Overview of NLPCC Shared Task 4: Stance Detection in Chinese Microblogs

Conference Paper · December 2016

DOI: 10.1007/978-3-319-50496-4_85

Overview of NLPCC Shared Task 4: Stance Detection in Chinese Microblogs

Ruifeng Xu[1,2(✉)], Yu Zhou[1], Dongyin Wu[1], Lin Gui[1], Jiachen Du[1], and Yun Xue[2]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, China
xuruifeng@hitsz.edu.cn, zhouyu.nlp@gmail.com, wudongyin@foxmail.com, guilin.nlp@gmail.com, dujiachen@stmail.hitsz.edu.cn

[2] Guangdong Provincial Engineering Technology Research Center for Data Science, Guangzhou, China
happyhuashida@sina.com.cn

Abstract. This paper presents the overview of the shared task, stance detection in Chinese microblogs, in NLPCC-ICCPOL 2016. The submitted systems are expected to automatically determine whether the author of a Chinese microblog is in favor of the given target, against the given target, or whether neither inference is likely. Different from regular evaluation tasks on sentiment analysis, the microblog text may or may not contain the target of interest, and the opinion expressed may or may not be towards to the target of interest. We designed two tasks. Task A is a mandatory supervised task which detects stance towards five targets of interest with given labeled data. Task B is an optional unsupervised task which gives only unlabeled data. Our shared task has had sixteen team participants for Task A and five results of Task B. The highest F-score obtained was 0.7106 for Task A and 0.4687 for Task B, respectively.

1 Introduction

With the rapidly growth of subjective text on the internet, sentiment analysis (Gui et al. 2016; Tang et al. 2014) or opinion mining (He et al. 2012; Gui et al. 2014) became a hot topic problem in nature language processing. The proposed techniques aim to detect the sentiment or opinion from the text. However, in many cases, we care more about the attitude of the author to a specific topic (Anand et al. 2011; Boltužić et al. 2014). For example, in the topic of American election, we may care about if the author of a text support Trump or not. We call this attitude as stance in a topic.

For more details, stance detection aims to determine the author's stance towards a certain target from given text. The target here may be an issue, a government policy, a social phenomenon, and a product. The target may or may not be mentioned explicitly in the text. Meanwhile, the stance could be in favor of, against, or neutral towards the target of interest (Mohammad and Kiritchenko 2016). For instance, it is reasonable for people to predicate that a woman is in favor of "Two-child policy" if she always wants the second kid. Nowadays people express their attitude towards almost everything through online websites or mobile Apps. Detecting the stance of the authors towards certain target should be helpful to many applications.

C.-Y. Lin et al. (Eds.): NLPCC-ICCPOL 2016, LNAI 10102, pp. 907–916, 2016.
DOI: 10.1007/978-3-319-50496-4_85

We formulate the task data as below:

<ID><Tab><Target><Tab><Text><Tab><Stance>

The IDs are unique for each microblog, and the Targets are specified. The Text contains everything that author has posted in the microblog, including pure text, emotion faces, address information, the original source, etc.

For example:

<ID> 1884
<Target>开放二胎 *Two child policy*
<Text>二胎了, 小伙伴们替我想个名字 *My second child is coming. Please think of name for me, my friends.*
<Stance>FAVOR

The automatic stance detecting system is required determine whether the author is in favor of or against the given target, or neither of those. To support the evaluation for stance detection, we manually labeled 4000 microblogs of targets of "iPhone SE", "春节放鞭炮 *Set off firecrackers in the Spring Festival*", "俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*", "开放二胎政策 *Two child policy*", and "深圳禁摩限电 *Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*" for Task A. For each target, we use 75% as training data and 25% as testing data, respectively. For Task B, we provide 2000 unlabeled microblogs for "转基因食品 G*enetically modified food*" and "朝鲜核试验 *The Nuclear Test in DPRK*" as training data. The participants may use unlabeled data from other sources in Task B. For both Task A and Task B, we provide 300 sample data for each target as trial data.

Sixteen team participants submitted their result of Task A and the highest F-score achieved was 0.7106. The best classification system actually used five separate classifiers for each target. That should be a very important factor that improves the performance. Meanwhile, most teams bi-gram, TF-IDF, word vector and sentiment lexicons features. Some teams used ensemble learning frameworks while other teams used relatively simple classifiers due to the amount of dataset. Although this task is to determine stance towards certain target, only about half of the participants have considered the specific target related features.

Five teams submitted their results experiment result of Task B. The highest achieved F-score was 0.4687. Generally speaking, the achieved performances of Task B are lower than the performances of Task A. Furthermore, it seems that these five teams did not use extra data to train their models for Task B.

Stance detection is a new challenge task. In this evaluation, the participated system achieved acceptable performance. The dataset for NLPCC Stance Detection are published online now. More effective features and learning methods are expected to handle this problem in the future.

The rest of this paper is organized as follows. Section 2 describes the definition of stance in this evaluation. Section 3 presents the dataset preparation. Section 4 presents the evaluation setting and evaluation metrics. Section 5 provides the evaluation results and discussions. Finally, Sect. 6 concludes this paper.

2 Stance Detection

In this section we present the definition of stance detection in this evaluation and the relationship between stance detection and sentiment analysis.

2.1 Stance Detection

Stance detection can be formulated in different ways. In the context of this task, stance detection is defined as automatically determining whether the author is in favor of the given target, against the given target, or whether neither inference is likely from the text. Consider the following target-microblog pair:

> Target: 俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*
> Microblog: 9月30日开始至今，俄空爆叙利亚，共死亡1331人，其中403人是一般的民众……其中的三分之一是无辜平民陪葬。*Since September 30*th, *a total of 1331 persons dead in the Russian air strikes in Syria, of which 403 people are the general public…… 1/3 of them are innocent civilians.*

Humans can deduce from the microblog that the speaker is likely against the target, namely *Russia's anti terrorist operations in Syria*. The corresponding stance label is AGAINST.

The aim of this task is to evaluate the performance on the systems which deduces the stance of the microblog. To successfully detect stance, the automatic systems normally have to identify relevant bits of information that may not be present in the focus text. In the above example microblog, if one emphasizes the death of innocent civilians, and then he or she is likely against the air strikes by Russia. Thus, we provide microblogs corresponding to each of the targets, from which systems can gather information to help the detection of stance.

Automatically detecting stance has widespread applications in information retrieval, text summarization, and textual entailment. In fact, one can argue that stance detection can often bring complementary information to sentiment analysis, because we often care about the author's evaluative outlook towards specific targets and propositions rather than simply about whether the speaker was angry or happy.

2.2 Stance Detection and Sentiment Analysis

Stance detection has some shared points with sentiment analysis. Actually they are distinct from each other. Sentiment analysis always aims to either classify a piece of text into a label of "Positive", "Negative", or "Neutral". In stance detection tasks, a target of interest is given to a collection of related microblogs. Based on this target, the system detects the stance/favorability of the author stance towards this target. The stance could be "FAVOR", "AGAINST", or "NONE". Notice that the target may be mentioned in the microblogs and it also may not be mentioned. Furthermore, there could be another target (or entity) mentioned in microblogs. For example, consider the following target-microblog pairs:

E.g. 1 Target 春节放鞭炮 *Set off firecrackers in the Spring Festival*

Microblog: 今天是大年初一, 传统意义上的春节, 放鞭炮接财神吃饺子一样不能少。 *Today is the first day of lunar New Year. Setting off firecrackers, welcoming the god of wealth, eating dumplings, a traditional Chinese Spring Festival will not go without one of these.*

E.g. 2 Target: 深圳禁摩限电 *Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*

Microblog: 主题执法日迟迟不来, 笋岗片区多了好几部崭新的三轮。 *Law enforcement day has been slow to come. A few new tricycle appeared in Sungang.*

E.g. 3 Target: 转基因食品 *Genetically modified food*

Microblog: 崔永元那帮人一点科学证据也不拿, 问问美国民众就算证明了? *Cui Yongyuan guys do not provide a little bit of scientific evidence. Ask the American people to prove it*?

E.g. 4 Target: 深圳禁摩限电 (*Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*)

Microblog: 规范电瓶车的违章行驶是最合理的好事, 要不出事儿太大了, 人命关天。 *Correcting the illegal driving of electric vehicles is the most reasonable thing. Otherwise, traffic accident brings big trouble. People's life is serious.*

In E.g. 1, the target is explicit, and people can easily deduce that the author support the act of setting off fireworks in Spring Festival. In E.g. 2, the target is not mentioned, but people can deduce that the author is talking about the target because it contains a place name in Shenzhen and the transportation related to the target. In E.g. 3, the direct opinion target is “Cui Yongyuan”, and apparently the author is satirizing Cui, while we know that Cui is famous for his against genetically modified food (GM food). Thus, the author tends to be support GM food.

Now consider E.g. 4, it has the same target of interest as E.g. 2. In sentiment analysis task, E.g. 2 will be classified into Negative while E.g. 4 into Positive. The sentiment polarities in those two examples are different. However, in stance detection, they will both be classified into FAVOR for the given target “深圳禁摩限电 *Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*”. These examples show the difference between sentiment analysis and stance detection.

3 Dataset for Stance Detection in Chinese Microblogs

Sina Weibo is a popular microblogs platform in China where people express stance implicitly or explicitly. Thus, the dataset for stance detection is constructed based on microblogs from Sina Weibo. The target-microblog pairs are firstly identified. In the second step, the stance of the microblog author to the target is annotated.

3.1 Dataset Construction and Annotation

The target-microblog pairs selection followed the guideline as given below:

1. The discussions and opinions on the target are hot, that is to say there are enough amounts of stances of FAVOR, AGAINST, and NONE.
2. The microblogs consists of complete sentences.
3. The dataset must contains sufficient microblogs covering different kinds of target mention situations, namely target are explicitly mentioned in microblogs/target are not explicitly mentioned or referred to/and others.

Totally, six targets are selected. The corresponding raw microblogs are retrieved from Sina Weibo. These microblogs are manually selected to cover different kinds of target mention situations.

The annotation of this dataset is manually conducted by a group of research students. The major instructions in stance labeling including:

1. Possible stance labels are FAVOR, AGAINST, NONE.
2. If the target of interest is explicitly expressed in the microblog, then directly give stance towards the target of the author.
3. If the target of interest is not explicitly expressed or inferred to in microblog while this microblog indeed is talking about the target, then give stance towards the target of the author based on the combination of comprehension of the microblog and understanding of the issue related to this target.
4. If another target is mentioned in microblog, carefully consider what the opinion is and what the opinion towards to in the microblog, with sound reasoning we can give proper stance towards the target of interest of the author.

The stance of each target-microblog pair is duplicated annotated by two students individually. If these two students provide the same annotation, the stance of this microblog-target pair is then labeled. If the different annotation is detected, the third student will be assigned to annotate this pair. Their annotation results will be voted to obtain the final label.

Table 1. Statistics of NLPCC stance detection dataset

Target	Instances in training dataset				Instances in testing dataset			
	Total	FAVOR	AGAINST	NONE	Total	FAVOR	AGAINST	NONE
Data for Task A								
iPhone SE	600	245	209	146	200	75	104	21
春节放鞭炮	600	250	250	100	200	88	94	18
俄罗斯在叙利亚的反恐行动	600	250	250	100	200	94	86	20
开放二胎政策	600	260	200	140	200	99	95	6
深圳禁摩限电	600	160	300	140	200	63	110	27
Data for Task B								
转基因食品	1000	-	-	-	200	55	97	48
朝鲜核试验	1000	-	-	-	200	39	98	63

3.2 Statistics of the Dataset

Table 1 lists the statistics of instances in the training and the testing datasets for Task A and Task B, respectively.

The stance distribution of instances is roughly the same as in real situations. Since we empirically selected the hot targets, the discussions from different points of views which lead to different stances, are collected and annotated. Meanwhile, there are many discussions with neutral stance or even lack of concern.

4 Evaluation Settings

The stance detection evaluation consists of two sub-tasks: Task A (supervised framework) and Task B (unsupervised framework). Task A is a mandatory task. Each participant is required submit the results for this task. Task B is an optional task. Each participant is allowed to submit only one running result for each task. In the running result, each microblog should be classified into three classes:

FAVOR: The author is in favor of the target (e.g., directly or indirectly by supporting someone/something, by opposing or criticizing someone/something opposed to the target, or by echoing the stance of somebody else).
AGAINST: The author is against the target (e.g., directly or indirectly by opposing or criticizing someone/something, by supporting someone/something opposed to the target, or by echoing the stance of somebody else).
NONE: None of the above.

Notice that NONE could be either the cases that the author has a neutral stance towards the target or the cases that there is no clue about what stance the author holds.

4.1 Sub-tasks

The microblogs corresponding to five targets are selected in Task A, including “iPhone SE”, “春节放鞭炮 *Set off firecrackers in the Spring Festival*”, “俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*”, “开放二胎政策 *Two child policy*”, and “深圳禁摩限电 *Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*”. For each target, there are 600 labeled training data instances and 3000 testing data instances.

The two targets, “转基因食品 *Genetically Modified Foods*” and “朝鲜核试验 *The Nuclear Test of DPRK”* are selected for Task B. Participants were provided 6000 related instances without any label.

4.2 Evaluation Metrics

Macro-average of F-score (FAVOR) and F-score (AGAINST) is employed as the bottom-line evaluation metric for both Task A and Task B, as shown below:

$$F_{AVG} = \frac{F_{FAVOR} + F_{AGAINST}}{2} \tag{1}$$

where F_{FAVOR} and $F_{AGAINST}$ are calculated as follows, respectively:

$$F_{FAVOR} = \frac{2 * P_{FAVOR} * R_{FAVOR}}{P_{FAVOR} + R_{FAVOR}} \tag{2}$$

$$F_{AGAINST} = \frac{2 * P_{AGAINST} * R_{AGAINST}}{P_{AGAINST} + R_{AGAINST}} \tag{3}$$

where, P and R are the for precision and recall, respectively. Note that only 'FAVOR' class and 'AGAINST' class are considered in evaluation metrics, because we take 'NONE' class as the negative class in this information retrieval case. Although 'NONE' class was not shown in evaluation metric, it was not disregarded since it affects the scores of recalls in the evaluation metric if falsely labeled. The macro F-scores of F_{FAVOR} and $F_{AGAINST}$ is equivalent to the micro F-scores over all the targets. Alternatively, macro F-scores could be determined by the mean of the F_{AVG} scores for each target. To reveal the performance in the whole dataset, the former was chosen as the official evaluation metric.

5 Submission Results and Discussions

There are two tasks in evaluation. 16 participants submitted valid results for Task A. Among them, 5 participants submitted valid results for Task B.

5.1 Submission Result for Task A

As mentioned before, Task A is a supervised task which aims to detect stance towards five targets. Here, Target-1 to Target are in turn corresponding to "iPhone SE", "春节放鞭炮 *Set off firecrackers in the Spring Festival*", "俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*", "开放二胎政策 *Two child policy*", and "深圳禁摩限电 *Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen*", respectively. The achieved performances on the overall and each target are listed in Table 2, respectively.

It is shown that the highest performance achieved in Task A is submitted by RUC_MMC. The achieved F_{FAVOR}, $F_{AGAINST}$, F_{AVG} are 0.6969, 0.7243, 0.7106, respectively. This system trained five separate models corresponding to five targets. In their model, five types of features including unigram, TFIDF, synonym, word vector and character vectors are employed. These features are adopted in the classifier based on Support Vector Machine (SVMs) and Random Forest with grid search of parameters.

Furthermore, it is shown that, generally speaking, the achieved F-value performance on Target-2 (iPhone SE) and Target-3 (俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*) are much lower than other targets. This may partially due to the fact that target-2 and target-3 have little related Hashtags in the corpus.

Table 2. Evaluation results for Task A

Team ID	OVERALL			Target-1	Target-2	Target-3	Target-4	Target-5
	F_{FAVOR}	$F_{AGAINST}$	F_{AVG}	F_{AVG}	F_{AVG}	F_{AVG}	F_{AVG}	F_{AVG}
RUC_MMC	0.6969	0.7243	0.7106	0.7730	0.5780	0.5814	0.8036	0.7652
TopTeam	0.6601	0.7186	0.6894	0.7449	0.5764	0.5232	0.7661	0.7949
SDS	0.6758	0.6965	0.6861	0.7784	0.5852	0.5332	0.7948	0.6883
CBrain	0.6618	0.7094	0.6856	0.7604	0.5528	0.4787	0.8135	0.7855
nlp_polyu	0.6476	0.6870	0.6673	0.7354	0.5312	0.5584	0.7708	0.7090
Scau_SDCM*	0.6304	0.7027	0.6666	0.7033	0.5493	0.5780	0.7639	0.7138
NEUDM	0.6268	0.6858	0.6563	0.7173	0.5485	0.5240	0.7497	0.7052
Printf	0.6183	0.6702	0.6443	0.7048	0.5769	0.5547	0.7150	0.6417
CQUT_AC996	0.5897	0.6557	0.6227	0.7015	0.4646	0.5280	0.7661	0.5879
March*	0.5858	0.6244	0.6051	0.6950	0.5466	0.4906	0.6442	0.6169
BIT_NLP_FC*	0.5573	0.5833	0.5703	0.7444	0.3460	0.3769	0.5888	0.4195
HLJUNLP	0.4584	0.6729	0.5656	0.5281	0.4494	0.5126	0.7553	0.4355
CIST-BUPT	0.4660	0.6136	0.5398	0.4754	0.4579	0.5003	0.6867	0.5048
Lib1010	0.4636	0.4944	0.4790	0.4551	0.4420	0.4934	0.4946	0.5045
USCGreenTree*	0.3609	0.5904	0.4756	0.4799	0.4052	0.4586	0.5288	0.3871
SCHOOL	0.3329	0.4662	0.3995	0.3422	0.4222	0.3903	0.4613	0.3676

The team ID with * means late submission.

Meanwhile, more named entities in the microblogs of these two targets affected the stance detection.

Most participators used textual features such as TF-IDF, word Uni-grams, word Bigrams, word embedding vectors, and sentiment lexicons like Boson lexicon, HowNet sentiment lexicon, etc. We noticed that two participators constructed target related lexicons for stance detection. TopTeam constructed an Internet-Sentiment dictionary and a Domain-Sentiment dictionary. BIT_NLP_FC employed a domain-dictionary and sentiment patterns. It seems that domain related sentiment lexicon brings few improvement contributions.

Many participators employed multiple classifiers in this evaluation. RUC_MMC employed multiple target-dependent classifiers to handle target-dependent stance problem. CBrain and nlp_polyu adopted ensemble learning framework which incorporates multiple base classifiers for stance detection.

5.2 Submission Result for Task B

Task B is an unsupervised task which aims to detect stance towards two targets without manually labeled data. Target-6 and Target-7 are corresponding to "转基因食品 G*enetically modified food*" and "朝鲜核试验 *The Nuclear Test in DPRK*", respectively. The achieved performances on the overall and each target are listed in Table 3, respectively.

It is observed that the achieved performances in Task B are evidently lower than performances achieved in Task A. Even though the lower performance is foreseeable

Table 3. Evaluation results for Task B

Team ID	OVERALL			Target-6	Target-7
	F_{FAVOR}	$F_{AGAINST}$	F_{AVG}	F_{AVG}	F_{AVG}
March*	0.3707	0.5667	0.4687	0.5173	0.4165
BIT_NLP_FC*	0.2706	0.6137	0.4421	0.4485	0.4289
CQUT_AC996	0.2985	0.5455	0.4220	0.4562	0.3815
TopTeam	0.0000	0.6555	0.3277	0.3266	0.3289
NEUDM	0.2478	0.3987	0.3232	0.1730	0.3628

The team ID with * means late submission.

for no training data, the achieved performance is still lower than expectation. The system with highest performance is submitted by March. The achieved F_{FAVOR}, $F_{AGAINST}$, F_{AVG} are 0.3707, 0.5667 and 0.4687, respectively. Such F-value is even lower than the second lowest performance in Task A.

Another phenomenon is that the achieved $F_{AGAINST}$ performances are much higher than F_{FAVOR} performances for most participant systems. This partially attributes to the fact that the stance of favor is much less than that of against in the microblogs in Task B. March has gained the highest F_{FAVOR} of 0.3707.

5.3 Discussions

Generally speaking, it is observed that the performances for both Task A and Task B have a lot of room for improvement, especially for Task B. The analysis on the evaluation results show that a good sentiment analysis system cannot ensure a high performance in stance detection task. It means that determining the stance side of mentioned named entity is important to stance detection. For example, for Target 3 (俄罗斯在叙利亚的反恐行动 *Russia's anti terrorist operations in Syria*), the named entities like 俄罗斯 *Russia*, 俄国 *Russia*, 老毛子 *Russian*, 普京 *Putin*, 战斗民族 *Fighting nation*, 医生 *Doctor* (*referred to President of Syria*) and 政府军 *Government troops* are in the side of Russia while the named entities like 美国 *U.S.A*, 米国 *U.S.A*, 反政府军 *Rebel forces* and ISIS are on the opposite side. Thus, the sentiment polarity of the sentence "反政府军真汉子 *The rebel forces are true men*" is positive but the stance of this microblog is AGAINST because it opposite the Russia's anti terrorist operations. How to obtain such knowledge is a big challenge.

Furthermore, in many cases, the microblogs do not mention the target of the interest explicitly at all. For an example sentence, "七十二个处女也不管用 *Seventy-two virgin are not helpful*", its stance determination is hard. If we know the saying that a dead Islam martyr will get seventy-two virgins as wives, we will know that this sentence opposite the Islam fighters. It means that the stance for Russia's anti terrorist operations described in this sentence is FAVOR. This example shows that the collection, understanding and utilization of background knowledge are important to stance detection. Meanwhile, the reasoning based on natural language understanding is helpful.

We hope the works in the future bring inspirations and show more possibilities in stance detection.

6 Conclusions

In this paper, we present the overview of the NLPCC shared task on stance detection in Chinese microblogs. The microblogs related to seven pre-defined targets were collected from Sina Weibo. The stance corresponding to the targets are then annotated to construct a corpus for stance detection task, namely whether the author is in favor of the given target, against the given target, or whether neither inference is likely. This corpus is divided to training dataset and testing dataset for the NLPCC stance detection evaluation. In this evaluation, we designed two tasks. Task A is a mandatory supervised task and Task B is an optional unsupervised task. Sixteen team participants and five team participants submitted their results for Task A and Task B, respectively. The highest F-score achieved was 0.7106 for Task A and 0.4687 for Task B, respectively. This shared task is the first attempt on stance detection in Chinese microblogs. It is expected to promote the research on stance detection.

Acknowledgment. We thank the hard work of Feng Shi, Qikang Wei, Jiannan Hu, Yao Yan, and Zhishan Zhao for preparing the dataset. This work was supported by the National Natural Science Foundation of China 61370165, 61632011, National 863 Program of China 2015AA015405, Shenzhen Peacock Plan Research Grant KQCX20140521144507925 and Shenzhen Foundational Research Funding JCYJ20150625142543470, Guangdong Provincial Engineering Technology Research Center for Data Science 2016KF09.

References

Gui, L., Xu, R., He, Y., Lu, Q., Wei, Z.: Intersubjectivity and sentiment: from language to knowledge. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 2789–2795 (2016)

Tang, D., Wei, F., Yang, N., et al.: Learning sentiment-specific word embedding for Twitter sentiment classification. In: Proceedings Annual Meeting of the Association for Computational Linguistics, pp. 1555–1565, 23–25 June 2014

He, Y., Saif, H., Wei, Z., Wong, K.-F.: Quantising opinion for political tweets analysis. In: LREC 2012, Eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey, pp. 3901–3906 (2012)

Gui, L., Xu, R., Lu, Q., Xu, J., et al.: Cross-lingual opinion analysis via negative transfer detection. In: Proceedings Annual Meeting of the Association for Computational Linguistics, pp. 860–865, 23–25 June 2014

Anand, P., Walker, M., Abbott, R., Tree, J.E.F., Bowmani, R., Minor, M.: Cats rule and dogs drool!: classifying stance in online debate. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 1–9, June 2011

Boltužić, F., Šnajder, J.: Back up your stance: recognizing arguments in online discussions. In: Proceedings of the First Workshop on Argumentation Mining, pp. 49–58, 26 June 2014

Mohammad, S.M., Kiritchenko, S.: SemEval-2016 task 6: detecting stance in tweets. In: International Workshop on Semantic Evaluation (2016)