Claim Retrieval in Twitter

[1]Wenjia Ma, [1]Wenhan Chao, [2]Zhunchen Luo (✉), [1]Xin Jiang

[1]School of Computer Science and Engineering, Beihang University
[2]Information Research Center of Military Science
PLA Academy of Military Science
[1]{mawenjia,chaowenhan,xinjiang}@buaa.edu.cn
[2]zhunchenluo@gmail.com

Abstract. Controversial topics, especially the new emerging ones are widely discussed and searched in social medias like Twitter. When people are interested in topics and search on Twitter, high quality tweets are expected to appear at the top. Since it is only argumentation that truly reasons things out, we believe that high quality tweets are those with argumentation that consists of claim and evidence. Moreover, claim is the heart of argumentation, we concentrate on claim retrieval in Twitter. Based on a learning-to-rank framework, we integrate Twitter structural information and topic-independent claim-related lexicon to re-rank the relevant tweet list pre-retrieved by BM25 scores. We also automatically construct topic-dependent claim-oriented lexicons to further elevate the retrieval performance. Additionally, our model can be easily adapted to new topics without any manual process or external information, which guarantees the practicability of our model.

Keywords: Claim Retrieval, Twitter Structural Information, Claim-Oriented Lexicon, Topic Adaptable.

1 Introduction

Since controversial topics, especially the new ones, are widely discussed in Twitter, the search tool of Twitter is frequently used by people. However, the retrieved tweets which only reflect tweeters' opinions or just general support or oppose these controversial topic are not meaningful enough. Argumentation is known as the most convincing structure, which is often used in law, persuasive essay, and debate domain and has been researched for decades. Among diverse argumentation definitions [18, 3, 15, 4, 9], a widespread one is claim and evidence [11]. Due to the short texts, S. and G. [15] point out that argumentation structure is rare, or likely to be incomplete in social media. It means that some tweets may contain only claims, while others may contain only evidences or both claims and evidences. Specifically, the heart of every argumentation lies in a single claim, which is a assertion the argumentation aims to prove [5]. Moreover, only when the claim is confirmed, can the evidences make sense. To help users swiftly obtain many pre-eminent claims about the query topic, there is a pressing need for tools that can automatically retrieve claim-oriented tweets.

Hence, given a topic, our task aims to retrieve a list of claim-oriented tweets. We assume a claim-oriented tweet should meet three criteria: 1) the tweet should be topic-related; 2) the tweet clearly supports or opposes the topic; 3) the tweet provides an

Topic: Abortion (should abortion be allowed)		
T1	RT @nelsonhardiman: Supreme Court strikes down Texas abortion restrictions: https://t.co/xsRz8IHiIK#SCOTUS #abortion#SupremeCourt #Texas	N
T2	@patrickmadrid she support abortion I say abortion is murder. before they were even born	Y
T3	@okeyjames i.e therapeutic abortion is allowed in Nigeria.	N
T4	like omfg how does someone else getting an abortion affect you in any way. if you're pregnant; want an abortion, get an abortion +	N
Topic: Animal testing (should animal testing be allowed)		
T5	I've just watched a disgusting video about animal testing and tomorrow I'm throwing all my none cruelty free makeup out.	Y

Table 1: Examples for tweets separately relevant to two topics, "abortion" and "animal testing". "Y" means it contains a claim and "N" means it does not.

arguable reason[1] for its stance. For examples, as shown in Table 1: **T1** is a piece of news which contains no stance; **T2** is clearly against the topic, and contains an explicit disputable reason, *"abortion is murder"*; **T3** is a objective truth which is not in dispute (seems like an evidence); **T4** just has an opposing stance without showing a reason; **T5** contains an implicit claim, *"animal testing used by cosmetics is cruel"*. Consequently, **T2** and **T5** are claim-oriented tweets that we need to retrieve.

Previous studies of predicting whether a document contains claims use supervised learning approaches [5, 14], parse tree measures [6], and more recent works concentrating on neural networks [2]. There are two major challenges rendering these approaches not suitable for our task.

Chaotic Twitter. Tweets are short and often contain specific conventions. For instance, in the first sample in Table 1, tweet contains hashtags, URLs, and re-tweet (RT@), while the textual content are really short. Cleaning these Twitter specific conventions using NLP techniques will cause incomplete sematic of the tweet. Therefore, these chaotic elements in Twitter represent an open challenge for standard claim detection approaches.

Vague Claim. In fact, the majority of online users do not really need to present a well-formed argumentation or their proposition. As a consequence, claims made by the users will often be unclear, ambiguous, vague, or simply poorly worded [16]. For example, people need background knowledge *"cosmetics often use animal testing"* to recognize that **T5** in Table 1 contains an implicit claim *"animal testing used by cosmetics is cruel"*, which is clearly challenging.

In this paper, we explore both Twitter structural information and claim-oriented information to address the above issues. Twitter structural information refers to hashtags, URLs, re-tweet (RT@), etc. And the claim-oriented information denotes indicative words whose appearances represents that the tweet is likely to contain claim. First, We

[1] This is to distinguish from "evidence" or "data" which is essential prerequisite for world knowledge [18].

utilize a learning-to-rank framework to learn a ranking function that uses both Twitter structural information and topic-independent claim-related information[2] in addition to traditional topic-related information and stance information. And then we elevate the performance by automatically generate topic-dependent claim-oriented lexicons and use them in a lexicon-based approach. Additionally, since the topic-dependent claim-oriented lexicon can be constructed using unlabeled topic-relevant tweets, our model can be easily adapted to new topics which guarantees the practicability of our model. The contributions of this work can be summarized as follows:

1) We define a novel claim-oriented tweet retrieval task. We construct a real-world dataset for this task.
2) Our method integrates both topic-independent and topic-dependent claim-oriented information and achieves portability to all controversial topics.
3) Experimental results show that best performance of our ranking model is significantly better than baselines.

2 Related Work

The task of automatic claim-oriented document detection was first introduced by Levy et al. [5] who used a supervised learning approach to detect context dependent claims in Wikipedia articles. Lippi and Torroni [6] focused on the rhetoric structure of claims and relied on the ability of Partial Tree Kernels to generate the feature set. More recently, Roitman et al. [14] proposed a two-step retrieval approach to do claim-oriented document retrieval task, and they concentrated on retrieving as many relevant claims as possible from wikipedia corpus. Our experimental results show that claim-oriented document retrieval features do not perform well in Twitter.

Our task shares relationship with argument mining in Twitter or online forum [10, 1, 17, 19]. Theodosis et al. [17] did not distinguish between domain entities and claims, since they thought the claims are not expressed literally. However, in our opinion, both explicit and implicit claims are contained in tweets, and only when the claim is confirmed, can the evidences make sense. Other examples often considered argument as evidence. Addawood and Bashir [1] used a supervised classifier trained with different kinds of features to capture the evidence types in social media. To conclude, none of the work mentioned above concentrated on claim mining in Twitter.

Since we define the claim-oriented tweet should contain a clear stance, stance detection in Twitter is also important for our task. M. et al. [8] proposed a state-of-art stance detection system using a SVM classifier along with distant supervision techniques. We use their features to measure whether there are stances in tweets.

3 Methodology

To generate a good function which ranks the tweets according to our principle for finding claim-oriented tweets, we investigate the features concerning topic relevance, stance

[2] Claim-related information refers to words whose appearance can make information gain for detecting whether a tweet contains claim.

existence and arguable reason inclusion of a tweet. In general, we use a learning-to-rank framework to integrate topic-related feature, stance detection features, Twitter structural features and topic-independent claim-related features. To further elevate the retrieval performance, we use a topic-dependent claim-oriented lexicon to score whether each tweet contains arguable reasons.

3.1 Learning to Rank Method

Learning-to-rank is a data driven approach that effectively incorporates a bag of features into the retrieval process. To generate a general model for all kinds of controversial topics, we develop topic-independent features into a learning-to-rank scenario. In the remainder of this section, we will focus on these topic-independent features.

Relevance Feature We use the Okapi BM25 [13] to measure the relevance between topics and tweets.

Stance Features Since the claim-oriented tweets need to express a clear stance toward the given controversial topic, we use a feature set ***TwitStan*** integrated in a state-of-art classifier which is proposed by M. et al. [8] to address the SemEval-2016 task on stance detection in Twitter. The features used for our method include n-grams, sentiment, target, POS, encodings, and word embeddings trained on large collections of tweets in November 2015 using Glove [12].

Twitter Structural Features Compared to traditional media data, Twitter has many specific structural information, such as URLs, hashtags, ect. Some of them have been proved to have significant influence on Twitter retrieval [7]. However, most argument mining works in Twitter treat tweets as plain texts by removing them [1]. This may lead to the information loss of tweets. To explore the relationship between Twitter structural information and claim-oriented tweets, we use them as binary features.

"***RT*** **@**" indicates copying and rebroadcasting of the original tweet, we assume that persuasive tweets containing clear propositions are more likely to be broadcasted. ***URL*** indicates the links to out side content. Observationally, advertisements and news that are unlikely to contain a claim in Twitter often contain a URL. Inspired by the assumption that high quality claims arise in debates or quarrels, we use "***reply***" which describes whether this tweet is a comment or a reply.

Topic-Independent Claim-Related Features Some claim-oriented tweets expressed arguable reasons explicitly, and they often express in general patterns, for instance,

1) @mmfa Abortion is not a choice, abortion is the killing of an innocent life
2) RT @hailey stiegel: MAKING ABORTION ILLEGAL IS NOT GETTING RID OF ABORTION, IT IS GETTING RID OF SAFE ABORTION

	t	$\neg t$	Row total
Claim_Oriented. set	C_{11}	C_{12}	C_{1*}
Non_Claim. set	C_{21}	C_{22}	C_{2*}
Col. total	C_{*1}	C_{*2}	C

Table 2: Table for information gain. $C_{1*} = C_{11} + C_{12}$; $C_{2*} = C_{21} + C_{22}$; $C_{*1} = C_{11} + C_{21}$; $C_{*2} = C_{12} + C_{22}$; $C = C_{11} + C_{12} + C_{21} + C_{22}$.

"`A is not B, it is C`" pattern appears in these explicit claim-oriented tweets. In order to capture these claim-oriented patterns, which involve be verbs, modal verb, we utilize an information gain based method to calculate the claim score of each word.

C_{ij} in Table 2 indicates the number of tweets having / not-having term *t* in the claim-oriented / non-claim set respectively. For example, C_{11} is the number of claim-oriented tweets which contain term *t*. Then, we give definitions of some concepts: $H(X)$ is the entropy of X. For each topic, the total claim entropy is called $H(C) = -\sum_{i=1}^{2} p_{i*} \log_2 p_{i*}$, where $p_{i*} = \frac{C_{i*}}{C}$ is the probability of the C_{i*}. For each term *t*, we compute the entropy of claim on the term *t* $H(C|t)$ as follows:

$$H(C|t) = -p_t \sum_{i=1}^{2} p(C_i|t) \log_2 p(C_i|t) - p_{(\neg t)} \sum_{i=1}^{2} p(C_i|\neg t) \log_2 p(C_i|\neg t) \quad (1)$$

$IG(C,t) = H(C) - H(C|t)$ calculates the information gain about claim of term *t*. The number of claim-oriented tweets varies from topics. For example, there are 40 tweets containing claims in topic **"abortion"**, but only 2 tweets contain claims on topic "**Trump**". Therefore, tweets about topic **"abortion"** are more likely to contain claims. If term scores are calculated without considering the topic, insignificant topic words will score higher and be seen as claim-oriented words. For instance, "abortion", "woman" (high frequency words on topic "**abortion**") etc. To avoid this situation, term scores are calculated separately according to topics. For each term *t*, we use $H(t|K) = \sum_{i=1}^{n} p_{k_i} H(t|K = k_i)$ to represent *t*'s distribution under the topic set K.

If term *t* is a topic-independent claim indicator, it should be evenly distributed under various topics. And this situation will cause $H(t|K)$ to increase. Therefore, *t*'s score $Claim_{TI}(t)$ which used to indicate claim relatedness is calculated as follows:

$$Claim_{TI}(t) = \sum_{k \in K} \frac{IG_k(C,t) \cdot H(t|K)}{TN_k} \quad (2)$$

where TN_k is the number of tweets about topic k. The highest score terms are selected to form the **T**opic-**I**ndependent **C**laim-**R**elated **Lex**icon ***TICRLex*** and will be used as topic-independent claim-related features.

3.2 Lexicon Method

Some arguable reasons in claim-oriented tweets are expressed implicitly. For instance, there are 2 tweets of topic **"death penalty"**:

> *1) @mmellmmar because death penalty treats you better if you are rich and guilty than if you are poor and innocent..*

2) Death penalty should not exist, esp because it is against those who are poor.#deathpenalty

They expressed the claim that *"the death penalty for the poor and the rich is different"*, which requires background knowledge to identify. We find that these implicit claim-oriented tweets often contain some topic-dependent words, like "poor", "rich" with topic "**death penalty**". To capture these words, we develop a approach to automatically generate topic-dependent claim-oriented lexicons using unlabeled topic-related tweets.. Additionally, since it is impossible to train a supervised model for every topic, we use topic-dependent claim-oriented lexicons in a lexicon-based method. We estimate the claim-oriented score of each tweet by calculating the average claim-oriented score over certain terms.

Topic-Dependent Claim-Oriented Lexicon We suppose that if term *t* often appear with topic-independent claim-oriented words simultaneously, then term *t* is likely to be a claim-oriented word. In the above two examples, we suppose that term "because" is a topic-independent claim-oriented word. The term "poor" appear with "because" twice in these two tweets. Since we suppose that topic-dependent claim-oriented and topic-independent claim-oriented words are often united, term "poor" can be seen as a claim-oriented word of topic "**death penalty**".

First, suppose we have already got the topic-independent claim-related lexicon ***TICRLex***. To distinguish claim-oriented terms in the claim-related lexicon, we introduce a signal function $Sgn(t)$ for each term *t*:

$$Sgn(t) = \begin{cases} -1 & \frac{C_{11}}{C_{*1}} \leq \frac{C_{1*}}{C} \\ 1 & \frac{C_{11}}{C_{*1}} > \frac{C_{1*}}{C} \end{cases} \tag{3}$$

$Claim_{TI}(t)$ is the term *t*'s claim score in ***TICRLex***. Then we compute the new score $Claim_{TI}(t)^+ = Claim_{TI}(t) \cdot Sgn(t)$ of each term *t* in ***TICRLex***. If $Claim_{TI}(t)^+ > 0$, means term *t* is **pos**itively related to claim, we add *t* to a new **Lex**icon called ***posLex***.

$CoT(w_i, t)$ represents the co-occurrence frequency of term *t* in topic-related tweet set TS with the term w_i in ***posLex***. TN_t is the number of tweets containing term t. *t*'s topic-dependent claim-oriented score $Claim_{TD}(t)$ is then defined as the weighted sum of $CoT(w_i, t)$:

$$Claim_{TD}(t) = \sum_{w_i \in \textbf{\textit{posLex}}} \frac{Claim_{TI}(w_i)^+ \cdot CoT(w_i, t)}{TN_t} \tag{4}$$

The highest score terms are selected to form the **T**opic-**D**ndependent **C**laim-**O**riented **Lex**icon ***TDCOLex***.

4 Experiments

4.1 Datasets

We construct a real-world dataset for our claim-oriented tweet retrieval task[3]. We crawled and indexed about 90 million tweets using the Twitter API in 2016 and reserve the

[3] `https://sourceforge.net/projects/claimretrieval/files/corpus/download`

English tweets. Using these tweets we implemented a search engine based on ElasticSearch[4]. We collected 30 debate topics from debate webisite[5] as the queries. Given a query the search engine would present a list of relevant tweets ranked based on the Okapi BM25 [13] score. A native English speaker and two experienced annotators with NLP background were hired to identify whether the tweet contains a claim following the criteria we proposed (in Section 1) by assigning binary labels to every tweet. The inter-annotator agreement was 90.1% for topic-relevance, 78.2% for clear stance and 75.2% for arguable reason[6]. The high consistency of the annotation proves our claim-oriented criteria are easy to convey to human labelers. We marked an instance with a claim only if at least 2 annotators labeled them as containing claim.Totally, 2520 tweets were selected for study and 586 tweets were identified as containing claims.

4.2 Experimental Settings

For learning to rank, SVM light[7] which implements the ranking algorithm is used. To avoid overfitting, we perform 10 fold cross-validation in our dataset. We use Mean Average Precision (MAP), Precision@5, and Precision@10 as evaluation metrics.

4.3 Baselines

We investigate the features used by previous similar tasks, and separately develop these bags of features into a learning-to-rank scenario as our baselines.

BM25 Similarity. We use BM25 similarity as a basic measure. The Okapi BM25 scoring shows the relevance between query topic and the tweet.

TwitStan. TwitStan is a feature set used in a state-of-art stance classifier for tweets [8]. We combine the BM25 as the relevance feature.

WikiClaim. WikiClaim is a claim-discovery feature list from Roitman et al. [14]. Considering tweets do not have title or headers, we only use the content features. We combine the BM25 as the relevance feature.

TwitArgument. Since claim and evidence are all argumentive components, we also use TwitArgument which is a feature set used by argument identification tasks in Twitter [17]. We combine the BM25 as the relevance feature.

4.4 Results

[4] `https://www.elastic.co/products/elasticsearch`

[5] `www.procon.org`

[6] The overall inter-annotator agreement was calculated by averaging the agreements on all tweets in the dataset. For each tweet, the inter-annotator agreement was calculated as the number of annotators who agree over the majority label divided by the total number of annotators for that tweet.

[7] `http://www.cs.cornell.edu/people/tj/svm\_light/svm\_rank.html`

Experiment I: Baselines Table 3 gives the performance of the baselines. Due to the particularity of corpus, $LTR_{WikiClaim}$ which is effective on Wikipedia corpus do not perform well. The results also show that $LTR_{TwitArgument}$ is much worse than $LTR_{TwitStan}$. Because argument mining in Twitter tends to find different types of evidence, which is usually described objectively and it is difficult to see the stance of tweeter. However, the claim needs the tweeter to clearly express his stance. So our following experiment is on the basis of $LTR_{TwitStan}$.

id	Baselines	MAP	P@5	P@10
1	$BM25$	0.299	0.253	0.260
2	$LTR_{TwitStan}$	**0.500**$^{\blacktriangle}$	**0.513**$^{\blacktriangle}$	**0.436**$^{\blacktriangle}$
3	$LTR_{WikiClaim}$	0.291	0.280	0.283
4	$LTR_{TwitArgument}$	0.328$^{\triangle}$	0.313	0.336$^{\triangle}$

Table 3: Results for baselines. A significant improvement over the $BM25$ with $^{\triangle}$ and $^{\blacktriangle}$ (for p < 0.05 and p < 0.01).

Experiment II: Topic-independent Features The first column of Table 4 presents the effect of using Twitter structural features and topic-independent claim-related features. Each feature is combined with the $LTR_{TwitStan}$ and evaluated separately. Among these Twitter features, **re-tweet** ("RT @"), **reply**, **structure** (re-tweet+URLs+reply) intuitively perform better than others, which serve as useful proofs to conceive that some Twitter specific features really have correlation with claims. The improvement of ranking result using **re-tweet** feature is very possible because of the high forward frequency of valuable claim. As for the **reply**, it is probably because the argumentation always occurs during the discuss or quarrel. Besides, some features' combination may greatly improve the performance. For example, News in Twitter presents a specific structure as it contains both **re-tweet** and **URLs**, and it rarely contains a claim. For comparison, we use a controversy lexicon (CL) that has been proved useful for document claim-oriented retrieval [14]. However, the 7th case in Table 4 shows that CL is not very effective in Twitter. This may be because the text of tweets is different from documents.

id	Twitter Features	MAP	P@5	P@10	id	Claim-oriented Lexicons	MAP	P@5	P@10
1	$LTR_{TwitStan}$	0.500	0.513	0.436	1	$LTR_{TwitStan}$	0.500	0.513	0.436
2	$+re-tweet$	0.557$^{\blacktriangle}$	0.513	0.436	9	***+[TDCOLex]***	0.542$^{\triangle}$	0.520$^{\blacktriangle}$	0.443$^{\triangle}$
3	$+URLs$	0.530$^{\triangle}$	0.526$^{\triangle}$	0.446$^{\triangle}$	10	$LTR_{TI} + [TD]$	**0.585**$^{\blacktriangle}$	**0.533**$^{\blacktriangle}$	**0.480**$^{\blacktriangle}$
4	$+reply$	0.536$^{\blacktriangle}$	0.531$^{\blacktriangle}$	0.473$^{\blacktriangle}$					
5	$+structure$	0.550$^{\blacktriangle}$	**0.540**$^{\blacktriangle}$	**0.480**$^{\blacktriangle}$					
6	$+$***TICRLex***	0.533$^{\blacktriangle}$	0.533$^{\blacktriangle}$	0.450$^{\blacktriangle}$					
7	$+CL$	0.514	0.513	0.436					
8	LTR_{TI}	**0.558**$^{\blacktriangle}$	0.532$^{\blacktriangle}$	0.450$^{\blacktriangle}$					

Table 4: Experiment results (structure:re-tweet+URLs+reply, TI: structure + ***TICRLex***, TD: ***TDCOLex***). A significant improvement over the $LTR_{TwitStan}$ with $^{\triangle}$ and $^{\blacktriangle}$ (for p < 0.05 and p < 0.01).

Experiment III: Topic-dependent Lexicon Table 5 gives claim-related terms in the ***TICRLex*** and the claim-oriented terms in ***TDCOLex*** of topic "abortion". Apparently, the terms in ***TICRLex*** are some modal verbs, linking verbs, conjunction, negative words and punctuation which often do not have an exact meaning but are used to form a sentence pattern. However, words in ***TDCOLex*** are tend to be content words. For example, when it comes to **Abortion**, "rights", "murder","control" are included. Part of the reason can be that abortion supporters often think that abortion is part of women rights, while *"abortion is murder"*, *"abortion is not birth control"* are claims widely accepted by opponents. The 8th case in Table 4 shows that topic-dependent lexicons provide further boost to a model on the basis of topic-independent features. It shows that our lexicon does capture important topic-dependent claim-oriented information.

Finally, both effective topic-independent and topic-dependent elements including BM25, features in $TwitStan$, Re-tweet, Reply, Urls, ***TICRLex***(best), ***TDCOLex***(best) have been added to build our best model $LTR_{TI} + [TD]$ which improved the MAP by 95.7% compared with solely BM25, and 17% compared with $LTR_{TwitStan}$.

TICRLex	:, is, will, a, ,, ..., if, were, more, and, in, are, who, even, be, have, ?, they, ;, would, you, this, but, all, on, we, no, want, than, that, !, because, those, thus, was
TDCOLex	murder, cheerleader, supported, failed, dangerous, excuses, LGBTQ, stop, healthyLiving, rights, control, catholic, proabortion

Table 5: Comparison of the claim terms in ***TICRLex*** and ***TDCOLex*** of topic "abortion".

5 Conclusion and Future Work

We define a novel claim-oriented tweet retrieval task which will be certainly helpful in the development of public opinion research. We utilize the Twitter structural information to deal with the chaotic Twitter problem, and leverage claim-oriented lexicons to solve the vague claim problem. The topic-dependent claim-oriented lexicon can be generated using a large number of unlabeled topic-related tweets. Hence, our model can be easily adapted to new emerging topics in Twitter. We construct a real-world dataset. The best performance of our model improves the MAP by 95.7% compared with $BM25$ baseline, and 17% compared with $LTR_{TwitStan}$ baseline.

The main future work is threefold: first, we plan to use our automatic method to get an extended corpus and leverage deep learning techniques to learn more claim-oriented features. Second, we will diversify the searched claims and detect the relevant evidence of the known claim to generate a complete argumentation structure in Twitter. Third, we will study how to assess the quality of a claim.

Acknowledgments. We appreciate the comments from anonymous reviewers. This work is supported by National Key Research and Development Program of China (Grant No.2017YFB1402400) and National Natural Science Foundation of China (No.61602490).

References

1. A. Addawood and M. Bashir. "what is your evidence?" a study of controversial topics on social media. In *The Workshop on Argument Mining*, pages 1–11, 2016.
2. S. Eger, J. Daxenberger, and I. Gurevych. Neural end-to-end learning for computational argumentation mining. In *ACL*, 2017.
3. A. Freeley and D. Steinberg. *Argumentation and Debate*. 2008.
4. I. Habernal and I. Gurevych. *Argumentation mining in user-generated web discourse*. MIT Press, 2016.
5. R. Levy, Y. Bilu, D. Hershcovich, E. Aharoni, and N. Slonim. Context dependent claim detection. In *COLING*, pages 1489–1500, 2014.
6. M. Lippi and P. Torroni. Context-independent claim detection for argument mining. In *International Conference on Artificial Intelligence*, pages 185–191, 2015.
7. Z. Luo, M. Osborne, S. Petrovi, and T. Wang. Improving twitter retrieval by exploiting structural information. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
8. Saif M., Parinaz S., and Svetlana K. Stance and sentiment in tweets. *ACM Trans. Internet Techn.*, 17(3):26:1–26:23, 2017. doi: 10.1145/3003433.
9. W. Ma, W. Chao, Z. Luo, and Jiang X. Crst: A claim retrieval system in twitter. In *COLING*, 2018.
10. D. Mihai, C. Elena, and V. Serena. Argument mining on twitter: Arguments, facts and sources. In *EMNLP*, pages 2307–2312, 2017.
11. R. Palau and M. Moens. Argumentation mining: the detection, classification and structure of arguments in text. In *International Conference on Artificial Intelligence and Law*, pages 98–107, 2009.
12. J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
13. S. Robertson, S. Walker, M. Hancock-Beaulieu, A. Gull, and M. Lau. Okapi at trec. In *Text Retrieval Conference*, pages 21–30, 1992.
14. Haggai Roitman, Shay Hummel, Ella Rabinovich, Benjamin Sznajder, Noam Slonim, and Ehud Aharoni. On the retrieval of wikipedia articles containing claims on controversial topics. In *International Conference Companion on World Wide Web*, pages 991–996, 2016.
15. Christian S. and Iryna G. Parsing argumentation structures in persuasive essays. *CoRR*, abs/1604.07370, 2016.
16. Jan S. Social media argumentation mining: The quest for deliberateness in raucousness. *CoRR*, abs/1701.00168, 2017.
17. G. Theodosis, L. Christos, P. Georgios, and Vangelis K. Argument extraction from news, blogs, and social media. In *International Journal on Artificial Intelligence Tools*, pages 287–299, 2015.
18. S. Toulmin. The uses of argument. *Ethics*, 10(1):251–252, 1958.
19. Z. Wei, Y. Liu, and Y. Li. Is this post persuasive? ranking argumentative comments in online forum. In *Meeting of the Association for Computational Linguistics*, pages 195–200, 2016.