

Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research

Soujanya Poria^{†*}, Devamanyu Hazarika[‡], Navonil Majumder[†], Rada Mihalcea[Ⓢ]

[†]Information Systems Technology and Design, Singapore University of Technology and Design, Singapore

[‡] School of Computing, National University of Singapore, Singapore

[Ⓢ] Electrical Engineering and Computer Science, University of Michigan, Michigan, USA

WE DEDICATE THIS PAPER TO THE MEMORY OF **PROF. JANYCE WIEBE**,
WHO HAD ALWAYS BELIEVED IN THE FUTURE OF THE FIELD OF SENTIMENT ANALYSIS.

Abstract—Sentiment analysis as a field has come a long way since it was first introduced as a task nearly 20 years ago. It has widespread commercial applications in various domains like marketing, risk management, market research, and politics, to name a few. Given its saturation in specific subtasks — such as sentiment polarity classification — and datasets, there is an underlying perception that this field has reached its maturity. In this article, we discuss this perception by pointing out the shortcomings and under-explored, yet key aspects of this field that are necessary to attain *true* sentiment understanding. We analyze the significant leaps responsible for its current relevance. Further, we attempt to chart a possible course for this field that covers many overlooked and unanswered questions.

Index Terms—Natural Language Processing, Sentiment Analysis, Emotion Recognition, Aspect Based Sentiment Analysis.



1 INTRODUCTION

SENTIMENT analysis, also known as opinion mining, is a research field that aims at understanding the underlying sentiment of unstructured content. E.g., in this sentence, “John dislikes the camera of iPhone 7”, according to the technical definition (Liu, 2012) of sentiment analysis, John plays the role of the *opinion holder* exposing his negative sentiment towards the *aspect* – camera of the *entity* – iPhone 7. Since its early beginnings (Pang et al., 2002; Turney, 2002), sentiment analysis has established itself as an influential field of research with widespread applications in industry. The ever increasing popularity and demand stem from the interest of individuals, businesses, and governments in understanding people’s views about products, political agendas, or marketing campaigns. Public opinion also stimulates market trends, which makes it relevant for financial predictions. Furthermore, education and healthcare sectors make use of sentiment analysis for behavioral analysis of students and patients.

Over the years, the scope for innovation and commercial demand have jointly driven research in sentiment analysis. However, over the past few years, there has been an emerging perception that the problem of sentiment analysis is merely a text/content categorization task – one that requires content to be classified into two or three categories of sentiments: positive, negative, and/or neutral. This has led

to a belief among researchers that sentiment analysis has reached its saturation. Through this work, we set forth to address this misconception.

Figure 1 shows that many benchmark datasets on the polarity detection subtask of sentiment analysis, like IMDB or SST-2, have reached saturation points, as reflected by the near perfect scores achieved by many modern data-driven methods. However, this does not imply that sentiment analysis is solved. Rather, we believe that this perception of saturation has manifested from excessive research publications focusing only on shallow sentiment understanding, such as, k-way text classification whilst ignoring other key un- and under-explored problems relevant to this field of research.

Liu (2015) presents sentiment analysis as *mini-NLP*, given its reliance on topics covering almost the entirety of NLP. Similarly, Cambria et al. (2017) characterize sentiment analysis as a big suitcase of subtasks and subproblems, involving open syntactic, semantic, and pragmatic problems. As such, there remains a number of open research directions to be extensively studied, such as understanding motive and cause of sentiment; sentiment dialogue generation; sentiment reasoning; and so on. At its core, effective inference of sentiment requires understanding of multiple fundamental problems in NLP. These include assigning polarities to aspects, negation handling, resolving co-references, and identifying syntactic dependencies to exploit sentiment flow. Sentiment analysis is also influenced by the figurative nature of language which is often exploited using linguistic devices, such as, sarcasm and irony. This complex composition of multiple tasks makes sentiment analysis a challenging yet interesting research space.

* Corresponding author (e-mail: sporia@sutd.edu.sg)

- S. Poria can be contacted at sporia@sutd.edu.sg
- D. Hazarika can be contacted at hazarika@comp.nus.edu.sg
- N. Majumder can be contacted at navonil_majumder@sutd.edu.sg
- R. Mihalcea can be contacted at mihalcea@umich.edu

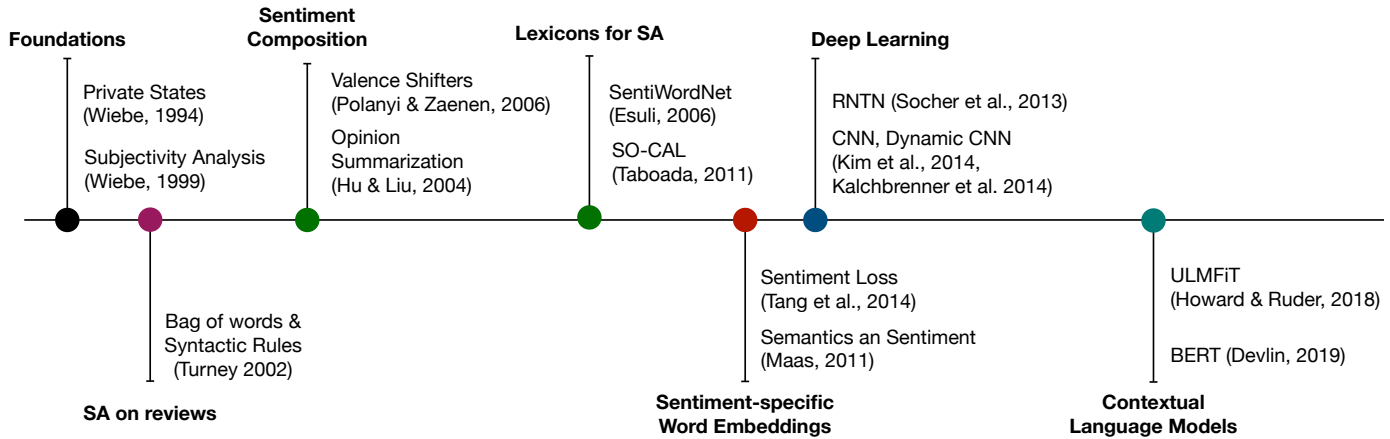


Fig. 3: A non-exhaustive illustration of some of the milestones of sentiment analysis research.

analysis, in particular, saw an increase in complex neural architectures that involve attention mechanisms (Wang et al., 2016), memory networks (Tang et al., 2016b) and adversarial learning (Karimi et al., 2020; Chen et al., 2018). For a comprehensive review of modern deep learning architectures, please refer to (Zhang et al., 2018a).

Although the majority of the works employing deep networks rely on automated feature learning, their heavy reliance on annotated data is often limiting. As a result, providing inductive biases via syntactic information, or external knowledge in the form of lexicons as additional input has seen a resurgence (Tay et al., 2018b).

As seen in Figure 1, the recent works based on neural architectures (Le & Mikolov, 2014; Dai & Le, 2015; Johnson & Zhang, 2016; Miyato et al., 2017; McCann et al., 2017; Howard & Ruder, 2018; Xie et al., 2019; Thongtan & Phientrakul, 2019) have dominated over traditional machine learning models (Maas et al., 2011; Wang & Manning, 2012). Similar trends can be observed in other benchmark datasets such as Yelp, SST (Socher et al., 2013), and Amazon Reviews (Zhang et al., 2015). Within neural methods, much like other fields of NLP, present trends are dominated by the contextual encoders, which are pre-trained as language models using the Transformer architecture (Vaswani et al., 2017). Models like BERT, XLNet, RoBERTa, and their adaptations have achieved the state-of-the-art performances on multiple sentiment analysis datasets and benchmarks (Hoang et al., 2019; Munikar et al., 2019; Raffel et al., 2019). Despite this progress, it is still not clear as to whether these new models learn the composition semantics associated to sentiment or simply learn surface patterns (Rogers et al., 2020).

Sentiment-Aware Word Embeddings: One of the critical building blocks of a deep-learning architecture is its word embeddings. It is known that word representations rely on the task it is being used for (Labutov & Lipson, 2013), however, most sentiment analysis-based models rely on generic word representations. Tang et al. (2014) proposed an important work in this direction that provided word representations tailored for sentiment analysis. While general embeddings mapped words with similar syntactic context into nearby representations, this work incorporated sentiment information into the learning loss to account for sentiment

regularities. Although the community has proposed some approaches in this topic (Maas et al., 2011; Bespalov et al., 2011), promising traction has been limited (Tang et al., 2015). Further, with the popularity of contextual models such as BERT, it remains to be seen how can sentiment information be incorporated into its embeddings.

Sentiment Analysis in Micro-blogs: Sentiment analysis in micro-blogs, such as Twitter, require different processing techniques compared to traditional text pieces. Being limited in length, one of the positives is that user’s tend to express their opinion in a straightforward manner. However, cases of sarcasm and irony often challenge these systems. Tweets are rife with internal slangs, abbreviations, and emoticons – which adds to the complexity for mining the opinions in them. Moreover, the limited length restricts the presence of contextual cues normally present in dialogues or documents (Kharde & Sonawane, 2016).

From a data point of view, opinionated data is found in abundance in these micro-blogs. Reflections of this has been observed in the recent benchmark shared tasks that has been mostly based on Twitter data. These include Semeval shared tasks for sentiment analysis, aspect based sentiment analysis and figurative language in Twitter ^{1, 2, 3, 4}.

A new trend amongst users in Twitter is the concept of daisy-chaining multiple tweets to compose a longer piece of text. Existing research, however, has not addressed this phenomena to acquire additional context. Future work on twitter sentiment analysis could be benefited from analyzing personality of the users based on the their historical tweets.

3 OPTIMISTIC FUTURE: UPCOMING TRENDS IN SENTIMENT ANALYSIS

The previous section highlighted some of the milestones in sentiment analysis research, which helped developing the field into its present state. Despite the progress, we believe, the problems are far from solved along with the emergence of new problems and applications. In this section, we take an optimistic view on the road ahead in sentiment analysis

1. <http://alt.qcri.org/semeval2015/task10/>

2. <http://alt.qcri.org/semeval2015/task12/>

3. <http://alt.qcri.org/semeval2015/task11/>

4. <https://sites.google.com/view/figlang2020/>

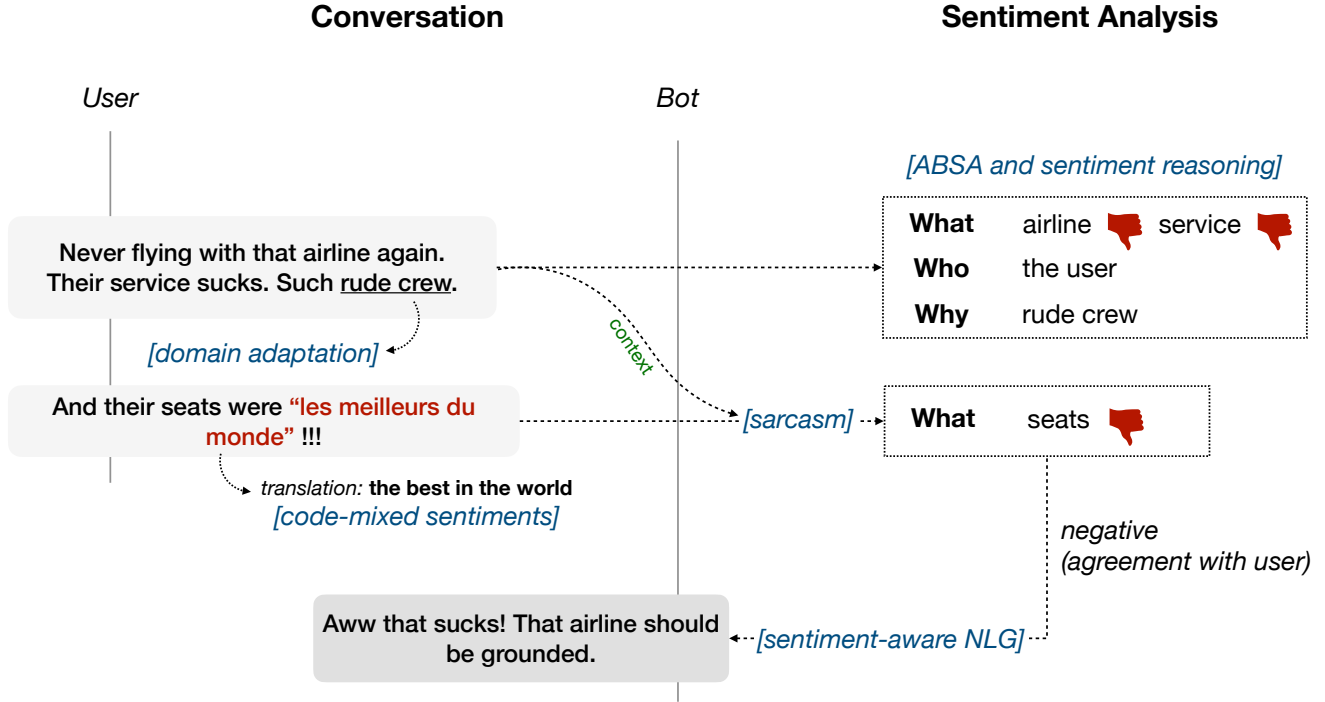


Fig. 4: The example illustrates the various challenges and applications that holistic sentiment analysis depends on.

research and highlight several applications rife with open problems and challenges.

Applications of sentiment analysis take form in many ways. Section 2.3 presents one such example where a user is chatting with a chit-chat style chatbot. In the conversation, to come up with an appropriate response, the bot needs an understanding of the user’s opinion. This involves multiple sub-tasks that include 1) extracting aspects like service, seats for the entity airline, 2) aspect-level sentiment analysis along with knowing 3) *who* holds the opinion and *why* (sentiment reasoning). Added challenges include analyzing code-mixed data (e.g. “les meilleurs du monde”), understanding domain-specific terms (e.g., rude crew), and handling sarcasm – which could be highly contextual and detectable only when preceding utterances are taken into consideration. Once the utterances are understood, the bot now has to determine appropriate response-styles and perform controlled-NLG based on the decided sentiment. The overall example demonstrates the dependence of sentiment analysis on these applications and sub-tasks, some of which are new and still at early stages of development. We discuss these applications next.

3.1 Aspect-Based Sentiment Analysis

Although sentiment analysis provides an overall indication of the author or speaker’s sentiments, it is often the case when a piece of text comprises of multiple aspects with varied sentiments associated to them. Take for example the following sentence “*This actor is the only failure in an otherwise brilliant cast.*”. Here, the opinion is attached to two particular entities, *actor* (negative opinion) and *cast* (positive opinion). Additionally, there is also an absence of an overall opinion that could be assigned to the full sentence.

Aspect-based Sentiment Analysis (ABSA) takes such fine-grained view and aims to identify the sentiments towards each entity (and/or their aspects) (Liu, 2015; Liu & Zhang, 2012). The problem involves two major sub-tasks, 1) *Aspect-extraction*, which identifies the aspects⁵ mentioned within a given sentence or paragraph (actor and cast in the above example) 2) *Aspect-level Sentiment Analysis* (aspect-level sentiment analysis), which determines the sentiment orientation associated with the corresponding aspects/ opinion targets (actor \mapsto negative and cast \mapsto positive) (Hu & Liu, 2004a). Proposed approaches for aspect extraction include rule-based strategies (Qiu et al., 2011; Liu et al., 2015), topic models (Mei et al., 2007; He et al., 2011), and more recently, sequential models such as CRFs (Shu et al., 2017). For aspect-level sentiment analysis, the algorithms primarily aim to model the relationship between the opinion targets and their context. To achieve this, models based on CNNs (Li & Lu, 2017), memory networks (Tay et al., 2017), and so on have been explored. Primarily, the associations have been learnt through attention mechanism (Wang et al., 2016).

Despite the advances in this field, there remain many factors which are open for research and hold the potential to improve performances further. We discuss them below.

3.1.1 Aspect-Term Auto-Categorization

Aspect-terms extraction is the first step towards aspect-level sentiment analysis. This task has been studied rigorously in the literature (Poria et al., 2016). Thanks to the advent of deep sequential learning, the performance of this task on the benchmark datasets (Hu & Liu, 2004b; Pontiki et al.,

5. In the context of aspect-based sentiment analysis, *aspect* is the generic term utilized for topics, entities, or their attributes/features. They are also known as opinion targets.

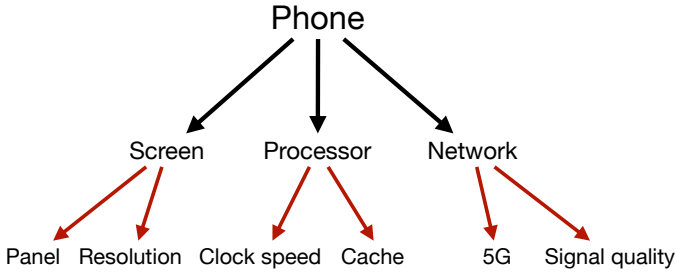


Fig. 5: An example of the aspect term auto-categorization.

2016) has reached a new level. Aspect terms are needed to be categorized into aspect groups to present a coherent view of the expressed opinions. We illustrate this categorization in Fig. 5. Approaches to aspect-term auto-categorization are mostly supervised and unsupervised topic classification based and lexicon driven. All these three types of approaches succumb to scalability issues when subjected to new domains with novel aspect categories. We believe that entity linking-based approaches, coupled with semantic graphs like Probase (Wu et al., 2012), should be able to perform reasonably while overcoming scalability issues. For example, the sentence “With this phone, I always have hard time getting signal indoors.” contains one aspect term *signal*, that can be passed to an entity linker — on a graph containing a tree shown in Fig. 5 — with the surrounding words as context to obtain aspect category *phone:signal-quality*.

3.1.2 Implicit Aspect-Level Sentiment Analysis

Sentiment can be expressed implicitly on aspects. Although under-studied, the importance of detecting implicit aspect level sentiment can not be ignored as they represent a unique nature of the natural language. For example, in the sentence “Oh no! Crazy Republicans voted against this bill”, the speaker expresses her/his negative sentiment on the *republicans* explicitly. By doing so, we can infer that the speaker’s sentiment towards the *bill* is positive. In the work by Deng et al. (2014), it is called as *opinion-oriented implicatures*. Approaches (Deng et al., 2014; Deng & Wiebe, 2014) to this problem primarily focus on linear programming, logic rules, and belief propagation in the graph network. Major attention from the neural network community to address this research problem is yet to be witnessed.

3.1.3 Aspect Term-Polarity Co-Extraction

Most existing algorithms in this area consider aspect extraction and aspect-level sentiment analysis as sequential (pipelined) or independent tasks. In both these cases, the relationship between the tasks is ignored. Efforts towards joint learning of these tasks have gained traction in recent trends. These include hierarchical neural networks (Lakkaraju et al., 2014), multi-task CNNs (Wu et al., 2016), and CRF-based approaches by framing both the sub-tasks as sequence labeling problems (Li et al., 2019a; Luo et al., 2019). The notion of joint learning opens up several avenues for exploring the relationships between both the sub-tasks and also, possible dependencies from other tasks. This strategy is adopted by transfer learning approaches, which we discuss next.

3.1.4 Transfer Learning in Aspect-Based Sentiment Analysis (ABSA)

Much like the recent trends in the overall field of NLP, transfer learning approaches such as BERT have shown potential in aspect-based sentiment analysis too (Huang & Carley, 2019). Simple baselines utilizing BERT has demonstrated competitive performances against sophisticated state-of-the-art methods (Li et al., 2019b) and also in out-of-domain settings (Hoang et al., 2019). These trends indicate the role of semantic understanding for the task of aspect-based sentiment analysis. What remains to be seen is the future role of BERT-based networks working in conjunction with the task-dependent designs existing as the present state of the arts in this area (Sun et al., 2019).

Knowledge can also be transferred from one sentiment task to another. E.g., aspect extraction can be utilized as a scaffolding for aspect-based sentiment analysis as these two tasks are correlated. It would also be interesting to transfer knowledge from textual to multimodal ABSA system.

3.1.5 Exploiting Inter-Aspect Relations for Aspect-Level Sentiment Analysis

The primary focus of algorithms proposed for aspect-level sentiment analysis has been to model the dependencies between opinion targets and their corresponding opinionated words in the context (Tang et al., 2016a). Besides, modeling the relationships between aspects also holds potential in this task (Hazarika et al., 2018c). For example, in the sentence “my favs here are the tacos pastor and the tostada de tinga”, the aspects “*tacos pastor*” and “*tostada de tinga*” are connected using conjunction “and” and both rely on the sentiment bearing word “*favs*”. Understanding such inter-aspect dependency can significantly aid the aspect-level sentiment analysis performance and remains to be researched extensively.

3.1.6 Quest for Richer and Larger Datasets

The two widely used publicly available datasets for aspect-based sentiment analysis are Amazon product review (Hu & Liu, 2004b) and Semeval 2017 (Pontiki et al., 2016) datasets. Both these datasets are quite small in size that hinders any statistically significant improvement in performance when comparing the methods that utilize these datasets.

3.2 Multimodal Sentiment Analysis

The majority of research works on sentiment analysis have been conducted using only textual modality. However, with the increasing number of user-generated videos available on online platforms such as YouTube, Facebook, Vimeo, and others, multimodal sentiment analysis has emerged at the forefront of sentiment analysis research. The commercial interests fuel this rise as the enterprises tend to make business decisions on their products by analyzing user sentiments in these videos. Figure 6 presents examples where the presence of multimodal signals in addition to the text itself is necessary in order to make correct predictions of their emotions and sentiments. Multimodal fusion is at the heart of multimodal sentiment analysis with an increasing number of works proposing new fusion techniques. These include Multiple Kernel Learning, tensor-based non-linear



Utterance: “Become a drama critic!”

Emotion: Joy **Sentiment:** Positive

Text	Audio	Visual
Ambiguous	Joyous tone	Smiling Face



Utterance: “Great, now he is waving back”

Emotion: Disgust **Sentiment:** Negative

Text	Audio	Visual
Positive/Joy	Flat tone	Frown

Fig. 6: Importance of multimodal cues. Green shows primary modalities responsible for sentiment and emotion.

fusion (Zadeh et al., 2017), memory networks (Zadeh et al., 2018a), amongst others. The granularity at which such fusion methods are applied also varies – from word-level to utterance-level.

Below, we identify three key directions that can aid future research:

3.2.1 Complex Fusion Methods vs Simple Concatenation

Multimodal information fusion is a core component of multimodal sentiment analysis. Although several fusion techniques (Zadeh et al., 2018c,a, 2017) have been recently proposed, in our experience, a simple concatenation-based fusion method performs at par with most of these methods. We believe these methods are unable to provide significant improvements in the fusion due to their inability to model correlations among different modalities and handle noise. Reliable fusion remains as a major future work.

3.2.2 Lack of Large Datasets

The field of multimodal sentiment analysis also suffers from the lack of larger datasets. The available datasets, such as MOSI (Zadeh et al., 2016), MOSEI (Zadeh et al., 2018b), MELD (Poria et al., 2018) are not large enough and carry suboptimal inter-annotator agreement that impedes the performance of complex deep learning frameworks.

3.2.3 Fine-Grained Annotation

The primary goal of multimodal fusion is to accumulate the contribution from each modality. However, measuring that contribution is not trivial as there is no available dataset that annotates the individual role of each modality. We show one such example in Figure 6, where each modality is labeled with the sentiment it carries. Having such rich fine-grained annotations should better guide multimodal fusion methods and make them more interpretable. This fine-grained annotation can also open the door to the new types of multimodal fusion approaches.

3.3 Contextual Sentiment Analysis

3.3.1 Influence of Topics

The usage of sentiment words varies from one topic to another. Words that sound neutral on the surface can bear

sentiment when conjugated with other words or phrases. For example, the word *big* in *big house* can carry positive sentiment when someone intends to purchase a *big house* for leisure. The same word, however, could evoke negative sentiments when used in the context – *A big house is hard to clean*. Unfortunately, research in sentiment analysis has not focused much on this aspect. The sentiment of some words can be vague and specified only when seen in context, e.g., the word *massive* in the context of *massive earthquake* and *massive villa*. In the future, a dataset composed of such contextual sentiment bearing phrases would be a great contribution to the research community.

This research problem is also related to word sense disambiguation. Below we present an example, borrowed from the work by Choi et al. (2017):

- The Federal Government *carried* the province for many years.
- The troops *carried* the town after a brief fight.

In the first sentence, the sense of *carry* has a positive polarity. However, in the second sentence, the same word has negative connotations. Hence, depending on the context, sense of words and their polarities can change. In (Choi et al., 2017), the authors adopted topic models to associate word senses with sentiment. As this particular research problem widens its scope to the task of word sense disambiguation, it would be useful to employ contextual language models to decipher word senses in contexts and assign the corresponding polarity.

3.3.2 Sentiment Analysis in Monologues and Conversational Context

Context is at the core of NLP research. According to several recent studies (Peters et al., 2018; Devlin et al., 2019), contextual sentence and word-embeddings can improve the performance of the state-of-the-art NLP systems by a significant margin.

The notion of context can vary from problem to problem. For example, while calculating word representations, the surrounding words carry contextual information. Likewise, to classify a sentence in a document, other neighboring sentences are considered as its context. Poria et al. (2017) utilize surrounding utterances in a video as context and experimentally show that contextual evidence indeed aids in classification.

There have been very few works on inferring implicit sentiment (Deng & Wiebe, 2014) from context. This is crucial for achieving true sentiment understanding. Let us consider this sentence “*Oh no. The bill has been passed*”. As there are no explicit sentiment markers present in the sentence – “*The bill has been passed*”, it would sound like a neutral sentence. Consequently, the sentiment behind ‘bill’ is not expressed by any particular word. However, considering the sentence in the context – “*Oh no*”, which exhibits negative sentiment, it can be inferred that the opinion expressed on the ‘bill’ is negative. The inferential logic that one requires to arrive at such conclusions is the understanding of sentiment flow in the context. In this particular example, the contextual sentiment of the sentence – “*Oh no*” flows to the next sentence and thus making it a negative opinionated sentence. Tackling such tricky and fine-grained cases require bespoke

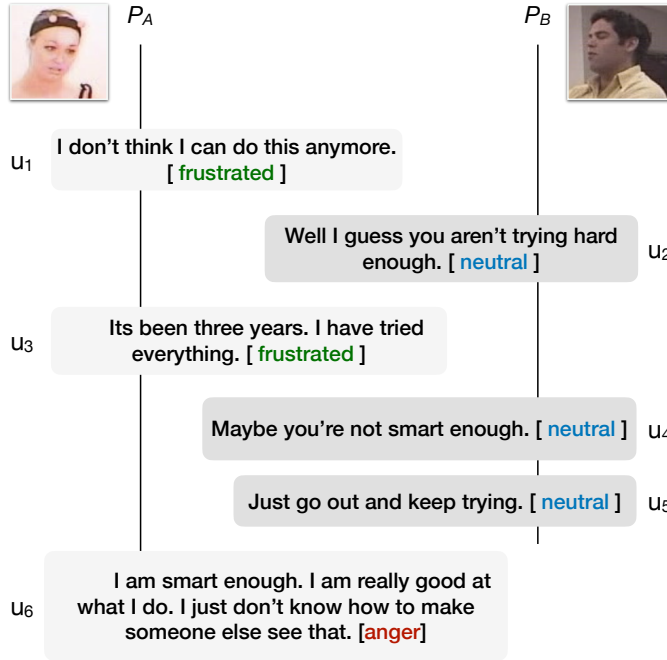


Fig. 7: An abridged dialogue from the IEMOCAP dataset (Busso et al., 2008).

modeling and datasets containing an ample quantity of such non-trivial samples. Further, commonsense knowledge can also aid in making such inferences. In the literature (Poria et al., 2017), the use of LSTMs to model such sequential sentiment flow has been ineffectual. We think it would be fruitful to utilize logic rules, finite-state transducers, belief, and information propagation mechanisms to address this problem. We also note that contextual sentences may not always help. Hence, one can ponder the use of a gate or switch to learn and further infer when to count on contextual information.

In conversational sentiment-analysis, to determine the emotions and sentiments of an utterance at time t , the preceding utterances at time $< t$ can be considered as its context. However, computing this context representation can often be difficult due to complex sentiment dynamics.

Sentiments in conversations are deeply tied with emotional dynamics consisting of two important aspects: *self* and *inter-personal dependencies* (Morris & Keltner, 2000). Self-dependency, also known as *emotional inertia*, deals with the aspect of influence that speakers have on themselves during conversations (Kuppens et al., 2010). On the other hand, inter-personal dependencies relate to the sentiment-aware influences that the counterparts induce into a speaker. Conversely, during the course of a dialogue, speakers also tend to mirror their counterparts to build rapport (Navarretta et al., 2016). This phenomenon is illustrated in Figure 7. Here, P_a is frustrated over her long term unemployment and seeks encouragement (u_1, u_3). P_b , however, is pre-occupied and replies sarcastically (u_4). This enrages P_a to appropriate an angry response (u_6). In this dialogue, self-dependencies are evident in P_b , who does not deviate from his nonchalant behavior. P_a , however, gets sentimentally influenced by P_b . Modeling self and inter-personal relation-

ships and dependencies may also depend on the topic of the conversation as well as various other factors like argument structure, interlocutors personality, intents, viewpoints on the conversation, attitude towards each other, and so on. Hence, analyzing all these factors is key for a true self and inter-personal dependency modeling that can lead to enriched context understanding.

The contextual information can come from both local and distant conversational history. As opposed to the local context, distant context often plays a lesser important role in sentiment analysis of conversations. Distant contextual information is useful mostly in the scenarios when a speaker refers to earlier utterances spoken by any of the speakers in the conversational history.

The usefulness of context is more prevalent in classifying short utterances, like *yeah, okay, no*, that can express different sentiment depending on the context and discourse of the dialogue. The examples in Figure 8 explain this phenomenon. The sentiment expressed by the same utterance “*Yeah*” in both these examples differ from each other and can only be inferred from the context.

Leveraging such contextual clues is a difficult task. Memory networks, RNNs, and attention mechanisms have been used in previous works, e.g., HRLCE (Huang et al., 2019a) or DialogueRNN (Majumder et al., 2019), to grasp information from the context. However, these models fail to explain the situations where contextual information is needed. Hence, finding contextualized conversational utterance representations is an active area of research.

3.3.3 User, Cultural, and Situational Context

Sentiment also depends on the user, cultural, and situational context.

Individuals have subtle ways of expressing emotions and sentiments. For instance, some individuals are more sarcastic than others. For such cases, the usage of certain words would vary depending on if they are being sarcastic. Let’s consider this example, P_a : *The order has been cancelled.*, P_b : *This is great!*. If P_b is a sarcastic person, then his response would express negative emotion to the order being canceled through the word *great*. On the other hand, P_b ’s response, *great*, could be taken literally if the canceled order is beneficial to P_b (perhaps P_b cannot afford the product he ordered). As necessary background information is often missing from the conversations, speaker profiling based on preceding utterances often yields improved results.

The underlying emotion of the same word can vary from one person to another. E.g., the word *okay* can bear different sentiment intensity and polarity depending on the speaker’s character. This incites the need to do user profiling for fine-grained sentiment analysis, which is a necessary task for e-commerce product review understanding.

Understanding sentiment also requires cultural and situational awareness. A *hot* and *sunny* weather can be treated as a good weather in USA but certainly not in Singapore. Eating *ham* could be accepted in one religion and prohibited by another.

A basic sentiment analysis system that only relies on distributed word representations and deep learning frameworks are susceptible to these examples if they do not encompass rudimentary contextual information.

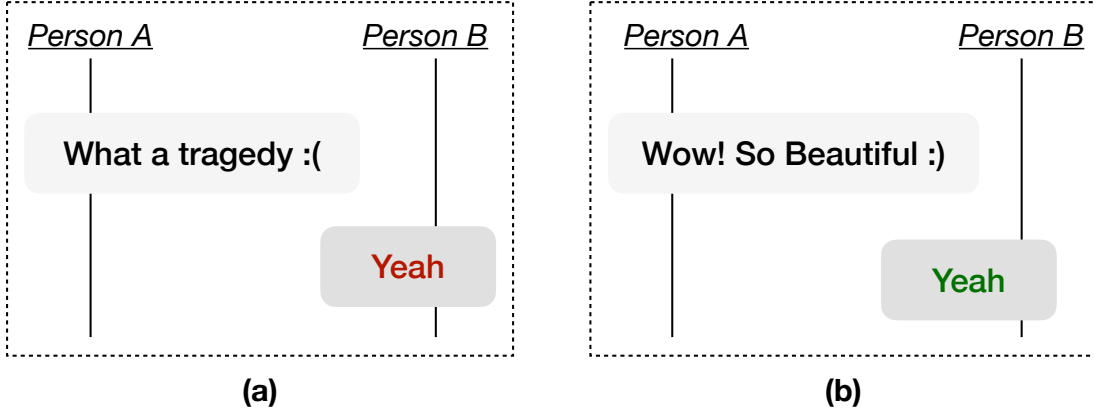


Fig. 8: Role of context in sentiment analysis in conversation.

3.3.4 Role of Commonsense Knowledge in Sentiment Analysis

In layman's term, commonsense knowledge consists of facts that all human beings are expected to know. Due to this characteristic, humans tend to ignore expressing commonsense knowledge explicitly. As a result, word embeddings trained on the human-written text do not encode such trivial yet important knowledge that can potentially improve language understanding. The distillation of commonsense knowledge, thus, has become a new trend in modern NLP research. We show one such example in the Fig. 9 which illustrates the latent commonsense concepts that humans easily infer or discover given a situation. In particular, the present scenario informs that David is a good cook and will be making pasta for some people. Based on this information, commonsense can be employed to infer related events such as, dough for the pasta would be available, people would eat food (pasta), the pasta is expected to be good (David is good cook), etc. These inferences would enhance the text representation with many more concepts that can be utilized by neural systems in diverse downstream tasks.

In the context of sentiment analysis, utilizing commonsense for associating aspects with their sentiments can be highly beneficial for this task. Commonsense knowledge graphs connect the aspects to various sentiment-bearing concepts via semantic links (Ma et al., 2018). Additionally, semantic links between words can be utilized to mine associations between the opinion target and the opinion-bearing word. What is the best way to grasp commonsense knowledge is still an open research question.

Commonsense knowledge is also required to understand *implicit* sentiment of the sentences that do not accommodate any explicit sentiment marker. E.g., the sentiment of the speaker in this sentence, "We have not seen the sun since last week" is negative as not catching the sight of the sun for a long time is generally treated as a negative event in our society. A system not adhering to this commonsense knowledge would fail to detect the underlying sentiment of such sentences correctly.

With the advent of commonsense modeling algorithms such as Comet (Bosselut et al., 2019), we think, there will be a new wave of research focusing on the role of commonsense knowledge in sentiment analysis in the near future.

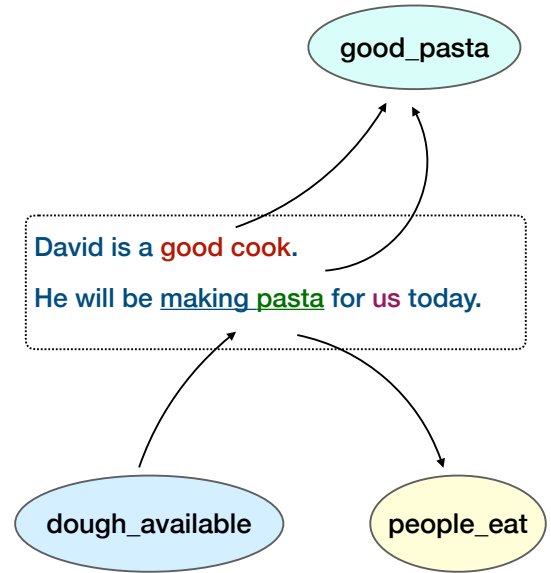


Fig. 9: An illustration of commonsense reasoning and inference.

3.4 Sentiment Reasoning

Apart from exploring the *what*, we should also explore the *who* and *why*. Here, the *who* detects the entity whose sentiment is being determined, whereas *why* reveals the stimulus/reason for the sentiment.

3.4.1 Who? The Opinion Holder

While analyzing opinionated text, it is often important to know the opinion holder. In most of the cases, the opinion holder is the person who spoke/wrote the sentence. Yet, there can be situations where the opinion holder is an entity (or entities) mentioned in the text (Mohammad, 2017). Consider the following two lines of opinionated text:

- The movie was too slow and boring.
- Stella found the movie to be slow and boring.

In both the sentences above, the sentiment attached to the movie is negative. However, the opinion holder for the first sentence is the speaker while in the second sentence it is Stella. The task could be further complex with the need to

map varied usage of the same entity term (e.g. Jonathan, John) or the use of pronouns (he, she, they) (Liu, 2012).

Many works have studied the task of opinion-holder identification – a subtask of opinion extraction (opinion holder, opinion phrase, and opinion target identification). These works include approaches that use named-entity recognition (Kim & Hovy, 2004), parsing and ranking candidates (Kim & Hovy, 2006), semantic role labeling (Wiegand & Ruppenhofer, 2015), structured prediction using CRFs (Choi et al., 2006), multi-tasking (Yang & Cardie, 2013), amongst others. The MPQA corpus (Deng & Wiebe, 2015) provided supervised annotations for this task. However, with respect to deep learning approaches, this topic has been understudied (Zhang et al., 2019; Quan et al., 2019).

3.4.2 Why? The Sentiment Stimulus

The majority of the sentiment analysis research works to date are about classifying contents into positive, negative and neutral. This oversimplification of the sentiment analysis task has resulted in the saturation of any major breakthrough. The future research in sentiment analysis should focus on what drives a person to express positive or negative sentiment on a topic or aspect.

To reason about a particular sentiment of an opinion-holder, it is important to understand the target of the sentiment (Deng & Wiebe, 2014), and whether there are implications of holding such sentiment. For instance, when stating “I am sorry that John Doe went to prison.”, understanding the target of the negative sentiment is “John Doe goes to prison”, and knowing that “go to prison” has negative implications on the actor John Doe, it implies positive sentiment toward John Doe.⁶ Moreover, it is important to understand what caused the sentiment. Li & Hovy (2017) discuss two possible reasons that give rise to opinions. Firstly, an opinion-holder might have an emotional bias towards the entity/topic in question. Secondly, the sentiment could be borne out of mental (dis)satisfaction towards a goal achievement.

The ability to reason is necessary for any explainable AI system. In the context of sentiment analysis, it is often desired to understand the cause of an expressed sentiment by the speaker. E.g. in a review on a smartphone, the speaker might dislike it because the battery drains so fast. While it is important to detect the negative sentiment expressed on *battery*, digging into the detail that causes this sentiment is also of prime importance (Liu, 2012). Till date, there is not much work exploring this aspect of the sentiment analysis research.

Grasping the cause of sentiment is also very important in dialogue systems. As an example, we can refer to Figure 10, Joey expresses *anger* once he ascertains Chandler’s *deception* in the previous utterance.

It is hard to define a taxonomy or tagset for the reasoning of both emotions and sentiments. At present, there is no available dataset which contains such rich annotations. Building such dataset would enable future dialogue systems to frame meaningful argumentation logic and discourse structure, taking one step closer to human-like conversation.

6. Example provided by Jan Wiebe (2016), personal communication.

3.5 Domain Adaptation

Most of the state-of-the-art sentiment analysis models enjoy the privilege of having in-domain training datasets. However, this is not a viable scenario as curating large amounts of training data for every domain is impractical. Domain adaptation in sentiment analysis solves this problem by learning the characteristics of the unseen domain. Sentiment classification, in fact, is known to be sensitive towards domains as mode of expressing opinions across domains vary. Also, valence of affective words may vary based on different domains (Liu, 2012).

Diverse approaches have been proposed for cross-domain sentiment analysis. One line of work models domain-dependent word embeddings (Sarma et al., 2018; Shi et al., 2018; K Sarma et al., 2019) or domain-specific sentiment lexicons (Hamilton et al., 2016), while others attempt to learn representations based on either co-occurrences of domain-specific with domain-independent terms (pivots) (Blitzer et al., 2007; Pan et al., 2010; Ziser & Reichart, 2018; Sharma et al., 2018) or shared representations using deep networks (Glorot et al., 2011).

One of the major breakthroughs in domain adaption research employs adversarial learning that trains to fool a domain discriminator by learning domain-invariant representations (Ganin et al., 2016). In this work, the authors utilize bag of words as the input features to the network. Incorporating bag of words features limits the network to get access to any external knowledge about the unseen words of the target domain. Hence, the performance improvement can be completely attributed to the efficacy of the adversarial network. However, in recent works, researchers tend to utilize distributed word representations such as Glove, BERT. These representations, aka word embeddings, are usually trained on huge open-domain corpora and consequently contain domain invariant information. Future research should explain whether the gain in domain adaptation performance comes from these word embeddings or the core network architecture.

In summary, the works in domain adaptation lean towards outshining the state of the art on benchmark datasets. What remains to be seen is the interpretability of these methods. Although some of the works claim to learn the domain-dependent sentiment orientation of the words during domain invariant training, there is barely any well-defined analysis to validate such claims.

3.5.1 Use of External Knowledge

The key idea that most of the existing works encapsulate is to learn domain-invariant shared representations as a means to domain adaptation. While global or contextual word embeddings have shown their efficacy in modeling domain-invariant and specific representations, it might be a good idea to couple these embeddings with multi-relational external knowledge graphs for domain adaptation. Multi-relation knowledge graphs represent semantic relations between concepts. Hence, they can contain complementary information over the word embeddings, such as Glove, since these embeddings are not trained on explicit semantic relations. Semantic knowledge graphs can establish relationships between domain-specific concepts of several

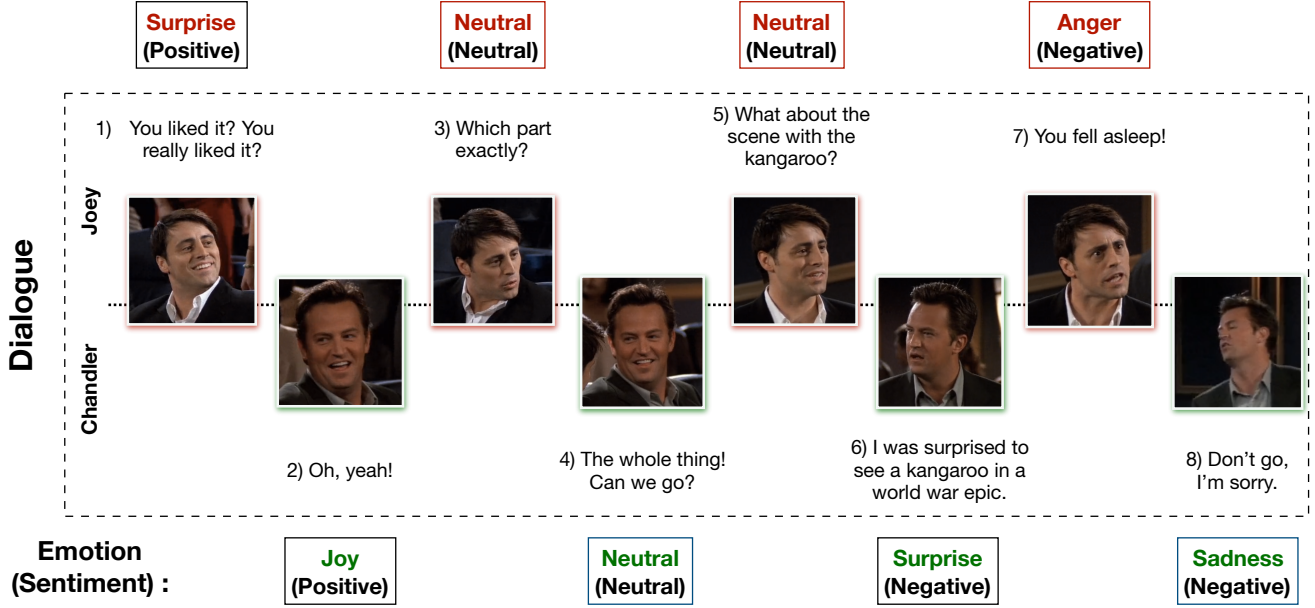
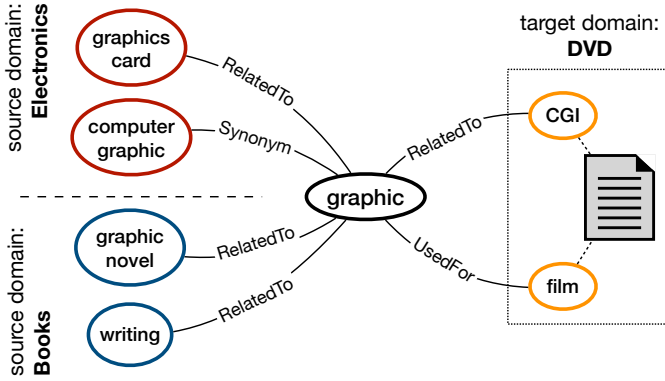


Fig. 10: Sentiment cause analysis.

domains using domain-general concepts – providing vital information that can be exploited for domain adaptation. One such example is presented in Fig. 11. Researchers are encouraged to read these early works (Alam et al., 2018; Xiang et al., 2010) on exploiting external knowledge for domain adaptation.

Fig. 11: Domain-general term *graphic* bridges the semantic knowledge between domain specific terms in Electronics, Books and DVD.

3.5.2 Scaling Up to Many Domains

Most of the present works in this area use the setup of a *source* and *target* domain pair for training. Although appealing, this setup requires retraining as and when the target domain changes. The recent literature in domain adaptation goes beyond single-source-target (Zhao et al., 2018) to multi-source and multi-target (Gholami et al., 2020b,a) training. However, in sentiment analysis, these setups have not been fully explored and deserve more attention (Wu & Huang, 2016).

3.6 Multilingual Sentiment Analysis

Majority of the research work on sentiment analysis has been conducted using English datasets. However, the advent of social media platforms has made multilingual content available via platforms such as Facebook and Twitter. Consequently, there is a recent surge in works with diverse languages (Dashtipour et al., 2016). The NLP community, in general, is now also vocal to promote research on languages other than English.⁷

In the context of sentiment analysis, despite the recent surge in multilingual sentiment analysis, several directions need more traction:

3.6.1 Language Specific Lexicons

Today's rule-based sentiment analysis system, such as Vader, works great for the English language, thanks to the availability of resources like sentiment lexicons. For other languages such as Hindi, French, Arabic, not many well-curated lexicons are available.

3.6.2 Sentiment Analysis of Code-Mixed Data

In many cultures, people on social media post content that are a mix of multiple languages (Lal et al., 2019; Guptha et al., 2020; Gambäck & Das, 2016). For example, "*Itna izzat diye aapne mujhe !!!* Tears of joy. :(:'", in this sentence, the bold text is in Hindi with roman orthography and the rest is in English. Code-mixing poses a significant challenge to the rule- and deep learning-based methods. A possible future work to combat this challenge would be to develop language models on code-mixed data. How and where to mix languages are a person's own choice, which is one of the main hardships. Another critical challenge associated with this task is to identify the deep compositional semantic

7. Because of a now widely known statement made by Professor Emily M. Bender on Twitter, we now use the term *#BenderRule* to require that the language addressed by research projects by explicitly stated, even when that language is English <https://bit.ly/3alqS0C>

Utterance

1) Chandler :

Oh my god! You almost gave me a heart attack!

- **Text** : suggests fear or anger.
- **Audio** : animated tone
- **Video** : smirk, no sign of anxiety



2) Sheldon :

Its just a *privilege* to watch your mind at work.

- **Text** : suggests a compliment.
- **Audio** : neutral tone.
- **Video** : straight face.



Fig. 12: Incongruent modalities in sarcasm present in the MUSTARD dataset (Castro et al., 2019).

that lies in the code mixed data. Unfortunately, only a little research has been carried out on this topic (Lal et al., 2019).

3.7 Sarcasm Analysis

The study of sarcasm analysis is highly integral to the development of sentiment analysis due to its prevalence in opinionated text (Maynard & Greenwood, 2014). Detecting sarcasm is highly challenging due to the figurative nature of text, which is accompanied by nuances and implicit meanings (Jorgensen et al., 1984). Over recent years, this field of research has established itself as an important problem in NLP with many works proposing different solutions to address this task (Joshi et al., 2017b). Broadly, the main contributions have emerged from the speech and text community. In speech, existing works leverage different signals such as prosodic cues (Bryant, 2010; Woodland & Voyer, 2011), acoustic features including low-level descriptors and spectral features (Cheang & Pell, 2008). Whereas in textual systems, traditional approaches consider rule-based (Khatti et al., 2015) or statistical patterns (González-Ibáñez et al., 2011b), stylistic patterns (Tsur et al., 2010), incongruity (Joshi et al., 2015; Tay et al., 2018a), situational disparity (Riloff et al., 2013), and hashtags (Maynard & Greenwood, 2014). While stylistic patterns, incongruity, and valence shifters are some of the ways that humans use to express sarcasm, it is also highly contextual. In addition, sarcasm also depends on a person’s personality, intellect and the ability to reason over commonsense. In the literature, these aspects of sarcasm remain under-explored.

3.7.1 Leveraging Context in Sarcasm Detection

Although the research for sarcasm analysis has primarily dealt with analyzing the sentence at hand, recent trends have started to acquire contextual understanding by looking beyond the text.

User Profiling and Conversational Context: Two types of contextual information have been explored for providing additional cues to detect sarcasm: *authorial context* and *conversational context*. Leveraging authorial context delves with analyzing the author’s sarcastic tendencies

(user profiling) by looking at their historical and meta data (Bamman & Smith, 2015). Similarly, the conversational context uses the additional information acquired from surrounding utterances to determine whether a sentence is sarcastic (Ghosh et al., 2018). It is often found that sarcasm is apparent only when put into context over what was mentioned earlier. For example, when tasked to identify whether the sentence *He sure played very well* is sarcastic, it is imperative to look at prior statements in the conversation to reveal facts (*The team lost yesterday*) or gather information about the speakers sincerity in making the current statement (*I never imagined he would be gone in the first minute*).

Given this contextual dependency, the question remains – *how can we model context efficiently?* The most popular approaches are based on sequential models e.g., LSTM (Poria et al., 2017) and doc2vec (Hazarika et al., 2018a). However, the results reported in these papers show only a minor improvement under the contextual setting. The quest for better contextual modeling is thus open – one that can explicitly understand facts and incongruity across sentences. These models are also not interpretable; hence, they fail to explain when and how they rely on the context.

Multimodal Context: Apart from gathering essential cues from the author and conversational context, we also identify *multimodal signals* to be important for sarcasm detection. Sarcasm is often expressed without linguistic markers, and instead, by using verbal and non-verbal cues. Change of tone, overemphasis on words, straight face, are some such cues that indicate sarcasm. There have been very few works that adopt multimodal strategies to determine sarcasm (Schifanella et al., 2016). Castro et al. (2019) recently released a multimodal sarcasm detection dataset that takes conversational context into account. Other works that consider multimodality focus on sarcasm perceived by the reader/audience. These works utilize textual features along with cognitive features such as gaze-behavior of readers (Mishra et al., 2016), electro/magneto-encephalographic (EEG/MEG) signals (Filik et al., 2014; Thompson et al., 2016). Figure 12 presents two cases where sarcasm is expressed through the incongruity between modalities. In the first case, the language modality indicates fear or anger. In contrast, the facial modality lacks any visible sign of anxiety that would agree with the textual modality. In the second case, the text is indicative of a compliment, but the vocal tonality and facial expressions show indifference. In both cases, the incongruity between modalities acts as a strong indicator of sarcasm. The only publicly available multimodal sarcasm detection dataset, MUSTARD, contains only 500 odd instances, posing a significant challenge to training deep networks on this dataset.

3.7.2 Annotation Challenges: Intended vs. Perceived Sarcasm

Sarcasm is a highly subjective tool and poses significant challenges in curating annotations for supervised datasets. This difficulty is particularly evident in *perceived sarcasm*, where human annotators are employed to label text as sarcastic or not. Sarcasm recognition is known to be a difficult task for humans due to its reliance on pragmatic factors such as common ground (Clark, 1996). This difficulty is also observed through the low annotator agreements across

the datasets curated for perceived sarcasm (González-Ibáñez et al., 2011a; Castro et al., 2019). To combat such perceptual subjectivity, recent approaches in emotion analysis utilize *perceptual uncertainty* in their modeling (Zhang et al., 2018b; Gui et al., 2017; Han et al., 2017).

In our experience of curating a multimodal sarcasm detection dataset (Castro et al., 2019), we observed poor annotation quality, which occurred mainly due to the hardships associated with this task. Hovy et al. (2013) noticed that people undertaking such tasks remotely online are often guilty of *spamming*, or providing careless or random responses.

One solution to this problem is to rely on self annotated data collection. While convenient, obtaining labeled data from hashtags has been found to introduce both noises (incorrectly-labeled examples) and bias (only certain forms of sarcasm are likely to be tagged (Davidov et al., 2010), and predominantly by certain types of Twitter users (Bamman & Smith, 2015)).

Recently, Oprea & Magdy (2019) proposed the iSarcasm dataset, which annotates labels by the original writers for the sarcastic posts. This kind of annotation is promising as it circumvents the issues mentioned above while capturing the intended sarcasm. To combat annotations for perceived sarcasm, Best-Worst Scaling (MaxDiff) (Kiritchenko & Mohammad, 2016c) could be employed to alleviate the effect of subjectivity in annotations.

3.7.3 Target Identification in Sarcastic Text

Identifying the target of ridicule within a sarcastic text – a new concept recently introduced by Joshi et al. (2018) – has important applications. It can aid chat-based systems better understand user frustration, and help aspect-based sentiment analysis tasks to assign the sarcastic intent with the correct target in general. Though similar, there are differences from the vanilla aspect extraction task (Section 3.1) as the text might contain multiple aspects/entities with only a subset being a sarcastic target (Patro et al., 2019). When expressing sarcasm, people tend not to use the target of ridicule explicitly, which makes this task immensely challenging to combat.

3.7.4 Style Transfer between Sarcastic and Literal Meaning

Translations between sarcastic and literal forms of text has many applications. We discuss about some of the promising directions in this topic below.

Figurative to Literal Meaning Conversion: Converting a sentence from its figurative meaning to its honest and literal form is an exciting application. It involves taking a sarcastic sentence such as “*I loved sweating under the sun the whole day*” to “*I hated sweating under the sun the whole day*”. It has the potential to aid opinion mining, sentiment analysis, and summarization systems. These systems are often trained to analyze the literal semantics, and such a conversation would allow for accurate processing. Present approaches include converting a full sentence using monolingual machine translation techniques (Peled & Reichart, 2017), and also word-level analysis, where target words are disambiguated into their sarcastic or literal meaning (Ghosh et al., 2015). This application could also help in 1) performing data augmentation and 2) generating adversarial examples as both the

forms (sarcastic and literal) convey the same meaning but with different lexical forms.

Generating Sarcasm from Literal Meaning: The ability to generate sarcastic sentences is an important yardstick in the development of Natural Language Generation (NLG). The goal of building socially-relevant and engaging interactive systems demand such creativity. Sarcastic content generation can also be beneficial for content/media generation that find applications in fields like advertisements. Mishra et al. (2019b) recently proposed a modular approach to generate sarcastic text from negative sentiment-aware scenarios. End-to-end counterparts to this approach have not been well studied yet. Also, most of the works here rely on a particular type of sarcasm – one which involves incongruities within the sentence. The generation of other flavors of sarcasm (as mentioned before) has not been yet studied. Detailed research on this topic with an emphasis on end-to-end learning is demanding yet lucrative.

3.8 Sentiment-Aware Natural Language Generation (NLG)

Language generation is considered one of the major components of the field of NLP. Historically, the focus of statistical language models has been to create syntactically coherent text using architectures such as n-grams models (Stolcke, 2002) or auto-regressive recurrent architectures (Bengio et al., 2003; Mikolov et al., 2010; Sundermeyer et al., 2012). These generative models have important applications in areas including representation learning, dialogue systems, amongst others. However, present-day models are not trained to produce affective content that can emulate human communication. Such abilities are desirable in many applications such as comment/review generation (Dong et al., 2017), and emotional chatbots (Zhou et al., 2018).

Early efforts in this direction included works that either focused on related topics such as personality-conditioned text generation (Mairesse & Walker, 2007) or pattern-based approaches for the generation of emotional sentences (Keshtkar & Inkpen, 2011). These works were significantly pipe-lined with specific modules for sentence structure and content planning, followed by surface realization. Such sequential modules allowed constraints to be defined based on personality/emotional traits, which were mapped to sentential parameters that include sentence length, vocabulary usage, or part-of-speech (POS) dependencies. Needless to say, such efforts, though well-defined, are not scalable to general scenarios and cross-domain settings.

3.8.1 Conditional Generative Models

We, human beings, count on several variables such as emotion, sentiment, prior assumptions, intent, or personality to participate in dialogues and monologues. In other words, these variables control the language that we generate. Hence, it is an overstatement to claim that a vanilla seq2seq framework can generate *near perfect* natural language. In recent trends, conditional generative models have been developed to address this task. Conditioning on attributes such as sentiment can be approached in several ways. One way to achieve this is by learning *disentangled representations*, where the key idea is to separate the textual content from

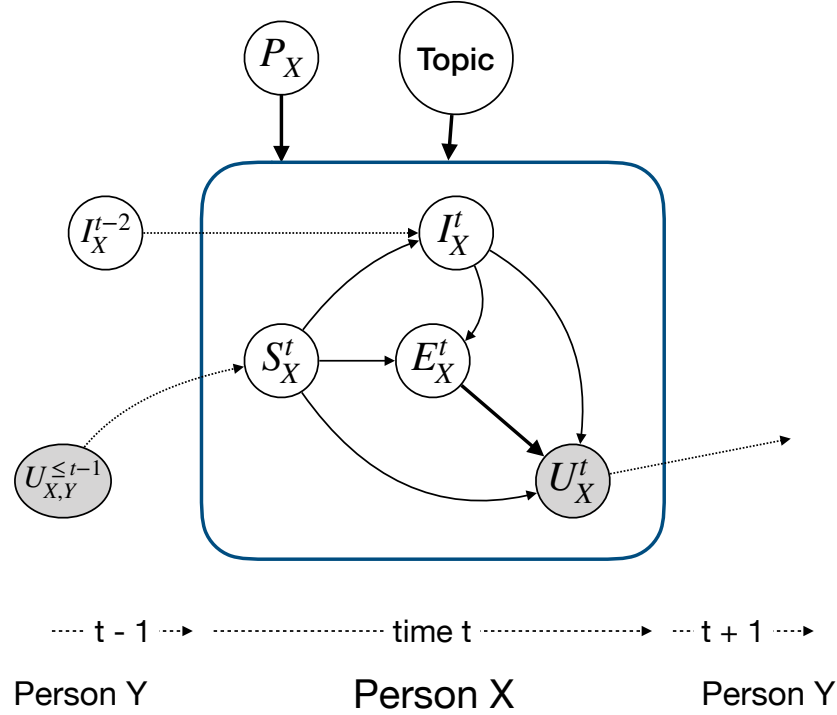


Fig. 13: Dyadic conversation—between person X and Y—are governed by interactions between several latent factors. Emotions are a crucial component in this generative process. In the illustration, P represents the personality of the speaker; S represents speaker-state; I denotes the intent of the speaker; E refers to the speaker’s emotional/sentiment-aware state, and U refers to the observed utterance. Speaker personality and the topic always condition upon the variables. At turn t , the speaker conceives several pragmatic concepts such as argumentation logic, viewpoint, and inter-personal relationship - which we collectively represent using the speaker-state S (Hovy, 1987). Next, the intent I of the speaker gets formulated based on the current speaker-state and previous intent of the same speaker (at $t - 2$). These two factors influence the emotional feeling of the speaker, which finally manifests as the spoken utterance.

high-level attributes such as sentiment and tense in the hidden latent code. Present approaches utilize generative models such as VAEs (Hu et al., 2017), GANs (Wang & Wan, 2018) or Seq2Seq models (Radford et al., 2017). Learning disentangled representations is presently an open area of research. Enforcing independence of factors in the latent representation and presenting quantitative metrics to evaluate the factored hidden code are some of the challenges associated with these models.

An alternate method is to pose the problem as an *attribute-to-text* translation task (Dong et al., 2017; Zang & Wan, 2017). In this setup, desired attributes are encoded into hidden states which condition upon a decoder tasked to generate the desired text. The attributes could include user’s preferences (including historical text), descriptive phrases (e.g. product description for reviews), and sentiment. Similar to general translation tasks, this approach demands parallel data and raises challenges in generalization, such as cross-domain generalization. Moreover, the attributes might not be available in desired formats. As mentioned, attributes might be embedded in conversational histories which would require sophisticated NLU capabilities similar to the ones used in task-oriented dialogue bots. They might also be in the form of structured data, such as Wikipedia tables or knowledge graphs, tasked to be translated into textual descriptions, i.e., data-to-text – an open area of research (Mishra et al., 2019a).

3.8.2 Sentiment-Aware Dialogue Generation

The area of controlled-text has also percolated into dialogue systems. The aim here is to equip emotional intelligence into these systems to improve user interest and engagement (Partala & Surakka, 2004; Prendinger & Ishizuka, 2005). Two key functionalities are important to achieve this goal (Hasegawa et al., 2013):

- 1) Given a user-query, anticipate the best emotional/sentiment response adhering to social rules of conversations.
- 2) Generate the response eliciting that emotion/sentiment.

Present works in this field either approach these two sub-problems independently (Ghosh et al., 2017) or in a joint manner (Gu et al., 2019). The proposed models range over various approaches, which include affective language models (Ghosh et al., 2017) or seq2seq models that are customized to generate emotionally-conditioned text (Zhou et al., 2018; Asghar et al., 2018). Kong et al. (2019) take an adversarial approach to generate sentiment-aware responses in the dialogue setup conditioned on sentiment labels. For a brief review of some of the recent works in this area, available corpora and evaluation metrics, please refer to Pamungkas (2019).

Despite the recent surge of interest in this application, there remains significant work to be done to achieve robust emotional dialogue models. Upon trying various emotional response generation models such as ECM (Zhou et al.,

2018), we surmise, these models lack the ability of conversational emotion recognition and tend to generate generic, emotionally incoherent responses. Better emotion modeling is required to improve contextual emotional understanding (Hazari et al., 2018b), followed by emotional anticipation strategies for the response generation. These strategies could be optimized to steer the conversation towards a particular emotion (Lubis et al., 2018) or be flexible by proposing appropriate emotional categories. For the generation stage, the quest for better text with diversity and coherence and fine-grained control over emotional intensity are still open problems. Also, automatic evaluation is a notorious problem that has plagued all applications of dialogue models.

To this end, following the work by Hovy (1987), we illustrate a sentiment and emotion-aware dialogue generation framework in Figure 13 that can be considered as the basis of future research. The model incorporates several cognitive variables i.e., intent, sentiment and interlocutor’s latent state for coherent dialogue generation.

3.8.3 Sentiment-Aware Style Transfer

Style transfer of sentiment is a new area of research. It focuses on flipping the sentiment of sentences by deleting or inserting new sentiment bearing words. E.g., to change the sentiment of “*The chicken was delicious*”, we need to find a replacement of the word *delicious* that carries negative sentiment.

Recent methods on sentiment-aware style transfer attempt to disentangle sentiment bearing contents from other non-sentiment bearing parts in the text by relying on rule-based (Li et al., 2018) and adversarial learning-based (John et al., 2019) techniques.

Adversarial learning-based methods to sentiment style transfer suffer from the lack of available parallel corpora which opens the door to a potential future work. Some initial works, such as (Shen et al., 2017), address non-parallel style transfer, albeit with strict model assumptions. We also think this research area should be studied together with the ALSA (aspect-level sentiment analysis) research to learn the association between topics/aspects and sentiment words. Considering the example above, learning better association between topics/aspects and opinionated words should aid a system to substitute *delicious* with *unpalatable* instead of another negative word *rude*.

3.9 Bias in Sentiment Analysis Systems

Exploring bias in machine learning has gained much traction recently. Studying bias in sentiment analysis is crucial, as the derived commercial systems are often shared by diverse demographics. Sentiment analysis systems are often used in such areas as healthcare, which deals with sensitive topics like counseling. Customer calls and marketing leads, from various backgrounds, are often screened for sentiment cues, and major decision-making is driven by the acquired analytics. Thus, understanding the presence of bias, especially for demographics, is critical. Unfortunately, the field is at its nascent stage and has received minimal attention. However, some developments have been observed in this area, which opens up numerous research directions. There

can be different types of bias, such as gender, race, age, etc. For the sake of brevity, in the following discussions, we use examples of gender bias.

3.9.1 Identifying Causes of Bias in Sentiment Analysis Systems

Bias can be introduced into the sentiment analysis models through three main sources:

- 1) *Bias in word embeddings*: Word embeddings are often trained on publicly available sources of text, such as Wikipedia. However, a survey by Collier & Bear (2012) found that less than 15% of contributions to Wikipedia come from women. Therefore, the resultant word embeddings would naturally under-represent women’s point of view.
- 2) *Bias in the model architecture*: Sentiment-analysis systems often use meta information, such as gender identifiers and indicators of demographics that include age, race, nationality, and geographical cues. Twitter sentiment analysis is one such application where conditioning on these variables is prevalent (Mitchell et al., 2013; Vosoughi et al., 2015; Volkova et al., 2013). Though helpful, such design choices can often lead to bias from these conditioned variables. A cogent solution to this issue could be to develop culture-specific sentiment analysis models rather than creating a generic one, albeit computationally inefficient.
- 3) *Bias in the training data*: There are different scenarios where a sentiment-analysis system can inherit bias from its training data. These include highly frequent co-occurrence of a sentiment phrase with a particular gender — for example, *woman* co-occurring with *nasty* —, over- or under-representation of a particular gender within the training samples, strong correlation between a particular demographic and sentiment label — for instance, samples from *female* subjects frequently belonging to *positive* sentiment category.

An author’s stylistic sense of writing can also be one of the many sources of bias in sentiment systems. E.g., one person uses strong sentiment words to express a positive opinion but prefers to use milder sentiment words in exhibiting negative opinions. A similar trend might prevail across races and genders, thereby making the task of identifying bias and de-biasing difficult.

3.9.2 Evaluating Bias

Recent works present corpora that curate examples, specifically to evaluate the existence of bias. The Equity Evaluation Corpus (EEC) (Kiritchenko & Mohammad, 2018) is one such example that focuses on finding *gender* and *racial* bias. The sentences in this corpus are generated using simple templates, such as <Person> made me feel <emotional state word>. While this is a good step, the work is limited to exploring bias that is related only to gender and race. Moreover, the templates utilized to create the examples might be too simplistic and identifying such biases and de-biasing them might be relatively easy. Future work should design more complex cases that cover a wider range of scenarios. Challenge appears when we have scenarios like *John told Monica that she lost her mental*

stability vs. John told Peter that he lost his mental stability. If the sentiment polarity in either of these two sentences is classified significantly different from the other, that would indicate a likely gender bias issue.

3.9.3 De-biasing

The primary approach to de-biasing is to perturb a text with word substitution to generate counterfactual cases in the training data. These generated instances can then be used to regularize the learning of the model, either by constraining the embedding spaces to be invariant to the perturbations or minimizing the difference in predictions between both the correct and perturbed instances. While recent approaches, such as (Huang et al., 2019b), have proposed these methods in language models, another direction could be to mask out bias contributing terms during training. However, such a method presents its own challenges since masking might cause semantic gaps.

In general, we observe that while many works demonstrate or discuss the existence of bias, and also propose bias detection techniques, there is a shortage of works that propose de-biasing approaches.

Apart from the traditional bias in models, bias can also exist at a higher level when making research choices. A simple example is the tendency of the community to resort to English-based corpora, primarily due to the notion of increased popularity and wider acceptance. Such trends diminish the research growth of marginalized topics and study of arguably more interesting languages – a gap which widens through time (Hovy & Spruit, 2016). As highlighted in Section 3.6, as a community, we should make conscious choices to help in the equality of under-represented communities within NLP and Sentiment Analysis.

4 CONCLUSION

Sentiment analysis is often regarded as a simple classification task to categorize contents into positive, negative, and neutral sentiments. In contrast, the task of sentiment analysis is highly complex and governed by multiple variables like human motives, intents, contextual nuances. Disappointingly, these aspects of sentiment analysis remain either un- or under-explored.

Through this paper, we strove to diverge from the idea that sentiment analysis, as a field of research, has saturated. We argued against this fallacy by highlighting several open problems spanning across subtasks under the umbrella of sentiment analysis, such as aspect level sentiment analysis, sarcasm analysis, multimodal sentiment analysis, sentiment-aware dialogue generation, and others. Our goal was to debunk, through examples, the common misconceptions associated with sentiment analysis and shed light on several future research directions. We hope this work would help reinvigorate researchers and students to fall in love with this immensely interesting and exciting field, again.

ACKNOWLEDGEMENTS

The authors are indebted to all the pioneers and researchers who contributed to this field.

This research is supported by A*STAR under its RIE 2020 Advanced Manufacturing and Engineering (AME) programmatic grant, Award No. - A19E2b0098, Project name - K-EMERGE: Knowledge Extraction, Modelling, and Explainable Reasoning for General Expertise.

NOTE

This paper will be updated periodically to keep the community abreast of any latest developments that inaugurate new future directions in sentiment analysis.

REFERENCES

- Alam, F., Joty, S. R., and Imran, M. Domain adaptation with adversarial training and graph embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 1077–1087, 2018.
- Asghar, N., Poupard, P., Hoey, J., Jiang, X., and Mou, L. Affective neural response generation. In *European Conference on Information Retrieval*, pp. 154–166. Springer, 2018.
- Bamman, D. and Smith, N. A. Contextualized sarcasm detection on twitter. In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*, pp. 574–577. AAAI Press, 2015.
- Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155, 2003.
- Bespalov, D., Bai, B., Qi, Y., and Shokoufandeh, A. Sentiment classification based on supervised latent n-gram analysis. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011, Glasgow, United Kingdom, October 24-28, 2011*, pp. 375–382. ACM, 2011.
- Blitzer, J., Dredze, M., and Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic*, 2007.
- Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., and Choi, Y. Comet: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4762–4779, 2019.
- Bryant, G. A. Prosodic contrasts in ironic speech. *Discourse Processes*, 47(7):545–566, 2010.
- Busso, C., Bulut, M., Lee, C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S. IEMOCAP: interactive emotional dyadic motion capture database. *Lang. Resour. Evaluation*, 42(4):335–359, 2008.
- Cambria, E., Poria, S., Gelbukh, A. F., and Thelwall, M. Sentiment analysis is a big suitcase. *IEEE Intell. Syst.*, 32(6):74–80, 2017.
- Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., and Poria, S. Towards multimodal sarcasm detection (an _obviously_ perfect paper). In *Proceedings of the 57th Conference of the Association for Computational*

- Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pp. 4619–4629. Association for Computational Linguistics, 2019.
- Cheang, H. S. and Pell, M. D. The sound of sarcasm. *Speech Commun.*, 50(5):366–381, 2008.
- Chen, X., Sun, Y., Athiwaratkun, B., Cardie, C., and Weinberger, K. Q. Adversarial deep averaging networks for cross-lingual sentiment classification. *Trans. Assoc. Comput. Linguistics*, 6:557–570, 2018.
- Choi, Y. and Cardie, C. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *2008 Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, Proceedings of the Conference, 25-27 October 2008, Honolulu, Hawaii, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 793–801. ACL, 2008.
- Choi, Y., Breck, E., and Cardie, C. Joint extraction of entities and relations for opinion recognition. In *EMNLP 2006, Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 22-23 July 2006, Sydney, Australia*, pp. 431–439. ACL, 2006.
- Choi, Y., Wiebe, J., and Mihalcea, R. Coarse-grained+/-effect word sense disambiguation for implicit sentiment analysis. *IEEE Transactions on Affective Computing*, 8(4): 471–479, 2017.
- Clark, H. H. *Using Language*. Cambridge University Press, 1996. ISBN 9780511620539.
- Collier, B. and Bear, J. Conflict, criticism, or confidence: An empirical examination of the gender gap in wikipedia contributions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW 12*, pp. 383392, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450310864.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12(Aug):2493–2537, 2011.
- Dai, A. M. and Le, Q. V. Semi-supervised sequence learning. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pp. 3079–3087, 2015.
- Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y. A., Gelbukh, A. F., and Zhou, Q. Erratum to: Multilingual sentiment analysis: State of the art and independent comparison of techniques. *Cognitive Computation*, 8(4):772–775, 2016.
- Davidov, D., Tsur, O., and Rappoport, A. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the fourteenth conference on computational natural language learning*, pp. 107–116. Association for Computational Linguistics, 2010.
- Deng, L. and Wiebe, J. Sentiment propagation via implicature constraints. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 377–385, Gothenburg, Sweden, April 2014. Association for Computational Linguistics.
- Deng, L. and Wiebe, J. MPQA 3.0: An entity/event-level sentiment corpus. In *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pp. 1323–1328. The Association for Computational Linguistics, 2015.
- Deng, L., Wiebe, J., and Choi, Y. Joint inference and disambiguation of implicit sentiments via implicature constraints. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 79–88, 2014.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics, 2019.
- Dong, L., Huang, S., Wei, F., Lapata, M., Zhou, M., and Xu, K. Learning to generate product reviews from attributes. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*, pp. 623–632. Association for Computational Linguistics, 2017.
- Esuli, A. and Sebastiani, F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation, LREC 2006, Genoa, Italy, May 22-28, 2006*, pp. 417–422. European Language Resources Association (ELRA), 2006.
- Filik, R., Leuthold, H., Wallington, K., and Page, J. Testing theories of irony processing using eye-tracking and erps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3):811, 2014.
- Gambäck, B. and Das, A. Comparing the level of code-switching in corpora. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pp. 1850–1855, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., and Lempitsky, V. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.
- Gavilanes, M. F., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., and González-Castaño, F. J. Unsupervised method for sentiment analysis in online texts. *Expert Syst. Appl.*, 58:57–75, 2016.
- Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K., and Pavlovic, V. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Trans. Image Processing*, 29:3993–4002, 2020a.
- Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K., and Pavlovic, V. Unsupervised multi-target domain adaptation: An information theoretic approach. *IEEE Trans. Image Processing*, 29:3993–4002, 2020b.
- Ghosh, D., Guo, W., and Muresan, S. Sarcastic or not: Word embeddings to predict the literal or sarcastic meaning of words. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lis-*

- bon, Portugal, September 17-21, 2015, pp. 1003–1012. The Association for Computational Linguistics, 2015.
- Ghosh, D., Fabbri, A. R., and Muresan, S. Sarcasm analysis using conversation context. *Computational Linguistics*, 44 (4), 2018.
- Ghosh, S., Chollet, M., Laksana, E., Morency, L.-P., and Scherer, S. Affect-lm: A neural language model for customizable affective text generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pp. 634–642, 2017.
- Glorot, X., Bordes, A., and Bengio, Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pp. 513–520, 2011.
- González-Ibáñez, R. I., Muresan, S., and Wacholder, N. Identifying sarcasm in twitter: A closer look. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers*, pp. 581–586. The Association for Computer Linguistics, 2011a.
- González-Ibáñez, R. I., Muresan, S., and Wacholder, N. Identifying sarcasm in twitter: A closer look. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA - Short Papers*, pp. 581–586. The Association for Computer Linguistics, 2011b.
- Gu, X., Xu, W., and Li, S. Towards automated emotional conversation generation with implicit and explicit affective strategy. In *Proceedings of the 2019 International Symposium on Signal Processing Systems*, pp. 125–130, 2019.
- Gui, L., Baltrušaitis, T., and Morency, L.-P. Curriculum learning for facial expression recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 505–511. IEEE, 2017.
- Guptha, V., Chatterjee, A., Chopra, P., and Das, A. Minority positive sampling for switching points - an anecdote for the code-mixing language modelling. In *Proceeding of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Association for Computational Linguistics, May 2020.
- Hamilton, W. L., Clark, K., Leskovec, J., and Jurafsky, D. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pp. 595–605. The Association for Computational Linguistics, 2016.
- Han, J., Zhang, Z., Schmitt, M., Pantic, M., and Schuller, B. From hard to soft: Towards more human-like emotion recognition by modelling the perception uncertainty. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 890–897. ACM, 2017.
- Hasegawa, T., Kaji, N., Yoshinaga, N., and Toyoda, M. Predicting and eliciting addressee’s emotion in online dialogue. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pp. 964–972. The Association for Computer Linguistics, 2013.
- Hatzivassiloglou, V. and McKeown, K. R. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*, pp. 174–181. Association for Computational Linguistics, 1997.
- Hatzivassiloglou, V. and Wiebe, J. M. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pp. 299–305. Association for Computational Linguistics, 2000.
- Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., and Mihalcea, R. CASCADE: contextual sarcasm detection in online discussion forums. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pp. 1837–1848. Association for Computational Linguistics, 2018a.
- Hazarika, D., Poria, S., Mihalcea, R., Cambria, E., and Zimmermann, R. ICON: interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pp. 2594–2604. Association for Computational Linguistics, 2018b.
- Hazarika, D., Poria, S., Vij, P., Krishnamurthy, G., Cambria, E., and Zimmermann, R. Modeling inter-aspect dependencies for aspect-based sentiment analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*, pp. 266–270. Association for Computational Linguistics, 2018c.
- He, Y., Lin, C., and Alani, H. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pp. 123–131. The Association for Computer Linguistics, 2011.
- Hoang, M., Bihorac, O. A., and Rouces, J. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics, NoDaLiDa 2019, Turku, Finland, September 30 - October 2, 2019*, pp. 187–196. Linköping University Electronic Press, 2019.
- Hovy, D. and Spruit, S. L. The social impact of natural language processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics, 2016.
- Hovy, D., Berg-Kirkpatrick, T., Vaswani, A., and Hovy, E. H. Learning whom to trust with MACE. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pp. 1120–1130. The Association for Computational Linguistics, 2013.

- Hovy, E. Generating natural language under pragmatic constraints. *Journal of Pragmatics*, 11(6):689–719, 1987.
- Howard, J. and Ruder, S. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 328–339. Association for Computational Linguistics, 2018.
- Hu, M. and Liu, B. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004*, pp. 168–177. ACM, 2004a.
- Hu, M. and Liu, B. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 168–177. ACM, 2004b.
- Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R., and Xing, E. P. Toward controlled generation of text. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, volume 70 of Proceedings of Machine Learning Research*, pp. 1587–1596. PMLR, 2017.
- Huang, B. and Carley, K. M. Syntax-aware aspect level sentiment classification with graph attention networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pp. 5468–5476. Association for Computational Linguistics, 2019.
- Huang, C., Trabelsi, A., and Zaïane, O. R. ANA at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and BERT. In *Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2019, Minneapolis, MN, USA, June 6-7, 2019*, pp. 49–53. Association for Computational Linguistics, 2019a.
- Huang, P., Zhang, H., Jiang, R., Stanforth, R., Welbl, J., Rae, J., Maini, V., Yogatama, D., and Kohli, P. Reducing sentiment bias in language models via counterfactual evaluation. *CoRR*, abs/1911.03064, 2019b.
- Hutto, C. J. and Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*, 2014.
- Irsoy, O. and Cardie, C. Opinion mining with deep recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIG-DAT, a Special Interest Group of the ACL*, pp. 720–728. ACL, 2014.
- John, V., Mou, L., Bahuleyan, H., and Vechtomova, O. Disentangled representation learning for non-parallel text style transfer. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pp. 424–434. Association for Computational Linguistics, 2019.
- Johnson, R. and Zhang, T. Supervised and semi-supervised text categorization using LSTM for region embeddings. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, volume 48 of JMLR Workshop and Conference Proceedings*, pp. 526–534. JMLR.org, 2016.
- Jorgensen, J., Miller, G. A., and Sperber, D. Test of the mention theory of irony. *Journal of Experimental Psychology: General*, 113(1):112, 1984.
- Joshi, A., Sharma, V., and Bhattacharyya, P. Harnessing context incongruity for sarcasm detection. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2: Short Papers*, pp. 757–762. The Association for Computer Linguistics, 2015.
- Joshi, A., Bhattacharyya, P., and Ahire, S. Sentiment resources: Lexicons and datasets. In *A Practical Guide to Sentiment Analysis*, pp. 85–106. Springer, 2017a.
- Joshi, A., Bhattacharyya, P., and Carman, M. J. Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, 50(5):73, 2017b.
- Joshi, A., Goel, P., Bhattacharyya, P., and Carman, M. J. Sarcasm target identification: Dataset and an introductory approach. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*. European Language Resources Association (ELRA), 2018.
- K Sarma, P., Liang, Y., and Sethares, W. Shallow domain adaptive embeddings for sentiment analysis. 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. 5548–5557, Hong Kong, China, November 2019. Association for Computational Linguistics.
- Karimi, A., Rossi, L., Prati, A., and Full, K. Adversarial training for aspect-based sentiment analysis with BERT. *CoRR*, abs/2001.11316, 2020.
- Keshtkar, F. and Inkpen, D. A pattern-based model for generating text to express emotion. In *International Conference on Affective Computing and Intelligent Interaction*, pp. 11–21. Springer, 2011.
- Kharde, V. A. and Sonawane, S. Sentiment analysis of twitter data : A survey of techniques. *CoRR*, abs/1601.06971, 2016.
- Khatti, A., Joshi, A., Bhattacharyya, P., and Carman, M. J. Your sentiment precedes you: Using an author’s historical tweets to predict sarcasm. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2015, 17 September 2015, Lisbon, Portugal*, pp. 25–30. The Association for Computer Linguistics, 2015.
- Kim, S. and Hovy, E. H. Determining the sentiment of opinions. In *COLING 2004, 20th International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2004, Geneva, Switzerland, 2004*.
- Kim, S.-M. and Hovy, E. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in*

- Text*, pp. 1–8. Association for Computational Linguistics, 2006.
- Kim, Y. Convolutional neural networks for sentence classification. In *EMNLP 2014*, pp. 1746–1751, 2014.
- Kiritchenko, S. and Mohammad, S. Happy accident: A sentiment composition lexicon for opposing polarity phrases. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016, Portorož, Slovenia, May 23–28, 2016*. European Language Resources Association (ELRA), 2016a.
- Kiritchenko, S. and Mohammad, S. The effect of negators, modals, and degree adverbs on sentiment composition. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@NAACL-HLT 2016, June 16, 2016, San Diego, California, USA*, pp. 43–52. The Association for Computer Linguistics, 2016b.
- Kiritchenko, S. and Mohammad, S. Examining gender and race bias in two hundred sentiment analysis systems. In Nissim, M., Berant, J., and Lenci, A. (eds.), *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, *SEM@NAACL-HLT 2018, New Orleans, Louisiana, USA, June 5–6, 2018*, pp. 43–53. Association for Computational Linguistics, 2018.
- Kiritchenko, S. and Mohammad, S. M. Capturing reliable fine-grained sentiment associations by crowdsourcing and best-worst scaling. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12–17, 2016*, pp. 811–817. The Association for Computational Linguistics, 2016c.
- Kong, X., Li, B., Neubig, G., Hovy, E., and Yang, Y. An adversarial approach to high-quality, sentiment-controlled neural dialogue generation. *arXiv preprint arXiv:1901.07129*, 2019.
- Kouloumpis, E., Wilson, T., and Moore, J. D. Twitter sentiment analysis: The good the bad and the omg! *Icwsn*, 11 (538–541):164, 2011.
- Ku, L., Liang, Y., and Chen, H. Opinion extraction, summarization and tracking in news and blog corpora. In *Computational Approaches to Analyzing Weblogs, Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-03, Stanford, California, USA, March 27–29, 2006*, pp. 100–107. AAAI, 2006.
- Kuppens, P., Allen, N. B., and Sheeber, L. B. Emotional inertia and psychological maladjustment. *Psychological Science*, 21(7):984–991, 2010.
- Labutov, I. and Lipson, H. Re-embedding words. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pp. 489–493, 2013.
- Lakkaraju, H., Socher, R., and Manning, C. Aspect specific sentiment analysis using hierarchical deep learning. In *NIPS Workshop on deep learning and representation learning*, 2014.
- Lal, Y. K., Kumar, V., Dhar, M., Shrivastava, M., and Koehn, P. De-mixing sentiment from code-mixed text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 371–377, 2019.
- Le, Q. and Mikolov, T. Distributed representations of sentences and documents. In *International conference on machine learning*, pp. 1188–1196, 2014.
- Li, H. and Lu, W. Learning latent sentiment scopes for entity-level sentiment analysis. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4–9, 2017, San Francisco, California, USA*, pp. 3482–3489. AAAI Press, 2017.
- Li, J. and Hovy, E. Reflections on sentiment/opinion analysis. In *A Practical Guide to Sentiment Analysis*, pp. 41–59. Springer, 2017.
- Li, J., Jia, R., He, H., and Liang, P. Delete, retrieve, generate: a simple approach to sentiment and style transfer. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pp. 1865–1874, 2018.
- Li, X., Bing, L., Li, P., and Lam, W. A unified model for opinion target extraction and target sentiment prediction. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 – February 1, 2019*, pp. 6714–6721. AAAI Press, 2019a.
- Li, X., Bing, L., Zhang, W., and Lam, W. Exploiting BERT for end-to-end aspect-based sentiment analysis. In *Proceedings of the 5th Workshop on Noisy User-generated Text, W-NUT@EMNLP 2019, Hong Kong, China, November 4, 2019*, pp. 34–41. Association for Computational Linguistics, 2019b.
- Liu, B. Sentiment analysis and subjectivity. In Indurkha, N. and Damerau, F. J. (eds.), *Handbook of Natural Language Processing, Second Edition*, pp. 627–666. Chapman and Hall/CRC, 2010.
- Liu, B. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
- Liu, B. *Sentiment Analysis - Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2015. ISBN 978-1-10-701789-4.
- Liu, B. and Zhang, L. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pp. 415–463. Springer, 2012.
- Liu, Q., Gao, Z., Liu, B., and Zhang, Y. Automated rule selection for aspect extraction in opinion mining. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25–31, 2015*, pp. 1291–1297. AAAI Press, 2015.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.
- Lloret, E., Balahur, A., Palomar, M., and Montoyo, A. Towards building a competitive opinion summarization system: Challenges and keys. In *Human Language Technolo-*

- gies: *Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, May 31 - June 5, 2009, Boulder, Colorado, USA, Student Research Workshop and Doctoral Consortium*, pp. 72–77. The Association for Computational Linguistics, 2009.
- Lubis, N., Sakti, S., Yoshino, K., and Nakamura, S. Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- Luo, H., Li, T., Liu, B., and Zhang, J. DOER: dual cross-shared RNN for aspect term-polarity co-extraction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pp. 591–601. Association for Computational Linguistics, 2019.
- Ma, Y., Peng, H., and Cambria, E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 5876–5883. AAAI Press, 2018.
- Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*, pp. 142–150. Association for Computational Linguistics, 2011.
- Mairesse, F. and Walker, M. Personage: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 496–503, 2007.
- Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A. F., and Cambria, E. Dialoguernn: An attentive RNN for emotion detection in conversations. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pp. 6818–6825. AAAI Press, 2019.
- Martineau, J. and Finin, T. Delta TFIDF: an improved feature space for sentiment analysis. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009*. The AAAI Press, 2009.
- Maynard, D. and Greenwood, M. A. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *LREC 2014 Proceedings*. ELRA, 2014.
- McCann, B., Bradbury, J., Xiong, C., and Socher, R. Learned in translation: Contextualized word vectors. In *Advances in Neural Information Processing Systems*, pp. 6294–6305, 2017.
- Mei, Q., Ling, X., Wondra, M., Su, H., and Zhai, C. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pp. 171–180. ACM, 2007.
- Mejova, Y. and Srinivasan, P. Exploring feature definition and selection for sentiment classifiers. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*. The AAAI Press, 2011.
- Mikolov, T., Karafiát, M., Burget, L., Černocký, J., and Khudanpur, S. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.
- Miller, G. A. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
- Mishra, A., Kanojia, D., Nagar, S., Dey, K., and Bhattacharyya, P. Harnessing cognitive features for sarcasm detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016.
- Mishra, A., Laha, A., Sankaranarayanan, K., Jain, P., and Krishnan, S. Storytelling from structured data and knowledge graphs : An NLG perspective. In *Proceedings of the 57th Conference of the Association for Computational Linguistics: Tutorial Abstracts, ACL 2019, Florence, Italy, July 28, 2019, Volume 4: Tutorial Abstracts*, pp. 43–48. Association for Computational Linguistics, 2019a.
- Mishra, A., Tater, T., and Sankaranarayanan, K. A modular architecture for unsupervised sarcasm generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pp. 6143–6153. Association for Computational Linguistics, 2019b.
- Mitchell, L., Harris, K. D., Frank, M. R., Dodds, P. S., and Danforth, C. M. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *CoRR*, abs/1302.3299, 2013.
- Miyato, T., Dai, A. M., and Goodfellow, I. J. Adversarial training methods for semi-supervised text classification. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.
- Moghaddam, S. and Ester, M. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, pp. 1825–1828. ACM, 2010.
- Mohammad, S. M. Challenges in sentiment analysis. In *A practical guide to sentiment analysis*, pp. 61–83. Springer, 2017.
- Moilanen, K. and Pulman, S. Sentiment composition. In *Proceedings of RANLP*, volume 7, pp. 378–382, 2007.
- Moraes, R., Valiati, J. F., and Neto, W. P. G. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Syst. Appl.*, 40(2):621–633, 2013a.
- Moraes, R., Valiati, J. F., and Neto, W. P. G. Document-level sentiment classification: An empirical comparison between svm and ann. *Expert Systems with Applications*, 40(2):621–633, 2013b.

- Morinaga, S., Yamanishi, K., Tateishi, K., and Fukushima, T. Mining product reputations on the web. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 341–349. ACM, 2002.
- Morris, M. W. and Keltner, D. How emotions work: The social functions of emotional expression in negotiations. *Research in organizational behavior*, 22:1–50, 2000.
- Munika, M., Shaky, S., and Shrestha, A. Fine-grained sentiment classification using BERT. *CoRR*, abs/1910.03474, 2019.
- Nakagawa, T., Inui, K., and Kurohashi, S. Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786–794. Association for Computational Linguistics, 2010.
- Nasukawa, T. and Yi, J. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pp. 70–77. ACM, 2003.
- Navaretta, C., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., and Maegaard, B. Mirroring facial expressions and emotions in dyadic conversations. In *LREC*, 2016.
- Oprea, S. and Magdy, W. sarcasm: A dataset of intended sarcasm. *CoRR*, abs/1911.03123, 2019.
- Pamungkas, E. W. Emotionally-aware chatbots: A survey. *CoRR*, abs/1906.09774, 2019.
- Pan, S. J., Ni, X., Sun, J., Yang, Q., and Chen, Z. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, pp. 751–760, 2010.
- Pang, B. and Lee, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, pp. 271. Association for Computational Linguistics, 2004.
- Pang, B., Lee, L., and Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79–86. Association for Computational Linguistics, 2002.
- Partala, T. and Surakka, V. The effects of affective interventions in human-computer interaction. *Interacting with computers*, 16(2):295–309, 2004.
- Patro, J., Bansal, S., and Mukherjee, A. A deep-learning framework to detect sarcasm targets. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pp. 6335–6341. Association for Computational Linguistics, 2019.
- Peled, L. and Reichart, R. Sarcasm SIGN: interpreting sarcasm with sentiment based monolingual machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pp. 1690–1700. Association for Computational Linguistics, 2017.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pp. 2227–2237. Association for Computational Linguistics, 2018.
- Polanyi, L. and Zaenen, A. Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications*, pp. 1–10. Springer, 2006.
- Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. Semeval-2014 task 4: Aspect based sentiment analysis. In Nakov, P. and Zesch, T. (eds.), *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014*, pp. 27–35. The Association for Computer Linguistics, 2014.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Mohammad, A.-S., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pp. 19–30, 2016.
- Poria, S., Cambria, E., Winterstein, G., and Huang, G. Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowl. Based Syst.*, 69:45–63, 2014.
- Poria, S., Cambria, E., and Gelbukh, A. F. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowl. Based Syst.*, 108:42–49, 2016.
- Poria, S., Cambria, E., Hazarika, D., Majumder, N., Zadeh, A., and Morency, L.-P. Context-dependent sentiment analysis in user-generated videos. In *ACL 2017*, volume 1, pp. 873–883, 2017.
- Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., and Mihalea, R. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*, 2018.
- Prendinger, H. and Ishizuka, M. The empathic companion: A character-based interface that addresses users’ affective states. *Applied Artificial Intelligence*, 19(3-4):267–285, 2005.
- Qiu, G., Liu, B., Bu, J., and Chen, C. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1):9–27, 2011.
- Quan, W., Zhang, J., and Hu, X. T. End-to-end joint opinion role labeling with BERT. In *2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, December 9-12, 2019*, pp. 2438–2446. IEEE, 2019.
- Radford, A., Jozefowicz, R., and Sutskever, I. Learning to generate reviews and discovering sentiment. *arXiv preprint arXiv:1704.01444*, 2017.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, abs/1910.10683, 2019.
- Rentoumi, V., Petrakis, S., Klenner, M., Vouros, G. A., and Karkaletsis, V. United we stand: Improving sentiment analysis by joining machine learning and rule based methods. In *LREC*, 2010.

- Riloff, E., Qadir, A., Surve, P., Silva, L. D., Gilbert, N., and Huang, R. Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pp. 704–714. ACL, 2013.
- Rogers, A., Kovaleva, O., and Rumshisky, A. A primer in bertology: What we know about how BERT works. *CoRR*, abs/2002.12327, 2020.
- Sarma, P. K., Liang, Y., and Sethares, B. Domain adapted word embeddings for improved sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pp. 37–42, 2018.
- Schifanella, R., de Juan, P., Tetreault, J. R., and Cao, L. Detecting sarcasm in multimodal social platforms. In *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, October 15-19, 2016*, pp. 1136–1145. ACM, 2016.
- Sharma, R., Bhattacharyya, P., Dandapat, S., and Bhatt, H. S. Identifying transferable information across domains for cross-domain sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 968–978, 2018.
- Shen, T., Lei, T., Barzilay, R., and Jaakkola, T. S. Style transfer from non-parallel text by cross-alignment. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 6830–6841, 2017.
- Shi, B., Fu, Z., Bing, L., and Lam, W. Learning domain-sensitive and sentiment-aware word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 2494–2504. Association for Computational Linguistics, 2018.
- Shu, L., Xu, H., and Liu, B. Lifelong learning CRF for supervised aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, pp. 148–154. Association for Computational Linguistics, 2017.
- Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1631–1642, 2013.
- Stolcke, A. Srilmm-an extensible language modeling toolkit. In *Seventh international conference on spoken language processing*, 2002.
- Sun, C., Huang, L., and Qiu, X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 380–385. Association for Computational Linguistics, 2019.
- Sundermeyer, M., Schlüter, R., and Ney, H. Lstm neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, 2012.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. D., and Stede, M. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307, 2011.
- Tai, K. S., Socher, R., and Manning, C. D. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pp. 1556–1566. The Association for Computer Linguistics, 2015.
- Tan, S., Cheng, X., Wang, Y., and Xu, H. Adapting naive bayes to domain adaptation for sentiment analysis. In *Advances in Information Retrieval, 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, 2009. Proceedings*, volume 5478 of *Lecture Notes in Computer Science*, pp. 337–349. Springer, 2009.
- Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pp. 1555–1565, 2014.
- Tang, D., Qin, B., and Liu, T. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1422–1432, 2015.
- Tang, D., Qin, B., Feng, X., and Liu, T. Effective lstms for target-dependent sentiment classification. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pp. 3298–3307. ACL, 2016a.
- Tang, D., Qin, B., and Liu, T. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pp. 214–224. The Association for Computational Linguistics, 2016b.
- Tay, Y., Tuan, L. A., and Hui, S. C. Dyadic memory networks for aspect-based sentiment analysis. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pp. 107–116. ACM, 2017.
- Tay, Y., Luu, A. T., Hui, S. C., and Su, J. Reasoning with sarcasm by reading in-between. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 1010–1020. Association for Computational Linguistics, 2018a.
- Tay, Y., Luu, A. T., Hui, S. C., and Su, J. Attentive gated lexicon reader with contrastive contextual co-attention for sentiment classification. In *Proceedings of the 2018 Confer-*

- ence on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018, pp. 3443–3453. Association for Computational Linguistics, 2018b.
- Thompson, D., Mackenzie, I. G., Leuthold, H., and Filik, R. Emotional responses to irony and emoticons in written language: evidence from eda and facial emg. *Psychophysiology*, 53(7):1054–1062, 2016.
- Thongtan, T. and Phienthrakul, T. Sentiment classification using document embeddings trained with cosine similarity. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop*, pp. 407–414. Association for Computational Linguistics, 2019.
- Tong, R. M. An operational system for detecting and tracking opinions in on-line discussion. In *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification*, volume 1, 2001.
- Tsur, O., Davidov, D., and Rappoport, A. ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*. The AAAI Press, 2010.
- Turney, P. D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pp. 417–424. ACL, 2002.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017.
- Volkova, S., Wilson, T., and Yarowsky, D. Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1815–1827, 2013.
- Vosoughi, S., Zhou, H., and Roy, D. Enhanced twitter sentiment classification using contextual information. In Balahur, A., der Goot, E. V., Vossen, P., and Montoyo, A. (eds.), *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EMNLP 2015, 17 September 2015, Lisbon, Portugal*, pp. 16–24. The Association for Computer Linguistics, 2015.
- Wang, K. and Wan, X. Sentigan: Generating sentimental texts via mixture adversarial networks. In Lang, J. (ed.), *Proceedings of the Twenty-Seventh International Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pp. 4446–4452. ijcai.org, 2018.
- Wang, S. I. and Manning, C. D. Baselines and bigrams: Simple, good sentiment and topic classification. In *The 50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, July 8-14, 2012, Jeju Island, Korea - Volume 2: Short Papers*, pp. 90–94. The Association for Computer Linguistics, 2012.
- Wang, Y., Huang, M., Zhu, X., and Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pp. 606–615. The Association for Computational Linguistics, 2016.
- Wiebe, J. Learning subjective adjectives from corpora. In Kautz, H. A. and Porter, B. W. (eds.), *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, July 30 - August 3, 2000, Austin, Texas, USA*, pp. 735–740. AAAI Press / The MIT Press, 2000.
- Wiebe, J. and Mihalcea, R. Word sense and subjectivity. In *ACL 2006, 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, Sydney, Australia, 17-21 July 2006*. The Association for Computer Linguistics, 2006.
- Wiebe, J. M. Tracking point of view in narrative. *Computational Linguistics*, 20(2):233–287, 1994.
- Wiebe, J. M., Bruce, R. F., and O’Hara, T. P. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 246–253. Association for Computational Linguistics, 1999.
- Wiegand, M. and Ruppenhofer, J. Opinion holder and target extraction based on the induction of verbal categories. In *Proceedings of the 19th Conference on Computational Natural Language Learning, CoNLL 2015, Beijing, China, July 30-31, 2015*, pp. 215–225. ACL, 2015.
- Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. Opinionfinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations*, pp. 34–35. Association for Computational Linguistics, 2005.
- Woodland, J. and Voyer, D. Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, 26(3):227–239, 2011.
- Wu, F. and Huang, Y. Sentiment domain adaptation with multiple sources. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics, 2016.
- Wu, H., Gu, Y., Sun, S., and Gu, X. Aspect-based opinion summarization with convolutional neural networks. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pp. 3157–3163. IEEE, 2016.
- Wu, W., Li, H., Wang, H., and Zhu, K. Probase: A probabilistic taxonomy for text understanding. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 05 2012.
- Xiang, E. W., Cao, B., Hu, D. H., and Yang, Q. Bridging domains using world wide knowledge for transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22(6):770–783, 2010.
- Xie, Q., Dai, Z., Hovy, E., Luong, M.-T., and Le, Q. V. Unsupervised data augmentation. *arXiv preprint arXiv:1904.12848*, 2019.

- Yang, B. and Cardie, C. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*, pp. 1640–1649. The Association for Computer Linguistics, 2013.
- Yu, H. and Hatzivassiloglou, V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pp. 129–136. Association for Computational Linguistics, 2003.
- Zadeh, A., Zellers, R., Pincus, E., and Morency, L. Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages. *IEEE Intelligent Systems*, 31(6): 82–88, 2016.
- Zadeh, A., Chen, M., Poria, S., Cambria, E., and Morency, L.-P. Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 1103–1114, 2017.
- Zadeh, A., Liang, P. P., Mazumder, N., Poria, S., Cambria, E., and Morency, L.-P. Memory fusion network for multi-view sequential learning. In *AAAI*, pp. 5634–5641, 2018a.
- Zadeh, A., Liang, P. P., Poria, S., Cambria, E., and Morency, L. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pp. 2236–2246. Association for Computational Linguistics, 2018b.
- Zadeh, A., Liang, P. P., Poria, S., Vij, P., Cambria, E., and Morency, L.-P. Multi-attention recurrent network for human communication comprehension. In *AAAI*, pp. 5642–5649, 2018c.
- Zang, H. and Wan, X. Towards automatic generation of product reviews from aspect-sentiment scores. In *Proceedings of the 10th International Conference on Natural Language Generation, INLG 2017, Santiago de Compostela, Spain, September 4-7, 2017*, pp. 168–177. Association for Computational Linguistics, 2017.
- Zhang, L., Wang, S., and Liu, B. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 8(4), 2018a.
- Zhang, M., Liang, P., and Fu, G. Enhancing opinion role labeling with semantic-aware word representations from semantic role labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 641–646. Association for Computational Linguistics, 2019.
- Zhang, X., Zhao, J. J., and LeCun, Y. Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pp. 649–657, 2015.
- Zhang, Z., Han, J., Coutinho, E., and Schuller, B. Dynamic difficulty awareness training for continuous emotion prediction. *IEEE Transactions on Multimedia*, 21(5):1289–1301, 2018b.
- Zhao, H., Zhang, S., Wu, G., Moura, J. M. F., Costeira, J. P., and Gordon, G. J. Adversarial multiple source domain adaptation. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada*, pp. 8568–8579, 2018.
- Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- Ziser, Y. and Reichart, R. Pivot based language modeling for improved neural domain adaptation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pp. 1241–1251. Association for Computational Linguistics, 2018.