



Capturing stance dynamics in social media: open challenges and research directions

Rabab Alkhalifa^{1,2}  · Arkaitz Zubiaga¹

Received: 29 March 2021 / Accepted: 19 January 2022
© The Author(s) 2022

Abstract

Social media platforms provide a goldmine for mining public opinion on issues of wide societal interest and impact. Opinion mining is a problem that can be operationalised by capturing and aggregating the stance of individual social media posts as supporting, opposing or being neutral towards the issue at hand. While most prior work in stance detection has investigated datasets that cover short periods of time, interest in investigating longitudinal datasets has recently increased. Evolving dynamics in linguistic and behavioural patterns observed in new data require adapting stance detection systems to deal with the changes. In this survey paper, we investigate the intersection between computational linguistics and the temporal evolution of human communication in digital media. We perform a critical review of emerging research considering dynamics, exploring different semantic and pragmatic factors that impact linguistic data in general, and stance in particular. We further discuss current directions in capturing stance dynamics in social media. We discuss the challenges encountered when dealing with stance dynamics, identify open challenges and discuss future directions in three key dimensions: utterance, context and influence.

Keywords Digital media · Human communication · Public opinion mining · Stance detection

✉ Rabab Alkhalifa
raalkhalifa@iau.edu.sa; r.a.a.alkhalifa@qmul.ac.uk

Arkaitz Zubiaga
a.zubiaga@qmul.ac.uk

¹ School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London, E1 4NS, UK <http://www.zubiaga.org/>

² College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O.Box 1982, Dammam 31451, Kingdom of Saudi Arabia

1 Introduction

With the proliferation of social media and blogs that enable anyone to post and share content, professional accounts from news organisations and governments aren't any longer the sole reporters of events of public interest (Kapoor et al., 2018). Posting a tweet or a video, or writing an article that goes viral and reaches millions of individuals is now more accessible to ordinary citizens (Mills, 2012). Where anyone can post their views on social media, the use of social media gains ground as a data source for public opinion mining. This data source provides a goldmine for nowcasting public opinion by aggregating the stances expressed by individual social media posts on a particular issue.

Research in stance detection has recently attracted an increasing interest (Küçük & Can, 2020), with two main directions. One of the directions includes determining the stance of posts as supporting, denying, querying or commenting on a rumour, which is used as a proxy to predict the likely veracity of the rumour in question (Zubiaga et al. 2016, 2018; Hardalov et al. 2021). The other direction, which is the focus of this paper, defines stance detection as a three-way classification task where the stance of each post is one of supporting, opposing or neutral (Augenstein et al., 2016), indicating the viewpoint of a post towards a particular issue. This enables mining public opinion as the aggregate of stances of a large collection of posts.

In using stance detection to mine public opinion, most research has been operationalised by evaluating on temporally constrained datasets. This presents important limitations when one wants to apply the models on temporally distant test datasets, as recent studies demonstrate. Due to the rapidly evolving nature of social media content, as well as the rapid evolution of people's opinions, a model trained on an old dataset may not perform at the same level on new data (Alkhalifa et al., 2021). This review paper discusses the different factors that impact evolving changes in public opinion and their impact on stance detection models, discussing previous work studying this problem.

This paper reviews research on stance detection from an interdisciplinary perspective focusing on the impact of time on model performance. We present current progress in addressing these factors, discuss existing datasets with their potential and limitations for investigating stance dynamics, as well as identify open challenges and future research directions. We focus particularly on the dynamic factors impacting stance, including variations across cultures and regions, but also temporal changes caused by events in the real world leading to changes in public opinion. We set forth directions for future research with the aim of furthering the consideration of dynamics surrounding the stance detection task. Our goal is to draw a framework which links the current trends in stance detection, from an interdisciplinary perspective covering computational challenges bridging with broader linguistics and social science angles.

1.1 Overview

Linguists are interested in understanding human language, which is often dependent on its context (Englebretson, 2007). The ethnographic definition of stance in everyday language may vary from the academic definition of stance given in the literature (Englebretson, 2007). Consequently, the definition of stance can be analysed from different perspectives, while most NLP work tends to focus on one of them. The prevalent definition of stance in NLP research stems from a usage-based perspective defined in the field of linguistics and is described by Englebretson (2007) in which stance is dependent on personal belief, evaluation or attitude. Additionally, stance can be seen as the expression of a viewpoint and it relates to the analysis and interpretation of written or spoken language using lexical, grammatical and phonetic characteristics (Cossette, 1998). For example, everyday words or phrases used by people during working hours or in performing specific tasks can express subjective features (Cossette, 1998) which can be used by NLP researchers in different applications. However, the stance term may appear and be used differently by researchers as it is strongly relevant to one's own interpretation of the concept.

1.2 Computational view of stance detection

Stance, as a message conveying the point of view of the communicator, is the opinion from whom one thing is discovered or believed. As a computational task, stance detection is generally defined as that in which a classifier needs to determine if an input text expresses a supporting, neutral or opposing view (Aldayel & Magdy, 2019). It is framed as a supervised classification task, where labelled instances are used to train a classification model, which is then applied on unseen test data.

While humans can easily infer whether an author is in favour or against a specific event, the task becomes more challenging when performed at scale, due to the need for automated NLP methods. Consequently, the stance detection task has attracted an increasing interest in the scientific community, including scholars from linguistics and communication as well as computational linguistics. However, the need to automate the task by means of NLP methods is still in its infancy with a growing body of ongoing research.

Understanding stance expressed in text is a critical, yet challenging task and it is the main focus of this review paper. Stance is often implicit and needs to be inferred rather than directly determined from explicit expressions given in the text; indeed the target may not be directly mentioned (Somasundaran & Wiebe, 2009; Mohammad et al., 2016b). However, given the scale of social media data, understanding attitudes and responses of people to different events becomes unmanageable if done manually. Current stance detection approaches leverage machine learning and NLP models to study political and other opinionated issues (Volkova et al., 2016; Al-Ayyoub et al., 2018; D'Andrea et al., 2019; Johnson & Goldwasser, 2016; Lai et al., 2017). However, using persuasive writing techniques and word choices (Burgoon et al., 1975) to convey a stance brings important challenges for current state-of-the-art models as there is a need to capture these features in a large-scale dataset. Recent research is increasingly considering pragmatic factors in texts, adopting stance dynamics and

the impact of language evolution. Research in this direction can shed light into other dimensions when defining and analysing stance. However, building representations for complex, shifting or problematic meanings is still an open problem that needs exploration.

1.3 Capturing dynamics in stance detection

The stance detection task overlaps with, and is closely related to, different classification tasks such as sentiment analysis (Chakraborty et al., 2020), troll detection (Tomaiuolo et al., 2020), rumour and fake news detection (Zubiaga et al., 2018; Rani et al., 2020; Collins et al., 2020), and argument mining (Lawrence & Reed, 2020). In addition, stance can be impacted by the discursive and dynamic nature of the task (Mohammad et al., 2017; Somasundaran & Wiebe, 2009; Simaki et al., 2017).

In reviewing the literature on stance dynamics, we break down our review into three different dimensions (see Fig. 1), which cover the different aspects impacting how stance is formed and how it evolves:

- **stance utterance**, referring to a single message conveying a particular stance towards a target.
- **stance context**, referring to the pragmatic, spatiotemporal and diachronic factors that make stance an evolving phenomenon.
- **stance influence**, referring to social factors including the author of a post, as well as reactions towards, and activity around, messages expressing a particular stance.

In what follows, we delve into each of these dimensions and associated literature.

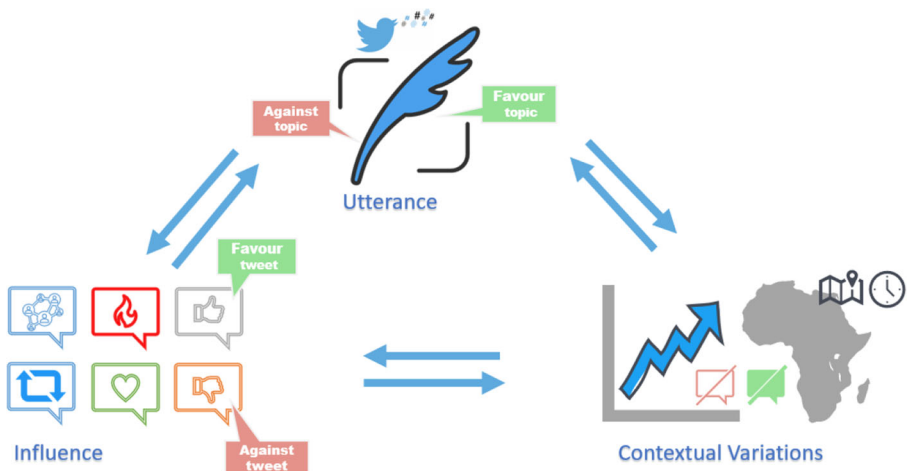


Fig. 1 Three dimensions around stance detection.

2 Stance utterance

Stance utterance refers to the stance expressed in a single message (Mohammad et al., 2016a), and reflects human interpretation of an event. It represents the features that form the textual viewpoint and they are essential for human inference and interpretation. Textual data can be analysed based on different features, which previous work have tackled by looking at a range of different challenges, which we discuss next.

One of the challenges in detecting the stance of a single utterance is **target identification**, i.e. determining who or what the stance is referring to. For example, in the utterance “I am supportive of A, but I’m totally against B”, the author expresses a supporting stance towards target A and an opposing stance towards target B. The **target** may be implicit and not always directly mentioned in the text (Schaefer & Stede, 2019). The target may be implicitly referred to Sobhani et al. (2019), or only aspects of it may be mentioned (Bar-Haim et al., 2017). These cases present the additional challenge of having to detect the target being referred to in a text prior to detecting the stance. Retrieval of messages likely referring to a target, as a first step to then do the target identification, is a challenge. Achieving high recall in relevant message retrieval can be difficult in the case of implicit messages (Mohammad et al., 2017). In addition, there is a risk of detecting false positives where a message may not be about the target at all, may not contain data expressing a stance, or may hold multiple stances in the same utterance (Lai et al., 2019; Simaki et al., 2017).

Nuances in the **wording** of an utterance can present another challenge in detecting stance. Opinions are not always explicitly expressed, and can also be implicit, explicit, ironic, metaphoric, uncertain, etc. Sun et al. (2016), Al-Ayyoub et al. (2018), and Simaki et al. (2018), which make stance detection more challenging. Moreover, surrounding words and symbols can alter the stance of an utterance (Sun et al., 2016), e.g. negating words or ironic emojis inverting the meaning of a text, which are especially challenging to detect. Recent models increasingly make use of more sophisticated linguistic and contextual features to infer stance from text. For example, looking at the **degree of involvement** by using special lexical terms, e.g. slang, jargon, specialist terms, and the informal lexicons associated with social intimacy (Tausczik & Pennebaker, 2010; Hamilton et al., 2016; Rumshisky et al., 2017; Liu et al., 2015). Also, use of **embeddings** where concept meanings can be biased and highly impacted by the cultural background and beliefs may lead to varying interpretations (Shoemark et al., 2019; Kutuzov et al., 2018; Hamilton et al., 2016a; Dubossarsky et al., 2017; Xu et al., 2019).

The **framing** of an utterance can also play an important role in the detection of stance. Framing refers to the adaptation of the wording to convey a specific interpretation of a story to a targeted audience (Walker et al., 2012) as in language and word choices, this can be seen in the following forms:

- **Reasoning/supporting evidence** about the target or aspects of it (Hasan & Ng, 2014; Addaood & Bashir, 2016; Bar-Haim et al., 2017; Simaki et al., 2017). For example, Bar-Haim et al. (2017) define the claim stance classification task as consisting of a target, a set of claims and the stance of these claims as either supporting or opposing the target. Further, they simplify the task by looking for

sentiment and contrast meaning between a given pair of target phrase and the topic candidate anchor phrase.

- **Attitude:** using single lexical terms holding polarity features related to the author sentiment and reaction to an event. Such word choices can be positive, negative, offensive, harmful, suspicious, aggressive, extremist (Blšták & Rozinajov, 2017). For instance, emotion and sentiment expressed in the text (Deitrick & Hu, 2013; Xu et al., 2011), which may express the author's view on the importance (or lack thereof) of the target.
- **Persuasion and quality of communication** (Hamilton, 2015; Aune & Kikuchi, 1993) through using grammatically correct sentences. For example, Lai et al. (2019) concluded that people connected to users taking cross-stance attitudes become less polarised and use neutral style when expressing their stance.

Another challenge in stance detection is determining if **stance is present** in an utterance, as the text may be neutral and not be opinionated towards the target. This involves determining whether or not an utterance expresses a stance, and subsequently determining the type of stance the author is taking (Mohammad et al., 2016b; Zubiaga et al., 2016; Simaki et al., 2017). For example, the most basic way of stance-taking could be the more extreme positions such as in favour or against to less extreme positions such as asserting, questioning, responding, commanding, advising, and offering which may lead to conversational and threaded stance context (Zubiaga et al., 2016). In such a context stance moves from its singular utterance structure to its augmenting component (see Section 4).

3 Stance context

Stance context refers to the impact of external factors in an utterance, including the collective viewpoint of a society in relation to the interpretation of a particular target. Context aims to capture the stance occurring in longitudinally evolving contexts, and can be impacted by shifts in opinions over time, locations, or cultures, among others. Any stance inference requires consideration of other points of view, potential stereotypes, as well as how public opinion evolved over time. This represents our understanding of the world dynamics and how stance may change over time (Lai et al., 2019; Volkova et al., 2016). Stance towards a topic may be considered stable only when it has the same polarity over time. Context involves factors causing changes in public opinion over time, such as real world events. Context constructs complex, shifting or problematic meanings which change the entire view of an event (Azaronyad et al., 2017; Stewart et al., 2017). We discuss the two main aspects that are considered when modelling stance context, which include **spatiotemporal changes** and **social changes**.

Collective and individual stance towards a target can be impacted by **spatiotemporal factors** (Volkova et al., 2016; Jackson et al., 2019). Events occurring in different locations/times get different attention depending on how likely they are to happen again and how unusual they are (Baly et al., 2018; Hamborg et al., 2019). Consequently, the audience judging the events have their own biases depending on

the cultural and ideological background, which leads to variations in stance across regions.

Even when we restrict geographical locations, there are other factors leading to **social changes** that have an impact on public opinion and stance. Social changes around a topic can lead to shifts in opinions (Volkova et al. 2016; D'Andrea et al. 2019; Lai et al., 2019). This can pose a significant challenge, particularly with the tendency in NLP to using distributed representations of words driven by co-occurrence frequency of words using sliding windows, and considering polysemy in more advanced language models. Words are treated based on their contextual similarity rather than solely based on their isolated frequencies. In order to build these models, one needs large collections of documents with a diverse vocabulary to produce high quality vector representations for different words. Consequently, these models rely on the amount of training data available, and the dimensionality of the word vectors (Mikolov et al., 2013). The emergence of out of vocabulary (OOV) words, not seen by these models, can be one of its main limitations. Different methods have been proposed to mitigate these limitations, for example through character-level representations in ELMo or FastText, and sub-word representations in BERT allowing models to incorporate segmented representations for unseen words (Ha et al., 2020). In prediction models, character-level and subword representations can lead to performance improvements with a trade-off on reduced model explainability; ongoing research is however investigating how to improve model explainability, exploiting for example attention scores produced by BERT (Bodria et al., 2020). Moreover, newly emerging words or words that shift their meaning over time would lead to outdated models. Challenges relating to social changes can be further broken down into the following:

- **Linguistic shift**, which is defined as slow and regular changes in the core meaning of a word. For example, “the word GAY shifting from meaning CARE-FREE to HOMOSEXUAL during the 20th century” (Kutuzov et al., 2018). This is also reflected in the semantic meaning of emoticons across different contexts, languages and cultures (Robertson et al., 2018). In multilingual settings, code-mixing of two languages in the same utterance (Khanuja et al., 2020), or borrowing a word from a different language due to influence from other languages, rather than internal changes in the same language.
- **Usage change**, which is the local change of a word’s nearest semantic neighbours from one meaning to another, as in the shift of word the “prison CELL to CELL phone” which is more of a cultural change than a semantic change (Hamilton et al., 2016a). Thus, different viewpoints allow collective stance to change especially when a story is viewed through different eyes and interpreted differently. For example, focusing on UK politics, Azarbondy et al. (2017) revealed that “The meaning given by Labours to MORAL is shifted from a PHILOSOPHICAL concept to a LIBERAL concept over time. In the same time, the meaning of this word is shifted from a SPIRITUAL concept to a RELIGIOUS concept from the Conservatives’ viewpoint. Moreover, two parties gave very different meanings to this word. Also, the meaning of DEMOCRACY is stable over time for both parties. However, Conservatives refer to democracy mostly as a UNITY concept, while Labours associate it with FREEDOM and SOCIAL JUSTICE.”

- **Changes in cultural associations**, which is measured as the distance between two words in the semantic space, as in “IRAQ or SYRIA being associated with the concept of WAR after armed conflicts had started in these countries” (Kutuzov et al., 2018). Also, the type of sentiment bore by a word can change over time. For example, “the word SLEEP acquiring more negative connotations related to sleep disorders, when comparing its 1960s contexts to its 1990s contexts” (Gulordava & Baroni, 2011). Moreover, studies have also looked at the relatedness of words over time, by looking at how the strength of the association between words changes. For example, Rosin et al. (2017) introduced a relationship model that supports the task of identifying, given two words (e.g. Obama and president), when they relate most to each other, having longitudinal data collections as input.

4 Stance influence

Stance influence refers to the aggregated importance surrounding an individual message expressing a particular stance, and can be measured by using different qualitative and quantitative metrics. These include the author’s profile and others’ reactions to a message.

Influence defines the quality of an utterance to make an impact, and can vary depending on the popularity and reputation of the author, as well as the virality of a post, among others. Next we discuss three aspects which are relevant to stance influence, i.e. **threading comments**, **network homophily** and **author profile**.

Social media platforms provide a place for conversations to develop, which lead to **threaded conversations** or **tree-structured conversations**. The formation of these conversations enables exchanging viewpoints on top of the initial author’s stance (Zubiaga et al., 2016; Guerra et al., 2017; Lai et al., 2019). For example, Lai et al. (2019) observed that users make use of replies for expressing diverging opinions. Research looking at whether retweeting a post indicates endorsement is so far inconsistent. Lai et al. (2018) observe that people tend to retweet what they agree on. Conversely, Guerra et al. (2017) argued that a retweet does not indicate supporting its underlying opinion.

There is evidence showing that social media users tend to connect and interact with other like-minded users (Lai et al., 2017; Conover et al., 2012), which is also known as the phenomenon of **network homophily**. Lai et al. (2019) looked at the impact of different characteristics of social media in sharing stance, showing for example that opposing opinions generally occur through replies as rather than through retweets or quotes, polarisation varies over time, e.g. increasing in the proximity of elections.

The identity of the person posting a piece of text expressing a stance, or the **author’s profile**, can also play an important role in the development of stance, for example if an influential user expresses an opinion. Two key factors of an author’s profile include:

- **Author’s ideology and background**, often inferred by observing the user’s profile (Elfardy & Diab, 2016; Conover et al., 2012; Yan et al., 2018; Lai et al.,

- 2019), can be used as additional features to determine the stance expressed by a user, rather than solely using the textual content of a post (Mohammad et al. 2016a, 2016b).
- **Author’s stance in the temporal space** (Garcia et al., 2015). For instance, media organisations may express viewpoints through different frames (Hasan & Ng, 2014), which takes time to be assessed (Zubiaga et al., 2016) and may also impact stance evolution and people’s stand points. This can also have an impact on how threaded conversations are developed.

5 Datasets

To study the stance detection task, different datasets covering various topics and pragmatic aspects have been created by researchers.

Table 1 shows the list of stance datasets available, along with their key characteristics; these include the time frame they cover, a key aspect in our focus on stance dynamics, as we are interested in identifying the extent to which existing datasets enable this analysis. For ten of the datasets we found, the time frame covered by the data is not indicated (marked in the table as N/A), which suggests that temporal coverage was not the main focus of these works. The rest of the datasets generally cover from a month to a maximum of 1 or 2 years; while the latter provides some more longitudinal coverage, we argue that it is not enough to capture major societal changes. The exception providing a dataset that covers a longer period of time is that by Conforti et al. (2020) and Addawood et al. (2018), with five years’ worth of data.

Despite the availability of multiple stance datasets and their ability to solve different generalisability problems (e.g. across targets, languages and domains), this analysis highlights the need for more longitudinal datasets that would enable persistence for temporal stance detection and temporal adaptation, ideally across cultures and languages. For the few datasets that contain some degree of longitudinal content, such as (Conforti et al., 2020) covering 57 months and (Addawood et al., 2018) covering 61 months, the available data is sparsely distributed throughout the entire time period. This again urges the need for more longitudinal datasets, which in turn provide more density for each time period. While data labelling is expensive and hard to afford at scale, possible solutions may include use of distant supervision (Purver & Battersby, 2012) for data collection and labelling or labelling denser datasets for specific time periods which are temporally distant from each other, despite leaving gaps between the time periods under consideration. Distant supervision has been widely used for other tasks such as sentiment analysis (Go et al., 2009), leading to datasets covering in some cases over 7 years (Yin et al., 2021), however its applicability to stance detection has not been studied as much.

In summary, we observe that existing datasets provide limited resources to capture language dynamics and leverage longitudinal analysis, which would then give rise to more research aiming to capture stance dynamics.

Table 1 Stance detection datasets, including the time frame covered

Ref.	Time frame	Months	topics (#)	Source	Language
Conforti et al. (2020)	Apr 2014 - Dec 2018	57	Finance (2)	Twitter	English
Addawood et al. (2018)	Jan 2012 - Jan 2017	61	Women to Drive Movement (1)	Twitter	Arabic
Rajadesingan and Liu (2014)	Apr 2013	1	US Issues (1)	Twitter	English
Volkova et al. (2016)	Sep 2014 - Mar 2015	7	Politics (1)	Vkontakte	Russian Ukrainian
Zubiaga et al. (2016)	Aug 2014 - Oct 2015	15	News events (9)	Twitter	English German
Mohammad et al. (2016b)	Jul 2015	1	US Issues (6)	Twitter	English
Schuff et al. (2017)	Jul 2015	1	US Issues (6)	Twitter	English
Simaki et al. (2017)	Jun - Aug 2015	3	Political blogs(1)	the BBC	English
Küçük and Can (2020)	Aug - Sep 2015	2	Sport (2)	Twitter	Turkish
Sobhani et al. (2019)	Oct 2015 - Feb 2016	5	US Issues (4)	Twitter	English
Addawood and Bashir (2016)	Jan - Mar 2016	3	Products (1)	Web	English
Lai et al. (2018, 2019)	Nov - Dec 2016	2	Italian Referend. (1)	Twitter	Italian
Yan et al. (2018)	Jan - Dec 2016	12	US Issues (2)	Twitter	English
D'Andrea et al. (2019)	Sep 2016 - Jun 2017	10	Health (1)	Twitter	Italian
Lozhnikov et al. (2018)	Nov 2017	1	Politics(1)	Twitter Meduza Russia Today	Russian
Baly et al. (2018)	Jan 2016 - Dec 2017	12	Middle East(1)	News articles	Arabic
Somasundaran and Wiebe (2009)	N/A	N/A	Products (1)	Convinceme	English
Anand et al. (2011)	N/A	N/A	Politics (12)	Convinceme	English
Walker et al. (2012)	N/A	N/A	US Issues (12)	4forums Createdebate	English
Skanda et al. (2017)	N/A	N/A	Indian Issues (4)	Facebook	Indian
Hercig et al. (2018)	N/A	N/A	Czech Issues (2)	News	Czech

Table 1 (continued)

Ref.	Time frame	Months	topics (#)	Source	Language
Xu et al. (2016)	N/A	N/A	Different topics (7)	Sina Weibo	Chinese
Ferreira and Vlachos (2016)	N/A	N/A	News Articles (1)	News articles	English
Hasan and Ng (2014)	N/A	N/A	US Issues (4)	Createdebate	English
Bar-Haim et al. (2017)	N/A	N/A	Open domain	IBM dataset	English
Taulé et al. (2018)	N/A	N/A	Catalan Independence (1)	Twitter	Spanish Catalan

6 Open challenges and future directions

In the previous sections we have discussed the three key factors relevant to stance and impacting its formation and temporal evolution, as well as existing datasets. In what follows, we discuss the main research challenges and set forth a number of future research directions. We first discuss core challenges, which are specific to stance detection, followed by general challenges, which are broader challenges that also have an impact on stance detection.

6.1 Core challenges

There are numerous open challenges that are specific to the stance detection task. To the best of our knowledge, few studies have specifically focused on the **evolving nature of topics** and its impact on stance detection models. Moreover, fluctuation of word frequencies and distributions over time highlight both the challenge and the importance of the task. Commonalities between the source and target tasks tend to be crucial for successful transfer (Vu et al., 2020). However, recent NLP models have shifted to transfer learning and domain adaptation where target tasks contain limited training data (Xu et al., 2019), source data pertains to a different domain (Zhang et al., 2020) or to a different language (Lai et al., 2020). We anticipate two main directions that would help extend this research: (1) furthering research in transfer learning that looks more into transferring knowledge over time, as opposed to the more widely studied subareas looking into domain adaptation (Ramponi & Plank, 2020) or cross-lingual learning (Lin et al., 2019), and (2) increasing the availability of longitudinal datasets that would enable further exploration of temporal transfer learning.

The majority of existing datasets are from the **domain** of politics and to a lesser extent business, and are hence constrained in terms of topics. Broadening the topics covered in stance datasets should be one of the key directions in future research.

In general, existing datasets cover **short time spans** in languages including English (Ferreira & Vlachos, 2016; Mohammad et al., 2016b; Simaki et al., 2017;

Somasundaran & Wiebe, 2009; 2009; Hercig et al., 2018; Walker et al., 2012; Anand et al., 2011; Conforti et al., 2020), Arabic (Baly et al., 2018; Addawood et al., 2018), Italian (Lai et al., 2018), Chinese (Xu et al., 2016), Turkish (Küçük & Can, 2020), Spanish and Catalan (Taulé et al., 2018), Kannada (Skanda et al., 2017), German (Zubiaga et al., 2016), Russian (Lozhnikov et al., 2018).

Recent efforts in multilingual stance classification have also published datasets including German, French and Italian (Mohtarami et al., 2019; Vamvas & Sennrich, 2020), and English, French, Italian, Spanish and Catalan (Lai et al., 2020), but are still limited in terms of the time frame covered. Longitudinal datasets annotated for stance would enable furthering research in this direction by looking into the temporal dynamics of stance.

The **quality and persistence of the data** are also important challenges that need attention. Annotation of stance is particularly challenging where a single post may contain multiple targets, or where users change their own stances towards a particular target, i.e. cross-stance attitude. These are challenges that lead to lower inter-annotator agreement and produce confusion even for humans (Lai et al., 2019; Sobhani et al., 2019). Moreover, relying on social media data under the terms of service of the platforms, reproducibility of some datasets is not always possible (Zubiaga, 2018). There is also a need for stance detection models that also consider context, for which suitable datasets are lacking. There are also cases where concepts including sentiment, stance and emotion are conflated, with few efforts to define stance (Mohammad et al., 2016b; Simaki et al., 2017) or to experimentally prove the difference between these concepts (Mohammad et al., 2017; Aldayel & Magdy, 2019).

In stance particularly, we can define these problems in four levels: (1) **utterance level** as changing stance from being in favour to being against, (2) **time level** as collective stance (Nguyen et al., 2012) of public pool change from highly in favour to highly against over time, (3) **domain level** where some words change its polarity from one domain to another (such as high prices indicating a favourable stance in the context of a seller but an opposing stance for customers), and (4) **cultural level** which represents stance shift between languages or various geographical locations. Indeed, use of a machine learning model training from old data may not be directly applicable to future datasets, e.g. due to suffering from domain bias, co-variate shift and concept drift. This can be cause by the nature of controversial topics and the impact of pragmatics such as time, location and ideology.

6.2 General challenges

We also identified gaps in the literature that are not exclusive to stance but have significant impact in stance prediction models such as the impact of **predefined lexicon** word resolution on the model's accuracy (Somasundaran & Wiebe, 2009). This is especially true when models dependent on a lexicon fail to capture the polarity of evolving words. Research in this direction has used pre-trained word embeddings such as GloVe (Pennington et al., 2014), FastText (Bojanowski et al., 2017), Elmo (Peters et al., 2018), and BERT (Devlin et al., 2019) among others which proved to mitigate the problem of polysemy though word vector representations. This is due to

the fact that these models are fed news articles and web data from different sources may be inherently biased (Ruder, 2017). Moreover, it has been shown that variations of architecture in state-of-the-art language models can significantly impact the performance of the model in downstream tasks. Other work focuses on flipped polarity and negation (Polanyi & Zaenen, 2006). Even though embedding models consider preceding and following words of a centre word for a given sentence (context), the temporal property of the word itself and its diachronic shift from one meaning to another has not been studied in the context of stance. The identification of diachronic shift of words has however been tackled as a standalone task (Fukuhara et al., 2007; Azarbyad et al., 2017; Tahmasebi et al., 2018; Shoemark et al., 2019; Dubossarsky et al., 2017; Stewart et al., 2017; Hamilton et al., 2016; Kutuzov et al., 2018; Hamilton et al., 2016a; Rumshisky et al., 2017). This is however yet to be explored in specific applications such as stance detection. This may also impact the models' performance across different domains and time frames.

The use of models developed in the field of **NLP** has been barely explored in the context of stance detection, which have been more widely studied for other tasks such as co-reference resolution (Somasundaran & Wiebe, 2009) and named entity recognition (Küçük & Can, 2020; Liu et al., 2013). Previous research has however highlighted problems in this direction (Lozhnikov et al. 2018; Küçük & Can 2020; Borges et al. 2019; Sobhani et al. 2019; Lai et al. 2018, 2019; Simaki et al. 2017), which suggests that further exploration and adaptation of NLP models may be of help. Similar to most approaches for social media data, pragmatic opinions (Somasundaran & Wiebe, 2009) including short opinions with few lexicon cues can negatively impact prediction performance, including hedging (Somasundaran & Wiebe, 2009), rhetorical questions (Hasan & Ng, 2014; Mohammad et al., 2017), inverse polarity (Mohammad et al., 2017; Skanda et al., 2017), sarcasm (Hasan & Ng, 2014; Skanda et al., 2017), all of which can have a significant impact in the classifier, especially in the case of two-way classification models.

We can summarise the challenges into two main categories:

1. Current deep learning models and the existence of large pre-trained embeddings can offer highly accurate results using training datasets. However, it can lead to biased results when applied to new, **unseen data**, e.g. data pertaining to a different point in time to the one seen during training. This highlights the difficulty of the task and the need to advance research in developing models that are independent of a specific use case and dataset, which can keep evolving as the data changes. Also, there is a need to develop data from different languages to mitigate the cultural biases in existing datasets. This can help detect and explain different perspectives while using specific topics to reason, compare and contrast a model's performance. This would also help further research in stance detection models that are more stable in performance. Moreover, in the case of certain languages, such as Arabic, the use of dialectal language instead of the modern standard language presents an additional challenge. More methods need to be investigated to improve a model's performance considering contextual variation (see Section 3).

2. The existence of social media accounts run by **bots** leads to fabricated viewpoints of events. These accounts may have been created to manipulate the true view and harm specific targets (for example, businesses or people). This manipulated information can in turn have an impact on specific points in time where the bots operate, and can jeopardise the applicability of stance detection models for certain points in time (e.g. during elections where bot participation may increase) if bots are not detected and removed from the dataset.

In summary, understanding and detecting **semantic shift** (Stewart et al., 2017; Rumshisky et al., 2017; Tahmasebi et al., 2018; Shoemark et al., 2019) in the meaning of words has been of much interest in linguistics and related areas of research, including political science, history. However, the majority of this literature focuses their efforts on uncovering language evolution over time, with a dearth of computational research assessing its impact in context-based prediction models such as those using embedding models. Moreover, combining a contextual knowledge using word embeddings in prediction models can help improve performance of stance detection models by leveraging their vector representations. However, current state-of-the-art research ignores the impact of contextual changes due to pragmatic factors such as social and time dimensions when building their models. This may impact a model's performance over time and can result in **outdated datasets and models**. This is due to the dependence of these models to use static data and pre-trained word embeddings to train models. While still training on data pertaining to a particular time period, models need to leverage the evolving nature of language in an unsupervised manner to keep stance detection performance stable. Temporal deterioration of models is however not exclusive to stance detection, and has been demonstrated to have an impact in other NLP tasks such as hate speech detection (Florio et al., 2020). While some social and linguistic changes may take time (Hamilton et al., 2016b) before they occur, recent literature proved that they may also occur in short periods of time (Shoemark et al., 2019; Azarbyad et al., 2017). Most importantly, unlike semantic changes which capture word fluctuations over time, temporal contextual variability may occur in corpus-based predictive models.

7 Conclusion

In this survey paper discuss the impact of temporal dynamics in the development of stance detection models, by reviewing relevant literature in both stance detection and temporal dynamics of social media. Our survey delves into three main factors affecting the temporal stability of stance detection models, which includes utterance, context and influence. We then discuss existing datasets and their limited capacity to enable longitudinal research into studying and capturing dynamics affecting stance detection. This leads to our discussion on research challenges needing to be tackled to further research in capturing dynamics in stance detection, which we split in two parts including core, stance-specific challenges and more general challenges.

Today's computational models are able to process big data beyond human scale, building on digital humanities and computational linguistics. This however poses a

number of challenges when dealing with longitudinally-evolving data. The changes produced by societal and linguistic evolution, among others, both of which are prominent in social media platforms, have significant impact on the shift of social beliefs by means of spreading ideas. With the proliferation of historical social media data and advanced tools, we argue for the need to build models that better capture this contextual change of stance. This necessitates furthering research in the modelling of temporal dynamics of human behaviour.

Current research is largely limited to datasets covering short periods of time. Such datasets are however of special importance when one is interested in monitoring the evolution of public stance about particular topic over time. Thus, future work should consider expanding existing datasets and adapting machine learning models by focusing on maintainability and long term performance.

Our review presents an interdisciplinary viewpoint by bridging the fields of linguistics, natural language processing and digital humanity. There is a need for bridging the efforts in the field of digital humanities by relying on large historical textual corpora and introducing large scale annotated datasets. This can then benefit longitudinal analyses and contribute to advancing stance detection models that capture linguistic and societal evolution into them.

Data Availability The authors confirms that all data generated or analysed during this study are included in this published article. Furthermore, primary and secondary sources and data supporting the findings of this study were all publicly available at the time of submission.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Addawood, A., Alshamrani, A., Alqahtani, A., Diesner, J., & Broniatowski, D. (2018). Women's driving in saudi arabia—analyzing the discussion of a controversial topic on twitter. In *2018 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation, BRiMS 2018*.
- Addawood, A., & Bashir, M. (2016). "What Is Your Evidence?" A Study of Controversial Topics on Social Media. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)* (pp. 1–11). Association for Computational Linguistics. <http://aclweb.org/anthology/W16-2801>.
- Al-Ayyoub, M., Rabab'ah, A., Jararweh, Y., Al-Kabi, M. N., & Gupta, B. B. (2018). Studying the controversy in online crowds' interactions. *Applied Soft Computing Journal*, *66*, 557–563. <https://doi.org/10.1016/j.asoc.2017.03.022>.

- Aldayel, A., & Magdy, W. (2019). Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–20. <https://doi.org/10.1145/3359307>, <https://dl.acm.org/citation.cfm?id=3359307>.
- Alkhalifa, R., Kochkina, E., & Zubiaga, A. (2021). Opinions are made to be changed: Temporally adaptive stance classification. In *Proceedings of the 2021 Workshop on Open Challenges in Online Social Networks* (pp. 27–32).
- Anand, P., Walker, M., Abbott, R., Tree, J. E. F., Bowmani, R., & Minor, M. (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 (WASSA 2.011)*, WASSA '11 (pp. 1–9). Association for Computational Linguistics. <http://www.aclweb.org/anthology/W11-1701>.
- Augenstein, I., Rocktäschel, T., Vlachos, A., & Bontcheva, K. (2016). Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 876–885).
- Aune, R. K., & Kikuchi, T. (1993). Effects of language intensity similarity on perceptions of credibility relational attributions, and persuasion. *Journal of Language and Social Psychology*, 12(3), 224–238. <https://doi.org/10.1177/0261927X93123004>.
- Azarbonyad, H., Dehghani, M., Beelen, K., Arkut, A., Marx, M., & Kamps, J. (2017). Words are malleable: Computing semantic shifts in political and media discourse. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 1509–1518).
- Baly, R., Karadzhev, G., Alexandrov, D., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3528–3539).
- Baly, R., Mohtarami, M., Glass, J., Márquez, L., Moschitti, A., & Nakov, P. (2018). Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 21–27).
- Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A., & Slonim, N. (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* (pp. 251–261).
- Blšák, M., & Rozinajov, V. (2017). Machine learning approach to the process of question generation. In *International Conference on Text, Speech, and Dialogue* (pp. 102–110). Springer.
- Bodria, F., Panisson, A., Perotti, A., & Piaggese, S. (2020). Explainability methods for natural language processing: Applications to sentiment analysis (discussion paper).
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.
- Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, 11(3), 1–26.
- Burgoon, M., Jones, S. B., & Stewart, D. (1975). Toward a message-centered theory of persuasion: Three empirical investigations of language intensity1. *Human Communication Research*, 1(3), 240–256.
- Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*, 7(2), 450–464.
- Collins, B., Hoang, D. T., Nguyen, N. T., & Hwang, D. (2020). Fake news types and detection models on social media a state-of-the-art survey. In *Asian Conference on Intelligent Information and Database Systems* (pp. 562–573). Springer.
- Conforti, C., Berndt, J., Pilehvar, M. T., Giannitsarou, C., Toxvaerd, F., & Collier, N. (2020). Will-they-won't-they: A very large dataset for stance detection on Twitter. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1715–1724). Association for Computational Linguistics. <https://www.aclweb.org/anthology/2020.acl-main.157>.
- Conover, M. D., Gonçalves, B., Flammini, A., & Menczer, F. (2012). Partisan asymmetries in online political activity. *EPJ Data Science*, 1(1), 6.
- Cossette, P. (1998). The study of language in organizations: A symbolic interactionist stance. *Human Relations*, 51(11), 1355–1377.
- D'Andrea, E., Ducange, P., Bechini, A., Renda, A., & Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116, 209–226. <https://doi.org/https://doi.org/10.1016/j.eswa.2018.09.009>, <http://www.sciencedirect.com/science/article/pii/S0957417418305803>.

- Deitrick, W., & Hu, W. (2013). Mutually Enhancing Community Detection and Sentiment Analysis on Twitter Networks. *Journal of Data Analysis and Information Processing*, 01(03), 19–29. <https://doi.org/10.4236/jdaip.2013.13004>, <http://dx.http://www.scirp.org/journal/jdaip>.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. 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)* (pp. 4171–4186).
- Dubossarsky, H., Grossman, E., & Weinshall, D. (2017). Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1136–1145).
- Elfardy, H., & Diab, M. (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)* (pp. 434–439).
- Englebretson, R. (2007). Stancetaking in discourse: An introduction. In *Rice Linguistics Symposium*. John Benjamins Publishing Company.
- Ferreira, W., & Vlachos, A. (2016). 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* (pp. 1163–1168). Association for Computational Linguistics. <http://aclweb.org/anthology/N16-1138>.
- Florio, K., Basile, V., Polignano, M., Basile, P., & Patti, V. (2020). Time of your hate: The challenge of time in hate speech detection on social media. *Applied Sciences*, 10(12), 4180.
- Fukuhara, T., Nakagawa, H., & Nishida, T. (2007). Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. In *ICWSM*.
- Garcia, D., Abisheva, A., Schweighofer, S., Serdült, U., & Schweitzer, F. (2015). Ideological and temporal components of network polarization in online political participatory media. *Policy & Internet*, 7(1), 46–79.
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12), 2009.
- Guerra, P. C., Souza, R. C. S. N. P., Assunç ao, R. M., & Meira, Jr, W. (2017). Antagonism also flows through retweets: The impact of out-of-context quotes in opinion polarization analysis. arXiv:1703.03895.
- Gulordava, K., & Baroni, M. (2011). A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.220.7680>.
- Ha, P., Zhang, S., Djuric, N., & Vucetic, S. (2020). Improving word embeddings through iterative refinement of word-and character-level models. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 1204–1213).
- Hamborg, F., Donnay, K., & Gipp, B. (2019). Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4), 391–415.
- Hamilton, M. A. (2015). Language intensity as an expression of power in political messages. In *The exercise of power in communication* (pp. 233–265). Springer.
- Hamilton, W. L., Clark, K., Leskovec, J., & Jurafsky, D. (2016). Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 595–605).
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2116–2121).
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1489–1501).
- Hardalov, M., Arora, A., Nakov, P., & Augenstein, I. (2021). A survey on stance detection for mis- and disinformation identification. arXiv:2103.00242.
- Hasan, K. S., & Ng, V. (2014). Why are you taking this stance? Identifying and classifying reasons in ideological debates. In *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* (pp. 751–762). Association for Computational Linguistics. <http://aclweb.org/anthology/D14-1083>.
- Hercig, T., Krejzl, P., & Král, P. (2018). Stance and Sentiment in Czech. *Computación y Sistemas*.

- Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science (New York, N.Y.)*, 366(6472), 1517–1522. <https://doi.org/10.1126/science.aaw8160>, <http://www.ncbi.nlm.nih.gov/pubmed/31857485>.
- Johnson, K., & Goldwasser, D. (2016). Identifying Stance by Analyzing Political Discourse on Twitter. In *Proceedings of the First Workshop on NLP and Computational Social Science* (pp. 66–75). Association for Computational Linguistics. <http://aclweb.org/anthology/W16-5609>.
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers*, 20(3), 531–558.
- Khanuja, S., Dandapat, S., Srinivasan, A., Sitaram, S., & Choudhury, M. (2020). Gluecos: An evaluation benchmark for code-switched nlp. arXiv:2004.12376.
- Kıřuçuk, D., & Can, F. (2020). Stance detection: A survey. *ACM Comput. Surv.*, 53(1), 12:1–12:37. <https://doi.org/10.1145/3369026>.
- Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. (2018). Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1384–1397).
- Lai, M., Cignarella, A. T., Farías, D. I. H., Bosco, C., Patti, V., & Rosso, P. (2020). Multilingual stance detection in social media political debates. *Computer Speech & Language*, 101075.
- Lai, M., Hernández Farías, D. I., Patti, V., & Rosso, P. (2017). Friends and enemies of clinton and trump: Using context for detecting stance in political tweets. In G. Sidorov, & O. Herrera-Alcántara (Eds.) *Advances in Computational Intelligence* (pp. 155–168). Cham: Springer International Publishing.
- Lai, M., Patti, V., Ruffo, G., & Rosso, P. (2018). Stance evolution and twitter interactions in an italian political debate. In M. Silberstein, F. Atigui, E. Kornysheva, E. Métais, & F. Meziane (Eds.) *Natural Language Processing and Information Systems* (pp. 15–27). Cham: Springer International Publishing.
- Lai, M., Tambuscio, M., Patti, V., Ruffo, G., & Rosso, P. (2019). Stance polarity in political debates: A diachronic perspective of network homophily and conversations on Twitter. *Data and Knowledge Engineering*, 124, 101738. <https://doi.org/10.1016/j.datak.2019.101738>.
- Lai, M., Tambuscio, M., Patti, V., Ruffo, G., & Rosso, P. (2017). Extracting graph topological information and users' opinion. In G. J. F. Jones, S. Lawless, J. Gonzalo, L. Kelly, L. Goeriot, T. Mandl, L. Cappellato, & N. Ferro (Eds.) *Experimental IR Meets Multilinguality, Multimodality, and Interaction* (pp. 112–118). Cham: Springer International Publishing.
- Lawrence, J., & Reed, C. (2020). Argument mining: A survey. *Computational Linguistics*, 45(4), 765–818.
- Lin, Y.-H., Chen, C.-Y., Lee, J., Li, Z., Zhang, Y., Xia, M., Rijhwani, S., He, J., Zhang, Z., Ma, X., & et al. (2019). Choosing transfer languages for cross-lingual learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3125–3135).
- Liu, S., Cheng, X., Li, F., & Li, F. (2015). TASC: Topic-adaptive sentiment classification on dynamic tweets. *IEEE Transactions on Knowledge and Data Engineering*, 27(6), 1696–1709. <https://doi.org/10.1109/TKDE.2014.2382600>.
- Liu, X., Wei, F., Zhang, S., & Zhou, M. (2013). Named entity recognition for tweets. *ACM Transactions on Intelligent Systems and Technology*, 4(1). <https://doi.org/10.1145/2414425.2414428>.
- Lozhnikov, N., Derczynski, L., & Mazzara, M. (2018). Stance prediction for russian: data and analysis. In *International Conference in Software Engineering for Defence Applications* (pp. 176–186). Springer.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781.
- Mills, A. J. (2012). Virality in social media: the spin framework. *Journal of Public Affairs*, 12(2), 162–169.
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. (2016). SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 31–41). San Diego: Association for Computational Linguistics. <https://www.aclweb.org/anthology/S16-1003>.
- Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., & Cherry, C. (2016). A dataset for detecting stance in tweets. In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016* (pp. 3945–3952). <http://alt.qcri.org/semEval2016/task6/>.
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3), 1–23.
- Mohtarami, M., Glass, J., & Nakov, P. (2019). Contrastive Language Adaptation for Cross-Lingual Stance Detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Lan-*

- guage Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 4433–4442). Association for Computational Linguistics.
- Nguyen, L. T., Wu, P., Chan, W., Peng, W., & Zhang, Y. (2012). Predicting collective sentiment dynamics from time-series social media. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining - WISDOM '12*, (Vol. 12 pp. 1–8). ACM Press. <http://dl.acm.org/citation.cfm?doi=2346676.2346682>.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 2227–2237).
- Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications* (pp. 1–10). Springer.
- Purver, M., & Battersby, S. (2012). Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 482–491).
- Rajadesingan, A., & Liu, H. (2014). Identifying users with opposing opinions in twitter debates. In W. G. Kennedy, N. Agarwal, & S. J. Yang (Eds.) *Social Computing, Behavioral-Cultural Modeling and Prediction* (pp. 153–160). Cham: Springer International Publishing.
- Ramponi, A., & Plank, B. (2020). Neural unsupervised domain adaptation in nlp survey. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6838–6855).
- Rani, N., Das, P., & Bharadwaj, A. (2020). Rumour detection in online social networks: Recent trends. Available at SSRN 3564070.
- Robertson, A., Magdy, W., & Goldwater, S. (2018). Self-representation on twitter using emoji skin color modifiers. arXiv:1803.10738.
- Rosin, G. D., Adar, E., & Radinsky, K. (2017). Learning word relatedness over time. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1168–1178).
- Ruder, S. (2017). Word embeddings in 2017: Trends and future directions. <https://ruder.io/word-embeddings-2017/>.
- Rumshisky, A., Gronas, M., Potash, P., Dubov, M., Romanov, A., Kulshreshtha, S., & Gribov, A. (2017). Combining network and language indicators for tracking conflict intensity. In *International Conference on Social Informatics* (pp. 391–404). Springer.
- Schaefer, R., & Stede, M. (2019). Improving implicit stance classification in tweets using word and sentence embeddings. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)* (pp. 299–307). Springer.
- Schuff, H., Barnes, J., Mohme, J., Padó, S., & Klinger, R. (2017). Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 13–23). Copenhagen: Association for Computational Linguistics. <https://www.aclweb.org/anthology/W17-5203>.
- Shoemark, P., Liza, F. F., Nguyen, D., Hale, S., & McGillivray, B. (2019). Room to glo: A systematic comparison of semantic change detection approaches with word embeddings. 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)* (pp. 66–76).
- Simaki, V., Paradis, C., & Kerren, A. (2018). Evaluating stance-annotated sentences from the Brexit Blog Corpus: A quantitative linguistic analysis. *ICAME Journal*, 42(1), 133–166. <https://doi.org/10.1515/icame-2018-0007>.
- Simaki, V., Paradis, C., Skeppstedt, M., Sahlgren, M., Kucher, K., & Kerren, A. (2017). Annotating Speaker Stance in Discourse: The Brexit Blog Corpus. *Corpus Linguistics and Linguistic Theory*. <https://doi.org/10.1515/cllt-2016-0060>.
- Skanda, V. S., Kumar, M. A., & Soman, K. P. (2017). Detecting stance in kannada social media code-mixed text using sentence embedding. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 964–969). IEEE.
- Sobhani, P., Inkpen, D., & Zhu, X. (2019). Exploring deep neural networks for multitarget stance detection. *Computational Intelligence*, 35(1), 82–97.

- Somasundaran, S., & Wiebe, J. (2009). Recognizing stances in online debates. In *ACL-IJCNLP 2009 - Joint Conf. of the 47th Annual Meeting of the Association for Computational Linguistics and 4th Int. Joint Conf. on Natural Language Processing of the AFNLP, Proceedings of the Conf.* (pp. 226–234). Association for Computational Linguistics. <http://www.cs.pitt.edu/mpqa>.
- Stewart, I., Arendt, D., Bell, E., & Volkova, S. (2017). Measuring, predicting and visualizing short-term change in word representation and usage in VKontakte social network. In *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017* (pp. 672–675). <https://pymorphy2.readthedocs.io/en/latest/>.
- Sun, Q., Wang, Z., Zhu, Q., & Zhou, G. (2016). Exploring various linguistic features for stance detection. In *Natural Language Understanding and Intelligent Applications* (pp. 840–847). Cham: Springer International Publishing.
- Tahmasebi, N., Borin, L., & Jatowt, A. (2018). Survey of computational approaches to lexical semantic change. arXiv:1811.06278.
- Taulé, M., Rangel, F., Antònia Martí, M., & Rosso, P. (2018). Overview of the task on multimodal stance detection in Tweets on catalan #1Oct referendum. In *CEUR Workshop Proceedings*, (Vol. 2150 pp. 149–166). <http://clic.uib.edu>.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. ISSN 0261927X. <http://jls.sagepub.com>.
- Tomaiuolo, M., Lombardo, G., Mordonini, M., Cagnoni, S., & Poggi, A. (2020). A survey on troll detection. *Future Internet*, 12(2), 31. <https://doi.org/10.3390/fi12020031>.
- Vamvas, J., & Sennrich, R. (2020). X-stance: A multilingual multi-target dataset for stance detection. arXiv:2003.08385.
- Volkova, S., Chetviorkin, I., Arendt, D., & Van Durme, B. (2016). Contrasting public opinion dynamics and emotional response during crisis. In *International Conference on Social Informatics* (pp. 312–329). Springer.
- Vu, T., Wang, T., Munkhdalai, T., Sordoni, A., Trischler, A., Mattarella-Micke, A., Maji, S., & Iyyer, M. (2020). Exploring and predicting transferability across nlp tasks. arXiv:2005.00770.
- Walker, M. A., Anand, P., Abbott, R., & Grant, R. (2012). Stance classification using dialogic properties of persuasion. In *NAACL HLT 2012 - 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference* (pp. 592–596).
- Walker, M. A., Anand, P., Tree, J. E. F., Abbott, R., & King, J. (2012). A corpus for research on deliberation and debate. In *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012* (pp. 812–817). <http://nlds.soe.ucsc.edu/software>.
- Xu, B., Mohtarami, M., & Glass, J. (2019). Adversarial domain adaptation for stance detection. arXiv:1902.02401.
- Xu, K., Li, J., & Liao, S. S. (2011). Sentiment community detection in social networks. In *ACM International Conference Proceeding Series* (pp. 804–805).
- Xu, R., Zhou, Y., Wu, D., Gui, L., Du, J., & Xue, Y. (2016). Overview of nlpcc shared task 4: Stance detection in chinese microblogs. In *Natural Language Understanding and Intelligent Applications* (pp. 907–916). Springer.
- Xu, Z., Li, Q., Chen, W., Cui, Y., Qiu, Z., & Wang, T. (2019). Opinion-aware knowledge embedding for stance detection. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data* (pp. 337–348). Springer.
- Yan, Y., Chen, J., & Shyu, M.-L. (2018). Efficient Large-Scale Stance Detection in Tweets. *International Journal of Multimedia Data Engineering and Management*, 9(3), 1–16. <https://doi.org/10.4018/ijmdem.2018070101>.
- Yin, W., Alkhalifa, R., & Zubiaga, A. (2021). The emojification of sentiment on social media: Collection and analysis of a longitudinal twitter sentiment dataset. arXiv:2108.13898.
- Zhang, B., Yang, M., Li, X., Ye, Y., Xu, X., & Dai, K. (2020). Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3188–3197). Association for Computational Linguistics. <https://www.aclweb.org/anthology/2020.acl-main.291>.
- Zubiaga, A. (2018). A longitudinal assessment of the persistence of twitter datasets. *Journal of the Association for Information Science and Technology*, 69(8), 974–984.
- Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, 51(2), 1–36.

- Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., Lukasik, M., Bontcheva, K., Cohn, T., & Augenstein, I. (2018). Discourse-aware rumour stance classification in social media using sequential classifiers. *Information Processing & Management*, *54*(2), 273–290.
- Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS one*, *11*(3), e0150989.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.