



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Stance detection on social media: State of the art and trends

Citation for published version:

ALDayel, A & Magdy, W 2021, 'Stance detection on social media: State of the art and trends', *Information Processing and Management*, vol. 58, no. 4, 102597. <https://doi.org/10.1016/j.ipm.2021.102597>

Digital Object Identifier (DOI):

<https://doi.org/10.1016/j.ipm.2021.102597>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Information Processing and Management

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Stance Detection on Social Media: State of the Art and Trends

ABEER ALDAYEL^{*}, School of Informatics. The University of Edinburgh, Edinburgh, UK

WALID MAGDY[†], School of Informatics. The University of Edinburgh, Edinburgh, UK

Stance detection on social media is an emerging opinion mining paradigm for various social and political applications in which sentiment analysis may be sub-optimal. There has been a growing research interest for developing effective methods for stance detection methods varying among multiple communities including natural language processing, web science, and social computing. This paper surveys the work on stance detection within those communities and situates its usage within current opinion mining techniques in social media. It presents an exhaustive review of stance detection techniques on social media, including the task definition, different types of targets in stance detection, features set used, and various machine learning approaches applied. The survey reports state-of-the-art results on the existing benchmark datasets on stance detection, and discusses the most effective approaches. In addition, this study explores the emerging trends and different applications of stance detection on social media. The study concludes by discussing the gaps in the current existing research and highlights the possible future directions for stance detection on social media.

Note

This is an early version of the paper. For the final version, please find it published at Elsevier IP&M in the following link: <https://doi.org/10.1016/j.ipm.2021.102597> .

Additional Key Words and Phrases: Stance, Stance detection

1 INTRODUCTION

Nowadays, social media platforms constitute a major component of an individual's social interaction. These platforms are considered robust information dissemination tools to express opinions and share views. People rely on these tools as the main source of news to connect with the world and get instant updates [1]. They are very beneficial because they allow individuals to explore various aspects of emerging topics, express their own points of view, get instant feedback, and explore the views held by the public. The huge dependency of users on these platforms as their main source of communication allows researchers to study different aspects of online human behavior, including the public stance toward various social and political aspects.

Stance is defined as the expression of the speaker's standpoint and judgement toward a given proposition [2]. Stance detection plays a major role in analytical studies measuring public opinion on social media, particularly on political and social issues. The nature of these issues is usually controversial, wherein people express opposing opinions toward differentiable points. Social issues such as abortion, climate change, and feminism have been heavily used as target topics for stance detection on social media [3]. Similarly, political topics, such as referendums and elections, have always been hot topics that have been used in stance detection to study public opinion [4]. Generally, the stance detection process is also known as perspective [5, 6] and viewpoint [7, 8] detection, in which perspectives are identified by expressing stances toward an object of a controversial topic [6].

The stance has been used in various research as a means to link linguistic forms and social identities that have the capability to better understand the background of people with polarized stances [9]. Consequently, the early work on stance detection emerged in analyzing political debates on online forums [10–13].

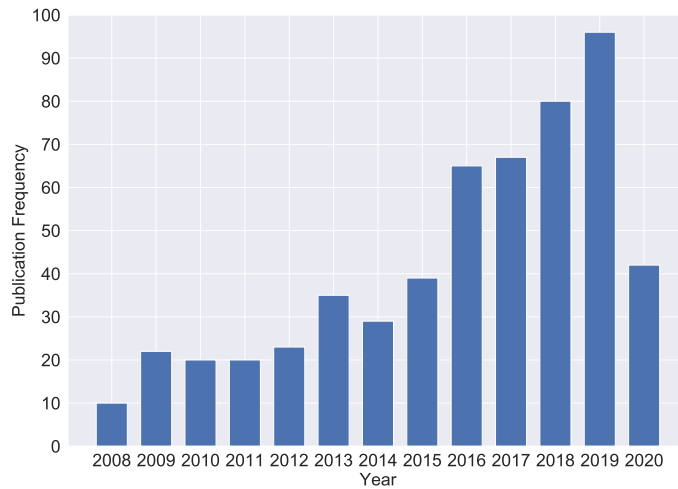


Fig. 1. Publications per year on stance detection as searched on Web of Science. The following keywords were used for search: “stance detection”, “stance prediction” and “stance classification”.

Recently, a growing body of work has used the concept of stance detection on social media as a new frontier to analyze various social and political issues across social media platforms. Figure 1 shows the number of publications from 2008 to 2020 on stance detection. As shown, there is a noticeable steady increase in the number of publications, with a sudden increase from 2016, which we relate to the release of the SemEval 2016 stance detection task [3], which provided the first benchmark dataset for stance detection on social media. Earlier work on stance detection focused on analysing of debates on online forums, which is distinguished from the more recent work that focused more on social media platforms, especially Twitter, since the former has a clear single context in comparison to social media platforms. In the online forums, the users debate in the form of a thread discussion where there is a flow of information related to the topic. In contrast, social media lacks such representations where the users participate solely with a topic and rarely with a context (i.e., thread).

As can be noticed from Figure 1, one of earliest initiatives to promote stance detection on social media is the "SemEval 2016 task 6" shared-task, which introduced a stance-specific benchmark dataset to help in evaluating stance detection on Twitter [3]. With a major focus on identifying stances toward social issues, SemEval-2016 has formulated the stance detection task around a few controversial issues, such as atheism, feminism, and opinions regarding some politicians. In addition, a new wave of stance detection applications has been triggered to handle issues that have infected social media lately, such as fake news and rumors [14–16], where the stance toward claims in a piece of news is utilized as a key feature for the detection of the credibility of that news.

The majority of work on stance detection has targeted the detection of the stance toward a given subject expressed in a given text. However, some works have studied the detection of the stances of users toward subjects without explicitly stating them, which is usually referred to as stance prediction. Thus, the work on stance detection can be categorized

into two main types, namely, detecting expressed views versus predicting unexpressed views. In the first type, the objective is to classify a user’s post and infer the current stance to be in favor of or against a given subject [3].

In the latter, the prediction is carried out to infer a user’s viewpoint on a given topic that the user has not discussed explicitly or toward an event that has not occurred yet. This type of stance detection has proven its effectiveness in predicting the future attitudes in the aftermath of an event [17, 18].

The work on stance detection can also be categorized based on the topic of the target of analysis, that is, it can be one specific target, multiple related targets, or a claim in a news article. Most existing designs of stance detection classifiers work to identify the user’s stance toward a specific topic. Sometimes, the classifier is built to detect the stance toward multiple related targets. This is a situation when the stance is detected toward two related entities that are typically opposed to against each other, such as the detection of the stance toward Clinton and Trump simultaneously, since if the stance is in favor one target, it will simply be against the other [19]. When the target is a claim in a news statement, the main objective is to identify if a given claim is supported by other posts. In this type of detection, the analyses are between two posts, such as in RumourEval [14, 15] or between news headers and other bodies of articles; these are mainly used as an initial step for fake news detection [20]

To our best knowledge, this is the first work to cover the current research directions on stance detection on social media and identify a framework for future research. Recently, there have been a few surveys related to stance detection. The first survey study [21] emphasized more on surveying the opinion mining methods in general, with a focus on stances toward products. Another study by [22] provides a comparative analysis of stance detection models with a sole focus on the textual features only. A more recent survey study [23] focused more on surveying the research work that modeled stance detection as a text entailment task using natural language processing (NLP) methods. While, the previous studies surveys the stance concerning one domain, in contract our main objective of this article is to survey the stance detection on social media platforms related to multiple research domains, including NLP, computational social science, and Web science. More-specifically, this study provides the following contributions:

- This study maps out the terrain of existing research on stance detection and synthesizes how this relates to existing theoretical orientations. It provides a through comparison of stance and sentiment for soci-political opinion mining and show the orthogonal relation of sentiment polarity and stance.
- This survey study provides a broader overview of stance detection methods, covering work that has been published in multiple research domains, including NLP, computational social science, and Web science.
- This study surveys the modeling of stances using text, networks, and contextual features that have been overlooked by previous surveys.
- Since this survey navigates the stance modeling across multiple domains, this study provides a comprehensive study investigates the stance detection applications on social media and provide a comprehensive discussion on the future trends and future directions of this field of study.

The rest of the paper is organized into several sections. Section 2 provides a theoretical definition of stance. Section 3 contrasts the differences between stance and sentiment. Sections 4 and 5 summarize the literature on stance prediction and categorize this work according to the types of detected stances (i.e., expressed versus unexpressed) and the target of the stance in the text. Sections 6 and 7 explore the different stance modeling and machine learning approaches for stance classification. Section 8 lists several applications of stance detection, such as social media analyses and fake-news detection. Stance detection resources are listed in Section 9. Finally, the current research trends on stance classification are discussed in Section 10, highlighting the gaps and suggesting the potential future work required in this area.

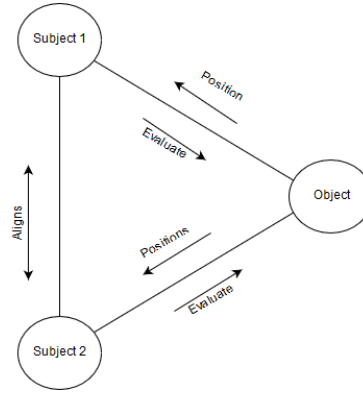


Fig. 2. The stance triangle, adapted from [Du Bois 20]

2 DEFINITIONS OF STANCE AND STANCE DETECTION

Biber and Finegan 1988, define a stance as the expression of the speaker's attitude, standpoint, and judgement toward a proposition [2].

[24] argues that stance-taking (i.e., a person taking a polarized stance toward a given subject) is a subjective and inter-subjective phenomenon in which the stance-taking process is affected by personal opinions and non-personal factors such as cultural norms. Stance taking is a complex process related to different personal, cultural, and social aspects. For instance, the political stance-taking process depends on experiential behavior, as stated by [25].

The process of detecting the stance of a given person on social media is still in its infancy, as it is still unclear what roles of language and social interaction play in inferring the user's stance.

Stance detection has a strong history in sociolinguistics, wherein the main concern is to study the writer's viewpoint through their text. Stance detection aims to infer the embedded viewpoint from the writer's text by linking the stance to three factors, namely, linguistic acts, social interactions, and individual identity. Using the linguistic features in stance detection is usually associated with attributes such as adjectives, adverbs, and lexical items [26].

It has been argued that stance taking usually depends on experiential behavior, which is based on previous knowledge about the object of evaluation [25]. This strengthens stance detection as a major component in various analytical studies.

Stance detection in social media is concerned with individuals' views toward objects of evaluation through various aspects related to the users' post and personality traits.

As stated in Due Bois's stance triangle shown in Fig 2, the process of taking a stance is based on three factors. These are 1) evaluating objects, 2) positioning subject (the self), and 3) aligning with other subjects (i.e., other social actors). For instance, "I am with the new legalization on climate Change" has a subject, "self," indicated by proposition "I" and further, the "with" indicates the favored position toward the object "Climate Change". In social media, identifying the stance subject (self) is mostly straightforward as each post is linked to the user.

Furthermore, from a sociolinguistic perspective [26], it has been argued that there is no completely neutral stance, as people tend to position themselves through their texts to be in favor of or against the object of evaluation. This creates a further complexity in identifying the stance of the social actors, since a stance is not usually transparent in the text, but sometimes must be inferred implicitly from a combination of interaction and historical context.

3 STANCE VS. SENTIMENT

In social media analysis studies, there is a noticeable misconception between sentiment and stance. Some of the studies use sentiment analyzers as the main method to measure the support of a given target, which is sub-optimal [27]. In this section, we highlight the main theoretical differences between stance and sentiment, and illustrate these differences with some examples. In addition, we apply a quantitative study on data labelled for both sentiment and stance to further demonstrate these differences.

3.1 Definition of the Tasks

Sentiment analysis is a well-known task in NLP to determine the polarity of emotion in a piece of text. Generally, this task could be defined as estimating the users' emotional polarity as positive, negative, or neutral [28, 29]. This can be seen in related work on SemEval sentiment tasks SemEval-/(2015, 2016, 2017 and 2020), where the aim of the sentiment analysis task is to determine the polarity toward an aspect of a product, such as a cell phone [30–34]. Unlike sentiment analysis, stance detection mainly focuses on identifying a person's standpoint or view toward an object of evaluation, either to be in favor of (supporting) or against (opposing) the topic. This is usually inferred by a mixture of signals besides the linguistic cues, such as the user's feelings, personality traits, and cultural aspects [2]. Sentiment analysis is mostly approached as a linguistic agnostic task that focuses on leveraging a given text's linguistic properties to identify the polarity of the text [35], while stance detection can take an additional paradigm by leveraging non-textual features such as networks and contextual features to infer the user's stance [36–38].

In general, sentiment analysis is concerned with detecting the polarity of the text, which can be inferred without the need to have a given target of interest; for example "*I am happy*". Thus, the sentiment analysis model can be represented as shown in equation 1, where T is a piece of text, and the outcome is usually one of three labels $\{positive, negative, neutral\}$. However, the main sentiment outcome can take different forms, such as binary polarity, multi-level polarity, and regression.

$$Sentiment(T) = \{Positive, Negative, Neutral\} \quad (1)$$

Another kind of sentiment analysis concerns inferring the polarity of a text by using a predefined set of targets. This kind of sentiment analysis is usually referred to as target-based sentiment analysis [33, 39–41]. In this kind of sentiment analysis, the classification task can be defined as follows:

$$Sentiment(T|G) = \{Positive, Negative, Neutral\} \quad (2)$$

In definition 2, the input G represents the target or the entity of evaluation. Still, in this kind of sentiment analysis, the dominant factor to gauge the polarity is the raw text.

For stance detection, a clear target G must be defined in advance, to assess the overall attitude toward this target.

The basic form of stance detection on social media can be formulated by using the attributes of the social actor. Thus, in this form of stance detection, the main two inputs are 1) text T or user U , and 2) given target G , as illustrated in equation 3.

$$Stance(T, U|G) = \{Favor, Against, None\} \quad (3)$$

There are three possible variations of this definition, which may include either the text, the social actor, or both as input for stance detection. The dependent factor in the stance detection task is the *target* of analysis. One of the

Table 1. sample of tweets illustrating the sentiment polarity of the expressed stance

#	Tweet	Target	Sentiment	Stance
1	It is so much fun having younger friends who are expecting babies. #beentheredonethat #chooselife .	Legalisation of Abortion	+	-
2	Life is sacred on all levels. Abortion does not compute with my philosophy. (Red on #OITNB)	Legalization of Abortion	0	-
3	The biggest terror threat in the World is climate change #drought #floods	Climate Change is the real concern	-	+
4	I am sad that Hillary lost this presidential race	Hillary Clinton	-	+

common practices of stance detection in social media is to infer the stance based on the raw text only, which maps the stance detection problem to a form of textual entailment task [3, 10]. According to a more relaxed definition, a text T entails a stance to a target G , ($T \rightarrow \text{stance to } G$), if the stance to a target can be inferred from the given text.

In social media analysis, this formulation has been further extended to include the social actor, which can be represented by a set of online behavioral features to infer the stance. Instead of the dependency on the raw text, the other representation of stance detection includes the social actor U as a salient factor to infer the stance. Reflecting on the origin of stance taking using the stance triangle as defined by Due Bois, as shown in Fig 2 [24], the social actor (subject) is the main element of this process. The structure of social media extends the possibility to represent the social actor by using a variety of network features such as profile information (including age and description) [18, 37].

To define the interplay between sentiment and stance, several studies have demonstrated that it is insufficient to use sentiment as the only dependent factor to interpret a user’s stance [11, 27, 42–44]. This is due to the complexity of interpreting stances from a given text, as they are not always directly aligned with the polarity of a given post. Table 1 shows examples of tweets that illustrate the orthogonal relationship between sentiment polarity and stance. For instance, in example 1, the tweet is against the legalization of abortion with a positive sentiment. In this example, the legalization of abortion is the target for which the opinion is expressed. While the text has a positive sentiment, it shows a negative or opposing stance. Furthermore, a text may not indicate any sentiment and still pose a clear stance toward a target. This is indicated in example 2, where the tweet has a neutral sentiment with an "against" stance toward the legalization of abortion. On the other hand, example 3 shows the opposite case, where a negative sentiment can be detected without a supporting stance toward the target. Finally, in example 4, applying target-dependent sentiment analysis toward Hillary Clinton leads to a clearly negative sentiment, while measuring the stance toward the same target is clearly supportive. These examples demonstrate that both phenomena (sentiment and stance) are orthogonal.

In the following section, a quantitative analysis on the alignment between stance and sentiment toward a set of targets is applied to quantify the sentiment polarity of the expressed stance.

3.2 Stance vs Sentiment Quantitatively

In order to show the overall agreement between sentiment and stance, we use a SemEval stance detection dataset, wherein the tweets are annotated with sentiment and stance labels. The dataset contains a set of 4,063 tweets covering five topics. These topics are 'Atheism' (A), 'Climate change is a real concern' (CC), 'Feminist movement' (FM), 'Hillary Clinton' (HC), and 'Legalization of abortion' (LA). Fig 3 illustrates the sentiment distribution on stances for the whole collection (favor, against, and none). The graph shows that the negative sentiment constitutes the major polarity over

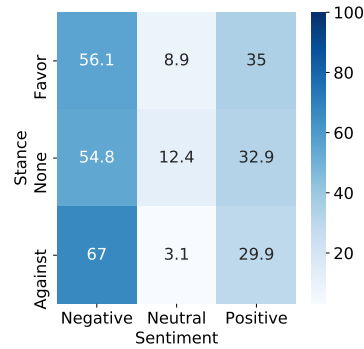


Fig. 3. Sentiment distribution over stance.

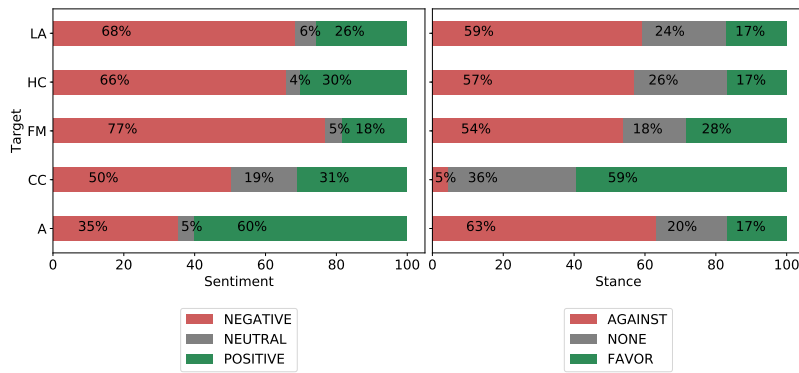


Fig. 4. Sentiment and stance distribution

the three stance categories, as the negative sentiment represents over 56% of the different stances. This reveals the tendency of using negative sentiments to express a viewpoint on a controversial topic.

The overall agreement between stance and sentiment is relatively minuscule. About 35% of the "favor" stance tweets have a positive sentiment. It can be observed that tweets with neutral sentiments fail to infer neutral stances, with only 12.4% of the tweets showing neither a positive nor negative polarity. Instead, 54.8% of the tweets have a negative polarity, while 32.9% of them carry a positive sentiment for an "against" stance. Thus, it is evident that sentiment does not simply represent stance.

A large amount of work uses sentiment as a proxy to interpret stances. Based on the assumption that the polarity of a text reveals the stance of the text's author, it has been shown that a stance could be inferred independently of the author's emotional state [43–46]. A recent study by [45] used sentiment to predict the stance through a multi-task learning model. In their study, the researchers used a sentiment lexicon along with to build a stance lexicon and incorporated it for the attention layer in the model. Another work by [47] used sentiment as an auxiliary task to predict the stance using a SemEval stance dataset.

It is not necessarily accurate to use sentiment polarity to analyze social media as a means to interpret the general stance toward a topic. Fig 4 shows the clear differences in detecting the users' attitudes toward five SemEval topics.

While 60% of the tweets show a positive sentiment toward atheism, the majority (63%) is opposed to this topic. Similarly, in analyzing the general attitude toward climate change, 50% of the tweets have negative sentiments, while 59% of the tweets actually support the idea that climate change is a real concern. Following the same trend, pure sentiment value fails to detect the accurate stance toward Hillary Clinton, the feminist movement, and the legalization of abortion, as these distributions show about 9% of a difference between the accurate "against" position and negative sentiment value for Hillary Clinton and the legalization of Abortion. In the feminist movement, the difference is laudable with a 23% misinterpretation of the real stance.

Further, we analyzed whether sentiment and stance are independent of each other. We used Cramer's test [48] to gauge the strength of relationship between sentiment and stance. The result from Cramer's test indicate that the variance (V) value ranges between 0 and 1, with 1 as an indication of a high association between the nominal variables [49]. By applying Cramer's test in the SemEval stance dataset, the resultant V value = 0.12, which is a strong indication of the independence of sentiment and stance from each other.

4 EXPRESSED VS UNEXPRESSED STANCE DETECTION

We can divide the literature on stance detection using social media data according to the status of the stance to be modeled, either as expressed in text or unexpressed. The first type of work is mainly referred to as stance classification, and it represents the majority of work in the existing literature. The other type of work is concerned with detecting the stance prior to an event, which is called stance prediction.

4.1 Stance Detection for Expressed Views

Many social media users express their support or opposition toward various topics online, which makes these platforms valuable sources to analyze public opinions toward these topics. This has motivated a rich body of research to focus on inferring the users' stances from their posts expressing their position.

The underlying method for this type is based on using a predefined set of keywords concerning the target of analysis. Most of the available stance datasets are concerned with studying the users' stance toward a post-event. For instance, a SemEval stance dataset [3] created three sets of hashtags for each event or entity to collect tweets concerning three stances. For example, for the topic "Hilary Clinton," the dataset used favor hashtags, such as #Hillary4President, against hashtags, such as #HillNo, and ambiguous hashtags. Ambiguous hashtags contain the target without a direct indication of the stance in the hashtag (e.g., #Hillary2016). In a study done by [50] to predict public opinion and what went viral during the 2016 US elections, tweets that were filtered out by using a set of 38 keywords related to the US elections for streaming relevant tweets to this event were used. Another study employed tweets generated by users to infer the opinion toward the independence of Catalonia and collected tweets by using two keywords related to the hashtags #Independencia and #27S [51]. Similarly, to study mass shooting events, a study by [52] made use of Gun Violence Archive and collected list of events between 2015 and 2018 to define a set of keywords that could be used to retrieve the tweets. Further, the study used the Twitter Firehose ¹ to collect tweets for these keywords.

After collecting the event-related data, it was annotated using predefined guidelines for labelling the data with the stance toward the given target; this method is applied for labelling the SemEval dataset [3, 44].

Most of the work on detecting the stance in the expressed views dataset applies NLP methods that model the task as text entailment to detect whether a piece of text is in favor of or against the target [53–60]. For instance, the work of

¹<https://developer.twitter.com/en/docs/tweets/compliance/api-reference/compliance-firehose>

[53] used a SemEval stance dataset which consisted of tweets and target entities to build a textual entailment model that predicted a target-specific stance. However, some studies showed that employing additional signals from the user’s network can further improve the stance detection performance, as will be discussed later in section 6.2.

4.2 Predicting the Unexpressed Views

Stance prediction aims to infer the stances of social media users with no explicit or implicit expression of these stances online. It is also used to predict a user’s view on events that have not occurred yet.

In stance predication, most studies tend to predict the users’ stances on two levels, that is, micro and macro. At the micro level, stance prediction means the estimation of an individual user’s standpoint toward a target or event in advance (pre-event), indicating the likelihood that the user will be in favour of or against the target event. This methodology is similar to recommending new items based on a user’s previous history of purchases. At the macro level, the public opinion toward an event is inferred, and the research tend to address this level of prediction as an aggregation of micro predictions[61]. Similarly, [62] designed a micro-level stance detection model based on users’ previous posts along with a logistic regression to predict stances for new users that were not included in the dataset. Their model produced word distributions for the against/support stances on the topic by using rules derived from user’s interactions. Another study [63] used matrix factorization to predict micro-level stances based on the voting patterns of the users and the topic model. The study by [64] predicted ideology using heterogeneous links between the users.

For instance, [37] investigated the best social media features to predict user attitudes toward new events[37]. This work differed from the SemEval-2016 task 6 (a target-specific task) in that the stance of the users was predicted based on each user’s historical collection of tweets, instead of a target-relevant single tweet. Rather than focusing on the tweet content (textual features), [37] used the similarity between users and inferred latent group beliefs as features in their model, and further employed a bipartite graph along with graph reinforcement. [36] followed the same concept as [37] to define network features, and used a joint matrix in which users and source contents were represented in the same latent space to balance message distribution across social media. A different method was proposed by [65] to predict an individual’s unexpressed stance by implementing a collaborative filtering approach and constructing a user ideology matrix that represented a user’s ideological stance for an event.

In macro-level stance prediction, few studies have addressed stance prediction. Most of these studies were customized to analyze specific cases, such as Islamophobia [17, 18]. Besides analyzing events, these studies carried out a further complementary study for the prediction of user’s stances in the aftermath of an event, based on previous tweets and user interaction. In debate forums,[61] have proposed a micro-level stance prediction model based on user behaviour toward new events in which they have not participated. In this study, this kind of user was referred to as a cold-start user, which is a well-known term commonly used in recommendation systems. In addition, they introduced a macro-level stance prediction model as an aggregation of each user’s stance (i.e., at the micro-level).

5 STANCE DETECTION ACCORDING TO TARGET

As discussed earlier, the stance detection task needs the presence of a defined target to detect the stance toward it. In the existing literature, stance detection can be categorized according to the type of the target of evaluation into three groups, namely, single defined targets, multi-related-targets, and claim-based targets. In the following sections, a further explanation of each type is provided.

5.1 Target-Specific Stance Detection

The basic form of stance detection on social media is the *target-specific* stance detection. Most of the previous studies focused on inferring the stance for a set of predefined targets [3, 57–60, 66]. In this type of stance detection, the text (T) or the user (U) is the main input to predict the stance toward a specific predefined single target (G).

$$Stance(T|U, G) = \{Favor, Against, None\} \quad (4)$$

The above equation means that a separate stance classification model must be built for each target (G) separately, unlike that in the case of sentiment, wherein a general model can be trained for target-independent sentiment analysis. This is the basic practice, even for benchmark datasets such as SemEval 2016, which covers multiple topics. Most of the published work on this dataset has trained a separate model for each topic (target) separately [3, 38, 56]. However, there have been recent trials to apply the transfer learning model between different targets, as will be discussed in more detail in section 7.2.

5.2 Multi-related-target stance detection

In multi-target stance detection, the goal is to jointly learn the social media users' orientation toward two or more targets for a single topic [19]. The main assumption behind this kind of stance detection is that when a person gives their stance for one target, it provides information on their stance toward other related targets.

$$Stance(T|U, G_n) = \{(Favour_{G_1}, Against_{G_{n+1}}), (Favourof_{G_{n+1}}, Against_{G_1})\} \quad (5)$$

For instance, a tweet could express a stance toward multiple US presidential candidates at the same time, so when users express their in-favor stance toward Hillary Clinton, it implies an against stance toward Trump [50, 67]. In [19] work, the first multi-target stance detection data-set is presented, containing 4,455 tweets related to the 2016 US elections. In order to detect the subjectivity toward two targets, [19] used an attention-based bidirectional recurrent neural network (RNN) to jointly learn the stances toward two related targets. The notion of multi-target stance detection has been usually used to analyze the relation between two political candidates, such as Clinton and Trump, by using domain knowledge about these targets to improve the classification performance [67]. Following the same approach, the study by [50] constructed a dataset with 3,450 tweets annotated with stance labels for the two US 2016 election candidates (Trump and Clinton) at the same time. Furthermore, the work of [68] proposed a memory-based algorithm focusing on jointly modeling multiple targets at the same time. Their memory-based model provides the current state-of-the-art result so far on the multi-target benchmark dataset.

5.3 Claim-based stance detection

In claim-based or open-domain stance detection, the target of the analysis is not an explicit entity, as is the case in the ones discussed earlier. This, however, is a claim in a piece of news. The first stage to detect a stance is to identify the main target claim from the sequence of conversation or given text. The main input to the claim-based stance detection model is the claim (C), which could be the rumor's post or based on an article headline. In the fake news task, the claim tends to be the article headline and the text is the article body. On the other hand, for the rumor's veracity task, the main input to be evaluated is the rumor's post, while the text is the replies to the rumors. The prediction label sets tend

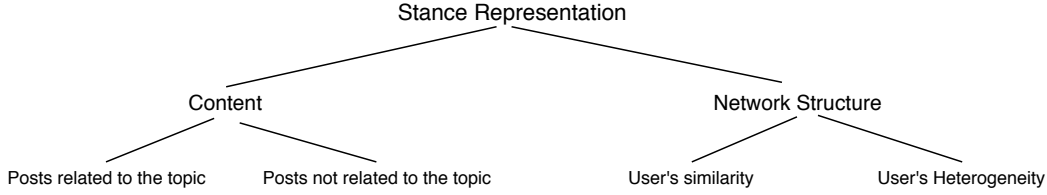


Fig. 5. Stance representation in social media

to take the form of confirming the claim or denying it.

$$Stance(T, C) = \{Confirming, Denying, Observing\} \quad (6)$$

Claim-based stance detection is considered a suitable method to analyze the veracity of the news. For that reason, claim-based stance detection has been heavily used for rumor resolution studies [14, 69–71]. In a study by [69], a supervised learning model was used, along with a new set of features known as "pragmatic features" which contained the named entity, event, sentiment, and emoticons. Interestingly, [70] concluded that problem-specific feature engineering outperformed the other state-of-the-art systems in rumor identification tasks. Their model, which used a random forest, outperformed the advanced LSTM-based sequential models in SemEval-2017 task 8 proposed by [72]. Along same lines, the study of [70] used the same feature engineering approach proposed by [42], in which lexical and semantic features were used to help with identifying the stance of a Twitter user. More recently, for the conversation-based tasks, the study by [71] showed that LSTM can outperform other sequential classifiers and feature-based models.

In another study by [73], multi-task learning was used with a stance detection layer to classify the stance of a given tweet as supporting or denying a given claim.

6 STANCE MODELING ON SOCIAL MEDIA

Stance on social media has been modeled using various online signals, which have been used as features for training the stance detection models. These signals can be categorized into two main types, namely, 1) content signals, such as the text of the tweets, and 2) network signals, such as the users' connections and interactions on their social networks. Figure 5 summarizes the main features used in each category to model stances on social media. In the following section, we describe each of these two categories of signals and how they have been used for detecting stances. Finally, we present a comparison of their effectiveness for detecting stances on multiple datasets.

6.1 Content Features

This section discusses stance representation, focusing on the textual features derived from the users' content online. As illustrated in Figure 5, the content can be collected based on the topic of the analysis. In this kind of feature representation, data is collected based on the range of keywords reflecting the topic. Another type of representation is concerned with collecting content that has no direct relation to the topic. The main objective in this case is to model the stance based on a user's behavioral data, rather than topic-level stance detection.

Furthermore, the content features can be categorized into two general types, namely, the linguistic features and users' vocabulary. The first type of features are related to the text's linguistic features that help in inferring the stance. The other type concerns modeling a user's stance based on the user's choice of vocabulary.

6.1.1 Linguistics Features. The majority of work on stance detection has been focused on utilizing the linguistic elements that capture a social media user’s stance. This massive dependency on linguistic cues is due to definition of stance detection as a textual entailment task, wherein the task is to detect a stance in a given piece of text (e.g., tweet [3]). In the existing literature, the stance detection work that is concerned with using textual cues to detect stances includes textual features, sentiment polarity, and latent semantics.

For instance, using textual features such as the n-gram modeling of the text has been heavily investigated in the existing literature [44, 74, 75]. Using the n-gram modeling of the text shows promising results. In the SemEval-2016 stance detection task, the use of word and character n-gram modeling obtained the best f-score among the other participating systems [3]. Another textual feature that has been used to infer the stance is the sentiment polarity of the post [6, 42, 46]. In general, the use of sentiment as a feature was not sufficient in predicting the stance, as some studies have concluded [6, 44]. Another kind of feature is the latent semantics feature that aims to reduce the dimension of a given input, such as mapping the sentences according to a predefined set of topics (i.e., topic modeling) [42]. Topic modeling has been applied by several different studies [42]. For instance, [42] used weighted textual matrix factorization (WTMF) and frame-semantic parsing to model a given tweet that had been used as feature by [76] to map unigrams to the topic sphere. Another study used a simple bag-of-topics to model target words.

6.1.2 Users’ Vocabulary Choice. There are a considerable number of studies that represent stance detection based on the users’ vocabulary. The hypothesis behind using this kind of user modeling is that individuals with same stance tend to use the same vocabulary choices to express their points of view [77].

The focus of these studies is mainly to disentangle the topic from the viewpoint where the vocabulary is not only linked to the topic but also individual attitudes and characteristics [5]. For instance, people with an against stance to abortion tend to use terms such as “pro-life” to express their opposing stance. The work of [5] built a user stance detection model based on users’ vocabulary choices through a generative model and discriminative model using Naive Bayes and SVM classifiers, respectively. Another work, [62] proposed a word-generative model based on the users’ interaction to build a set of word representations for each stance toward a topic. In addition, the work of [78] followed the same direction, wherein the users’ interactions helped in shaping the set of vocabulary used to identify the stance. Another study by [7] used a hierarchical topic for opinion discovery, based on the author of the text vocabulary. Lately, the work on vocabulary choice has been linked with the user behavior on social media to further improve the stance prediction model. The work of [62] used a generative model based on the users’ keyword choices for each standpoint, along with their interactions as regularization to predict the stance. In addition [79], introduced user interactions and post embedding by using an embedding vector for the pro-stance and a different vector for the con-stance.

6.2 Network Features

Social media provides a special kind of social data due to the structure of its platforms, wherein users can be characterized and/or analyzed based on their social connections and interactions.

Many existing works used network features to gauge the similarity between the users.

The network features that have been used to learn the users’ representations on social media can be grouped under two categories, namely, users’ behavioral data [37, 66, 80] and users’ meta-data attributes [81]. The application of users’ behavioral data to identify the stance is motivated by the notion of homophily, based on the social phenomenon according to which “individuals associate with similar ones” [82]. When it comes to social media, the similarity between users is considered a core property that helps in inferring stances.

The interaction elements have been used to define the similarity between the users. One of the elements that has been extensively used to infer Twitter users' stances is the retweet [17, 83–85]. Another element that has been heavily investigated is the hashtag; this element has been used in literature to infer similarity between users to predict the stance [37, 86]. The work of [86] used the soft cosine similarity feature to gauge the similarity between the users that post under the same hashtag. The work of [37] used graph reinforcement to calculate the similarity between the users that post under the same hashtag.

In a recent study by [38], three types of network features to model stances on social media were defined. These network features are the 1) interaction network, 2) preferences network, and 3) connection network. The interaction network represents the users' direct interactions with other users, in the sense of retweets, mentions, and replies. This type of network provides the best performance score of the stance detection model in comparison with two other networks. The preference network is the network of others users that post or are mentioned in the tweets the users like. This network allows the detection of stances for users who may have limited posting or interaction behaviors online. Finally, the connection network includes the friends and followers of the users. The three types of networks provide the best performance in comparison with content features 2.

Another study by [4] introduced a multi-layer graph model to represent profiles from different platforms and extract communities. By doing so, the model allowed a stance to diffuse from the set of known profiles and predict profiles' stances by using pairwise similarities between the users' profiles.

The use of network features has been beneficial in detecting social media users' stances and future attitudes in the aftermath of an event (i.e., stance prediction). For instance, in a study done by [17], a user's similarity components was utilized as the features. In this work, the similarity was calculated based on the interaction elements of a given tweet. These interaction elements included mentions, retweets, and replies, as well as website links (URLs) and hashtags used by users in their tweets. Similarly, the work of [80] used the retweet and reply features to define the users' social networks.

Another line of study has used heterophily to measure the dissimilarity between the users to model the stance [8]. For instance, the study by [8] was based on the tendency of a user to reply to the opposed viewpoint. The study used a rebuttal variable to model the users' interactions and denote if the replies attacked the previous author's parent post. The value of the rebuttal depended on the degree of opposition between the viewpoint of the parent post and parent tweet.

6.3 Comparative Analysis of Stance Modeling

Table 2 shows a comparative analysis of the stance-detection performance using the two types of modeling, content versus network. In addition, it reports the results of studies that used a combination of both these types. It must be noted that using network features to model the stance outperforms the content modeling of the stance. Textual modeling has the lowest performance as this type of modeling depends on the textual entailment task, while the network features provide the highest performance in comparison with the former. This may explain the fact that using network features puts the bigger picture of modeling the users' attitudes using their interactions and similarities into consideration. Using network features overcomes the limitations posed by the textual entailment modeling of stance. As demonstrated by [38]'s this study used a SemEval dataset to model stances on social media using the model with the best reporting results on stance datasets to compare the performance of the stance, using content and network features. This study provides a thorough comparative analysis of the stance's overall performance textual and network features are used. The use of network features provides the best performance in comparison with previous studies, wherein textual

Table 2. A comparison of stance detection models using network , content and both as features.

Dataset	NW	Content	Both	ML Algorithm	Study
Before Paris attacks	85.0	82.0	84.0	SVM	[18]
Islam Dataset	84.0	76.0	-	SVM	[37]
Islands Dataset	79.0	71.0	-	SVM	
Gun control, abortion and Obama care	81.9	60.6	82.0	Non-negative matrix factorization	[36]
SemEval stance dataset	71.6	69.8	72.5	SVM	[38]
SemEval stance dataset	61.8	62.8	65.9	SVM	[87]
LIAC dataset	67.8	54.1	-	Generative Model	[62]

modeling along with transfer learning methods were used, as reported by [3]. Furthermore, a study by [36] showed that the network features outperform the textual modeling of stance when using non-negative matrix factorization for stance detection. The study by [37] confirmed the same conclusion, wherein the use of network features outperformed the textual representation using two datasets (Islam and Island datasets). As the study by [18] indicated, the network modeling of stance outperformed the use of textual modeling for the expressed and non-expressed stances on social media.

7 STANCE DETECTION ALGORITHMS

In this section, the main machine learning (ML) algorithms used for stance detection are discussed. According to the existing literature, the ML algorithms used for stance detection on social media can be divided into three main approaches, namely, 1) supervised learning, 2) weakly-supervised and transfer-learning, and 3) unsupervised stance detection. In the following sections, each of these approaches are discussed in greater detail.

7.1 Supervised Learning

Supervised learning is the basic and most common approach for most of the work on stance detection [60, 65, 88–91]. In this approach, a stance dataset is annotated using a predefined set of labels, usually two (pro/against) or three (pro/against/none) labels. For example, the SemEval-2016 stance dataset uses three labels, "In-Favor," "Against," and "None" [3] for a set of five different topics. Many studies have been published on this dataset using different supervised ML algorithms such as classical algorithms, for instance, Naive Bayes (NB), SVM, and decision trees, and deep learning algorithms, for example, RNNs, and LSTMs, to detect the stance on this labelled dataset.

For example, the work of [44] used an SVM model with linguistic and sentiment features to predict the stance. This study showed that the use of content features only (i.e., n-gram) provided an F1 score equal to 69.0%, which surpassed the use of sentiment as a feature with an F1 score of about 66.8%. Another study by [90] used labelled data toward 14 topics to train an NB model to predict the stance. The work of [42] used an SVM model as well as lexical and semantic features to classify the stance in SemEval, which had an F1 score equal to 63.6%. Another study by [92] used a stacked classifier and syntactic features to classify the stance in a SemEval stance dataset. Their model showed a minuscule improvement on the overall stance detection performance with a 62% F1 score. The work of [53] proposed a neural ensemble model using bidirectional LSTM on a SemEval stance dataset along with a fast-text embedding layer. Their model indicated an improvement on the overall F score, raising it to 72.1%. A recent work by [45] used a bidirectional gated recurrent unit to build a multitask learning model that leveraged the sentiment of a tweet to detect the stance.

This model showed an improvement of the stance detection model's overall F score of the SemEval dataset to reach a score equal to 72.3%. A detailed comparison of each study's performance on the SemEval dataset is reported later in section 7.4.

7.2 Transfer Learning

To address the scarcity of the labelled data for each target in the stance detection task, some studies in this field attempted to incorporate transfer learning techniques to enrich the representation of targets in the dataset and enhance the overall performance of stance detection [59, 93]. In transfer learning, the knowledge that an algorithm has learned from one task is applied to a separate task, such that in the task in stance detection, transfer learning is applied across different targets. One of the well known stance detection datasets that motivated the work on transfer learning is the SemEval stance (Task B) [3]. This dataset contains 78,000 unlabelled tweets related to Trump and provides a good source for research, to be explored through various transfer-learning algorithms. For example, in the work of [93], the SemEval stance (Task B) was used to train a bag-of-words auto-encoder along with Hillary-labelled data to detect the stance. Another study by [59] used Hillary Clinton's labelled data along with Trump's unlabelled data to develop a rule-based algorithm to help in detecting the stance for the SemEval stance Task B. In the work of [94], a CNN model was used with Google news embedding on the SemEval stance Task B. For Subtask B, their model trained on two class datasets and further, by using a modified softmax layer, the classification of three classes with a voting scheme was performed. With these configurations, their model ranked among the top three models with an overall F-measure of 56.28 for Task B. Moreover, the study by [95] implemented distant supervision using SemEval (Task B) to learn the features from two unlabelled datasets by using hashtag prediction. Another study by [96] used topic transfer knowledge to leverage the shared topics between two targets and predict the stance for unlabelled data. Recently, many studies have incorporated the concept of transfer learning using new datasets other than the SemEval stance (Task B). A recent study by [97] used Wikipedia articles to build knowledge extraction for each topic on a dataset that contained seven topics from Wikipedia. Another work by [98] used three datasets, namely blogs, US elections, and Moral Foundations Twitter, and designed algorithms that incorporated label dependencies between them to train the stance detection model. Another study by [99] used weak supervised stance detection to predict the stance towards 7 political targets to analyze the politicians' approval from tweets. In their model they used distance supervision method (DSSD) to label the data using certain keywords and hashtags. They found that using targeted oriented method (DSSD) is not generalized for unseen targets, such as Macron and Putin. One reason behind that is due to the heuristics used to generate weakly labeled data to train (DSSD) may not hold in different time periods. They concluded that low resource methods, stance detection trained on topic specific data, is more robust in predicting the stance in comparison with off the shelf methods such as using sentiment analysis VADER.

7.3 Unsupervised Learning

Recently, attention has been devoted toward building unsupervised stance detection models. In these kinds of studies, clustering techniques are primarily used with a focus on the user and topic representation on the social media platform [8, 77, 100, 101]. The work of [8] proposed an unsupervised model using the clustering model at the author and topic levels. In this study, six topics collected from two online debate forums, namely, 4Forums and CreateDebate were used. Their clustering model leveraged the content and interaction networks of the users (i.e., retweets and replies).

A recent study by [77] used the clustering technique to create an initial set of stance partition for annotation. In their work, the researchers used unlabelled tweets related to three topics, Kavanaugh, Trump, and Erdogan. Their findings

Table 3. Comparing the stance detection models on SemEval stance dataset.

Algorithm	Model	Features	F1	Study
Supervised learning	SVM	NW(Mentions+urls)	71.56	[38]
	SVM	NW(Mentions+urls)+content	72.5	[38]
	RNN	NW (followee)	61.8	[87]
	RNN	NW (followee)+Content	65.9	[87]
	Nested LSTMs	Content	72.1	[53]
	bidirectional gated recurrent unit	Content	72.33	[45]
	Hierarchical Attention NN	Content	69.8	[102]
Transfer learning	SVM	Content	70.0	[56]
	BiGRU	Noisy labeling + topic modeling	60.8	[103]
	BiLSTM	Sentiment lexicon	65.3	[104]
	Deep averaging network	words embedding	35.2	[105]
	Deep averaging network	Glove	30.2	[105]

showed that using retweets as a feature provided the best performance score upon implementing clustering algorithm (DBSCAN), which surpassed the supervised method when using the fast-text and SVM models. Their findings are considered greatly motivational for the use unsupervised methods for stance classification in the future.

7.4 Best Performing Algorithms

Since the SemEval stance dataset was the most-used dataset for bench-marking the performance on stance detection by using multiple ML approaches, this dataset has been used in this section to compare different ML approaches discussed in the above sections. Table 3 shows the best performing models based on the type of algorithm. As is expected, the transfer learning models have a lower performance score compared to supervised learning models. However, the study by [104] shows promising performance, with an F-score of 0.653. Nevertheless, other trials, such as the work of [105], which used a deep averaging network (DAN) with GloVe word embedding, achieved an average F1 score of around 0.3%, which is close to random performance.

As is evident in Table 3, the supervised algorithms are more effective for stance detection when they achieve higher F-scores on the SemEval dataset than transfer-learning approaches. This can be seen in the models that have incorporated network features to detect the stance. It is interesting to see that simpler machine learning models, such as SVM are more effective than deep learning models. In addition, these models achieve even better performances when they incorporate network features along with content, such that using both with a simple linear SVM is more effective than using word embedding with RNNs and LSTMs.

There are no unsupervised algorithms that have been applied to the same SemEval dataset to date. The only reported results, as discussed earlier, are those by [66], wherein the model was applied on different datasets (Trump, Kavanaugh, and Erdogan datasets). The unsupervised model of this study used network features and outperformed the model with the use of fast-text word embedding with a clustering purity equal to 98%. Overall, it can be observed that the best performing models are the ones framed in user and social contexts with the use of user network features in the stance detection task.

8 STANCE DETECTION APPLICATIONS

Stance detection has been mainly used to identify the attitude toward an entity of analysis to allow the measurement of public opinion toward an event or entity. However, there are other applications for stance detection that are discussed in this section.

8.1 Analytical Studies

The use of stance detection has proven beneficial as a social sensing technique to measure public support related to social, religious, and political topics.

Some examples of the use of stance detection for analyzing political topics are the studies on analyzing public reaction toward Brexit based on Twitter data [55, 106–108]. Furthermore, most of the stance detection studies used data from the 2016 US elections to analyze the stance toward two candidates, Hillary Clinton and Donald Trump [50, 67]. The work of [109] analyzed the stance toward the political debates on Twitter regarding the Italian constitutional referendum held in December 2016. Another work by [110] studied the public stance toward Catalan independence. A more recent study by [106] used a multilingual dataset to study political debates on social media. The researchers collected entities and events related to politics in five languages, English, French, Italian, Spanish, and Catalan.

Another line of study uses stance detection to analyze the public viewpoint on social aspects. The public opinion toward immigration has been recently studied [111, 112]. The work of [111] examined the attitude toward the refugees using data from Twitter. They collected tweets using the keyword "refugees" in different languages. They used the sentiment of the discourse to analyze the public stance toward the refugees. Another work by [112] studied immigration in the United Kingdom by annotating the tweets related to immigration with sentiment polarity. They used negative polarity as an indication of the against stance and positive sentiment as an indication of the in-support viewpoint. They found that about 5% of the users were against immigration, while 23% were in support of the immigration. It worth mentioning that the last two studies mentioned are seen as sub-optimal in measuring the stance due to their heavy reliance on the sentiment of the text, which has been demonstrated in section 3.2 and previous studies [27, 75] to be sub-optimal. A recent study by [113] analyzed the users' images posted on Facebook to understand their ideological leaning and how political ideology is conveyed through images. In their work, the researchers used the scores generated by the CNN image classifier model to further explore the features that distinguish the liberals from the conservatives. The most distinct features predicting liberal images are related to economic equality, whereas conservative members tend to use images that have objects related to state and economic power, such as "court". Another study by [114] used social media users stance to predict Media Bias. They used users stance towards eight topics along with valence score of top users to predict media media leaning.

On the other hand, stance detection has been used to analyze the attitude related to disruptive events [17, 52]. For instance, the work of [17] studied the public attitude toward Muslims after the Paris terror attacks in 2015. In their work, the researchers collected tweets that mentioned keywords related to Islam. Subsequently, they analyzed users' interactions and used the historical tweets to predict their attitudes toward Muslims. The work of [52] analyzed the attitudes toward 21 mass shooting events. They first derived a list of mass shooting events between 2015 and 2018 from the Gun Violence Archive. For each event, they defined a list of keywords to collect tweets related to those events. Further, to study the polarization of opinion toward these events, they estimated the party of each user, based on the political accounts they followed.

8.2 Applications Related to Social-Based Phenomena

Stance detection has been used to solve algorithmic issues on social media platforms. These issues are a reflection of social phenomena on these platforms. The most common phenomena that have affected various social media platforms are echo-chambers and homophily [115]. Previous studies have concluded that social media are polarized in nature, which boosts homophilic behavior [37, 82, 116–118]. Homophily is the social phenomenon that concerns people's tendency to connect with "like-minded friends." An echo-chamber is the cascade of certain information among groups of people. This social behavior has been magnified in social media structures wherein certain beliefs are amplified within close circles of communication. Consequently, people are exposed to content in consent with the same opinions that they hold. As a result, this reinforces social media users' view biases and blinds the users from other sides of information. Therefore, stance detection has been used to help in measuring and alleviating the problems that result from the polarization on social media. For example, a study by [119] used stance as the main factor to expose users to contradicting views. Further, stance detection was used to identify and measure the controversy level on social media platforms [120–122]. In a study by [122] the stance summarization technique was used to re-rank controversial topics. This method collected arguments based on five topics and summarized the overall stance with respect to the top tweets for a given topic.

8.3 Veracity Checking Applications

The rapid dissemination of news on social media platforms encourages people to depend on these platforms as their main sources of information. This kind of information consumption triggers critical issues in social media related to the credibility of the exchanged information, namely fake news and rumor detection. Recently, a lot of attention has been devoted toward the classification of stances to help in taking the first step toward solving the veracity checking issue [14, 15, 20].

A rumor is defined as an uncertain piece of information that lacks a secure standard of evidence for determining whether it is true or false [123]. It has been shown that false rumors have obvious effects across various fields. For example, in the healthcare sphere, false rumors on the Web pose major public health concerns. In 2014, rumors about the Ebola epidemic in West Africa emerged on social media, which made it more difficult for healthcare workers to combat the outbreak [124]. Rumors also affect modern-day journalism, which depends on social media as the main platform for breaking news. Additionally, there have been numerous occasions where false rumors have yielded severe stock market consequences [125]. To this end, many initiatives have been introduced, such as the Emergent dataset, PHEME dataset, and RumourEval SemEval-2017 and -2019 datasets [14, 15, 72, 126]. These initiatives have been implemented to encourage the development of tools that can help with the verification of the misinformation in news articles.

For these task, stance detection has been used as a key feature for checking the credibility of a piece of news. As discussed in section 5.3, stance in comments toward the news is measured to detect if these comments are confirming or denying the news, which, in turn, is used to detect if the news is a rumor or authentic.

Unlike rumors, fake news aims to create misleading information and it is always false [127]. Data veracity refers to the truthfulness, trustworthiness, and accuracy of the content. Social media posts have massive amounts of user-generated content through which fake information can be easily spread. One particularly challenging task is verifying the truthfulness of social media information, as the content by nature tends to be short and limited in context; these characteristics make it difficult to estimate the truthfulness of social media information compared to other information resources, such as news articles. One of the proposed methods to address the fake news is based on detecting the

stances toward news organizations. This is due to the fact that understanding the stances of other organizations toward a news topic is helpful in inferring the truthfulness of the news article. The Fake News Challenge initiative (FNC-1) adopted this approach and proposed a stance detection task to estimate the stance of articles toward a given headline (i.e., claim). The best performing system at the FNC was proposed by [128]. In their study, they used gradient-boosted decision trees and a convolutional neural network (CNN), along with many textual features. Another study by [129] achieved relatively similar results to the best system with a feature-light memory network model enhanced with LSTM and CNN. A more recent study by [130] used three features extracted from users' interactions, news authors, and the contents of news article to better detect the fake news. In the study by [131], a combination of lexical knowledge, word embedding, and n-gram was to detect the stances in two datasets, FNC-1 and Emergent. The study by [132] proposed a new text representation that incorporated the first two sentences of the article along with the news headline and the entire article to train a bidirectional RNN model using an FNC-1 dataset.

9 STANCE DETECTION RESOURCES

This section lists the current available resources for stance detection tasks. Table 4 shows the available datasets with stance annotation data in a chronological order. This table categorises the datasets on a higher level as classification and prediction datasets, as defined in section 4. In classification tasks the datasets further categorized as: target-specific, multi-target and claim-based stance dataset. The Other type of taxonomy is stance prediction datasets, with macro and micro predictions. The current data sources that has been used for stance detection includes: Twitter [3] [133] [15], Wikipedia[134], International debate education association website [134] and various news sites such as: Snopes.com [133].

Target-specific datasets: Table 5 shows a list of available datasets for stance detection on social media data. There are two publicly available datasets that contains stance annotations for predefined targets on social media. The first dataset is the SemEval stance detection [3] provides two datasets to serve two frameworks: supervised framework (Task A) and weakly supervised framework (Task B). In the supervised framework, the annotation scheme evaluates the author of a tweet stance towards five targets (atheism, Hillary Clinton, climate change, feminist movement and legalization of abortion). In this dataset, the tweet-target pair is annotated for stance and sentiment labels. On the other hand, the weakly supervised framework dataset contains tweets related to one target (Donald Trump). In this dataset the training data contains 78,000 tweets related to (Donald Trump) without stance labels. Several studies proposed stance detection model for SemEval stance dataset.

Another work by [135] provides a dataset related to (*Me Too*) movement. This dataset contains around 9000 tweet annotated with stance, hate-speech relevance, stance dialogue act and sarcasm. Another work by [136] introduce a dataset with stance and name entity labels. The dataset contains 1065 tweets annotated with stance labels towards two entities Galatasaray and Fenerbahce. A recent study by [137] provides a dataset for stance detection that contains around 51K tweets covering financial domain. Another study by [138] provides a dataset consists of German, French and Italian comments from Smartvote. The dataset constricted to evaluate cross-lingual methods for stance detection towards 150 targets.

Claim-based datasets: In this kind of stance detection dataset the object of evaluation is the source of information instead of social actor. The Rumours dataset [139] is a claim-based stance detection dataset designed for Rumours resolutions. This dataset contains 10,417 tweets related to Obama, Airfrance, cellphone, Michelle and plain. In this dataset the rumour tweet is evaluated against set of other tweets to define the stance of these tweets in supporting or denying the source of the rumour. Additionally, the Emergent dataset [126] is a Claim-based stance detection dataset

for Fact checking. This dataset contains rumours from variety of sources such as rumour sites, e.g. snopes.com, and Twitter. Another dataset available for claim detection, is Fake-News dataset. This dataset contains news articles from the Emergent dataset where the news headline being evaluated against set of a body text.

The SemEval 2019 rumours detection dataset by [14], enriched the SemEval 2017 (rumours detection task) dataset by adding new data from Reddit and extend the language representation of this dataset to include Russia.

Multi-related-targets: The two datasets that have multi-related-targets stance annotations are the Trump vs. Hillary dataset[50] and Multi-targets dataset [19]. In Trump vs. Hillary dataset, each tweet is stance annotated for the two candidates in the same time such as (supporting Hillary and Against Trump). The same annotation technique has been used in Multi-targets dataset for an extended list of US presidential candidates. The Multi-target dataset [19, 54] contains three pairs of targets Clinton-Sander, Clinton-Trump and Cruz-Trump.

Stance prediction datasets: As a result of the lack of benchmarks in this kind of stance detection, the researchers tend to build their own datasets as illustrated in tables 4. These datasets are designed to predict the stance before the event time. There are two datasets constructed by [37], where they used twitter as a source for stance prediction data.

Table 4. data-set with stance annotations for prediction task (on Chronological order)

Data-set	Type	Data Source	Stance definition (annotation)	Size
[61]	micro-level + macro-level	Debate forums	The probability of user choosing a stance on an issue	users 4,994, issues 1,727 arguments 39,549 (average 23 per issue)
[62]	(micro-level)	Online-forums	Stance(issue,posts)=pro,con	news articles from CNN and 4fourms
[17]	(micro-level)	Twitter	Stance(previous tweets,profiles,topic)=POS,NEG	Two data sets (not available) Islands dataset(Arabic) [33,024 tweets 2,464Users] Islam dataset [7,203 tweets 3,412 Users]

Data-set	Type	Application	Source of Data	Topics	Annotation	Size
Rumours dataset [139]	Claim-based	Rumours resolutions	Twitter	Obama, Airfrance, cellphone, Michelle, palin	Stance(users tweets,rumour's claim)={Confirming, Deny, Doubtful}	10,417 tweets
SemEval-Stance Detection Task A[3]	Target specific	General	Twitter	Atheism, Climate Change is a Real Concern, Feminist Movement, Hillary Clinton, and Legalization of Abortion	Stance(target, tweet)={Favor, Against, Neither}	Training: 2,914 Test-ing: 1,249.
SemEval-Stance Detection Task B [3]	Target specific	General	Twitter	Donald Trump	Stance(target, tweet)={Favor, Against, Neither}	707 labeled tweets 78,000 unlabeled
Emergent dataset [126]	Claim-based	Fact checking	variety of sources	world and national U.S. news and technology stories.	Stance(claim Headline, article)={for, against, observing}	300 claims, 2,595 associated article headlines
SemEval Rumour stance classification -Subtask A - SDQC Support/ Rumour stance classification [15]	Claim-based	The Veracity of a rumour.	Twitter	charliehebdo, ebola-essien, ferguson, germanwings-crash, ottawa shooting, prince-toronto, putinmissing, sydneyseige	Stance(originating rumourous, tweets reply)={deny, support, commenting}	training: 4519 threads, Testing: 28 threads, 1049 threads
SemEval Rumour stance classification Subtask B - Veracity prediction [15]	Claim-based	The Veracity of a rumour.	Twitter	charliehebdo, ebola-essien, ferguson, germanwings-crash, ottawa shooting, prince-toronto, putinmissing, sydneyseige	Stance(originating rumourous, tweets reply)={deny, support, commenting}	training: 297 threads , Testing: 28 threads
Fake News challenge 1 (FNC-1) [1]	Target specific	Fake news detection	Derived from Emergent dataset	various news articles	Stance(article headline, a body text)={Agrees, Disagrees, Discusses, Unrelated}	49972 instances
Multi-Targets Stance Detection [19]	Multi-related-targets	General	Twitter	Hillary Clinton-Sanders, Hillary Clinton-Trump and Cruz-Trump	Stance (multi-targets, tweet)={ Target A Favor/Against Target B}	4,455 tweets
Trump vs. Hillary dataset[50]	Multi-related-targets	General	Twitter	Hillary Clinton-Donald Trump	Stance (multi-targets, tweet)={ Target A Favor/Against Target B}	3,450 tweets
The Claim Polarity Dataset [134]	Claim-based	Argument construction	Claims are from Wikipedia and motions from (IDEA) website	5 topics selected randomly	Stance(claim, motion)={Pro,Con}	2,394 claims.
RumourEval 2019-Task A	Claim-based	The Veracity of a rumour	Tweets	Various topics	Stance(originating rumourous, tweets reply)={deny, support, commenting}	5216 training tweets
Hyperpartisan News Detection SemEval2019	Claim-based	Allegiance of news	News articles	various topics	Stance(article, entity)={ prejudiced, non-hyperpartisan }	750,000 articles
X-stance	Target specific	Allegiance of news	comments from smartvote.ch	150 targets	Stance(comment, target)={ support, against }	67,000 comments

Table 5. Publicly available data-sets with stance annotations for stance classification

10 DISCUSSION

Social media provides a rich source for social studies to analyze public opinion, especially on controversial issue. While sentiment analysis has been used for decades to analyse public opinion on products and services, stance detection comes as the correspondent solution for analysing public opinion on political and social topics, where sentiment analysis fails to reflect support. Stance detection in social media presents an optimal technique for various applications and has been used heavily in interdisciplinary studies including politics, religion and humanities.

It is worth noticing that small amount of work used stance to analyze the social issues in comparison with political topics. This is due to the controversial nature of the political topics which facilitates the data collection for the stance detection.

The sentiment analysis has a minuscule effect on detecting the accurate stance on social media. This is due to the complexity of stance modeling on social media, which can be inferred based on various factors ranging from the textual cues to the social identity of the author. Nevertheless, there has been a noticeable misconception in the literature where the sentiment has been used solely to indicate the stance. These studies defined a linear relation between the against/in-favor stance and the negative/ positive sentiment. As the analysis in section 3.2 shows that the relation between sentiment and stance is orthogonal.

Stance detection has been mostly approached using classification-based algorithm. This is mostly applied using supervised learning algorithms with huge dependency on human-annotated data. Consequently, techniques such as transfer learning and unsupervised learning have been used to resolve the scarcity of the annotated data but with less attention from researchers compared to supervised methods. This scarcity reflected by the need to enrich the data with information related to the object of interest. For instance, to detect stances related to climate change, information related to global warming considered beneficial for stance detection in-order to cover the complete aspect of the topic. Other studies worked on overcoming this issue by using distant supervision to annotate the data without having to manually label the data.

On the other end of the spectrum, few studies have dealt with stance prediction, which is mostly handled in relation with specific events. The main goal behind this kind of stance detection is to predicting the unexpressed views and to infer people standpoints toward an event in advance (pre-event). Therefore, stance prediction suits the analytical studies to examine the temporal effects of various events on public opinion such as candidates elections [50] or social phenomena as Islamophobic [17]. In this kind of studies, the dataset contains pre-event and post-event posts annotated with the user's stance before and after the event consequently. Thereby, predicting the stances is based on the user's past behavior which can be extracted from network features along with the post's content [140].

While there has been a large interest from the NLP community on developing effective approaches for stance detection [23] that mainly modeled the task as a text-entailment task, there is also large amount of work from the social computing and computational social science communities that showed the effectiveness of using user's interactions and network feature for stance detection and prediction. This shows that the task of stance detection is a multidisciplinary task that has been of interest to multiple computer science communities.

10.1 Possible directions for moving forward

The work on stance detection on social media has been growing over the last few years. More research gets published from multiple communities on the task and more applications get created. We believe that there are some gaps in the

existing work and potential emerging directions that should receive more attention in the near future on this task. In the following, we discuss these directions.

Some of the published work on transfer learning techniques for stance detection has shown some promising performance. The current techniques which used a topic based transfer learning achieved lower performance score in comparison with other techniques such as supervised learning techniques, as shown in section 7.4, however still promising. The transfer learning techniques for stance detection need a further enhancement by adapting a non-topical aspects to enhance the current state of this methodology on detecting the stance on social media, and potentially reaching the point to have comparable results with supervised on-topic stance detection. This might lead at some point to a general stance-classifier that can adapt to the topic of interest without the need to manually label new training data.

Several studies showed that stance detection models show consistent improvement on the overall performance when using the network features instead of just using content of post only. This kind of stance modeling draws more emphasis on the crucial role of users online behaviour and social attributes in the stance detection models. While network features are more effective, but they still require more computationally expensive, since they require gather additional information about users, which might be not highly practical in some cases. The need for further investigation in this direction is required to understand how to reach more effective, but at the same time highly efficient stance detection models that utilise network information.

Currently, the SemEval 2016 stance dataset [3] is the most popular benchmark dataset for target specific stance detection, and has served perfectly in many research publications, especially within the NLP community. Nevertheless, this dataset contains an average of only 500 records for each topic in the training set, which is quite limited. In addition, it is becoming old overtime, which means getting some of its tweets deleted, and thus failing to retrieve network information of their users for future research [38]. There is a real need for a new benchmark dataset for stance detection that provides a sufficient amount of data to train and test the stance detection models. In addition, most of the existing datasets are mainly focusing on social, political and religious issues. Having new datasets that covers additional domains would be of importance to explore the stance detection task for these new domains. Finally, creating benchmark datasets of multiple languages would be of interest, especially for the NLP community, where they focus on content-based stance detection approaches.

In general, as stance detection have been used heavily as social sensing technique on social media to study societal aspects ranging from politic, religion and social topics, this urges the need to incorporate more robust modeling of the stance on social by using network features for a more accurate analysis. As the use of textual entailment modeling shown to be sub-optimal when compared with network representation of the stance on social media section 6.3. Despite the undeniable benefits of stance detection for various societal studies, privacy measures could be developed to better preserve the privacy of the social media users. These privacy methods can provide a green analysis methods for social sciences and humanities with minimal cost of redesigning the social media platforms.

11 CONCLUSION

This survey covered the work on stance detection on social media and provided an overview of the current approaches to handle stance modeling. We initially presented the main difference between stance detection and sentiment analysis to remove any confusion between both techniques when used for measuring public opinion online. Later, we described the different types of targets when applying stance detection, namely: target-specific, multi-related-targets, and claim-based. Then we showed the existing work on detecting expressed stance or predicting unexpressed user’s stance on future events. Later, the most used features for modeling stance and the different machine learning approaches used were

discussed and compared, showing that network features are superior to content (textual features) for most of the studies on stance detection. In addition, supervised methods using SVM were found to be the most effective for different datasets. Also, the recent attempts for applying transfer learning and unsupervised learning for stance detection have promising results and is expected to be one of the main research direction in this area in the future. We also discussed the different applications of stance detection, including fake news detection which is one of the most popular research topics in the recent couple of years. Finally, we summarise the existing resources on stance detection, including datasets and most effective approaches.

Our final discussion on the survey listed the potential future directions in this area, including the need for having new datasets and the importance of using NLP and social computing approaches for better stance detection on social media.

REFERENCES

- [1] N. Newman, Mainstream media and the distribution of news in the age of social discovery, Reuters Institute for the Study of Journalism, University of Oxford (2011).
- [2] D. Biber, E. Finegan, Adverbial stance types in english, *Discourse processes* 11 (1) (1988) 1–34.
- [3] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, C. Cherry, SemEval-2016 task 6: Detecting stance in tweets, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, Association for Computational Linguistics, San Diego, California, 2016, pp. 31–41. doi:10.18653/v1/S16-1003. URL <https://www.aclweb.org/anthology/S16-1003>
- [4] O. Fraiser, G. Cabanac, Y. Pitarch, R. Besançon, M. Boughanem, Stance classification through proximity-based community detection (2018).
- [5] B. B. Klebanov, E. Beigman, D. Diermeier, Vocabulary choice as an indicator of perspective, in: *Proceedings of the ACL 2010 Conference Short Papers*, Association for Computational Linguistics, 2010, pp. 253–257.
- [6] H. Elfardy, Perspective Identification in Informal Text, Columbia University, 2017.
- [7] L. Zhu, Y. He, D. Zhou, Hierarchical viewpoint discovery from tweets using bayesian modelling, *Expert Systems with Applications* 116 (2019) 430–438.
- [8] A. Trabelsi, O. R. Zaiane, Unsupervised model for topic viewpoint discovery in online debates leveraging author interactions, in: *Twelfth International AAAI Conference on Web and Social Media*, 2018.
- [9] R. Bassiouney, Stance-taking, *The International Encyclopedia of Language and Social Interaction* (2015).
- [10] W.-H. Lin, T. Wilson, J. Wiebe, A. Hauptmann, Which side are you on?: identifying perspectives at the document and sentence levels, in: *Proceedings of the tenth conference on computational natural language learning*, Association for Computational Linguistics, 2006, pp. 109–116.
- [11] S. Somasundaran, J. Wiebe, Recognizing stances in online debates, in: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, Association for Computational Linguistics, 2009, pp. 226–234.
- [12] A. Murakami, R. Raymond, Support or oppose?: classifying positions in online debates from reply activities and opinion expressions, in: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Association for Computational Linguistics, 2010, pp. 869–875.
- [13] M. A. Walker, J. E. F. Tree, P. Anand, R. Abbott, J. King, A Corpus for Research on Deliberation and Debate., in: *LREC*, 2012, pp. 812–817.
- [14] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, L. Derczynski, SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours, in: *Proceedings of the 13th International Workshop on Semantic Evaluation*, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019, pp. 845–854. doi:10.18653/v1/S19-2147. URL <https://www.aclweb.org/anthology/S19-2147>
- [15] L. Derczynski, K. Bontcheva, M. Liakata, R. Procter, G. W. S. Hoi, A. Zubiaga, SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours, *Proceedings of the 13th International Workshop on Semantic Evaluation*. 2019 (2017).
- [16] A. Aker, A. Zubiaga, K. Bontcheva, A. Kolliakou, R. Procter, M. Liakata, Stance Classification in Out-of-Domain Rumours: A Case Study Around Mental Health Disorders, in: *International Conference on Social Informatics*, Springer, 2017, pp. 53–64.
- [17] K. Darwish, W. Magdy, A. Rahimi, T. Baldwin, N. Abokhodair, Predicting Online Islamophobic Behavior after# ParisAttacks, *The Journal of Web Science* 3 (1) (2017).
- [18] W. Magdy, K. Darwish, N. Abokhodair, A. Rahimi, T. Baldwin, #isisisnotislam or# deportallmuslims?: Predicting unspoken views, in: *Proceedings of the 8th ACM Conference on Web Science*, ACM, 2016, pp. 95–106.
- [19] P. Sobhani, D. Inkpen, X. Zhu, A dataset for multi-target stance detection, *EACL 2017* (2017) 551.
- [20] H. Allcott, M. Gentzkow, Social Media and Fake News in the 2016 Election, *Tech. rep.*, National Bureau of Economic Research (2017).
- [21] R. Wang, D. Zhou, M. Jiang, J. Si, Y. Yang, A survey on opinion mining: from stance to product aspect, *IEEE Access* (2019).

- [22] S. Ghosh, P. Singhanian, S. Singh, K. Rudra, S. Ghosh, Stance detection in web and social media: A comparative study, in: F. Crestani, M. Braschler, J. Savoy, A. Rauber, H. Müller, D. E. Losada, G. Heinatz Bürki, L. Cappellato, N. Ferro (Eds.), *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Springer International Publishing, Cham, 2019, pp. 75–87.
- [23] D. Küçük, F. Can, Stance detection: A survey, *ACM Computing Surveys (CSUR)* 53 (1) (2020) 1–37.
- [24] J. W. Du Bois, The stance triangle, *Stancetaking in discourse: Subjectivity, evaluation, interaction* 164 (2007) 139–182.
- [25] D. McKendrick, S. A. Webb, Taking a political stance in social work, *Critical and Radical Social Work* 2 (3) (2014) 357–369.
- [26] A. Jaffe, *Stance: sociolinguistic perspectives*, OUP USA, 2009.
- [27] A. Aldayel, W. Magdy, Assessing Sentiment of the Expressed Stance on Social Media, in: *Socinfo 2019*, 2019.
- [28] B. Pang, L. Lee, et al., Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval* 2 (1–2) (2008) 1–135.
- [29] D. Jurafsky, J. H. Martin, *Speech and language processing*, Vol. 3, Pearson London, 2014.
- [30] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. PYKL, D. Garrette, B. Gambäck, T. Chakraborty, T. Solorio, A. Das, Semeval-2020 sentiment task 9: Overview of sentiment analysis of code-mixed tweets, in: *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval-2020)*, Association for Computational Linguistics, Barcelona, Spain, 2020.
- [31] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in: *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 2015, pp. 486–495.
- [32] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, V. Stoyanov, SemEval-2016 task 4: Sentiment analysis in twitter, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, Association for Computational Linguistics, San Diego, California, 2016, pp. 1–18. doi:10.18653/v1/S16-1001. URL <https://www.aclweb.org/anthology/S16-1001>
- [33] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, et al., Semeval-2016 task 5: Aspect based sentiment analysis, in: *10th International Workshop on Semantic Evaluation (SemEval 2016)*, 2016.
- [34] S. Rosenthal, N. Farra, P. Nakov, SemEval-2017 task 4: Sentiment analysis in twitter, in: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 502–518. doi:10.18653/v1/S17-2088. URL <https://www.aclweb.org/anthology/S17-2088>
- [35] F. Benamara, M. Taboada, Y. Mathieu, Evaluative language beyond bags of words: Linguistic insights and computational applications, *Computational Linguistics* (2017).
- [36] P. Lahoti, K. Garimella, A. Gionis, Joint non-negative matrix factorization for learning ideological leaning on twitter, in: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ACM, 2018, pp. 351–359.
- [37] K. Darwish, W. Magdy, T. Zanoouda, Improved Stance Prediction in a User Similarity Feature Space, in: *ASONAM'17*, 2017.
- [38] A. Aldayel, W. Magdy, Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media, in: *CSCW 2019*, 2019.
- [39] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm, in: *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [40] M. Karamibekr, A. A. Ghorbani, Sentiment analysis of social issues, in: *Social Informatics (SocialInformatics)*, 2012 International Conference on, IEEE, 2012, pp. 215–221.
- [41] P. K. Singh, S. K. Singh, S. Paul, Sentiment classification of social issues using contextual valence shifters, *International Journal of Engineering and Technology* (2018).
- [42] H. Elfardy, M. Diab, Cu-gwu perspective at semeval-2016 task 6: Ideological stance detection in informal text, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 434–439.
- [43] P. Sobhani, S. Mohammad, S. Kiritchenko, Detecting stance in tweets and analyzing its interaction with sentiment, in: *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, 2016, pp. 159–169.
- [44] S. M. Mohammad, P. Sobhani, S. Kiritchenko, Stance and sentiment in tweets, *ACM Transactions on Internet Technology (TOIT)* 17 (3) (2017) 26.
- [45] Y. Li, C. Caragea, Multi-task stance detection with sentiment and stance lexicons, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 6300–6306.
- [46] S. M. Mohammad, Challenges in sentiment analysis, in: *A Practical Guide to Sentiment Analysis*, Springer, 2017, pp. 61–83.
- [47] D. S. Chauhan, R. Kumar, A. Ekbal, Attention Based Shared Representation for Multi-task Stance Detection and Sentiment Analysis, in: *International Conference on Neural Information Processing*, Springer, 2019, pp. 661–669.
- [48] H. Cramér, *Mathematical methods of statistics (PMS-9)*, Vol. 9, Princeton university press, 2016.
- [49] A. M. Liebetrau, *Measures of association*, Vol. 32, Sage, 1983.
- [50] K. Darwish, W. Magdy, T. Zanoouda, Trump vs. hillary: What went viral during the 2016 us presidential election, in: *International Conference on Social Informatics*, Springer, 2017, pp. 143–161.
- [51] M. Taulé, M. A. Martí, F. M. Rangel, P. Rosso, C. Bosco, V. Patti, et al., Overview of the task on stance and gender detection in tweets on catalan independence at ibereval 2017, in: *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017*, Vol. 1881, CEUR-WS, 2017, pp. 157–177.
- [52] D. Demszky, N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, D. Jurafsky, Analyzing polarization in social media: Method and application to tweets on 21 mass shootings, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1* (2019).

- [53] U. A. Siddiqua, A. N. Chy, M. Aono, Tweet Stance Detection Using an Attention based Neural Ensemble Model, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019, pp. 1868–1873.
- [54] P. Sobhani, D. Inkpen, X. Zhu, Exploring deep neural networks for multitarget stance detection, *Computational Intelligence* 35 (1) (2019) 82–97.
- [55] V. Simaki, C. Paradis, A. Kerren, Stance classification in texts from blogs on the 2016 british referendum, in: *International Conference on Speech and Computer*, Springer, 2017, pp. 700–709.
- [56] U. A. Siddiqua, A. N. Chy, M. Aono, Stance detection on microblog focusing on syntactic tree representation, in: *International Conference on Data Mining and Big Data*, Springer, 2018, pp. 478–490.
- [57] I. Augenstein, T. Rocktäschel, A. Vlachos, K. Bontcheva, Stance detection with bidirectional conditional encoding, *EMNLP 2016* (2016).
- [58] C. Liu, W. Li, B. Demarest, Y. Chen, S. Couture, D. Dakota, N. Haduong, N. Kaufman, A. Lamont, M. Pancholi, IUCL at SemEval-2016 Task 6: An Ensemble Model for Stance Detection in Twitter., in: *SemEval@ NAACL-HLT*, 2016, pp. 394–400.
- [59] M. Dias, K. Becker, INF-UFRGS-OPINION-MINING at SemEval-2016 Task 6: Automatic Generation of a Training Corpus for Unsupervised Identification of Stance in Tweets, *Proceedings of SemEval (2016)* 378–383.
- [60] Y. Igarashi, H. Komatsu, S. Kobayashi, N. Okazaki, K. Inui, Tohoku at SemEval-2016 Task 6: Feature-based Model versus Convolutional Neural Network for Stance Detection., in: *SemEval@ NAACL-HLT*, 2016, pp. 401–407.
- [61] M. Qiu, Y. Sim, N. A. Smith, J. Jiang, Modeling User Arguments, Interactions, and Attributes for Stance Prediction in Online Debate Forums, in: *Proceedings of the 2015 SIAM International Conference on Data Mining*, SIAM, 2015, pp. 855–863.
- [62] D. Rui, S. Yizhou, W. Lu, G. Yupeng, Z. Yuan, Weakly-Guided User Stance Prediction via Joint Modeling of Content and Social Interaction, in: *CIKM17*, Singapore, 2017.
- [63] Y. Gu, Y. Sun, N. Jiang, B. Wang, T. Chen, Topic-factorized ideal point estimation model for legislative voting network, in: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2014, pp. 183–192.
- [64] Y. Gu, T. Chen, Y. Sun, B. Wang, Ideology Detection for Twitter Users with Heterogeneous Types of Links, *arXiv preprint arXiv:1612.08207* (2016).
- [65] S. Gottipati, M. Qiu, L. Yang, F. Zhu, J. Jiang, Predicting users political party using ideological stances, in: *International Conference on Social Informatics*, Springer, 2013, pp. 177–191.
- [66] D. Kareem, S. Peter, M. J. Aupetit, N. Preslav, Unsupervised user stance detection on twitter, *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14 (2019).
- [67] M. Lai, D. I. H. Farias, V. Patti, P. Rosso, Friends and enemies of clinton and trump: using context for detecting stance in political tweets, in: *Mexican International Conference on Artificial Intelligence*, Springer, 2016, pp. 155–168.
- [68] P. Wei, J. Lin, W. Mao, Multi-target stance detection via a dynamic memory-augmented network, in: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, ACM, 2018, pp. 1229–1232.
- [69] S. Hamidian, M. Diab, Rumor detection and classification for twitter data, in: *Proceedings of the Fifth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS)*, 2015, pp. 71–77.
- [70] A. Aker, L. Derczynski, K. Bontcheva, Simple open stance classification for rumour analysis, *RANLP (2017)* 31–39.
- [71] A. Zubiaga, E. Kochkina, M. Liakata, R. Procter, M. Lukasik, K. Bontcheva, T. Cohn, I. Augenstein, Discourse-aware rumour stance classification in social media using sequential classifiers, *Information Processing & Management* 54 (2) (2018) 273–290.
- [72] E. Kochkina, M. Liakata, I. Augenstein, Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM, *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (2017).
- [73] Q. Li, Q. Zhang, L. Si, Rumor detection by exploiting user credibility information, attention and multi-task learning, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1173–1179.
- [74] P. Anand, M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, M. Minor, Cats rule and dogs drool!: Classifying stance in online debate, in: *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, Association for Computational Linguistics, 2011, pp. 1–9.
- [75] P. Sobhani, Stance detection and analysis in social media, Ph.D. thesis, Université d’Ottawa/University of Ottawa (2017).
- [76] B. G. Patra, D. Das, S. Bandyopadhyay, Ju_nlp at semeval-2016 task 6: detecting stance in tweets using support vector machines, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 440–444.
- [77] K. Darwish, P. Stefanov, M. J. Aupetit, P. Nakov, Unsupervised user stance detection on twitter, *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14 (2019).
- [78] A. Benton, M. Dredze, Using author embeddings to improve tweet stance classification, in: *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, 2018, pp. 184–194.
- [79] C. Li, A. Porco, D. Goldwasser, Structured representation learning for online debate stance prediction, in: *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3728–3739.
- [80] T. Thonet, G. Cabanac, M. Boughanem, K. Pinel-Sauvagnat, Users are known by the company they keep: Topic models for viewpoint discovery in social networks, in: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ACM, 2017, pp. 87–96.
- [81] M. Pennacchiotti, A.-M. Popescu, Democrats, republicans and starbucks aficionados: user classification in twitter, in: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2011, pp. 430–438.
- [82] A. Bessi, F. Petroni, M. Del Vicario, F. Zollo, A. Anagnostopoulos, A. Scala, G. Caldarelli, W. Quattrociocchi, Homophily and polarization in the age of misinformation, *The European Physical Journal Special Topics* 225 (10) (2016) 2047–2059.

- [83] J. Borge-Holthoefer, W. Magdy, K. Darwish, I. Weber, Content and network dynamics behind egyptian political polarization on twitter, in: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 2015, pp. 700–711.
- [84] I. Weber, V. R. K. Garimella, A. Batayneh, Secular vs. islamist polarization in egypt on twitter, in: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ACM, 2013, pp. 290–297.
- [85] A. Rajadesingan, H. Liu, Identifying users with opposing opinions in twitter debates, in: *International conference on social computing, behavioral-cultural modeling, and prediction*, Springer, 2014, pp. 153–160.
- [86] K. Dey, R. Shrivastava, S. Kaushik, V. Mathur, Assessing the effects of social familiarity and stance similarity in interaction dynamics, in: *International Workshop on Complex Networks and their Applications*, Springer, 2017, pp. 843–855.
- [87] V. Lynn, S. Giorgi, N. Balasubramanian, H. A. Schwartz, Tweet classification without the tweet: An empirical examination of user versus document attributes, in: *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, 2019, pp. 18–28.
- [88] Q. Zhang, S. Liang, A. Lipani, Z. Ren, E. Yilmaz, From Stances’ Imbalance to Their Hierarchical Representation and Detection, in: *WEB CONFERENCE 2019: PROCEEDINGS OF THE WORLD WIDE WEB CONFERENCE (WWW 2019)*, Assoc Comp Machinery; Microsoft; Amazon; Bloomberg; Google; Criteo AI Lab; CISCO; NTENT; Spotify; Yahoo Res; Wikimedia Fdn; Baidu; DiDi; eBay; Facebook; LinkedIn; Megagon Labs; Mix; Mozilla; Netflix Res; NE Univ, Khoury Coll Comp Sci; Pinterest; Quora; Visa Res; Walmart Labs; Airbnb; Letgo; Gordon & Betty Moore Fdn; Webcastor, 2019, pp. 2323–2332. doi:10.1145/3308558.3313724.
- [89] M. Lai, A. T. Cignarella, D. I. H. Far as, C. Bosco, V. Patti, P. Rosso, Multilingual Stance Detection in Social Media Political Debates, *Computer Speech & Language* (2020) 101075.
- [90] M. A. Walker, P. Anand, R. Abbott, J. E. F. Tree, C. Martell, J. King, That is your evidence?: Classifying stance in online political debate, *Decision Support Systems* 53 (4) (2012) 719–729.
- [91] P. Krejzl, J. Steinberger, UWB at SemEval-2016 Task 6: Stance Detection., in: *SemEval@ NAACL-HLT*, 2016, pp. 408–412.
- [92] M. Wojatzki, T. Zesch, ltl. uni-due at semeval-2016 task 6: Stance detection in social media using stacked classifiers, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 428–433.
- [93] I. Augenstein, A. Vlachos, K. Bontcheva, Usfd at semeval-2016 task 6: Any-target stance detection on twitter with autoencoders, in: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 389–393.
- [94] W. Wei, X. Zhang, X. Liu, W. Chen, T. Wang, pkudblab at SemEval-2016 Task 6: A Specific Convolutional Neural Network System for Effective Stance Detection., in: *SemEval@ NAACL-HLT*, 2016, pp. 384–388.
- [95] G. Zarrella, A. Marsh, MITRE at semeval-2016 task 6: Transfer learning for stance detection, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (2016).
- [96] P. Wei, W. Mao, Modeling Transferable Topics for Cross-Target Stance Detection, in: *PROCEEDINGS OF THE 42ND INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR ‘19)*, Assoc Comp Machinery; ACM Special Interest Grp Informat Retrieval, 2019, pp. 1173–1176. doi:10.1145/3331184.3331367.
- [97] K. Hanawa, A. Sasaki, N. Okazaki, K. Inui, Stance detection attending external knowledge from wikipedia, *Journal of Information Processing* 27 (2019) 499–506.
- [98] W. Ferreira, A. Vlachos, Incorporating label dependencies in multilabel stance detection, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 6351–6355.
- [99] I. Sen, F. Fl ock, C. Wagner, On the reliability and validity of detecting approval of political actors in tweets, in: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 1413–1426.
- [100] A. Joshi, P. Bhattacharyya, M. Carman, Political issue extraction model: A novel hierarchical topic model that uses tweets by political and non-political authors, in: *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2016, pp. 82–90.
- [101] R. Ammar, K. Muc ahid, D. Kareem, E. Tamer, B. Cansin, Embeddings based clustering for target specific stances: The case of a polarized turkey, *ICWSM2021* (2021).
- [102] Q. Sun, Z. Wang, Q. Zhu, G. Zhou, Stance detection with hierarchical attention network, in: *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 2399–2409.
- [103] P. Wei, W. Mao, G. Chen, A topic-aware reinforced model for weakly supervised stance detection, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 2019, pp. 7249–7256.
- [104] Y. Li, C. Caragea, Multi-task stance detection with sentiment and stance lexicons, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 6299–6305. doi:10.18653/v1/D19-1657.
URL <https://www.aclweb.org/anthology/D19-1657>
- [105] S. Ebner, F. Wang, B. Van Durme, Bag-of-words transfer: Non-contextual techniques for multi-task learning, in: *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, 2019, pp. 40–46.
- [106] M. Lai, A. T. Cignarella, D. I. H. Farias, C. Bosco, V. Patti, P. Rosso, Multilingual stance detection in social media political debates, *Computer Speech & Language* (2020) 101075.

- [107] V. Simaki, C. Paradis, M. Skeppstedt, M. Sahlgren, K. Kucher, A. Kerren, Annotating speaker stance in discourse: the brexit blog corpus, *Corpus Linguistics and Linguistic Theory* (2017).
- [108] M. Grčar, D. Cherepnalkoski, I. Mozetič, P. K. Novak, Stance and influence of twitter users regarding the brexit referendum, *Computational social networks* 4 (1) (2017) 6.
- [109] M. Lai, V. Patti, G. Ruffo, P. Rosso, Stance evolution and twitter interactions in an italian political debate, in: *International Conference on Applications of Natural Language to Information Systems*, Springer, 2018, pp. 15–27.
- [110] M. Taulà, M. A. Martí, F. M. Rangel, P. Rosso, C. Bosco, V. Patti, Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017, in: *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017*, Vol. 1881, CEUR-WS, 2017, pp. 157–177.
- [111] E. Gualda, C. Rebollo Díaz, et al., The refugee crisis on twitter: A diversity of discourses at a european crossroads (2016).
- [112] J. Bartlett, R. Norrie, Immigration on twitter: understanding public attitudes online, *Demos* (2015).
- [113] N. Xi, D. Ma, M. Liou, Z. C. Steinert-Threlkeld, J. Anastasopoulos, J. Joo, Understanding the political ideology of legislators from social media images, in: *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14, 2020, pp. 726–737.
- [114] P. Stefanov, K. Darwish, A. Atanasov, P. Nakov, Predicting the topical stance and political leaning of media using tweets, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 527–537.
- [115] C. Fuchs, *Social media: A critical introduction*, Sage, 2017.
- [116] P. Barberà, J. T. Jost, J. Nagler, J. A. Tucker, R. Bonneau, Tweeting from left to right: Is online political communication more than an echo chamber?, *Psychological science* 26 (10) (2015) 1531–1542.
- [117] W. Quattrociocchi, A. Scala, C. R. Sunstein, Echo chambers on facebook (2016).
- [118] K. Garimella, et al., Polarization on social media (2018).
- [119] K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, Mary, mary, quite contrary: Exposing twitter users to contrarian news, in: *Proceedings of the 26th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee*, 2017, pp. 201–205.
- [120] M. AlAyyoub, A. Rababah, Y. Jararweh, M. N. AlKabi, Gupta, B. Brij, Studying the controversy in online crowds interactions, *Applied Soft Computing* 66 (2018) 557–563.
- [121] S. Dori-Hacohen, J. Allan, Automated controversy detection on the web, in: *European Conference on Information Retrieval*, Springer, 2015, pp. 423–434.
- [122] M. Jang, J. Allan, Explaining controversy on social media via stance summarization, *arXiv preprint arXiv:1806.07942* (2018).
- [123] D. Levinson, M. Ember, *Encyclopedia of cultural anthropology*, Vol. 4, Holt New York, 1996.
- [124] J. M. Shultz, F. Baingana, Y. Neria, The 2014 ebola outbreak and mental health: current status and recommended response, *Jama* 313 (6) (2015) 567–568.
- [125] D. Schmidt, Stock market rumors and credibility, *Document de travail*, HEC Paris (2015).
- [126] W. Ferreira, A. Vlachos, Emergent: a novel data-set for stance classification, in: *Proceedings of the 15th Conference of the North American Chapter of the ACL*, ACL, San Diego, California, 2016.
- [127] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, R. Procter, Detection and resolution of rumours in social media: A survey, *arXiv preprint arXiv:1704.00656* (2017).
- [128] S. Baird, D. Sibley, Y. Pan, Talos targets disinformation with fake news challenge victory. [online], *talosintelligence* (2017).
URL <https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html>
- [129] M. Mohtarami, R. Baly, J. Glass, P. Nakov, L. Márquez, A. Moschitti, Automatic stance detection using end-to-end memory networks, in: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), Vol. 1, 2018, pp. 767–776.
- [130] K. Shu, S. Wang, H. Liu, Beyond news contents: The role of social context for fake news detection, in: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, ACM, 2019, pp. 312–320.
- [131] B. Ghanem, P. Rosso, F. Rangel, Stance Detection in Fake News A Combined Feature Representation, in: *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 66–71. doi:10.18653/v1/W18-5510.
URL <https://www.aclweb.org/anthology/W18-5510>
- [132] L. Borges, B. Martins, P. Calado, Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News, *ACM JOURNAL OF DATA AND INFORMATION QUALITY* 11 (3) (Jul. 2019). doi:10.1145/3287763.
- [133] W. Ferreira, A. Vlachos, Emergent: a novel data-set for stance classification, in: *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, 2016, pp. 1163–1168.
- [134] R. Bar-Haim, I. Bhattacharya, F. Dinuzzo, A. Saha, N. Slonim, Stance classification of context-dependent claims, in: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Vol. 1, 2017, pp. 251–261.
- [135] A. Gautam, P. Mathur, R. Gosangi, D. Mahata, R. Sawhney, R. R. Shah, # metooma: Multi-aspect annotations of tweets related to the metoo movement, *Proceedings of the International AAAI Conference on Web and Social Media* 14 (1) (2020) 209–216.
URL <https://ojs.aaai.org/index.php/ICWSM/article/view/7292>
- [136] D. Küçük, F. Can, A tweet dataset annotated for named entity recognition and stance detection, *arXiv preprint arXiv:1901.04787* (2019).

- [137] C. Conforti, J. Berndt, M. T. Pilehvar, C. Giannitsarou, F. Toxvaerd, N. Collier, Will-they-won't-they: A very large dataset for stance detection on twitter, *ACL* (2020).
- [138] J. Vamvas, R. Sennrich, X-stance: A multilingual multi-target dataset for stance detection, *EMNLP 2020* (2020).
- [139] V. Qazvinian, E. Rosengren, D. R. Radev, Q. Mei, Rumor has it: Identifying misinformation in microblogs, in: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2011, pp. 1589–1599.
- [140] I. Himmelboim, S. McCreery, M. Smith, Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on twitter, *Journal of Computer-Mediated Communication* 18 (2) (2013) 154–174.
- [141] J. Ebrahimi, D. Dou, D. Lowd, A Joint Sentiment-Target-Stance Model for Stance Classification in Tweets., in: *COLING*, 2016, pp. 2656–2665.
- [142] Y. Zhou, A. I. Cristea, L. Shi, Connecting targets to tweets: Semantic attention-based model for target-specific stance detection, in: *International Conference on Web Information Systems Engineering*, Springer, 2017, pp. 18–32.
- [143] P. Vijayaraghavan, I. Sysoev, S. Vosoughi, D. Roy, DeepStance at SemEval-2016 Task 6: Detecting Stance in Tweets Using Character and Word-Level CNNs, *Proceedings of SemEval* (2016) 425–431.
- [144] U. A. Siddiqua, A. N. Chy, M. Aono, Tweet stance detection using multi-kernel convolution and attentive lstm variants, *IEICE TRANSACTIONS on Information and Systems* 102 (12) (2019) 2493–2503.
- [145] L. Derczynski, K. Bontcheva, M. Lukasik, T. Declerck, A. Scharl, G. Georgiev, P. Osenova, T. P. Lobo, A. Kolliakou, R. Stewart, et al., PHEME computing veracity the fourth challenge of big social data, in: *Proceedings of the Extended Semantic Web Conference EU Project Networking session ESCWPN*, 2015.
- [146] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, P. Tolmie, Analysing how people orient to and spread rumours in social media by looking at conversational threads, *PloS one* 11 (3) (2016) e0150989.

A APPENDIX A: STANCE DETECTION WORK

This appendix provides list of the most recent stance prediction and detection work on social media

Study	Features	ML	Dataset
[17]	Content Features (Hashtags, Text); Profile Features (Description, Name, Location); Network Features (Mention, Reply, Retweet)	SVM	Islamophobic dataset (Twitter) [Not available]
[18]	Content Features (Hashtags, Text); Profile Features (Description, Name, Location); Network Features (Mention, Reply, Retweet)	SVM	Islamophobic dataset (Twitter) [Not available]
[37]	Content Features(Text); Interaction Elements(retweeted accounts, used hashtags, mentioned accounts (MEN), shared URLs (URL)); User Similarity	SVM	Islands Dataset and Islamophobic dataset (Twitter) [Not available]
[36]	A combination of network and content	Non-negative matrix factorization	dataset covered Three controversial topics:gun control,abortion and obamacare (Twitter) [Not available]
[65]	similarity between users	Probabilistic Matrix Factorization	1000 user profile of Democrats and Republicans (debate.org) [Not available]
[62]	post level interaction and user level interaction	Stance-based Text Generative Model with Rule-based User-User Interaction Links	CNN dataset, 4Forums and IAC discussion forum [Not available]

Table 6. Work in stance prediction

Study	Task	Features	ML	Dataset
[38]	target-specific	NW features	SVM	SemEval-2016 shared task 6 [Available]
[87]	target-specific	NW (followee)	RNN	SemEval-2016 shared task 6 [Available]
[53]	target-specific	Content	Nested LSTMs	SemEval-2016 shared task 6 [Available]
[102]	target-specific	Content	Hierarchical Attention NN	SemEval-2016 shared task 6 [Available]
[56]	target-specific	Content	SVM Tree Kernel	SemEval-2016 shared task 6 [Available]
[103]	target-specific	Content+Sentiment lexicon	BiLSTM	SemEval-2016 shared task 6 [Available]

[103]	target-specific	Content+Noisy stance labeling + Topic Modeling	BiGRU	SemEval-2016 shared task 6 [Available]
[105]	target-specific	words embedding	Deep averaging network	SemEval-2016 shared task 6 [Available]
[58]	target-specific	bag-of-words and word vectors (GloVe and word2vec)	gradient boosting decision trees and SVM and merge all classifiers into an ensemble method	SemEval-2016 shared task 6 [Available]
[59]	target-specific	n-gram and sentiment	SVM	SemEval-2016 shared task 6 [Available]
[59]	target-specific	n-gram and sentiment	SVM	SemEval-2016 shared task 6 [Available]
[60]	target-specific	Reply, BagOfWord, BagOfDependencies, POS tags Sentiment WordNet, Sentiment Word Subject, Target Sentiment and Point-wise Mutual Information	CNN	SemEval-2016 shared task 6 [Available]
[60]	target-specific	Reply, BagOfWord, BagOfDependencies, POS tags Sentiment WordNet, Sentiment Word Subject, Target Sentiment and Point-wise Mutual Information	CNN	SemEval-2016 shared task 6 [Available]
[57]	target-specific	word2vec	Bidirectional LSTMs	SemEval-2016 shared task 6 [Available]
[91]	target-specific	hashtags, n-grams, tweet length, Part-of-speech, General Inquirer, entity-centered sentiment dictionaries, Domain Stance Dictionary	Maximum entropy classifier	SemEval-2016 shared task 6 [Available]
[141]	target-specific	n-gram and sentiments	discriminative and generative models	SemEval-2016 shared task 6 [Available]
[94]	target-specific	Google news word2vec and hash-tags	CNN	SemEval-2016 shared task 6 [Available]
[95]	target-specific	word2vec hash-tags	LSTM	SemEval-2016 shared task 6 [Available]
[85]	target-specific	unigrams, bigrams and trigrams	Naive Bayes	hotly contested gun reforms debate from April 15th, 2013 to April 18th, 2013. [Available]
[142]	target-specific	word embeddings	Bi-directional GRU-CNN	SemEval-2016 shared task 6 [Available]
[143]	target-specific	word embeddings	convolutional neural networks(CNN)	SemEval-2016 shared task 6 [Available]

[42]	target-specific	Lexical Features, Latent Semantics, Sentiment, Linguistic Inquiry , Word Count and Frame Semantics features	SVM	SemEval-2016 shared task 6 [Available]
[67]	target-specific	sentiment, Opinion target, Structural Features(Hashtags, Mentions, Punctuation marks), text-Based Features(targetByName, targetByPronoun, targetParty, targetPartyColleagues)	Gaussian Naive Bayes classifier	Hillary Clinton and Donald Trump dataset [Not available]
[19]	multi-target	word vectors	bidirectional recurrent neural network (RNN)	Multi-Target Stance dataset [Available]
[144]	multi-target	content of tweets	Multi-Kernel Convolution and Attentive LSTM	Multi-Target Stance dataset [Available]
[134]	claim-based	Contrast scores	random forest and SVM	The claim polarity dataset claims are from Wikipedia and motions f are from rom (IDEA)(on-line forums) [Available]
[70]	claim-based	Linguistic, message-based, and topic-based such as (Bag of words, POS tag, Sentiment, Named entity and others	Random Forest, Decision tree and Instance Based classifier (K-NN)	RumourEval and PHEME datasets [Available]
[69]	claim-based	tweet content, Unigram-Bigram Bag of Words, Part of Speech, Sentiment, Emoticon, Named-Entity Recognition , event, time, Reply, Re-tweet, User ID, Hashtag, URL	decision trees	[139] [Available]
[70]	claim-based	BOW,Brown Cluster, POS tag, Sentiment, Named entity, Reply, Emoticon, URL, Mood, Originality score, is User Verified(0-1),Number Of Followers, Role score, Engagement score, Favourites score and other tweets related features	decision tree, Random Forests and Instance Based classifier	RumourEval dataset [15] and the PHEME dataset [145] [Available]
[71]	claim-based	Word2Vec, POS, Use of negation, Use of swear words, Tweet length, Word count, Use of question mark, Use of exclamation mark, Attachment of URL and other contextualized features	Linear CRF and tree CRF , a Long Short-Term Memory (LSTM)	PHEME dataset [145] and Rmour dataset associated with eight events corresponding to breaking news events [146] [Available]

[72]	claim-based	word2vec, Tweet lexicon (count of negation words and count of swear words),Punctuation, Attachments,Relation to other tweets, Content length and Tweet role (source tweet of a conversation)	branch-LSTM , a neural network architecture that uses layers of LSTM units	Rumoureal dataset [15] [Available]
------	-------------	--	--	---------------------------------------

Table 7. Work in stance classification