Argument based Online Deceptive Review Spam Detection

Haijing Liu
Institute of Software Chinese Academy of Sciences
University of Chinese Academy of Sciences
haijing2015@iscas.ac.cn

Pin Lv
Institute of Software Chinese Academy of Sciences
lvpin@iscas.ac.cn

Shiqiang Geng
School of Automation, Beijing Information Science and Technology University
gsq_study@126.com

Xueyu Duan, Xinwei Zhao, Junxing Hu
Institute of Software Chinese Academy of Sciences
University of Chinese Academy of Sciences
{duanxueyu16,zhaoxinwei16,hujunxing17}@mails.ucas.ac.cn

ABSTRACT

Deceptive reviews, which are deliberately made to deceive the readers, sometimes can be very similar to genuine reviews, rendering the deceptive reviews detection a challenging task. Existing approaches mainly focus on designing linguistic features and few work studies this task from the *argumentation* perspective. We observe that the evidence-conclusion discourse relations, also known as arguments, often appear in online reviews, and our hypothesis is that some argument based features, e.g. the percentage of argumentative sentences, the conclusions-evidence ratios, are good indicators between deceptive and truthful reviews. We manually annotate arguments in 80 hotel reviews, and use the argument based features proposed by [24] to investigate our hypothesis. Empirical experiments show that, using the argument based features alone outperforms the state-of-the-art baseline features; in addition, when be used together with the argument based features, the state-of-the-art baseline features can also have a performance improvement.

KEYWORDS

Argumentat(ion); Deceptive Opinion Spam; Web Mining; Computational Linguistics; Machine Learning

ACM Reference format:
Haijing Liu, Pin Lv, Shiqiang Geng, and Xueyu Duan, Xinwei Zhao, Junxing Hu. 2017. Argument based Online Deceptive Review Spam Detection. In *Proceedings of The 2017 SIGIR Workshop On eCommerce, Tokyo Japan, August 2017 (SIGIR eCom),* 6 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Product reviews keep playing an important role in the decision process of online shopping, which can influence and form consumers opinion and subsequently affect sales [3, 4, 9]. Accordingly, there is a growing incentive for businesses to solicit and manufacture *deceptive opinion spam*–fictitious reviews that have been deliberately written to sound authentic and deceive the reader [31]. However, [31] shows that deceptive opinion spam is not easily identified by human readers. Therefore, automatically identifying online deceptive reviews becomes an important task, and has received considerable research attentions in recent years [13, 14, 18, 19, 22, 26, 30, 31].

The majority of existing work focus on designing efficient features for identifying the deceptive reviews. Roughly, the widely used features mainly include two categories: the *review centric* and *reviewer centric* features [8]. [19, 30, 31] extract review centric features from review text, such as *POS tag* features, *n-gram* features and *LIWC*. The *n-gram* features perform best among them [31]. While [14, 22, 26] explore the reviewer centric features, such as *reviewer ratings, percentage of positive reviews* and *maximum content similarity*. However, there is few work focuses on detecting deceptive reviews from the *argumentation* perspective.

An argument is a basic unit people use to persuade their audiences to accept a particular state of affairs [10]. An argument usually consists of a *claim* and some *premises* offered in support of the claim. For example, consider the following review excerpt: "*The staff was amazing, and went out of their way to help us*"; the text segments before the comma is a claim, and the texts after the comma is a premise supporting the claim. [24] solves the review helpfulness prediction problem by using argument based features they proposed and they claim that the argument based features are good intrinsic characteristic of helpful online reviews. In this work, we try to explore whether the argument based features are helpful to identify online deceptive reviews. [6] designs argumentative frameworks to construct argumentative features to detect deceptive reviews, however it mainly focuses on argument strength and yield a slight improvement. In this work, we pay attention to the argument components in the reviews and try to study the argument features based on components.

Following [24], we manually annotated arguments in 80 hotel reviews and made it publicly available, so as to use these "ground truth" arguments to extract argument based features. Similar to [19, 30, 31], we model the deceptive review identification task as a classification problem. Empirical results based on our corpus suggest that, the argument based

SIGIR eCom, August 2017, Tokyo Japan

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

features outperform each of four baseline features in accuracy, precision, recall, F1-score and AUC, by up to 5.30%, 4.38%, 5.30%, 5.67% and 5.30% on average, resp. Moreover, combined with the argument based features, each type of baseline features has an improvement in accuracy, precision, recall, F1-score and AUC on our annotated corpus, by up to 6.08%, 5.21%, 6.08%, 6.45% and 6.08% on average, resp. Furthermore, the dimension of the argument based features is lower than the best state-of-the-art features: n-grams features, which the dimension is huge and online learning is problematic. Finally, we use the effective argument based features to give some insights into the difference between deceptive and truthful reviews.

The rest of this paper is organized as follows. Section 2 presents the related work of argument based online deceptive review spam detection. Then we describe our corpus annotation in Section 3. Details about the argument based features are introduced in Section 4. The experiment and results are described in Section 5. Then, we give some discussion in Section 6. Finally, we conclude and present directions for future research in Section 7.

2 RELATED WORK

Two lines of researches are closely related to argument based online deceptive review spam detection problem we describe in this paper: deceptive review spam detection and argumentation mining application.

2.1 Deceptive Review Spam Detection

Web spam [1, 33] and email spam [5] are most widely studied spam activities [29]. Recently there has been increasing concern about deceptive opinion spam [13, 14, 18, 19, 22, 26, 30, 31]. [15] first studied the deceptive opinion problem and then much significant work has been done on this task [14, 16, 19, 22, 26–28, 30, 31]. Some literature [19, 30, 31] studied review centric features, especially linguistic features, such as *POS tag* features, *n-gram* features and *LIWC* features from review text. [31] created a gold-standard collection by employing Turkers to write fake reviews, and follow-up research was based on their data [19, 30].

Except the research based on review text above, some existing work [14, 22] explored the reviewer centric features, such as *reviewer ratings, percentage of positive reviews* and *maximum content similarity.* [26] studied reviewers' behavior to identify deceptive reviews and [27, 28] considered the group behavior to identify group of fake reviewers. However, reviewer centric features are ineffective at times, for example, when the deceptive reviewers change their account name or make a new account(cold start), blocking deceptive reviewers' access is no longer useful filtering method [16]. Hence, our paper pay attention to the review centric features, i.e. the *argument* based features.

2.2 Argumentation Mining Application

Argumentation mining [23, 25] is a branch of philosophy that studies the act or process of forming reasons and of drawing conclusions and applying them to various domains, such as legal texts[32], scientific articles [7], persuasive essays[34], online forums[38], online reviews [24, 36, 37], and many others. Recently, various applications in the field of argument mining have been gaining more attention, [35] used argument mining to assess the argumentation quality of essay, [38] explored argumentative features to rank comments in the online forum.

With the recent advances of argumentation mining application in multiple domains, there are growing attention the application of argumentation in the online user review. [36, 37] studied sentiment analysis using argumentation. [6] designed argumentative frameworks to construct argumentative features to detect deceptive reviews. [21] provided an argument corpus of Chinese hotel reviews via using crowdsourcing technique. [24] annotated another English hotel review corpus and solved the review helpfulness prediction problem using argument based features. In this paper, we annotate arguments in 80 hotel reviews sampled from existing gold standard corpus and use the argument based features to solve the detection of online deceptive reviews.

3 CORPUS

We choose the corpus from the gold standard corpus of deceptive reviews built by [30], which consists of 800 deceptive and 800 truthful hotel reviews from the Tripadvisor website. To annotate the "ground truth" argument structures, we randomly sampled 80 reviews of the Hilton hotel, including 40 deceptive and 40 truthful reviews, both of them have 20 positive and 20 negative reviews. We follow the annotated scheme in [24] to annotate arguments in these 80 hotel reivews [1]. Specifically, in line with [36], we viewed each subsentence in the review as a clause and asked three annotators independently to annotate each clause as one of the following seven categories of argument components:

Major Claim: a summary of the main opinion of a review. For instance, *"I have enjoyed the stay in the hotel", "I am sad to say that i am very disappointed with this hotel"*;

Claim: a subjective opinion on a certain aspect of a hotel. For example, *"The staff was amazing", "The room is spacious"*;

Premise: an objective reason/evidence supporting a claim. For instance, *"The staff went out of their way to help us"*, it supports the first example claim above; *"We had a sitting room as well as a balcony"*, it supports the second example claim above;

Premise Supporting an Implicit Claim (PSIC): an objective reason/evidence that supporting an *implicit claim*, which does appear in review. For instance, *"just five minutes' walk to the down town"* supports some implicit claims like *"the location of the hotel is good"*, although this implicit claims has never appeared in the review;

Background: an objective description that does not give direct opinions but provides some background information. For example, *"We checked into this hotel at midnight", "I*

[1] Corpus available by request from the first author

stayed five nights at this hotel because i was attending a conference at the hotel";

Recommendation: a positive or negative recommendation for the hotel. For instance, *"I would definitely come to this hotel again the next time I visit London", "Do not come to this hotel if you look for some clean places to live"*;

Non-argumentative: for all the other clauses.

	Fleiss Kappa
MajorClaim	0.955
Claim	0.788
Premise	0.600
PSIC	0.944
Background	0.960
Recommendation	1.000
Non-argumentative	0.877

Table 1: The IRA scores for each type of argument components in the hotel review corpus.

Similar to [24], we use the Fleiss kappa metric [12] to evaluate the reliability and quality of the obtained annotations, and the results are presented in Table 1. We can see that the lowest Kappa scores for type *premise* is still 0.6, suggesting that the quality of the annotations are substantial [17]; this indicates there exist little noises even in the ground truth argument structures. In line with [24], we aggregate the annotations using majority voting.

4 FEATURES

Due to the argument based features are extracted from the review text, therefore we consider the review centric features in previous work as our baseline features.

4.1 Baseline Features

We choose four types of review centric features as our baseline features as follows.

POS tag. Previous work [19, 31] indicate the *POS tag* features has a good performance in the deceptive review spam detection task. Similar to [19, 31], we consider the frequencies of each *POS tag* for each review as *POS tag* baseline features.

Unigram. The *Unigram* features has been used widely in the detecting deceptive reviews [19, 30, 31]. Following [19, 30, 31], we use *Unigram* features with lowercased and unstemmed as *Unigram* baseline features.

Bigram$^+$. Existing work [14, 30, 31] suggest that the *Bigram* features has the best performance in all n-gram features. In line with [14, 30, 31], we choose the *Bigram* features as one of the baseline features. Specifically, *Bigram* baseline features subsumes the preceding *Unigram* baseline features.

Trigram$^+$. In addition, we also test the *Trigram* features used in [31]. The *Trigram* baseline features subsumes the preceding *Bigram* baseline features.

4.2 Argument based Features

Following [24], we consider the argument components in each review, and build the argument based features. We consider four granularity of argument features, detailed as follows.

Component-level argument features. A natural feature that we believe to be useful is the ratio of different argument component numbers. For example, we may be interested in the ratio between the number of premises and that of claims; a high ratio suggests that there are more premises supporting each claim, indicating that the review gives many evidences. To generalise this component ratio feature, we propose *component-combination ratio* features: we compute the ratios between any two argument components combinations. For example, we may be interested in the ratio between the number of MajorClaim+Claim+Premise and that of Background+Non-argumentative. As there are 7 types of labels, the number of possible combinations is $2^7 - 1 = 127$, and thus the possible number of combination ratio pairs is $127 \times 126 = 16002$. In other words, the component-level feature is a 16002-dimensional real vector.

Token-level argument features. In a finer-granularity, we consider the number of tokens in argument components to build features: for example, suppose a review has only two claims, one has 10 words and the other has 5 words; we may want to know the average number of words contained in each claim, the total number of words in claims, etc. In total, for each argument component type, we consider 5 types of token-level statistics: the total number of words in the given component type, the length (in terms of word) of the shortest/longest component of the given type, and the mean/variance of the number of words in each component of the given type. Thus, there are in total $7 \times 5 = 35$ features to represent the token-level statistics.

In addition, the ratio of some token-level statistics may also be of interests: for example, given a review, we may want to know the ratio between the number of words in Claims+MajorClaims and that in Premises. Thus, the combination ratio can also be applied here. We consider only the combination ratio for two statistics: the total number of words and the average number of words in each component-combination; hence, there are $16002 \times 2 = 32004$ dimensions for the combination ratio for the statistics. In total, there are $32004 + 35 = 32039$ dimensions for the token-level argument features.

Letter-level argument features. In the finest-granularity, we consider the letter-level features, which may give some information the token-level features do not contain: for example, if a review has a big number of letters and a small number of words, it may suggests that many long and complex words are used in this review, which, in turn, may suggests that the linguistic complexity of the review is relative high and the review may gives some very professional opinions. Similar to the token-level features above, we design 5 types of statistics and their combination ratios. Thus, the dimension for the letter-level features is the same to that of the token-level features.

FeaType	Accuracy	Precision	Recall	F1-score	AUC
AF	78.3%	78.8	78.3	78.3	78.3
POS tag	65.4%	67.1	65.4	64.4	65.4
POS tag + AF	76.9%	77.9	76.9	76.7	76.9
Unigram	75.8%	77.1	75.8	75.5	75.8
Unigram + AF	79.9%	80.3	79.9	79.8	79.9
Bigram$^+$	76.5%	77.8	76.5	76.2	76.5
Bigram$^+$ + AF	**80.1%**	**80.6**	**80.1**	**80.0**	**80.1**
Trigram$^+$	74.2%	75.7	74.2	73.8	74.2
Trigram$^+$ + AF	79.3%	79.8	79.3	79.2	79.3

Table 2: Online deceptive reviews detection performances using argument based features and/or baseline features. AF stands for argument based features.

Position-level argument features. Another dimension to consider argument features is the positions of argument components: for example, if the major claims of a review are all at the very beginning, we may think that readers can more easily grasp the main idea of the review and, thus, the review is more likely to be helpful. For each component, we use a real number to represent its position: for example, if a review has 10 sub-sentences (i.e. clauses) in total and the first sub-sentence the component overlaps is the second sub-sentence, then the position for this component is $2/10 = 0.2$. For each type of argument component, we may be interested in some statistics for its positions: for example, if a review has several premises, we may want to know the location of the earliest/latest appearance of premises, the average position of all premises and its variance, etc. Similar to the token- and letter-level features, we design the same number of features for position-level features.

5 EXPERIMENT AND RESULTS

Similar to [19, 30, 31], we model the deceptive review identification task as a classification problem and thus use accuracy, precision, recall, F1-score and area under the curve (AUC) as evaluation metrics. To determine those argument based features which are good indicators between deceptive and truthful reviews, we choose the argument based features by the information gain value. Finally, we use the argument based features of top 50 information gain value. In line with most existing works on deceptive review identification [19, 30, 31], we use LibSVM [2] as our classifier and conducted the 10-fold cross validation. The experiment results of each feature set are listed in the Table 2.

From Table 2, we can see that the argument based features outperform each of four baseline features in accuracy, precision, recall, F1-score and AUC, by up to 5.30%, 4.38%, 5.30%, 5.67% and 5.30% on average, resp. Moreover, combined with the argument based features, each type of baseline features has an improvement in accuracy, precision, recall, F1-score and AUC on our annotated corpus, by up to 6.08%, 5.21%, 6.08%, 6.45% and 6.08% on average, resp. In line with [31], Bigram$^+$ feature has the best performance in all kinds of n-gram features in all metrics and the performance of POS tag features is lower than those of all three types of n-gram features. Combined with the best review centric baseline features, i.e., Bigram$^+$ baseline features, we achieve the highest performance of deceptive review detection task. These results validate that our hypothesis which argument based features are good indicators between deceptive and truthful reviews is right. Furthermore, due to its fixed and more lower dimensions compared to the n-gram features, the argument based features has a promising application vision.

6 DISCUSSION

To detect the difference between deceptive and truthful reviews from an argumentation perspective, we analyse the argument based features of top 50 information gain. We find that more than half (52%) features are from the token-level argument features. This indicate that the statistics of tokens in the review is an important factor for identifying deceptive reviews, which is in line with the claim that about 80% of spammers have no reviews longer than 135 words while more than 92% of reliable reviewers have an average review length of greater than 200 words by [8]. Moreover, 42% features are from the component-level argument features and this indicates that the component-level argument features also have an effect on identifying deceptive reviews. Furthermore, we find that there are more recommendation components in deceptive reviews than truthful reviews. Compared to the truthful reviews, the deceptive reviews tend to persuade readers by giving more recommendation information. We also observe that in some truthful reviews there are more non-argumentative components which are obviously not related to the opinions about hotel than those deceptive reviews. In some truthful reviews, there are a big part of content about telling reviewer' story in the hotel, which does not contain obvious opinion about hotel. We also find that compared to deceptive reviews, there are more background components in the truthful reviews, which give specific information of their travel, such as specific arrival time, accompanying partners and the purpose of travel. This is in line with the conclusion in [19, 31] that truthful reviews encode more spatial details.

7 CONCLUSION AND FUTURE WORK

In this work, we novelly use the argument based features to detect online deceptive reviews. We manually annotate an argument corpus of 80 hotel reviews and make it publicly available. Based on the corpus, we use the argument based features, and compare the performance of argument based features with that of the state-of-the-art features based on review text. Empirical results suggest that our proposed argument based features outperform the best review centric baseline features, and combined with the argument based features, the performance of each review centric baseline features improves significantly. Furthermore, compared to the n-gram features with best performance in the previous works, the dimension of the argument based features is fixed and more lower, which has a promising application vision. In addition, we analyse the effective argument based features so as to explain the difference between deceptive and truthful reviews from an argumentation perspective.

In future work, we intend to extract the argument based features automatically. Recently, deep-learning based argument mining [11, 20] has produced some promising results , and we plan to investigate the performance of automatically extracting argument based features on deceptive review spam detection. Furthermore, we will try to expand our corpus and annotate some online reviews in other domains to test the effectiveness of the argument based features.

REFERENCES

[1] Carlos Castillo, Debora Donato, Aristides Gionis, Vanessa Murdock, and Fabrizio Silvestri. 2007. Know your neighbors: Web spam detection using the web topology. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 423–430.

[2] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[3] Patrali Chatterjee. 2001. Online Reviews: Do Consumers Use Them? *Advances in Consumer Research* 28 (2001).

[4] Pei-Yu Chen, Shin-yi Wu, and Jungsun Yoon. 2004. The impact of online recommendations and consumer feedback on sales. *ICIS 2004 Proceedings* (2004), 58.

[5] Paul-Alexandru Chirita, Jörg Diederich, and Wolfgang Nejdl. 2005. MailRank: using ranking for spam detection. In *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 373–380.

[6] Oana Cocarascu and Francesca Toni. 2016. Detecting deceptive reviews using Argumentation. In *Proceedings of the 1st International Workshop on AI for Privacy and Security*. ACM, 9.

[7] Danish Contractor, Yufan Guo, and Anna Korhonen. 2012. Using Argumentative Zones for Extractive Summarization of Scientific Articles.. In *COLING*, Vol. 12. 663–678.

[8] Michael Crawford, Taghi M Khoshgoftaar, Joseph D Prusa, Aaron N Richter, and Hamzah Al Najada. 2015. Survey of review spam detection using machine learning techniques. *Journal of Big Data* 2, 1 (2015), 23.

[9] Chrysanthos Dellarocas, Neveen Awad, and Xiaoquan Zhang. 2004. Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning. *ICIS 2004 Proceedings* (2004), 30.

[10] Judith Eckle-Kohler, Roland Kluge, and Iryna Gurevych. 2015. On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse.. In *EMNLP*. 2236–2242.

[11] Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural End-to-End Learning for Computational Argumentation Mining. *arXiv preprint arXiv:1704.06104* (2017).

[12] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[13] ASA Hammad and Alaa El-Halees. 2013. An approach for detecting spam in arabic opinion reviews. *The International Arab Journal of Information Technology* 12 (2013).

[14] Nitin Jindal and Bing Liu. 2007. Analyzing and detecting review spam. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*. IEEE, 547–552.

[15] Nitin Jindal and Bing Liu. 2008. Opinion spam and analysis. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 219–230.

[16] Seongsoon Kim, Hyeokyoon Chang, Seongwoon Lee, Minhwan Yu, and Jaewoo Kang. 2015. Deep semantic frame-based deceptive opinion spam analysis. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1131–1140.

[17] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[18] Fangtao Li, Minlie Huang, Yi Yang, and Xiaoyan Zhu. 2011. Learning to identify review spam. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22. 2488.

[19] Jiwei Li, Myle Ott, Claire Cardie, and Eduard H Hovy. 2014. Towards a General Rule for Identifying Deceptive Opinion Spam.. In *ACL (1)*. 1566–1576.

[20] Minglan Li, Yang Gao, Hui Wen, Yang Du, Haijing Liu, and Hao Wang. 2017. Joint RNN Model for Argument Component Boundary Detection. *arXiv preprint arXiv:1705.02131* (2017).

[21] Mengxue Li, Shiqiang Geng, Yang Gao, Haijing Liu, and Hao Wang. 2017. Crowdsourcing Argumentation Structures in Chinese Hotel Reviews. *arXiv preprint arXiv:1705.02077* (2017).

[22] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 939–948.

[23] Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)* 16, 2 (2016), 10.

[24] Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, and Hao Wang. 2017. Using Argument-based Features to Predict and Analyse Review Helpfulness. (2017). arXiv:1707.07279 arXiv:1707.07279v1.

[25] Marie-Francine Moens. 2013. Argumentation Mining: Where are we now, where do we want to be and how do we get there?. In *Post-Proceedings of the 4th and 5th Workshops of the Forum for Information Retrieval Evaluation*. ACM, 2.

[26] Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 632–640.

[27] Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 191–200.

[28] Arjun Mukherjee, Bing Liu, Junhui Wang, Natalie Glance, and Nitin Jindal. 2011. Detecting group review spam. In *Proceedings of the 20th international conference companion on World wide web*. ACM, 93–94.

[29] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013. What yelp fake review filter might be doing?. In *ICWSM*.

[30] Myle Ott, Claire Cardie, and Jeffrey T Hancock. 2013. Negative Deceptive Opinion Spam.. In *HLT-NAACL*. 497–501.

[31] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 309–319.

[32] Raquel Mochales Palau and Marie-Francine Moens. 2009. Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*. ACM, 98–107.

[33] Nikita Spirin and Jiawei Han. 2012. Survey on web spam detection: principles and algorithms. *ACM SIGKDD Explorations Newsletter* 13, 2 (2012), 50–64.

[34] Christian Stab and Iryna Gurevych. 2014. Identifying Argumentative Discourse Structures in Persuasive Essays.. In *EMNLP*. 46–56.

[35] Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2016. Using Argument Mining to Assess the Argumentation Quality of

Essays. In *Proceedings of the 26th International Conference on Computational Linguistics*.

[36] Henning Wachsmuth, Johannes Kiesel, and Benno Stein. 2015. Sentiment Flow-A General Model of Web Review Argumentation.. In *EMNLP*. 601–611.

[37] Henning Wachsmuth, Martin Trenkmann, Benno Stein, and Gregor Engels. 2014. Modeling Review Argumentation for Robust Sentiment Analysis.. In *COLING*. 553–564.

[38] Zhongyu Wei12, Yang Liu, and Yi Li. 2016. Is This Post Persuasive? Ranking Argumentative Comments in the Online Forum. In *The 54th Annual Meeting of the Association for Computational Linguistics*. 195.