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

Abeer AlDayel, School of Informatics/ The University of Edinburgh

a.aldayel@ed.ac.uk

Walid Magdy, School of Informatics/ The University of Edinburgh

wmagdy@inf.ed.ac.uk

Stance detection on social media is an emerging opinion mining paradigm for various social and political applications where sentiment analysis might be sub-optimal. This paper surveys the work on stance detection and situates its usage within current opinion mining techniques in social media. An exhaustive review of stance detection techniques on social media is presented, including the task definition, the different types of targets in stance detection, the features set used, and the various machine learning approaches applied. The survey reports the 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 the different applications of stance detection on social media. The study concludes by providing discussion of the gaps in the current existing research and highlighting the possible future directions for stance detection on social media.

Additional Key Words and Phrases: Stance detection, Opinion, Stance prediction, viewpoints, perspectives, Sentiment and Stance

1. INTRODUCTION

Nowadays, social media platforms constitute a major component of an individual's social interaction. These platforms are considered a robust information dissemination tool to express opinion and share views. People rely on these tools as the main source of news to connect with the world and to get instant updates [Newman 2011]. They have a major beneficial side that allows individuals to explore various aspects of an emerging topics, express their own points of view, get instant feedback and exploring the public views. The huge dependency of users on these platforms as the main source of communication allowed researchers to study different aspects of online human behavior, including public stance towards various social and political aspects.

Stance is defined as the expression of the speaker's standpoint and judgment toward a given proposition [Biber and Finegan 1988]. 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, where people express the opposing opinions towards differentiable points. Social issues such as abortion, climate changes and feminism have been heavily used as target topics in stance detection in social media [Mohammad et al. 2016]. Similarly, political topics, such as referendums and elections, have been always as hot topics that used stance detection to study public opinion [Fraisier et al. 2018]. Generally, stance detection process is also known as perspective [Klebanov et al. 2010; Elfardy 2017] and viewpoint [Zhu et al. 2019; Trabelsi and Zaiane 2018] detection. In which, perspective is identified by expressing stances toward an object of a controversial topic [Elfardy 2017].

The stance has been used in various research as a mean to link between linguistic forms and social identities which has the capability tobetter understand the background of people with a polarised stance [Bassiouney 2015]. Consequently, the early work on stance detection was emerged in analyzing political debates in online forums [Lin et al. 2006; Somasundaran and Wiebe 2009; Murakami and Raymond 2010; Walker et al. 2012b].

Recently, a growing body of work has used the concept of stance detection on social media as new frontier to analyze various social and political issues in social media platforms. What distinguish the earlier work on

analysing stances of debates on online forums from the recent work on social media is that the former has a clear single context compared with social media platforms, such as Twitter. In the online forums, the users debate in form of thread discussion where there is a flow of information related to the topic. In contrast, social media lacks such representation where the user's participate to a topic solely, and rarely with a context (i.e.thread).

One of earliest initiatives to promote stance detection on social media is the 'SemEval 2016 task 6' shared-task, where it introduced a stance specific benchmark dataset to help in evaluating stance detection on Twitter [Mohammad et al. 2016]. With a major focus on identifying stance towards social issues, SemEval 2016 has formulated the stance detection task around few controversial issues, such as atheism, feminism, and some politicians. In addition, a new wave of stance detection applications has been triggered to handle some issues that have infected the social media lately, such as fake news and rumours [Gorrell et al. 2019; Derczynski et al. 2017; Aker et al. 2017b], where the stance towards claims in a piece of news is utilized as a key feature for the detection of the credibility of this news.

The majority of work on stance detection has targeted the detection of the stance towards a given subject expressed in a given text. However, some work have studied the detection of stance of users towards 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: detecting expressed views vs 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 or against a given subject [Mohammad et al. 2016].

In the later one, the prediction is carried out to infer user's viewpoint on a given topic that the user did not discuss explicitly or towards an event that has not occur yet. This type of stance detection has proven its effectiveness in predicting the future attitudes in the aftermath of an event [Darwish et al. 2017c; Magdy et al. 2016].

The work on stance detection can be also categorized based on the topic of the target of analysis, where it can be one specific target, multiple-related targets, or a claim in a news article. Most of the existing work designs stance detection classifiers to identify the user's stance towards one given specific topic. Sometimes the classifier is built to detect the stance towards multiple-related targets. This is the situation when stance is detected towards two related entities that are typically opponents, such as detecting the stance towards Clinton and Trump simultaneously, since if stance is in favor one target, it would be simply against the other [Sobhani et al. 2017]. When the target is a claim in a news statement, the main objective is to identify if a given claim supported by other posts. In this type of detection, the analyses is between two posts such as in RumourEval [Gorrell et al. 2019; Derczynski et al. 2017] or between news header and other body of articles, which is mainly used as an initial step for fake news detection [Allcott and Gentzkow 2017]

This article surveys the stance detection on social media platforms. It maps out the terrain of existing research on stance detection and synthesizes how it relates to existing theoretical orientations. To our best knowledge, this is the first work to cover the current research directions on stance detection in social media and to identify a framework for future research. Recently, there has been a couple of surveys that relates to stance detection. The first survey study by [Wang et al. 2019] focused more on surveying the opinion mining methods in general with focus on stances towards products. A more recent survey study by [Küçük and Can 2020] focused more on surveying the research work that modeled stance detection as a text entailment task using natural language processing (NLP) methods. This survey study provides a broader coverage to stance detection methods covering the work that has been published in multiple research domains including NLP, Computational social science, and Web science. It surveys the modeling of stance using text, network, and behavioural features; which is overlooked by previous surveys.

The rest of the paper is organized as follows:Section 2 provides theoretical definition of stance. Section 3 contrasts the differences between stance and sentiment. Section 4 and 5 summaries the literature on stance prediction and categorizes this work according to the type of detected stance (expressed vs unexpressed) and
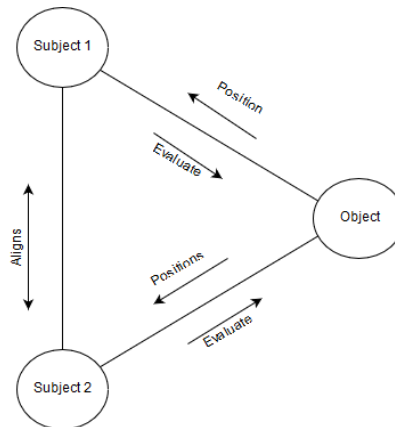
Fig. 1. The stance triangle, adapted from [Du Bois 20]

the target of the stance in the text. Section 6 and 7 explore with the different stance modeling and machine learning approaches for stance classification. Section 8 lists several application for stance detection, such as social media analysis and fake-news detection. Stance detection resources are listed in section 9. Finally, the current research trends on stance classification is discussed in Section 10 while highlighting the gaps and suggesting the potential future work required in this area.

2. STANCE DETECTION

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

Du Bois [2007] argues that stance-taking (i.e. a person taking a polarised stance towards a given subject) is a subjective and inter-subjective phenomenon in which stance-taking process is affected by personal opinion and non-personal factors such as cultural norms. Stance taking is a complex process relates to different personal, cultural and social aspects. For instance, the political stance taking depends on experiential behavior as stated by [McKendrick and Webb 2014].

The process of detecting stance of a given person on social media is still in its infancy as it is not yet clear what the role of language and social interaction plays in inferring the user's stance.

Stance detection has a strong history in sociolinguistic, where 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 the stance detection is usually associated with attributes such as adjectives, adverbs and lexical items [Jaffe 2009].

It has been argued that stance taking usually depends on experiential behavior which is based on previous knowledge about the object of evaluation [McKendrick and Webb 2014]. This strengthen the stance detection as a major component in various analytical studies.

Stance detection on social media concerns with individual's views towards an object of evaluation by using various aspects related to the user's post and personality traits.

As stated in Due Bois's stance triangle shown in Fig 1, the process of taking stance is based on three factors: 1) Evaluating objects; 2) Positioning subject (the self); and 3) Aligning with other subjects (other social actor). For instance, "I am with the new legalization on Climate Change" has a subject "self" indicated by proposition "I" and the "with" indicates the favor position towards the object "'Climate Change". In social media, identifying the stance object is mostly straightforward as each post is linked to the user.

Furthermore, from sociolinguistic perspective [Jaffe 2009], it has been argued that there is no complete neutral stance as people tends to position themselves through their texts as in favor or against the object of evaluation. This casts a further complexity in identifying stance of the social actors, since stance is not usually transparent in the text, but sometimes needs to 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 main method to measure support of a given target, which is sub-optimal [Aldayel and Magdy 2019a]. In this section, we highlight the main theoretical differences between the stance and sentiment, and illustrate these differences with some examples. In addition, we apply a quantitative study on data labeled 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 user emotional polarity, either being positive, negative or neutral [Pang et al. 2008; Jurafsky and Martin 2014]. This can be seen in related work on SemEval sentiment tasks SemEval-/(2015, 2016, 2017 and 2020), where the aim of sentiment analysis task is to determine the polarity towards an aspect of a product such as cell phone [Patwa et al. 2020; Pontiki et al. 2015; Nakov et al. 2016; Pontiki et al. 2016; Rosenthal et al. 2017]. Unlike sentiment analysis, stance detection mainly focus on identifying a person's standpoint or view towards an object of evaluation either being in-favor (supporting) or against (opposing) the topic . This is usually inferred by a mixture of signals beside the linguistic cues such as user's feelings, personality traits and cultural aspects [Biber and Finegan 1988]. Sentiment analysis is mostly approached as a linguistic agnostic task that focuses on leveraging given text linguistic properties to identify the polarity of text [Benamara et al. 2017], while stance detection can take an additional paradigm by leveraging non-textual features such as network and contextual features to infer user's stance [Lahoti et al. 2018; Darwish et al. 2017a; Aldayel and Magdy 2019b].

In general, sentiment analysis concerns with detecting the polarity of the text, which can be inferred without the need of having a given target of interest; for example *"I am happy"*. Thus, 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.

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

Another kind of sentiment analysis work concerns with inferring the polarity of a text by using predefined set of targets, this kind of sentiment analysis is usually refereed to as target-based sentiment analysis [Ma et al. 2018; Pontiki et al. 2016; Karamibekr and Ghorbani 2012; Singh et al. 2018]. In this kind of sentiment analyses, the classification task can be defined as follows:

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

In the 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 needs to be defined in advance to assess the overall attitude towards 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

Table I. 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	-	+

G as illustrated in equation 3.

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

There are three possible variations of this definition which may include as input either the text, the social actor or both for stance detection. The dependent factor in stance detection task is the *target* of analysis. One of the common practice in the stance detection on 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 [Lin et al. 2006; Mohammad et al. 2016]. From a more relaxed definition, a text T entails a stance to a target G, ($T \longrightarrow$ 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 could 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's stance triangle shown in Fig 1 [Du Bois 2007], where 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 (age, description,..etc) [Magdy et al. 2016; Darwish et al. 2017a].

To define the interplay between sentiment and stance, several studies have demonstrated that using sentiment as the only dependent factor to interpret user's stance is not sufficient [Somasundaran and Wiebe 2009; Elfardy and Diab 2016; Sobhani et al. 2016; Mohammad et al. 2017; Aldayel and Magdy 2019a]. This is due to the complexity of interpreting stances from a given texts as it is not always directly aligned with the polarity of a given post. Table I shows tweets examples that illustrates the orthogonal relation between sentiment polarity and the stance. For instance, in example 1 the tweet is against legalisation of abortion with a positive sentiment. In this example legalisation of abortion is the target for which the opinion is expressed. While the text has a positive sentiment, yet it shows an against stance. Furthermore, a text may not indicates any sentiment and still posing a clear stance to a target. Such as example 2, where the tweet has a neutral sentiment with an against stance to Legalization of Abortion. While example 3 shows the opposite case, where a negative sentiment can be detected but without a supporting stance to the target. Finally, in example 4, applying target-dependent sentiment analysis towards Hillary Clinton will lead to a clear negative sentiment, while measuring the stance towards 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 towards 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 SemEval stance detection dataset, where 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),
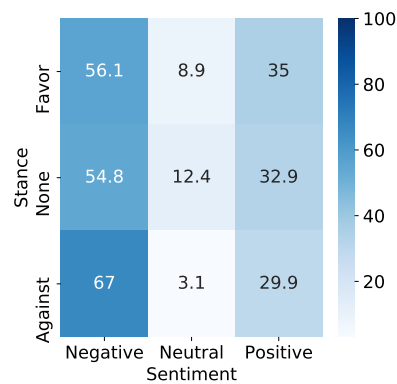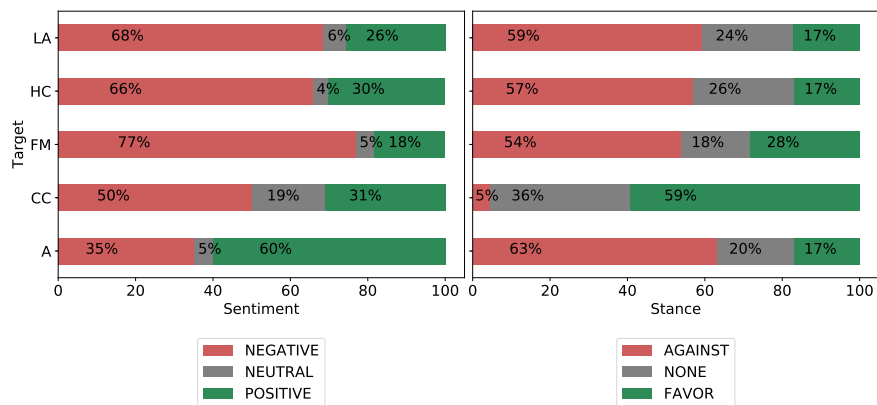
Fig. 2. Sentiment distribution over stance.

Fig. 3. Sentiment and stance distribution

'Feminist Movement' (FM), 'Hillary Clinton' (HC), and 'Legalisation of Abortion' (LA). Fig 2 illustrates the sentiment distribution over stances for the whole collection (Favor, Against, None). The graph shows that the negative sentiment constitutes the major polarity over the three stances categories. As negative sentiment represents over 56% of different stances. This reveals the tendency of using negative sentiments to express a viewpoint in a controversial topic.

The overall agreement between stance and sentiment is relatively minuscule. About 35% of Favor stance tweets have a positive sentiment. It can be observed that tweets with neutral sentiment fail to infer the neutral stances with only 12.4% of tweets with neither polarity. Instead 54.8% of the tweets have negative polarity, while 32.9% of tweets carry positive sentiment for a against stance. It is clear that sentiment does not simply represent stance.

Large amount of work use sentiment as a proxy to interpret stances. Based on the assumption that polarity of a text leads to reveal the stance of the text author. It has been shown that stance could be inferred independently from author's emotion state [Li and Caragea 2019a; Sobhani et al. 2016; Mohammad et al. 2017; Mohammad 2017]. A recent study by [Li and Caragea 2019a], used sentiment to predict the stance by using a multi-task learning model. In their study they used a sentiment lexicon along with to build

a stance lexicon and incorporated it for the attention layer in the model. Another work by [Chauhan et al. 2019] used sentiment as auxiliary task to predict the stance using SemEval stance dataset.

Using sentiment polarity to analyze the social media as a mean to interpret the general stance toward a topic is not necessarily accurate. Fig 3 shows the clear differences in detecting the users attitudes towards five SemEval topics. While 60% of tweets show a positive sentiment toward Atheism, the majority (63%) is opposing this topic. Similarly, in analyzing the general attitude to climate change, 50% of tweets have negative sentiment, while actually 59% of tweets support the idea that climate change is a real concern. Following the same trend, the pure sentiment value fails to detect the accurate stance toward the Hillary Clinton, Feminist Movement and legalization of Abortion. As these distribution show about 9% differences between the accurate against position and negative sentiment value in Hillary Clinton and Legalization of Abortion. In Feminist Movement the difference is laudable, by 23% misinterpretation of the real stance.

Further, we analyzed whether the sentiment and stance are independent of one another. We used Cramer's test [Cramér 2016] to gauge the strength of relationship between sentiment and stance. The result from Cramer's test is variance (V) value ranged between 0 and 1, with 1 as an indication of high association between the nominal variables [Liebetrau 1983]. By applying Cramer's test on SemEval stance dataset the resulted V value = 0.12, which is a strong indication of the in-dependency between sentiment and stance.

4. EXPRESSED VS UNEXPRESSED STANCE DETECTION

We can divide the literature work of stance detection using social media data according to the status of the stance to be modeled; either being 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 literature. The other type of work concerns with detecting the stance prior to an event which called stance prediction.

4.1 Stance Detection for Expressed Views

Many social media users express their support or opposition towards various topics online, which makes these platforms a valuable source to analyze public opinions towards these topics. This motivated a rich body of research to focus on inferring 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 concern with studying the users stance towards a post-event. For instance, SemEval stance dataset [Mohammad et al. 2016] created three sets of hashtags for each event/entity to collect tweets concerning three stances. For example, for the topic "Hilary Clinton", they used favor hashtags, such as #Hillary4President, against hashtags: #HillNo, and ambiguous hashtags. The ambiguous hashtags contains the target without the direct indication of the stance in the hashtag (eg: #Hillary2016). In study done by [Darwish et al. 2017b] to predict the public opinion and what went viral during 2016 US elections, the study used tweets that have been filtered out by using set of 38 keywords related to the US elections for streaming relevant tweets to this event. Another study used the tweets generated by users to infer the opinion toward independence of Catalonia and collected tweets by using two keywords related to hashtags #Independencia and #27S [Taulé et al. 2017]. Similarly, to study the mass shooting events, the study by [Demszky et al. 2019] used GunViolence Archive and collected list of events between 2015 and 2018 to define the set of keywords that can be used to retrieve the tweets. Then the study used Twitter firehose [1] to collect tweets for these keywords.

After collecting the events related data, it gets annotated using a predefined guidelines for labeling the data with the stance towards the given target, which is the method applied for labeling the SemEval dataset [Mohammad et al. 2017; Mohammad et al. 2016].

---

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

Most of the work on detecting the stance in expressed-views dataset applies NLP methods that model the task as text entailment to detect whether a piece of text is in favor or against the target [Siddiqua et al. 2019a; Sobhani et al. 2019; Simaki et al. 2017a; Siddiqua et al. 2018; Augenstein et al. 2016a; Liu et al. 2016; Dias and Becker 2016; Igarashi et al. 2016]. For instance, the work of [Siddiqua et al. 2019a] used SemEval stance dataset which consists of tweets and targets entities to build a textual entailment model that predicts a target-specific stance. However, some work showed that using additional signals from the user's network can further improve the stance detection performance as will be discussed later in section 6.2 [Aldayel and Magdy 2019b; Lynn et al. 2019].

4.2 Predicting the Unexpressed View

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

In stance predication, most of studies tend to predict users' stances on two levels, micro and macro. At the micro level, stance prediction means the estimation of the individual user's standpoint towards a target/event in advance (pre-event), indicating the likelihood that a user will be (in favour or against) the target event. This methodology is similar to recommending new items based on a user's previous history of purchases. The macro level infers public opinion toward an event, and the research tend to address this level of prediction as an aggregation of micro predictions[Qiu et al. 2015]. Similarly, [Rui et al. 2017] designed a micro-level stance detection model by using users' previous posts along with logistic regression to predict stances for new users who are not included in the dataset. Their model produces word distributions for the against/support stances on the topic by using rules derived from user's interactions. Another study [Gu et al. 2014] used matrix factorization to predict the micro-level stance based on voting patterns of the users and the topic model. The study of [Gu et al. 2016] predict the ideology using heterogeneous links between the users.

For instance [Darwish et al. 2017a], investigated the best social media features to predict user attitudes toward new events[Darwish et al. 2017a]. This work differs from SemEval2016 task 6 (target-specific) in that the stance of the user is predicted by using each user historical collection of tweets instead of using a target-relevant single tweet. Instead of focusing on the tweet content (textual features), [Darwish et al. 2017a] used similarity between users and inferred latent group beliefs as features in their model, and they used a bipartite graph along with graph reinforcement. [Lahoti et al. 2018] followed the same concept used by [Darwish et al. 2017a] in defining network features, and they used a joint matrix in which users and source contents are represented with the same latent space to balance message distribution in social media. Different method has been proposed by [Gottipati et al. 2013] to predict individual's unexpressed stance by implementing a collaborative filtering approach and constructing a user ideology matrix representing a user's ideological stance for an event.

In the macro-level stance prediction, few studies have addressed stance prediction. Most of these studies were customized to analyze specific case studies, such as Islamophobia [Magdy et al. 2016; Darwish et al. 2017c]. 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,[Qiu et al. 2015] have proposed a micro-level stance prediction model using user behaviour toward new events in which they have not participated. In their study, they referred to this kind of user as a cold-start user, which is a well-known term used commonly in recommendation systems. In addition, they introduced a macro-level stance prediction model as an aggregation of each user stance (micro-level).

5. STANCE DETECTION ACCORDING TO TARGET

As discussed earlier, stance detection task needs the presence of a defined target to detect the stance towards it. In the literature, stance detection can be be categorised according to the type of the target of evaluation

into three categorizes, namely: single defined target, multi-related-targets, and claim-based. In the following, a further explanation to 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 [Mohammad et al. 2016; Kareem et al. 2019; Augenstein et al. 2016a; Liu et al. 2016; Dias and Becker 2016; Igarashi et al. 2016]. In this type of stance detection the text (T) or the user (U) will be the main input to predict the stance toward a specific predefined single target (G).

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

The above equation means that a separate stance classification model needs to be built for each target (G) separately, unlike sentiment where 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 is covering multiple topics, most of the published work on this dataset trained a separate model for each topic (target) separately [Mohammad et al. 2016; Aldayel and Magdy 2019b; Siddiqua et al. 2018]. However, there have been recent trials to apply transfer learning model between different targets as will be discussed in more details in section 7.2.

5.2 Multi-related-targets stance detection

In multi-target stance detection the goal is to jointly learn the social media user orientation towards two or more targets for a single topic[Sobhani et al. 2017]. The main assumption behind this kind of stance detection is that when a person gives his stance for one target this provides an information about his stance towards the other related targets.

$$Stance(T|U, G_n) = \{(Favour G_1, Against G_{n+1}), (Favour of G_{n+1}, Against G_1)\} \qquad (5)$$

For instance, a tweet could express a stance towards multiple US presidential candidates at the same time, so when users express their in-favor stance to Hillary Clinton that's implies an against stance toward Trump [Darwish et al. 2017b; Lai et al. 2016]. In [Sobhani et al. 2017] work, they present the first multi targets stance detection data-set which contains 4,455 tweets related to the 2016 US elections. In order to detect the subjectivity towards two targets, [Sobhani et al. 2017] used an attention-based bidirectional recurrent neural network (RNN) to jointly learn the stance towards two related targets. The notion of multi-targets 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 [Lai et al. 2016]. Following the same approach, the study [Darwish et al. 2017b] constructed a dataset with 3,450 tweets annotated with stance labels for the two US 2016 election candidates (Trump and Hillary) at the same time. Furthermore, the work of [Wei et al. 2018] 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, also known as open-domain stance detection, the target of the analysis is not an explicit entity as the ones discussed earlier; however, it is a claim in a piece of news. The first stage to detect 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 rumour's post or an article headline based. In the fake news task, the claim tends to be the article headline and the text is the article body. On
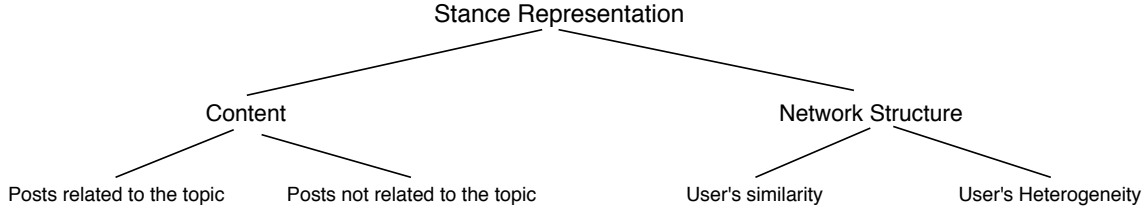
Stance Representation

Content

Network Structure

Posts related to the topic

Posts not related to the topic

User's similarity

User's Heterogeneity

Fig. 4. Stance representation in social media

the other hand for the rumour's veracity task, the main input to be evaluated is the rumour's post and the text is the replies to the rumours. The prediction label sets tend to take the form of confirming the claim or denying it.

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

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[Hamidian and Diab 2015; Aker et al. 2017a; Zubiaga et al. 2018; Gorrell et al. 2019]. In study by [Hamidian and Diab 2015], they used a supervised learning model along with new set of features called "pragmatic features" which contains: named entity, event, sentiment and emoticons. Interestingly, [Aker et al. 2017a] concluded that problem-specific features engineering outperformed the other state-of-the-art systems in rumour identification tasks. Their model, which used a random forest outperformed the advanced LSTM-based sequential model in SemEval 2017 task 8 proposed by [Kochkina et al. 2017]. Within the same line, the study of [Aker et al. 2017a] used the same feature engineering approach proposed by [Elfardy and Diab 2016], in which lexical and semantic features used to help with identifying the stance of a Twitter user. More recently, for the conversation-based tasks, the study by [Zubiaga et al. 2018] showed that using LSTM can outperform other sequential classifiers and feature-based models.

In another study by [Li et al. 2019] they used multi-task learning with stance detection layer to classify the stance of a given tweet as supporting/ 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. Those signals can be categorised into two main categories: 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 4 summarises the main features used in each category to model stance on social media. In the following we describe each of these two categories of signals and how they have been used for detecting stance. Finally, we present a comparison to their effectiveness for detecting stance on multiple datasets.

6.1 Content Features

This section discusses the stance representation focusing on the textual features derived from the user's content online. As illustrated in Figure 4, the content can be collected based on the topic of the analysis. In this kind of feature representation, the data is collected based on range of keywords reflecting the topic. Another type of representation concerns with collecting content that has no direct relation to the topic. The main objective in this case is to model the stance based on the user's behavioural data rather than topic level stance detection.

Furthermore, the content-features can be categorized into two general types: the linguistics features and user's vocabulary. The first type of features concerns with the text linguistic features that helps in inferring the stance. The other type concerns with modeling user's stance based on the user's choice of vocabulary.

6.1.1 *Linguistics Features.* The majority of work on stance detection focused on utilizing the linguistic elements that capture the social media user's stance. This massive dependency on linguistic cues is due to defining the stance detection as textual entailment task, where the task is to detect stance in a given piece of text (e.g. tweet [Mohammad et al. 2016]). In the literature, the stance detection work that concerns with using textual cues to detect stances includes: textual features, sentiment polarity, and latent semantics.

For instance, using textual features such as n-gram modeling of the text has been heavily investigated in the literature [Anand et al. 2011; Mohammad et al. 2017; Sobhani 2017]. Using the n-gram modeling of the text shows promising result. In the SemEval 2016 stance detection task, using word and char n-gram modeling managed to get the best f-score among the other participating systems [Mohammad et al. 2016]. Another textual feature that have been used to infer the stance is the sentiment polarity of the post [Mohammad 2017; Elfardy and Diab 2016; Elfardy 2017]. In general, the use of sentiment as feature was not sufficient in predicting the stance as studies concluded [Mohammad et al. 2017; Elfardy 2017]. Another kind of features is the latent semantics features which aims to reduce the dimension of a given input such as mapping the sentences to predefined set of topics (topic modeling) [Elfardy and Diab 2016]. Topic modeling has been applied by different studies [Elfardy and Diab 2016]. For instance [Elfardy and Diab 2016] used Textual Weighted Textual Matrix Factorization (WTMF) and frame-semantic parser to model a given tweet has been used as feature by [Patra et al. 2016] to map unigrams to topic sphere. Another work used a simple bag of topic to model targets words.

6.1.2 *User's Vocabulary Choice.* There is a considerable number of studies that represent the stance based on the user's vocabulary. The hypothesise behind using this kind of user modeling is that the individuals with same stance tend to use the same vocabulary choice to express their point of view [Darwish et al. 2019].

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 to the individual attitude and characteristics [Klebanov et al. 2010]. For instance, people with against stance to abortion tend to use vocabulary such as pro-life to express their opposing stance. The work of [Klebanov et al. 2010] built a user stance detection model based on users vocabulary choice using a generative model and discriminative model using Naive Bayes and SVM classifiers respectively. Another work, [Rui et al. 2017] proposed a word generative model based on the users interaction to build a set of words representation for each stance toward a topic. In addition, the work of [Benton and Dredze 2018] follows the same direction where the users' interactions help in shaping the set of vocabulary used to identify the stance. Another study by [Zhu et al. 2019] 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 behaviour in the social media to further improve the stance prediction model. The work of [Rui et al. 2017] used a generative model based on the user's keywords choice for each standpoint along with users interactions as regularization to predict the stance. In addition [Li et al. 2018], introduced a users interaction and posts embedding by using an embedding vector for the Pro-stance and another vector for the Con-stance.

6.2 Network Features

Social media provides a special kind of the social data due to the structure of these platforms, where users can be characterised and/or analysed based on their social connections and interactions.

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

The network features that have been used to learn the users representations in social media can be grouped under two categories: users behavioral data [Thonet et al. 2017; Darwish et al. 2017a; Kareem et al. 2019],

and user meta-data attributes [Pennacchiotti and Popescu 2011]. Using users behavioral data to identify the stance is motivated by the notion of homophily, which is based on the social phenomenon "individuals associate with similar ones" [Bessi et al. 2016]. In social media, the similarity between users considered a core property that helps in inferring stances.

The interaction elements have been used to define the similarity between the users. One of theses elements that has been extensively used to infer twitter user's stance is the retweet [Borge-Holthoefer et al. 2015; Darwish et al. 2017c; Weber et al. 2013; Rajadesingan and Liu 2014]. Another element that has been heavily investigated is hashtags, this element has been used in literature to infer similarity between users in order to predict the stance [Darwish et al. 2017a; Dey et al. 2017]. The work of [Dey et al. 2017] used soft cosine similarity to gauge the similarity between the users who post on the same hashtags. The work of [Darwish et al. 2017a] used graph reinforcement to calculate the similarity between the users who post on the same hashtags.

In a recent study by [Aldayel and Magdy 2019b] they defined three types of network features to model the stance on social media. Those network features are: 1) interaction network, 2) preferences network and 3) connection network. The interaction network represent the users direct interactions with other users in sense of retweet, mention and reply. This type of network provides the best performance score of stance detection model in compression with other two networks. The preference network is the network of others users that post or are mentioned in the tweets the user likes. This network allows detecting stance for users who might have limited posting or interaction behaviour 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 II.

Another study by [Fraisier et al. 2018] introduced a multi-layer graph model to represent profiles from different platforms and to extract communities. By doing this the model allows stance to diffuse from the set of known profiles and to predict profiles' stance by using pairwise similarities between users' profiles.

The use of network features has shown its benefit in detecting social media user stance and future attitudes in the aftermath of an event (stance prediction). For instance, in study done by [Darwish et al. 2017c] they used the user's similarity components as features. In their work, the similarity calculated based on the interaction elements of a given tweet. These interaction elements are: mentions, re-tweets, and replies; Website links (URLs) and hashtags used by users in their tweets. Similarly, the work of [Thonet et al. 2017] used the retweet and reply to define the user's social network.

Another line of study, use heterophily to measure the dissimilarity between the users to model the stance[Trabelsi and Zaiane 2018]. For instance the study of [Trabelsi and Zaiane 2018] used 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 reply attacks the previous author parent post. The value of the rebuttal depends on the degree of opposition between the viewpoint of parent post and the parent tweet.

6.3 Comparative analysis of the stance modeling

Table II shows a comparative analysis of the stance detection performance using the two types of modeling, content vs network. In addition, it reports the results of studies that used a combination of both. It can be noticed that using Network features to model the stance outperforms the content modeling of the stance. The Textual modeling has the lowest performance as this type of modeling depends on textual entailment task, while the network features provide the highest performance in comparison with the later. This can be as a reason of the fact that using network features puts into consideration the bigger picture of modeling the user's attitude using the users' interactions and similarity. Using Network features overcomes the limitations poses by textual entailment modeling of stance. As demonstrated by the [Aldayel and Magdy 2019b]'s study, where they used stance SemEval dataset to model stance on social media using the model with best reporting results on stance dataset to compare the performance of stance when using content and network features. Their study

Table II. 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	[Magdy et al. 2016]
Islam Datase	**84.0**	76.0	-	SVM	[Darwish et al. 2017a]
Islands Datase	**79.0**	71.0	-	SVM	
Gun control, abortion and Obama care	**81.9**	60.6	82.0	Non-negative matrix factorization	[Lahoti et al. 2018]
SemEval stance dataset	71.6	69.8	**72.5**	SVM	[Aldayel and Magdy 2019b]
SemEval stance dataset	61.8	62.8	**65.9**	SVM	[Lynn et al. 2019]
LIAC dataset	**67.8**	54.1	-	Generative Model	[Rui et al. 2017]

provides a thorough comparative analysis of the stance overall performance when using textual and network features. The use of network features provides the best performance in compression with previous studies, where they used the textual modeling along with transfer learning methods as reported by [Mohammad et al. 2016]. Furthermore, in study by [Lahoti et al. 2018] they show that the network features outperform the textual modeling of stance when using non-negative matrix factorization for stance detection. The study of [Darwish et al. 2017a] confirms the same conclusion where the use of network features outperforms the textual representation using two datasets (Islam and Island data-sets). As the study by [Magdy et al. 2016] shows that network modeling of stance outperforms the use of textual modeling for the expressed and non-expressed stance on social media.

7. STANCE DETECTION ALGORITHMS

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

7.1 Supervised learning

This is the basic and most common approach for most of the work on stance detection [Zhang et al. 2019; Lai et al. 2020b; Walker et al. 2012a; Krejzl and Steinberger 2016; Igarashi et al. 2016; Gottipati et al. 2013]. With 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' [Mohammad et al. 2016] for a set of five different topics. Many studies have been published on this dataset used different supervised ML algorithms such as classical algorithms, such as (Naive Bayes) NB, SVM, and decision trees; and deep learning algorithms, such as RNNs and LSTMs, to detect the stance on this labeled dataset.

For example, the work of [Mohammad et al. 2017] used a SVM model with linguistic and sentiment features to predict the stance. Their study showed that the use of content features only (n-gram) provides an F1 score equal to (69.0%) which surpassed the use of sentiment as feature with about 66.8% F1 score. Another work by [Walker et al. 2012a] used an labeled data toward 14 topics to train a NB model to predict the stance.The work of [Elfardy and Diab 2016] used SVM model and lexical and semantic features to classify the stance in SemEval stance which has F1 score equal to 63.6%. Another study by [Wojatzki and Zesch 2016] used stacked classifier and syntactic features to classify the stance in SemEval stance dataset. Their model show a minuscule improvement on the overall stance detection performance with about (62%) F1 score. The work of [Siddiqua et al. 2019a] proposed neural ensemble model using bidirectional LSTM on SemEval stance dataset along with fast-text embedding layer. Their model shows an improvement on the overall F score to 72.1%. A recent work by [Li and Caragea 2019a] used bidirectional gated recurrent unit to build a multitask learning model that leverages the sentiment of a tweet to detect the stance. This model

shows an improvement of stance detection model overall F score of Semeval dataset to reach a score equal to 72.3%. A detailed comparison of each study performance on Semeval dataset is reported later in section 7.4.

7.2   Transfer learning

To address the scarcity of the labeled data for each target in stance detection task, some studies in this field attempted to incorporate transfer learning techniques to enrich the representation of target in the dataset and enhance the overall performance of stance detection [Dias and Becker 2016; Augenstein et al. 2016b]. In transfer learning, the knowledge that an algorithm has learned from one task is applied to a separate task, where 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 SemEval stance (Task B) [Mohammad et al. 2016]. This dataset contains 78,000 unlabeled tweets related to Trump which provides a good source for research to explore with various transfer-learning algorithms. For example in the work of [Augenstein et al. 2016b], they used the SemEval stance (task B) to train a bag-of-word auto-encoder along with Hillary labeled data to detect the stance. Another work by [Dias and Becker 2016] used Hillary Clinton labeled data along with the Trump unlabeled data to develop rules-based algorithm to help in detecting the stance for task B of SemEval stance task. In the work of [Wei et al. 2016], they used CNN model with Google news embedding on the SemEval stance task B. For subtask B, their model trained on 2 classes dataset and then by using a modified softmax layer they performed 3 classes classification with voting scheme. With these configurations their model ranked on the top three models with overall F-measure 56.28 for task B. Moreover, the study by [Zarrella and Marsh 2016] implemented distant supervision using SemEval (Task B) to learn features from two unlabelled datasets by using hashtag prediction. Another study by [Wei and Mao 2019] used topic transfer knowledge to leverage the shared topics between two targets and predict the stance for unlabeled data. Recently, many studies Incorporated the concept of transfer learning using new dataset other than SemEval stance (task B). A recent study by [Hanawa et al. 2019] used Wikipedia articles to build a knowledge extraction for each topic on dataset that contains seven topics from Wikipedia. Another work by [Ferreira and Vlachos 2019] used three datasets, namely (Blogs, US Elections and Moral Foundations Twitter) and design algorithm for incorporates label dependencies between them to train stance detection model.

7.3   Unsupervised Learning

Recently the attention has been devoted toward building unsupervised stance detection model. In this kind of studies they mostly used clustering techniques with focus on the user and topic representation on the social media platform [Darwish et al. 2019; Joshi et al. 2016; Trabelsi and Zaiane 2018]. The work of [Trabelsi and Zaiane 2018] proposed unsupervised model using clustering model at the author and topic levels. In their study they used six topics collected from two online debate forums 4Forums and Create-Debate. Their clustering model leverages the content and interaction network of the users (retweets and replies).

A recent study by [Darwish et al. 2019] used clustering technique to create an initial set of stance partition for the annotation. In their work they used unlabeled tweets related to three topics: Kavanaugh, Trump and Erdogan. Their findings show that using retweet as feature provides the best performance score when implementing clustering algorithm (DBSCAN) which surpass the supervised method when using fast-text and SVM model. Their finding is considered a large motivation for using 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 is used in this section to compare different ML

Table III. Comparing the Stance detection models on SemEval stance dataset.

Algorithm	Model	Features	F1	Study
Supervised learning	SVM	NW(Mentions+urls)	71.56	[Aldayel and Magdy 2019b]
	SVM	NW(Mentions+urls)+content	**72.5**	[Aldayel and Magdy 2019b]
	RNN	NW (followee)	61.8	[Lynn et al. 2019]
	RNN	NW (followee)+Content	65.9	[Lynn et al. 2019]
	Nested LSTMs	Content	72.1	[Siddiqua et al. 2019a]
	bidirectional gated recurrent unit	Content	72.33	[Li and Caragea 2019a]
	Hierarchical Attention NN	Content	69.8	[Sun et al. 2018]
	SVM	Content	70.0	[Siddiqua et al. 2018]
Transfer learning	BiGRU	Noisy labeling + topic modeling	60.8	[Wei et al. 2019]
	BiLSTM	Sentiment lexicon	**65.3**	[Li and Caragea 2019b]
	Deep averaging network	words embedding	35.2	[Ebner et al. 2019]
	Deep averaging network	Glove	30.2	[Ebner et al. 2019]

approaches discussed in the above sections. Table III shows the best performing models based on the type of algorithm. As can be expected, the transfer learning models have the lower performance score compared to supervised learning models. However, the performance by [Li and Caragea 2019b] shows promising performance with an F-score of 0.653. However, some other trials, such as the work of [Ebner et al. 2019], which uses the Deep averaging network (DAN) with the Glove word embedding, achieves an average F1 score is around 0.3%, which is close to random performance.

As can be noticed clearly from Table III, the supervised algorithms are more effective for stance detection, where they achieve higher F-scores on the SemEval dataset than transfer-learning approaches. This can be seen in the models that 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 performance when incorporate network features along with content, where using both with a simple linear SVM is more effective than using word-embedding with RNNs and LSTMs.

There is no unsupervised algorithms that have been applied to the same SemEval dataset till date. The only reported results as discussed earlier are the ones by [Kareem et al. 2019], where the model applied on different dataset (Trump, Kavanaugh and Erdogan datasets). The unsupervised model of this study used the network features and outperformed the model with the use of fast text word embedding with clustering purity equal to 98%. Overall, it can be noticed that the best performing models are the one framed in user and social context with use of user network features in the stance detection task.

8. STANCE DETECTION APPLICATIONS

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

8.1 Analytical Studies

Using stance detection has proven its benefit as social sensing technique to measure public support related to social, religion and political topics.

As examples of using stance detection for analysing political topics are the studies on analysing public reaction towards Brexit using twitter data [Lai et al. 2020a; Simaki et al. 2017b; Grčar et al. 2017; Simaki et al. 2017a]. Furthermore, most of the stance detection studies used US 2016 elections data to analyze the stance toward two candidates Hillary Clinton and Trump [Lai et al. 2016; Darwish et al. 2017b]. The work of [Lai et al. 2018] analysed the stance towards the political debates in Twitter about the Italian constitutional referendum held on December 2016. Another work by [Taulĳ et al. 2017] studied the public

stance toward Catalan Independence. A more recent study by [Lai et al. 2020a] used a multilingual dataset to study political debate on social media. They collected entities and events related to politics in five languages: English, French, Italian,Spanish and Catalan.

Another line of studies used stance detection to analyse the public viewpoint towards social aspects. The public opinion towards immigration has been recently studied [Gualda et al. 2016; Bartlett and Norrie 2015]. The work of [Gualda et al. 2016] studied the attitude towards the refugees using twitter. They collected tweets using the keyword "refugees" in different languages. They used the sentiment of the discourse to analyze the public stance towards the refugees. Another work by [Bartlett and Norrie 2015] studied the immigration in the United Kingdom by annotating the tweet related to immigration with sentiment polarity. They used the negative polarity as an indication of the against stance and the positive sentiment as an indication of the support viewpoint. They found that about 5% of the users were against the immigration, while 23% were supporting the immigration. It worth mentioning that the last two mentioned studies are seen sub-optimal in measuring stance due to relying heavily on sentiment of the text, which has been demonstrated in section 3.2 and previous studies [Aldayel and Magdy 2019a; Sobhani 2017] to be sub-optimal. A recent study by [Xi et al. 2020] analyzed the users posted images in Facebook to understand the ideological leaning and how political ideology is conveyed through images. In their work they used the scores generated by 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. While for conservative members tend to use images that has objects related to state and economic power such as âĂIJcourtâĂİ.

On the other hand, stance detection has been used to analyze the attitude related to disruptive events [Demszky et al. 2019; Darwish et al. 2017c]. For instance, the work of [Darwish et al. 2017c] studied the public attitude towards Muslims after Paris terrorist attacks in 2015. In their work they collected tweets that mentions keywords related to Islam. After that they analyzed users' interactions and used the historical tweets to predict their attitudes towards Muslims. The work of [Demszky et al. 2019] analyzed the attitudes about 21 mass shooting event. They first derived list of mass-shootings events between 2015 and 2018 from the Gun-Violence Archive. For each event they defined list of key words to collect tweets related to these events. Then to study the polarization of opinion toward these events they estimated the party for each user based on the political accounts they follow.

8.2 Applications related to social-based phenomena

Stance detection has been used to solve algorithmic issues on the social media platforms. These issues are a reflection of social phenomena on these platforms. The most common phenomena that affected various social media platforms are echo-chambers and homophily [Fuchs 2017]. Previous studies have concluded that social media are polarized in nature which boost homophily behavior [Barberá et al. 2015; Bessi et al. 2016; Quattrociocchi et al. 2016; Darwish et al. 2017a; Garimella et al. 2018]. Homophily is the social phenomenon that concerns with people tendency to connect with "like minded friends". Echo-chamber is the cascade of a certain information among group of people. This social behavior has been magnified in social media structure where certain beliefs are amplified within close circle of communication. Consequently, people are exposed to content with consent to the same opinion that they hold. As result, this reinforce social media users' views biases and blinds the users from other sides of information. Therefore, stance detection has been used to help in measuring and alleviating the problems resulted from polarization on social media. For example, the study by [Garimella et al. 2017] uses stance as the main factor to expose user to contradicting views. Also stance detection was used to identify and measure the controversy level on social media platform [Al-Ayyoub et al. 2018; Dori-Hacohen and Allan 2015; Jang and Allan 2018]. In study by [Jang and Allan 2018] they used stance summarization technique to re-rank controversial topics. Their method collects arguments toward five topics and summarize the overall stance with respect to 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 the main source 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 rumour detection. Recently, a surge of attention has been devoted to classifying stances to help in setting the first step towards solving veracity checking issue [Gorrell et al. 2019; Derczynski et al. 2017; Allcott and Gentzkow 2017].

A rumour is defined as an uncertain piece of information that lacks a secure standard of evidence for determining whether it is true or false [Levinson and Ember 1996]. It has been shown that false rumours have obvious effect across various fields. For example, in the healthcare sphere, false rumours on Web pose a major public health concern. In 2014, rumours about the Ebola epidemic in West Africa emerged on social media, which made it more difficult for healthcare workers to combat the outbreak [Shultz et al. 2015]. Rumours also affect modern-day journalism, which depends on social media as main platform for breaking news. Additionally, there have been numerous occasions where false rumours have yielded severe stock market consequences [Schmidt 2015]. To this end, many initiatives, such as Emergent dataset, PHEME dataset and RumourEval SemEval-2017 and 2019 datasets [Ferreira and Vlachos 2016a; Derczynski et al. 2017; Kochkina et al. 2017; Gorrell et al. 2019]. These initiatives have been conducted to encourage the development of tools that can help with verifying 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 towards the news is measured to detect if these comments are confirming or denying the news, which is used later to detect if the news is a rumour or authentic.

Unlike rumours, fake news aims to create a misleading information and it always false [Zubiaga et al. 2017]. Data veracity refers to the truthfulness, trustworthiness and accuracy of the content. Social media posts have massive amounts of user-generated contents 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 towards news organizations. This is due to the fact that understanding other organizations stances toward a news topic helps in inferring the truthfulness of news article. Fake News Challenge initiative (FNC-1) adopted this approach and proposed a stance detection task to estimate the stance of articles to a given headline (claim). The best performing system at the Fake News Challenge was proposed by [Baird et al. 2017]. In their study they used gradient-boosted decision trees and convolutional neural network (CNN) along with many textual features. Another recent study by [Mohtarami et al. 2018] achieved relatively similar results to the best system with feature-light memory network model enhanced with (LSTM and CNN). A more recent study by [Shu et al. 2019] used three features extracted from user's interactions, news author and contents of news article to better detect the fake-news. In the study by [Ghanem et al. 2018], they used a combination of lexical, word embedding and n-gram to detect stance in two datasets (FNC-1) and Emergent. A recent study by [Borges et al. 2019], proposed a new text representation that incorporates the first two sentences of the article along with news headline and the entire article to train a bi-directional RNN model using (FNC-1) dataset.

9. STANCE DETECTION RESOURCES

This section lists the current available resources for stance detection tasks. Table IV 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 [Mohammad et al. 2016] [Ferreira and Vlachos 2016b] [Derczynski et al. 2017], Wikipedia[Bar-Haim et al. 2017], International debate education association website [Bar-Haim et al. 2017] and various news sites such as: Snopes.com [Ferreira and Vlachos 2016b].

Target-specific datasets: Table V 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 [Mohammad et al. 2016] 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 [Gautam et al. 2019] 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 [Küçük and Can 2019] introduce a dataset with stance and name entity labels. The dataset contains 1065 tweets annotated with stance labels towards two entities Galatasaray and Fenerbahc. A recent study by [Conforti et al. 2020] provides a dataset for stance detection that contains around 51K tweets covering financial domain.

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 [Qazvinian et al. 2011] 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 [Ferreira and Vlachos 2016a] 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 [Gorrell et al. 2019], 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[Darwish et al. 2017b] and Multi-targets dataset [Sobhani et al. 2017]. 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 [Sobhani et al. 2017; Sobhani et al. 2019] 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 IV. These datasets are designed to predict the stance before the event time. There are two datasets constructed by [Darwish et al. 2017a], where they used twitter as a source for stance prediction data.

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

Data-set	Type	Data Source	Stance definition (annotation)	Size
[Qiu et al. 2015]	micro-level + macro-level	Debate forums	The probability of user choosing a stance on an issue	users 4,994, issues 1,727arguments 39,549 (average 23 per issue)
[Rui et al. 2017]	(micro-level)	Online-fourms	Stance(issue,posts)=pro,con	news articles from CNN and 4fourms
[Darwish et al. 2017c]	(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 [Qazvinian et al. 2011]	Claim-based	Rumours resolutions	Twitter	Obama, Airfrance, cellphone, Michelle, palin	Stance(users claim)={Confirm, Deny, Doubtful} tweets,rumour's claim)	10,417 tweets
SemEval-Stance Detection Task A[Mohammad et al. 2016]	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 Testing: 1,249.
SemEval-Stance Detection Task B [Mohammad et al. 2016]	Target specific	General	Twitter	Donald Trump	Stance(target, tweet)={Favor, Against, Neither}	707 labeled tweets 78,000 unlabeled
Emergent dataset [Ferreira and Vlachos 2016a]	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 [Derczynski et al. 2017]	Claim-based	The Veracity of a rumour.	Twitter	charliehebdo, ebola-essien, ferguson, germanwings-crash, ottawa shooting, prince-toronto, putinmissing, sydneysiege	Stance(originating rumourous, tweets reply)={deny, support, commenting}	training: 4519 threads, Testing: 1049 threads
SemEval Rumour stance classification Subtask B - Veracity prediction [Derczynski et al. 2017]	Claim-based	The Veracity of a rumour.	Twitter	charliehebdo, ebola-essien, ferguson, germanwings-crash, ottawa shooting, prince-toronto, putinmissing, sydneysiege	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 [Sobhani et al. 2017]	Multi-related-targets	General	Twitter	Hillary Clinton-Sanders, Clinton-Trump and Cruz-Trump	Stance (multi-targets, tweet)={ Target A Favor—Against Target B}	4,455 tweets
Trump vs. Hillary dataset[Darwish et al. 2017b]	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 [Bar-Haim et al. 2017]	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

Table V. : 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 [Darwish et al. 2017b] or social phenomena as Islamophobic [Darwish et al. 2017c]. 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 [Himelboim et al. 2013].

While there has been a large interest from the NLP community on developing effective approaches for stance detection [Küçük and Can 2020] 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 [Mohammad et al. 2016] 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 [Aldayel and Magdy 2019b]. 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

Ahmet Aker, Leon Derczynski, and Kalina Bontcheva. 2017a. Simple Open Stance Classification for Rumour Analysis. (2017), 31–39.

Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, Anna Kolliakou, Rob Procter, and Maria Liakata. 2017b. Stance Classification in Out-of-Domain Rumours: A Case Study Around Mental Health Disorders. In *International Conference on Social Informatics*. Springer, 53–64.

Mahmoud Al-Ayyoub, Abdullateef RababâĂŹah, Yaser Jararweh, Mohammed N Al-Kabi, and Brij B Gupta. 2018. Studying the controversy in online crowdsâĂŹ interactions. *Applied Soft Computing* 66 (2018), 557–563.

Abeer Aldayel and Walid Magdy. 2019a. Assessing Sentiment of the Expressed Stance on Social Media. In *Socinfo 2019*.

Abeer Aldayel and Walid Magdy. 2019b. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. In *CSCW 2019*.

Hunt Allcott and Matthew Gentzkow. 2017. *Social Media and Fake News in the 2016 Election.* Technical Report. National Bureau of Economic Research.

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. 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, 1–9.

Isabelle Augenstein, Tim RocktÃďschel, Andreas Vlachos, and Kalina Bontcheva. 2016a. Stance detection with bidirectional conditional encoding. *EMNLP 2016* (2016).

Isabelle Augenstein, Andreas Vlachos, and Kalina Bontcheva. 2016b. 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)*. 389–393.

Sean Baird, Doug Sibley, and Yuxi Pan. 2017. Talos Targets Disinformation with Fake News Challenge Victory. [online]. *talosintelligence* (2017). https://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html

Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. 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. 251–261.

Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26, 10 (2015), 1531–1542.

Jamie Bartlett and Richard Norrie. 2015. Immigration on Twitter: understanding public attitudes online. *Demos* (2015).

Reem Bassiouney. 2015. Stance-Taking. *The International Encyclopedia of Language and Social Interaction* (2015).

Farah Benamara, Maite Taboada, and Yannick Mathieu. 2017. Evaluative language beyond bags of words: Linguistic insights and computational applications. *Computational Linguistics* (2017).

Adrian Benton and Mark Dredze. 2018. 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*. 184–194.

Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Homophily and polarization in the age of misinformation. *The European Physical Journal Special Topics* 225, 10 (2016), 2047–2059.

Douglas Biber and Edward Finegan. 1988. Adverbial stance types in English. *Discourse processes* 11, 1 (1988), 1–34.

Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. 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, 700–711.

Luis Borges, Bruno Martins, and Pavel Calado. 2019. 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 (July 2019). DOI:http://dx.doi.org/10.1145/3287763

Dushyant Singh Chauhan, Rohan Kumar, and Asif Ekbal. 2019. Attention Based Shared Representation for Multi-task Stance Detection and Sentiment Analysis. In *International Conference on Neural Information Processing*. Springer, 661–669.

Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. Will-They-Won't-They: A Very Large Dataset for Stance Detection on Twitter. *ACL* (2020).

Harald Cramér. 2016. *Mathematical methods of statistics (PMS-9)*. Vol. 9. Princeton university press.

Kareem Darwish, Walid Magdy, Afshin Rahimi, Timothy Baldwin, and Norah Abokhodair. 2017c. Predicting Online Islamophopic Behavior after# ParisAttacks. *The Journal of Web Science* 3, 1 (2017).

Kareem Darwish, Walid Magdy, and Tahar Zanouda. 2017a. Improved Stance Prediction in a User Similarity Feature Space. In *ASONAM'17*.

Kareem Darwish, Walid Magdy, and Tahar Zanouda. 2017b. Trump vs. Hillary: What Went Viral During the 2016 US Presidential Election. In *International Conference on Social Informatics*. Springer, 143–161.

Kareem Darwish, Peter Stefanov, Michaël J Aupetit, and Preslav Nakov. 2019. Unsupervised User Stance Detection on Twitter. *arXiv preprint arXiv:1904.02000* (2019).

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings. *arXiv preprint arXiv:1904.01596* (2019).

Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. *arXiv preprint arXiv:1704.05972* (2017).

Leon Derczynski, Kalina Bontcheva, Michal Lukasik, Thierry Declerck, Arno Scharl, Georgi Georgiev, Petya Osenova, Toms Pariente Lobo, Anna Kolliakou, Robert Stewart, and others. 2015. PHEME: Computing VeracityâĂŤthe Fourth Challenge of Big Social Data. In *Proceedings of the Extended Semantic Web Conference EU Project Networking session (ESCW-PN)*.

Kuntal Dey, Ritvik Shrivastava, Saroj Kaushik, and Vaibhav Mathur. 2017. Assessing the Effects of Social Familiarity and Stance Similarity in Interaction Dynamics. In *International Workshop on Complex Networks and their Applications*. Springer, 843–855.

Marcelo Dias and Karin Becker. 2016. 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.

Shiri Dori-Hacohen and James Allan. 2015. Automated controversy detection on the web. In *European Conference on Information Retrieval*. Springer, 423–434.

John W Du Bois. 2007. The stance triangle. *Stancetaking in discourse: Subjectivity, evaluation, interaction* 164 (2007), 139–182.

Seth Ebner, Felicity Wang, and Benjamin Van Durme. 2019. 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)*. 40–46.

Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2016. A Joint Sentiment-Target-Stance Model for Stance Classification in Tweets.. In *COLING*. 2656–2665.

Hebatallah Elfardy. 2017. *Perspective Identification in Informal Text*. Columbia University.

Heba Elfardy and Mona Diab. 2016. 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)*. 434–439.

William Ferreira and Andreas Vlachos. 2016a. 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.

William Ferreira and Andreas Vlachos. 2016b. 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*. 1163–1168.

William Ferreira and Andreas Vlachos. 2019. 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)*. 6351–6355.

Ophélie Fraisier, Guillaume Cabanac, Yoann Pitarch, Romaric Besançon, and Mohand Boughanem. 2018. Stance Classification through Proximity-based Community Detection. (2018).

Christian Fuchs. 2017. *Social media: A critical introduction*. Sage.

Kiran Garimella and others. 2018. Polarization on Social Media. (2018).

Kiran Garimella, Gianmarco De Francisc iMorales, Aristides Gionis, and Michael Mathioudakis. 2017. 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, 201–205.

Akash Gautam, Puneet Mathur, Rakesh Gosangi, Debanjan Mahata, Ramit Sawhney, and Rajiv Ratn Shah. 2019. # MeTooMA: Multi-Aspect Annotations of Tweets Related to the MeToo Movement. *arXiv preprint arXiv:1912.06927* (2019).

Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2018. 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, 66–71. DOI:http://dx.doi.org/10.18653/v1/W18-5510

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. 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, 845–854. `DOI:`http://dx.doi.org/10.18653/v1/S19-2147

Swapna Gottipati, Minghui Qiu, Liu Yang, Feida Zhu, and Jing Jiang. 2013. Predicting userâĂŹs political party using ideological stances. In *International Conference on Social Informatics*. Springer, 177–191.

Miha Grčar, Darko Cherepnalkoski, Igor Mozetič, and Petra Kralj Novak. 2017. Stance and influence of Twitter users regarding the Brexit referendum. *Computational social networks* 4, 1 (2017), 6.

Yupeng Gu, Ting Chen, Yizhou Sun, and Bingyu Wang. 2016. Ideology Detection for Twitter Users with Heterogeneous Types of Links. *arXiv preprint arXiv:1612.08207* (2016).

Yupeng Gu, Yizhou Sun, Ning Jiang, Bingyu Wang, and Ting Chen. 2014. 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, 183–192.

Estrella Gualda, Carolina Rebollo Díaz, and others. 2016. The refugee crisis on Twitter: A diversity of discourses at a European crossroads. (2016).

Sardar Hamidian and Mona Diab. 2015. Rumor detection and classification for twitter data. In *Proceedings of the Fifth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS)*. 71–77.

Kazuaki Hanawa, Akira Sasaki, Naoaki Okazaki, and Kentaro Inui. 2019. Stance Detection Attending External Knowledge from Wikipedia. *Journal of Information Processing* 27 (2019), 499–506.

Itai Himelboim, Stephen McCreery, and Marc Smith. 2013. 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.

Yuki Igarashi, Hiroya Komatsu, Sosuke Kobayashi, Naoaki Okazaki, and Kentaro Inui. 2016. Tohoku at SemEval-2016 Task 6: Feature-based Model versus Convolutional Neural Network for Stance Detection.. In *SemEval@ NAACL-HLT*. 401–407.

Alexandra Jaffe. 2009. *Stance: sociolinguistic perspectives*. OUP USA.

Myungha Jang and James Allan. 2018. Explaining Controversy on Social Media via Stance Summarization. *arXiv preprint arXiv:1806.07942* (2018).

Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman. 2016. 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*. 82–90.

Dan Jurafsky and James H. Martin. 2014. *Speech and language processing*. Vol. 3. Pearson London.

Mostafa Karamibekr and Ali A Ghorbani. 2012. Sentiment analysis of social issues. In *Social Informatics (SocialInformatics), 2012 International Conference on*. IEEE, 215–221.

Darwish Kareem, Stefanov Peter, Michael J. Aupetit, and Nakov Preslav. 2019. Unsupervised User Stance Detection on Twitter. *arXiv preprint arXiv:1904.02000* (2019).

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2010. Vocabulary choice as an indicator of perspective. In *Proceedings of the ACL 2010 Conference Short Papers*. Association for Computational Linguistics, 253–257.

Elena Kochkina, Maria Liakata, and Isabelle Augenstein. 2017. Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM. *arXiv preprint arXiv:1704.07221* (2017).

Peter Krejzl and Josef Steinberger. 2016. UWB at SemEval-2016 Task 6: Stance Detection.. In *SemEval@ NAACL-HLT*. 408–412.

Dilek Küçük and Fazli Can. 2019. A Tweet Dataset Annotated for Named Entity Recognition and Stance Detection. *arXiv preprint arXiv:1901.04787* (2019).

Dilek Küçük and Fazli Can. 2020. Stance Detection: A Survey. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–37.

Preethi Lahoti, Kiran Garimella, and Aristides Gionis. 2018. 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, 351–359.

Mirko Lai, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020a. Multilingual Stance Detection in Social Media Political Debates. *Computer Speech & Language* (2020), 101075.

Mirko Lai, Alessandra Teresa Cignarella, Delia IrazÃž HernÃądez FarÃŋas, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020b. Multilingual Stance Detection in Social Media Political Debates. *Computer Speech & Language* (2020), 101075.

Mirko Lai, Delia Irazú Hernández Farías, Viviana Patti, and Paolo Rosso. 2016. Friends and enemies of clinton and trump: using context for detecting stance in political tweets. In *Mexican International Conference on Artificial Intelligence*. Springer, 155–168.

Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2018. Stance Evolution and Twitter Interactions in an Italian Political Debate. In *International Conference on Applications of Natural Language to Information Systems*. Springer, 15–27.

David Levinson and Melvin Ember. 1996. *Encyclopedia of cultural anthropology*. Vol. 4. Holt New York.

Chang Li, Aldo Porco, and Dan Goldwasser. 2018. Structured Representation Learning for Online Debate Stance Prediction. In *Proceedings of the 27th International Conference on Computational Linguistics*. 3728–3739.

Quanzhi Li, Qiong Zhang, and Luo Si. 2019. 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*. 1173–1179.

Yingjie Li and Cornelia Caragea. 2019a. 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)*. 6300–6306.

Yingjie Li and Cornelia Caragea. 2019b. 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, 6299–6305. DOI:http://dx.doi.org/10.18653/v1/D19-1657

Albert M Liebetrau. 1983. *Measures of association*. Vol. 32. Sage.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. 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, 109–116.

Can Liu, Wen Li, Bradford Demarest, Yue Chen, Sara Couture, Daniel Dakota, Nikita Haduong, Noah Kaufman, Andrew Lamont, and Manan Pancholi. 2016. IUCL at SemEval-2016 Task 6: An Ensemble Model for Stance Detection in Twitter.. In *SemEval@ NAACL-HLT*. 394–400.

Veronica Lynn, Salvatore Giorgi, Niranjan Balasubramanian, and H Andrew Schwartz. 2019. 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*. 18–28.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Thirty-second AAAI conference on artificial intelligence*.

Walid Magdy, Kareem Darwish, Norah Abokhodair, Afshin Rahimi, and Timothy Baldwin. 2016. # isisisnotislam or# deportallmuslims?: Predicting unspoken views. In *Proceedings of the 8th ACM Conference on Web Science*. ACM, 95–106.

David McKendrick and Stephen A Webb. 2014. Taking a political stance in social work. *Critical and Radical Social Work* 2, 3 (2014), 357–369.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. 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, 31–41. DOI:http://dx.doi.org/10.18653/v1/S16-1003

Saif M Mohammad. 2017. Challenges in sentiment analysis. In *A Practical Guide to Sentiment Analysis*. Springer, 61–83.

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)* 17, 3 (2017), 26.

Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. 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. 767–776.

Akiko Murakami and Rudy Raymond. 2010. 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, 869–875.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. 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, 1–18. DOI:http://dx.doi.org/10.18653/v1/S16-1001

Nic Newman. 2011. Mainstream media and the distribution of news in the age of social discovery. *Reuters Institute for the Study of Journalism, University of Oxford* (2011).

Bo Pang, Lillian Lee, and others. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* 2, 1–2 (2008), 1–135.

Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay. 2016. 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)*. 440–444.

Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Dan Garrette, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. SemEval-2020 Sentimix 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.

Marco Pennacchiotti and Ana-Maria Popescu. 2011. Democrats, republicans and starbucks afficionados: user classification in twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 430–438.

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, and others. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *10th International Workshop on Semantic Evaluation (SemEval 2016)*.

Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*. 486–495.

Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1589–1599.

Minghui Qiu, Yanchuan Sim, Noah A. Smith, and Jing Jiang. 2015. 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, 855–863.

Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. 2016. Echo chambers on Facebook. (2016).

Ashwin Rajadesingan and Huan Liu. 2014. Identifying users with opposing opinions in Twitter debates. In *International conference on social computing, behavioral-cultural modeling, and prediction*. Springer, 153–160.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. 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, 502–518. DOI:http://dx.doi.org/10.18653/v1/S17-2088

Dong Rui, Sun Yizhou, Wang Lu, Gu Yupeng, and Zhong Yuan. 2017. Weakly-Guided User Stance Prediction via Joint Modeling of Content and Social Interaction. In *CIKMâĂŹ17*. Singapore.

Daniel Schmidt. 2015. Stock market rumors and credibility. *Document de travail, HEC Paris* (2015).

Kai Shu, Suhang Wang, and Huan Liu. 2019. 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, 312–320.

James M Shultz, Florence Baingana, and Yuval Neria. 2015. The 2014 Ebola outbreak and mental health: current status and recommended response. *Jama* 313, 6 (2015), 567–568.

Umme Aymun Siddiqua, Abu Nowshed Chy, and Masaki Aono. 2018. Stance Detection on Microblog Focusing on Syntactic Tree Representation. In *International Conference on Data Mining and Big Data*. Springer, 478–490.

Umme Aymun Siddiqua, Abu Nowshed Chy, and Masaki Aono. 2019a. 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)*. 1868–1873.

Umme Aymun Siddiqua, Abu Nowshed Chy, and Masaki Aono. 2019b. Tweet Stance Detection Using Multi-Kernel Convolution and Attentive LSTM Variants. *IEICE TRANSACTIONS on Information and Systems* 102, 12 (2019), 2493–2503.

Vasiliki Simaki, Carita Paradis, and Andreas Kerren. 2017a. Stance classification in texts from blogs on the 2016 British referendum. In *International Conference on Speech and Computer*. Springer, 700–709.

Vasiliki Simaki, Carita Paradis, Maria Skeppstedt, Magnus Sahlgren, Kostiantyn Kucher, and Andreas Kerren. 2017b. Annotating speaker stance in discourse: the Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory* (2017).

Prakash Kumar Singh, Shailendra Kumar Singh, and Sanchita Paul. 2018. Sentiment Classification of Social Issues Using Contextual Valence Shifters. *International Journal of Engineering and Technology* (2018).

Parinaz Sobhani. 2017. *Stance Detection and Analysis in Social Media*. Ph.D. Dissertation. Université d'Ottawa/University of Ottawa.

Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2017. A dataset for multi-target stance detection. *EACL 2017* (2017), 551.

Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. 2019. Exploring deep neural networks for multitarget stance detection. *Computational Intelligence* 35, 1 (2019), 82–97.

Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*. 159–169.

Swapna Somasundaran and Janyce Wiebe. 2009. 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, 226–234.

Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou. 2018. Stance detection with hierarchical attention network. In *Proceedings of the 27th International Conference on Computational Linguistics*. 2399–2409.

Mariona Taulé, M Antonia Martí, Francisco M Rangel, Paolo Rosso, Cristina Bosco, Viviana Patti, and others. 2017. 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, 157–177.

Mariona TaulĂ, M. Antonia MartĂŋ, Francisco M. Rangel, Paolo Rosso, Cristina Bosco, and Viviana Patti. 2017. 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, 157–177.

Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat. 2017. 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, 87–96.

Amine Trabelsi and Osmar R Zaiane. 2018. Unsupervised model for topic viewpoint discovery in online debates leveraging author interactions. In *Twelfth International AAAI Conference on Web and Social Media*.

Prashanth Vijayaraghavan, Ivan Sysoev, Soroush Vosoughi, and Deb Roy. 2016. DeepStance at SemEval-2016 Task 6: Detecting Stance in Tweets Using Character and Word-Level CNNs. *Proceedings of SemEval* (2016), 425–431.

Marilyn A Walker, Pranav Anand, Rob Abbott, Jean E Fox Tree, Craig Martell, and Joseph King. 2012a. That is your evidence?: Classifying stance in online political debate. *Decision Support Systems* 53, 4 (2012), 719–729.

Marilyn A. Walker, Jean E. Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012b. A Corpus for Research on Deliberation and Debate.. In *LREC*. 812–817.

Rui Wang, Deyu Zhou, Mingmin Jiang, Jiasheng Si, and Yang Yang. 2019. A Survey on Opinion Mining: from Stance to Product Aspect. *IEEE Access* (2019).

Ingmar Weber, Venkata R Kiran Garimella, and Alaa Batayneh. 2013. 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, 290–297.

Penghui Wei, Junjie Lin, and Wenji Mao. 2018. Multi-Target Stance Detection via a Dynamic Memory-Augmented Network. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 1229–1232.

Penghui Wei and Wenji Mao. 2019. 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, 1173–1176. DOI:http://dx.doi.org/10.1145/3331184.3331367

Penghui Wei, Wenji Mao, and Guandan Chen. 2019. A Topic-Aware Reinforced Model for Weakly Supervised Stance Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7249–7256.

Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengjiao Wang. 2016. pkudblab at SemEval-2016 Task 6: A Specific Convolutional Neural Network System for Effective Stance Detection.. In *SemEval@ NAACL-HLT*. 384–388.

Michael Wojatzki and Torsten Zesch. 2016. 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)*. 428–433.

Nan Xi, Di Ma, Marcus Liou, Zachary C Steinert-Threlkeld, Jason Anastasopoulos, and Jungseock Joo. 2020. Understanding the Political Ideology of Legislators from Social Media Images. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 726–737.

Guido Zarrella and Amy Marsh. 2016. MITRE at semeval-2016 task 6: Transfer learning for stance detection. *arXiv preprint arXiv:1606.03784* (2016).

Qiang Zhang, Shangsong Liang, Aldo Lipani, Zhaochun Ren, and Emine Yilmaz. 2019. 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, 2323–2332. DOI:http://dx.doi.org/10.1145/3308558.3313724

Yiwei Zhou, Alexandra I. Cristea, and Lei Shi. 2017. Connecting targets to tweets: Semantic attention-based model for target-specific stance detection. In *International Conference on Web Information Systems Engineering*. Springer, 18–32.

Lixing Zhu, Yulan He, and Deyu Zhou. 2019. Hierarchical viewpoint discovery from tweets using Bayesian modelling. *Expert Systems with Applications* 116 (2019), 430–438.

Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2017. Detection and resolution of rumours in social media: A survey. *arXiv preprint arXiv:1704.00656* (2017).

Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. 2018. Discourse-aware rumour stance classification in social media using sequential classifiers. *Information Processing & Management* 54, 2 (2018), 273–290.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. 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
[Darwish et al. 2017c]	Content Features (Hashtags, Text); Profile Features (Description, Name, Location); Network Features (Mention, Reply, Retweet)	SVM	Islamophobic dataset (Twitter) [Not available]
[Magdy et al. 2016]	Content Features (Hashtags, Text); Profile Features (Description, Name, Location); Network Features (Mention, Reply, Retweet)	SVM	Islamophobic dataset (Twitter) [Not available]
[Darwish et al. 2017a]	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]
[Lahoti et al. 2018]	A combination of network and content	Non-negative matrix factorization	dataset covered Three controversial topics:gun control,abortion and obamacare (Twitter) [Not available]
[Gottipati et al. 2013]	similarity between users	Probabilistic Matrix Factorization	1000 user profile of Democrats and Republicans (debate.org) [Not available]
[Rui et al. 2017]	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 VI. : Work in stance prediction

Study	Task	Features	ML	Dataset
[Aldayel and Magdy 2019b]	target-specific	NW features	SVM	SemEval-2016 shared task 6 [Available]
[Lynn et al. 2019]	target-specific	NW (followee)	RNN	SemEval-2016 shared task 6 [Available]
[Siddiqua et al. 2019a]	target-specific	Content	Nested LSTMs	SemEval-2016 shared task 6 [Available]
[Sun et al. 2018]	target-specific	Content	Hierarchical Attention NN	SemEval-2016 shared task 6 [Available]
[Siddiqua et al. 2018]	target-specific	Content	SVM Tree Kernel	SemEval-2016 shared task 6 [Available]
[Wei et al. 2019]	target-specific	Content+Sentiment lexicon	BiLSTM	SemEval-2016 shared task 6 [Available]
[Wei et al. 2019]	target-specific	Content+Noisy stance labeling + Topic Modeling	BiGRU	SemEval-2016 shared task 6 [Available]
[Ebner et al. 2019]	target-specific	words embedding	Deep averaging network	SemEval-2016 shared task 6 [Available]

[Liu et al. 2016]	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]
[Dias and Becker 2016]	target-specific	n-gram and sentiment	SVM	SemEval-2016 shared task 6 [Available]
[Dias and Becker 2016]	target-specific	n-gram and sentiment	SVM	SemEval-2016 shared task 6 [Available]
[Igarashi et al. 2016]	target-specific	Reply, BagOfWord, BagOfDependencies, POS tags Sentiment WordNet, Sentiment Word Subject, Target Sentiment and Pointwise Mutual Information	CNN	SemEval-2016 shared task 6 [Available]
[Igarashi et al. 2016]	target-specific	Reply, BagOfWord, BagOfDependencies, POS tags Sentiment WordNet, Sentiment Word Subject, Target Sentiment and Pointwise Mutual Information	CNN	SemEval-2016 shared task 6 [Available]
[Augenstein et al. 2016a]	target-specific	word2vec	Bidirectional LSTMs	SemEval-2016 shared task 6 [Available]
[Krejzl and Steinberger 2016]	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]
[Ebrahimi et al. 2016]	target-specific	n-gram and sentiments	discriminative and generative models	SemEval-2016 shared task 6 [Available]
[Wei et al. 2016]	target-specific	Google news word2vec and hashtags	CNN	SemEval-2016 shared task 6 [Available]
[Zarrella and Marsh 2016]	target-specific	word2vec hash-tags	LSTM	SemEval-2016 shared task 6 [Available]
[Rajadesingan and Liu 2014]	target-specific	unigrams, bigrams and trigrams	Naive Bayes	hotly contested gun reforms debate from April 15th, 2013 to April 18th, 2013. [Available]
[Zhou et al. 2017]	target-specific	word embeddings	Bi-directional GRU-CNN	SemEval-2016 shared task 6 [Available]
[Vijayaraghavan et al. 2016]	target-specific	word embeddings	convolutional neural networks(CNN)	SemEval-2016 shared task 6 [Available]
[Elfardy and Diab 2016]	target-specific	Lexical Features, Latent Semantics, Sentiment, Linguistic Inquiry , Word Count and Frame Semantics features	SVM	SemEval-2016 shared task 6 [Available]

[Lai et al. 2016]	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]
[Sobhani et al. 2017]	multi-target	word vectors	bidirectional recurrent neural network (RNN)	Multi-Target Stance dataset [Available]
[Siddiqua et al. 2019b]	multi-target	content of tweets	Multi-Kernel Convolution and Attentive LSTM	Multi-Target Stance dataset [Available]
[Bar-Haim et al. 2017]	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]
[Aker et al. 2017a]	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]
[Hamidian and Diab 2015]	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	[Qazvinian et al. 2011] [Available]
[Aker et al. 2017a]	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 [Derczynski et al. 2017] and the PHEME dataset [Derczynski et al. 2015] [Available]
[Zubiaga et al. 2018]	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 [Derczynski et al. 2015] and Rmour dataset associated with eight events corresponding to breaking news events [Zubiaga et al. 2016] [Available]

[Kochkina et al. 2017]	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	Rumoureval dataset [Derczynski et al. 2017] [Available]

Table VII. : Work in stance classification