frequency bins[9]. Each bin corresponds to a position in the vector.
- **Oracle (O):** One-hot vector for gold labels of the other task (e.g. sentiment label for stance detection).
- **Word n-grams (WN$_n$):** Separate binary feature for each word $n$-gram in the utterance text. We do it separately for different orders $n \in \{1, 2, 3\}$ and remove n-grams with frequency $f \leq 2$.

[7] `":-("`, `";-("`, `":-/"`, `"Rv"`

[8] We use edu.stanford.nlp.process.WordShapeClassifier with the WORDSHAPECHRIS1 setting available in Standford CoreNLP library [8].

[9] The frequencies from the training data are split into three equal-size bins according to 33% quantiles.

Table 6: Experiment results on the Czech stance detection in %.

| Features | Stance | | | Sentiment | | |
|---|---|---|---|---|---|---|
| | $F1_{ma3}$ | $F1_{ma2}$ | *Acc* | $F1_{ma3}$ | $F1_{ma2}$ | *Acc* |
| Random Class | 32.1 | 33.4 | 32.9 | 29.6 | 23.1 | 33.2 |
| Majority Class | 21.6 | 32.4 | 47.9 | 25.1 | 00.0 | 60.6 |
| Best results from Hercig et al. [3] | **51.3** | **56.4** | **54.9** | – | – | – |
| O | 34.0 | 51.1 | 52.5 | 36.7 | 21.9 | 56.2 |
| $WN_1$ | 48.1 | 52.0 | 50.6 | 55.1 | 47.5 | 60.9 |
| $WN_1$ + O | 51.7 | 56.2 | 54.3 | 59.1 | 52.4 | 64.3 |
| $WN_1$ + TP | 50.7 | 55.1 | 53.4 | 58.7 | 51.9 | 64.2 |
| $WN_1$ + TZ | 51.5 | 55.8 | 54.1 | 58.9 | 52.2 | 64.0 |
| $WN_1$ + TP + TZ | 51.5 | 55.9 | 54.2 | 59.1 | 52.3 | 64.4 |
| $WN_1$ + $ChN_{5,7}$ | 50.3 | 55.2 | 53.9 | 56.4 | 47.1 | 65.1 |
| $WN_1$ + $ChN_{5,7}$ + O | **54.3** | **59.0** | **57.1** | 58.8 | 50.2 | **67.4** |
| $WN_1$ + $WN_{2,3}$ | 50.8 | 55.8 | 53.9 | 57.6 | 49.8 | 64.1 |
| $WN_1$ + $WN_{2,3}$ + O | 53.7 | 58.5 | 56.6 | **59.9** | **52.8** | 65.7 |
| Feature set* | 54.2 | 58.8 | 57.3 | 60.1 | 51.8 | **68.3** |
| Feature set – $ChN_{5,7}$ | 54.3 | 58.4 | 57.6 | **61.3** | **54.4** | 67.2 |
| Feature set – E | 54.4 | 58.9 | 57.4 | 59.7 | 51.3 | 68.2 |
| Feature set – FW | **54.8** | **59.2** | **57.8** | 60.4 | 52.3 | **68.3** |
| Feature set – LW | 54.5 | 58.9 | 57.5 | 58.7 | 49.8 | 67.8 |
| Feature set – TL | 54.2 | 59.1 | 57.4 | 59.7 | 51.3 | 68.0 |
| Feature set – Sh | 54.2 | 58.8 | 57.3 | 59.0 | 50.5 | 67.4 |
| Feature set – $WN_{1,2,3}$ | 54.5 | 58.5 | 57.4 | 58.2 | 49.4 | 67.1 |
| Feature set – O | 54.0 | 58.7 | 57.2 | 60.3 | 52.0 | 68.4 |
| Feature set – TP | 54.3 | 58.9 | 57.5 | 60.0 | 51.8 | 68.2 |
| Feature set – TZ | 54.2 | 58.8 | 57.4 | 60.0 | 51.7 | 68.0 |
| Best combination† Stance | **56.2** | **60.3** | **59.1** | 59.4 | 51.0 | 67.7 |
| Best combination‡ Sentiment | 54.8 | 58.9 | 57.7 | **62.0** | **54.6** | **68.9** |

* $ChN_{5,7}$ + E + FW + LW + TL + Sh + $WN_{1,2,3}$ + O + TP + TZ

† $ChN_7$ + E + Sh + $WN_1$ + O + TP + TZ

‡ $ChN_5$ + E + LW + TL + Sh + $WN_{1,2,3}$ + O + TP + TZ

## 6 Experiments

For all experiments we report the macro-averaged F1-score of two classes $F1_{ma2}$ (*In favor* and *Against*) – the official metric for the SemEval-2016 stance detection task[9], accuracy, and the macro-averaged F1-score of all three classes ($F1_{ma3}$).

Table 6 shows results of all our experiments. We performed experiments with using the gold sentiment labels as features for stance detection and using the gold stance labels as features for sentiment analysis (i.e. using the Oracle feature). The results show that the Oracle feature improves results in all cases. The Oracle feature combined with unigrams and character n-grams also outperforms the previous state-of-the-art results for stance detection by 3.0% $F1_{ma3}$, 2.6% $F1_{ma2}$, and 2.2% *Acc.*

Another experiment included using features that indicate the presence of the "Miloš Zeman" entity and the "president" entity combined with the gold labels as in Oracle feature. Our expectation was that this should improve the results (as it did in English), however the results show that in fact the information about the presence of the target entity does not lead to better results.

We further performed an ablation study for the combination of features ($ChN_{5,7}$ + E + FW + LW + TL + Sh + $WN_{1,2,3}$ + O + TP + TZ). In Table 6 the bold numbers denote the best results for the given column.

The ablation study shows that the FW feature present little to no information gain for the classifier. We further experimented with combinations of features and that lead to the best feature sets for both stance detection and sentiment analysis (see the last two lines in Table 6). Both of these sets contain emoticons, word shape, oracle and target entities.

## 7 Conclusion

We presented the first Czech dataset annotated for both stance and sentiment labels including the presence of target entities. We have shown that stance and sentiment can be mutually beneficial and confirmed our initial hypothesis. Moreover, we have outperformed the state-of-the-art results for stance detection in Czech and set a new state-of-the-art results for the sentiment analysis part of the dataset.

Our best result outperformed the previous stance detection state of the art by 4.9% $F1_{ma3}$, 3.9% $F1_{ma2}$, and 4.2% *Acc.* The sentiment analysis unigram baseline was outperformed by 6.9% $F1_{ma3}$, 7.1% $F1_{ma2}$, and 8.0% *Acc.*

In the future we plan to extend this analysis on other target entities and explore the usefulness of labels assigned by trained models instead of using gold labels for the Oracle feature.

## Acknowledgments

This publication was supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports under the program NPU I and by university specific research project SGS-2016-018 Data and Software Engineering for Advanced Applications.

# Bibliography


[1] Bikel, D. M., Miller, S., Schwartz, R., and Weischedel, R. (1997). Nymble: a high-performance learning name-finder. In *Proceedings of the fifth conference on Applied natural language processing*, pages 194–201. Association for Computational Linguistics.

[2] Hercig, T., Brychcín, T., Svoboda, L., and Konkol, M. (2016). UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 342–349. Association for Computational Linguistics.

[3] Hercig, T., Krejzl, P., Hourová, B., Steinberger, J., and Lenc, L. (2017). Detecting stance in czech news commentaries. In Hlaváčová, J., editor, *Proceedings of the 17th ITAT: Slovenskočeský NLP workshop (SloNLP 2017)*, volume 1885 of *CEUR Workshop Proceedings*, pages 176–180, Bratislava, Slovakia. Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, CreateSpace Independent Publishing Platform.

[4] Hourová, B. (2017). Automatic detection of argumentation. Master's thesis, University of West Bohemia, Faculty of Applied Sciences.

[5] Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

[6] Konkol, M. (2014). Brainy: A Machine Learning Library. In Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L., and Zurada, J., editors, *Artificial Intelligence and Soft Computing*, volume 8468 of *Lecture Notes in Computer Science*, pages 490–499. Springer International Publishing.

[7] Krejzl, P., Hourová, B., and Steinberger, J. (2016). Stance detection in online discussions. In Bieliková, M. and Srba, I., editors, *WIKT & DaZ 2016 11th Workshop on Intelligent and Knowledge Oriented Technologies 35th Conference on Data and Knowledge*, pages 211–214. Vydateľstvo STU, Vazovova 5, Bratislava, Slovakia.

[8] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

[9] Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.

[10] Mohammad, S. M., Sobhani, P., and Kiritchenko, S. (2017). Stance and sentiment in tweets. *ACM Trans. Internet Technol.*, 17(3):26:1–26:23.

[11] Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N.,

Jiménez-Zafra, S. M., and Eryiğit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, San Diego, California. Association for Computational Linguistics.

[12] Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado*, pages 486–495.

[13] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). SemEval-2014 Task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

[14] Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., and Stoyanov, V. (2015). SemEval-2015 Task 10: Sentiment Analysis in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 451–463, Denver, Colorado. Association for Computational Linguistics.

[15] Straka, M., Hajič, J., and Straková, J. (2016). UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Paris, France. European Language Resources Association (ELRA).

[16] Wei, W., Zhang, X., Liu, X., Chen, W., and Wang, T. (2016). pkudblab at SemEval-2016 Task 6 : A Specific Convolutional Neural Network System for Effective Stance Detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 384–388, San Diego, California. Association for Computational Linguistics.