SCIENCE CHINA
Technological Sciences

Special Topic: Natural Language Processing Technology
• Review •

October 2020 Vol. 63 No. 10: 1947–1970
https://doi.org/10.1007/s11431-020-1634-3

Recent advances in deep learning based sentiment analysis

YUAN JianHua[1], WU Yang[1], LU Xin[1], ZHAO YanYan[1*], QIN Bing[1,2] & LIU Ting[1]

[1]*Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, Harbin 150001, China;*
[2]*Pengcheng Lab, Shenzhen 518066, China*

Received March 9, 2020; accepted May 11, 2020; published online September 15, 2020

Sentiment analysis is one of the most popular research areas in natural language processing. It is extremely useful in many applications, such as social media monitoring and e-commerce. Recent application of deep learning based methods has dramatically changed the research strategies and improved the performance of many traditional sentiment analysis tasks, such as sentiment classification and aspect based sentiment analysis. Moreover, it also pushed the boundary of various sentiment analysis task, including sentiment classification of different text granularities and in different application scenarios, implicit sentiment analysis, multimodal sentiment analysis and generation of sentiment-bearing text. In this paper, we give a brief introduction to the recent advance of the deep learning-based methods in these sentiment analysis tasks, including summarizing the approaches and analyzing the dataset. This survey can be well suited for the researchers studying in this field as well as the researchers entering the field.

coarse-grained, fine-grained, implicit, multi-modal, generation

Citation: Yuan J H, Wu Y, Lu X, et al. Recent advances in deep learning based sentiment analysis. Sci China Tech Sci, 2020, 63: 1947–1970, https://doi.org/10.1007/s11431-020-1634-3

1 Introduction

Sentiment floods our everyday life: people tweet and express opinions on their daily life and hot topics, consumer read product review before shopping and write reviews after experiencing product or service, retailer and manufactures improve their products or service through surveying and analyzing people's opinions, etc. Thus, mining the sentiment exists in these user-generated contents has been a hot research topic in natural language processing, data mining, web mining and social media analysis, drawing growing attention from both research and industry communities during the past two decades.

Before the era of deep learning, machine learning based methods with rich hand-craft features dominate the field of natural language processing as well as sentiment analysis. Since the past decades, neural network-based feature extractors [1, 2] have become predominant for their superior ability in automatically extracting meaningful and abstract semantic features for sentiment analysis, which take place of previous sparse and high-dimensional feature engineering based methods. In this paper, we introduce common and practical structures used for deep learning-based sentiment analysis tasks as well as specialty of sentiment analysis problems in recent advance of this field.

Twenty years ago, the first sentiment analysis task was to determine whether a word is positive or negative. And this simple task started the intense research of sentiment analysis. There are so many researches and applications in this field, and in this paper we focus on the most fundamental tasks such as sentiment classification and aspect based sentiment analysis, and some recently promising tasks such as implicit sentiment analysis, multimodal sentiment analysis and sentiment generation, which are listed in Figure 1. We will give a

*Corresponding author (email: yyzhao@ir.hit.edu.cn)

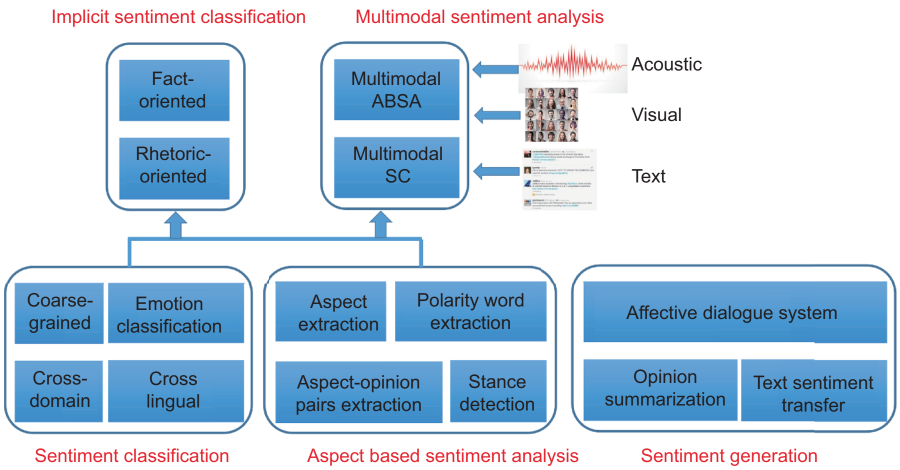

Figure 1 (Color online) The framework of sentiment analysis.

discussion about these tasks.

(1) Sentiment classification is a typical kind of classification task. According to different text granularities, we can obtain two general tasks, coarse-grained and emotion classification. Since the deep learning techniques can dramatically improve the classification problems, the deep learning models especially the recent BERT [3] can be helpful for both coarse-grained and emotion classification, even if we do not use other sentiment features. Therefore, in this filed, we can shift our attentions to the meaningful cross-domain and cross-lingual tasks, due to the fact that some domains or some languages do not have enough labeled data.

(2) Aspect based sentiment analysis (ABSA) aims to change an unstructured piece of review into a structured form. It includes many fundamental tasks, such as aspect extraction, polarity word extraction, aspect- polarity term co-extraction. Besides, it also includes an aspect-based sentiment classification (ASC) task, which is used to determine the sentiment polarity of a sentence towards an aspect. ABSA tasks along with sentiment classification are very useful and can be widely used in many different applications. We can explore many extensions of ABSA tasks, and stance detection is a typical one.

(3) Implicit sentiment mainly refers to a review without obvious polarity words. These implicit sentiment reviews are difficult to be processed. For instance, “There is a layer of dust on the table” expresses a negative sentiment towards the sanitary condition in a hotel review. This is a very promising research point, since solving this problem can further improve the performances of sentiment classification and the ABSA. We summarize that implicit sentiment analysis includes two parts: fact oriented and rhetoric oriented. There are many interesting ideas for both parts.

(4) Multimodal sentiment analysis (MSA) aims to analyzing the sentiment using multiple modalities. Multimodality is a brand new way to analyze sentiments. Sentiment, especially emotion is very rich, some of them cannot be reflected from just one modality. Besides the text, the commonly used modalities are acoustic and visual. Recently, a number of multimodal corpora are emerged, and it can promote the development of multimodal sentiment analysis. And the achievements from MSA can improve the traditional sentiment analysis.

(5) Sentimental text generation aims to generate a new text with sentiment according to different application scenarios. In this field, dialogue based sentiment content generation is an important task, since in dialogue system especially talking to a chatting robot, we need to identify the person’s emotion and generate the response with proper emotion. Besides, there are other sentiment generation tasks, such as sentiment summary, and text sentiment transfer. In general, sentiment generation is tightly linked to specific applications.

As the research field of sentiment analysis is inextricably bound to psychological sciences that attempt to understand human emotions, deep understanding of sentiments may need with more brain-inspired and psychologically motivated reasoning methods [4].

In the following, we will describe the above five sentiment analysis tasks at different granularity and different application scenarios. We mainly discuss how to use the neural networks, summarizing the models and providing the datasets. We finally conclude this paper and discuss future directions.

2 Sentiment classification

Task definition Sentiment classification is one of the fundamental tasks in sentiment analysis, which focuses on classifying the sentiment orientation of a given text sequence (Figure 2). A text sequence could be a short phrase, a sen-

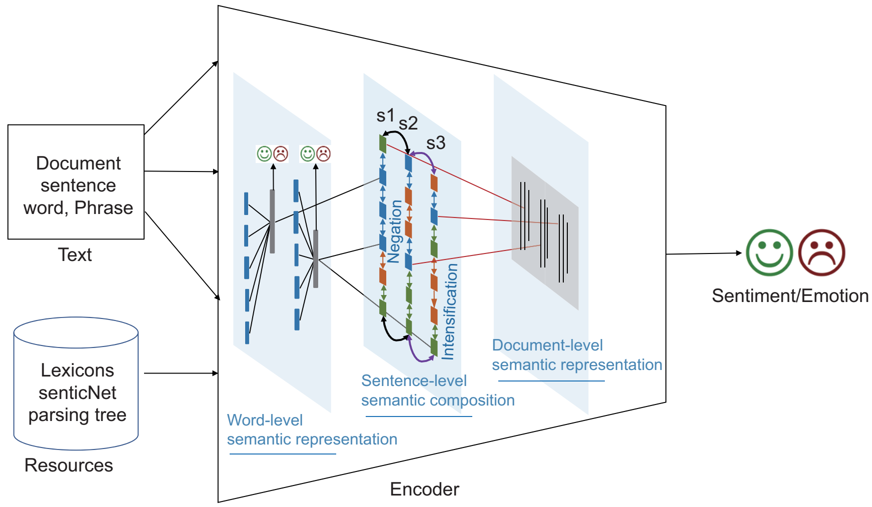

Figure 2 (Color online) A general architecture of sentiment classification model.

tence, or a long document. Such variance in text length results in different complexity of semantic composition. Besides, the output label could be one of sentiment labels (positive, negative, neutral), star ratings or emotion labels (happy, sad, angry, surprise, etc.).

2.1 Sentiment classification of different granularities

2.1.1 Sentence-level sentiment classification

Sentence-level sentiment classification studies classifying the polarity of a given sentence. For instance in Table 1, a classifier should classify the second sentence as negative for it expressing an overall negative opinion. Roughly, existing models can be divided into the following categories.

NN w/o attention Kim [5] leveraged convolutional neural network to model sentence with convolution and pooling operations. Wang [6] proposed a disconnected recurrent neural network which limited the information flow to fixed steps in vanilla RNN.

NN with attention Lin et al. [7] applied self-attention during sentence modeling. Wang et al. [8] adopted capsule network based method where a positive and a negative capsule were employed to capture corresponding sentiment features.

NN with external resources Many external resources, which include sentiment lexicons, POS tagging tools, syntactic and dependency parsing tools, SenticNet, etc., can usually provide complementary information for sentiment classification. As sentiment lexicons contain general and accurate sentiment scores for sentiment words, Teng et al. [9] combined them with context-sensitive weights to decide final sentiment score. Different from this, Tay et al. [10] treated sentiment lexicons as a word-level prior and utilize attention mechanism over positive and negative lexicons to improve sentiment feature extraction. Socher et al. [11] applied recursive auto-encoder and recursive neural tensor network on top of a parsing tree. Tai et al. [12] and Zhu et al. [13] extended the basic LSTM with tree topology. Qian et al. [14] used linguistic rules like negation and intensification to regularize the hidden feature of a LSTM network. Zhang et al. [15] designed an information distilled LSTM (ID-LSTM) which delete unnecessary words and a hierarchical structure LSTM (HS-LSTM) which learned to compose words into phrases with deep reinforcement learning based method.

Table 1 Examples of sentiment classification

Text	Sentiment
This cake tasted pretty good.	Positive
The movie is quite boring.	Negative

2.1.2 Document-level sentiment classifications

Similar to sentence-level sentiment classification, document-level sentiment classification focuses on identifying the sentiment polarity of a given document. While it is hard for RNN-based sequence models to learn long range semantic dependency in long documents, most existing work adopts a hierarchical structure: first model word-to-sentence composition and then sentence-to-document composition.

Hierarchical modeling Tang et al. [16] proposed GRNN with hierarchical composition structure, which first modeled each sentence with a CNN or LSTM and then fed these sentence representations into a bi-directional gated neural network. To better capture the most salient semantics in both word-to-sentence and sentence-to-document composition phases, Yang et al. [17] utilized a hierarchical attention structure.

Incorporate user and product information While many documents come from product reviews, the user preference and product detail information are ignored if only using the above mentioned methods. To this end, Tang et al. [18] proposed to use user and product matrix to encode their information respectively. They first captured user-text and product-text representations through matrix product. Then they linearly concatenated two representations and apply a "tanh" non-linearity, a convolution and pooling operation to get the final sentence feature capturing user and product information. Chen et al. [19] replaced user and product matrix with user and product vectors respectively. Then they used concatenated vectors to attend the text from both word and sentence level. Dou [20] attacked the problem using a deep memory network where reviews of the same product and of same user are stored in the memory and a multi-hop attention was applied over the constructed memory. Wu et al. [21] extended Chen et al. [19] by introducing two extra loss for user-weighted document representation and product-weighted document representation respectively.

Apart from RNN-based models, recent application of graph neural networks has also improved the performance of sentiment classifiers. Yao, Mao and Luo [22] proposed a text graph convolutional network to joint learn embeddings for both words and documents which outperforms state-of-the-art methods without any external word embeddings and knowledge.

Discussion Though previous work achieved great results on benchmark datasets, most of them focused on designing architecture to learn better sentiment semantic representation from text alone. Some used external features like dependency parsing and lexicons that provided supplementary information to original text. Recent advances in deep contextualized word representations and language models have provide better use of context modeling of large-scale unlabeled data. We hope to see successful application of sentiment-specific pre-trained language models just like sentiment-specific word embeddings (SSWE [23]).

2.2 Sentiment classification with domain transfer

Due to the discrepancy existing among different domains, sentiment classifier that trained one domain may not perform ideally on new domains, thus requesting more sophisticate methods to bridge such feature gap among domains. To solve this problem, two tasks are well-studied, which are cross-domain and multi-domain sentiment classification. Cross-domain sentiment classification focuses on learning a transferable feature extractor without using target domain data or only using limited amount of target domain data, while multi-domain sentiment classification task aims to develop a model that is trained using data from multiple domains to achieve best average performance on all these domains. In Figure 3, we give an example of cross domain sentiment classification from electronic domain to kitchen domain.

2.2.1 *Cross-domain sentiment classification*

Adversarial shared-private network Li et al. [24] introduced an end-to-end adversarial memory network (AMN) to learn domain-invariant pivot sentiment features through multi-hop attention over memory network.

New pivots Yu and Jiang [25] proposed to use auxiliary tasks of whether an input sentence contains a positive or negative domain independent sentiment words to learn robust sentence embeddings across domains. Furthermore, Li et al. [26] took detecting domain-specific sentiment words as another auxiliary task and proposed a hierarchical attention transfer network (HATN), which facilitated automatic extraction of pivots and non-pivots compared to previous work. Zhang et al. [27] proposed an interactive attention transfer network (IATN) that incorporated additional information from shared aspects as pivots for domain transfer.

2.2.2 *Multi-domain sentiment classification*

Multi-domain sentiment classification task is useful for the

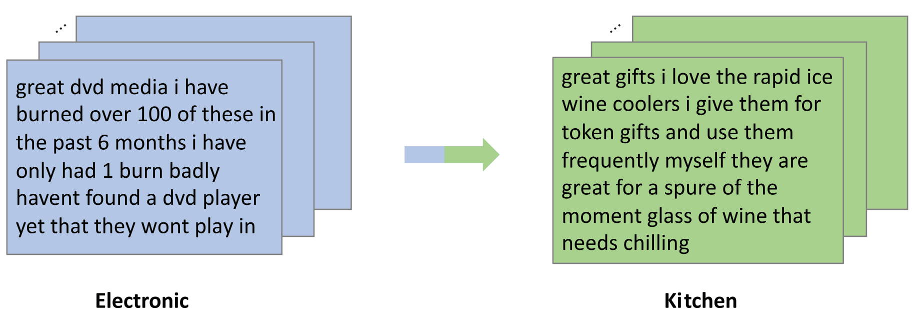

Figure 3 (Color online) A example of cross domain sentiment classification: from electronic domain to kitchen domain.

scenario where we have many different domains while each domain only has limited amount of labeling data. For multi-domain sentiment classification task, existing methods can be divided into several groups.

Shared-private network One major line is to exploit the shared-private framework, where domain-agnostic features are captured at the shared networks (usually stacked BLSTMs) and domain-specific representations by feature extractor of each domain. Most model simply combine these two parts and feed them into a shared sentiment classifier for final sentiment labels.

Adversarial shared-private model Liu et al. [28] found that shared representations acquired from simple parameter sharing may contain domain-specific feature which could contaminate performance. They applied adversarial domain classification into the shared part to force the learning of domain-agnostic sentiment features. Chen and Cardie [29] extended this framework with theoretical justifications.

Domain attention-based Another line of work [30–32] implicitly utilized such share-private ideas where they first learn domain-specific query vectors and then used these to compose domain-specific representation by attending the outputs from shared sentence encoder. Liu et al. [30] used domain descriptor to attend the general sentence representations and enhanced that representation with external domain knowledge memory. Zheng et al. [31] first used a sentence encoder with domain classifier on top to extract domain specific query vector and then used that vector to attend outputs of the other encoder. Similar to this, Cai and Wan [32] treated the domain-specific query vector as domain embeddings and added that to input of every word in the other encoder.

Discussion Existing work has tried various methods to learn domain-invariant representations for domain-transfer tasks for cross-domain sentiment classification and extra domain-aware sentence representation for multi-domain sentiment classification. For future work, we would recommend the application of concatenating both general pre-trained language models and light-weighted revised domain-specific pre-trained language models.

2.3 Sentiment classification with language transfer

As deep learning based methods rely heavily on annotated data and sentiment resources differs among languages, it is necessary to adapt sentiment resources in resource-rich languages to resource-poor (compared to English) languages. Thus, the task of cross-lingual sentiment classification tackles the problems of learning a sentiment classifier that works well on target language without any labeled data of that language. Similarly, multi-lingual sentiment classification focuses on building a sentiment classifier that works well on average in all the languages of interest.

Zhou et al. [33] first translated training data into target language and then developed a hierarchical attention based BLSTM model for joint training on the source language data and corresponding translated data. In this way, they learned a robust sentiment classifier for the target language, which was verified on a benchmark dataset using English as the source language and Chinese as the target language. Chen et al. [34] proposed adversarial deep averaging networks where they applied an adversarial language discriminator to push the shared deep averaging network based feature extractor to learn language-independent sentiment feature. The model was trained using the labeled data from source language and unlabeled data from both languages, which greatly outperformed a strong pipeline method based on Google Translation.

Different from previous work that only has one source language to transfer, Chen et al. [35] studied the problem of cross-lingual sentiment classification using multiple sources. Apart from learning language-invariant sentiment features, they also exploited the Mixture-of-Expert structure to learn specific features shared by certain languages. Akhtar et al. [36] leveraged bilingual word embeddings learned through a parallel corpus and achieved good performance in both cross-lingual and multi-lingual settings.

Discussion Similar to domain transfer problems in sentiment classification, it would be interesting to see better leverage of text encoders earned from large-scale unlabeled corpus for each language to find more robust pivots for language transfer tasks in sentiment analysis.

2.4 Emotion classification

To depict a person's diverse emotion state on social media, the task of emotion classification is introduced. This task is generally considered as a multi-label classification problem where a classifier need to tag a text snippet with multiple labels, e.g., "happy", "sad", "anger", "disgust", "surprise" or "fear". As each sentence may contain one or more emotions, the multi-label emotion classification task is either tackled as several binary classification problems or solved by selecting labels with probabilities over a preset threshold [37].

Felbo et al. [38] proposed an attention-based BLSTM model and Kim et al. [39] develop an attention-based CNN model for emotional classification. Abdul-Mageed and Ungar [40] improved the performance of emotion classification by building a very large dataset for fine-grained emotions for training deep learning models. Wang et al. [41] explored the problem of code-switch emotion classification, where the input text may contain multiple languages. They design a bilingual attention network (BAN) model to select informative words both mono-lingually and bilingually to generate document representations and combine these feature vectors for final emotion prediction.

2.5 Datasets and evaluation metrics

We summarize the benchmark datasets for coarse-grained sentiment classification in Table 2. On most of the above datasets, models are evaluated using classification accuracy. For Tang2015, MAE and RSME are also calculated.

- **MR** [42] This dataset is introduced in 2002 and has been widely adopted since then. It is constructed from IMDB reviews with apparent ratings, which are transformed into three categories: positive, negative and neutral.
- **CR** [43] This dataset is collected by Hu and Liu [43] in 2004 with product reviews of camera, dvd, mp3 and cell phone.
- **SST** [11] This dataset contains 215154 phrases with fine-grained sentiment labels in the parse trees of 11855 sentences in movie reviews. It is usually used for binary or five-way sentiment evaluation.
- **Twitter** [44] This dataset is used in one of the shared tasks of SemEval2017 Task 4 Subtask A. It is used for ternary sentiment classification on Twitter.
- **IMDB** [45] This dataset is also extracted from IMDB website. It is a binary sentiment classification dataset consisting of 50000 balanced positive and negative reviews.
- **Yelp** [46] This dataset is obtained from Yelp Dataset Challenge in 2015. Both binary and five-way classification tasks are provided.
- **Tang2015** [18] This dataset differs from previous dataset in that it has additional provided the user who wrote the text and the product the text is written for. Reviews came from IMDB and Yelp.

Table 2 Statistics of several sentiment classification datasets

Task	Granularity	Dataset	Domain	Size
Sentiment	Sentence	MR	Product review	2k
		CR	Movie	3k
		SST-2	Movie	11k
		SST-5	Movie	94k
		Twitter	Twitter	10k
	Document	IMDB	Movie	50k
		Yelp-2	Yelp review	500k
		Yelp-5	Yelp review	3k
		Tang2015	Yelp, IMDB	394k
	Document, Sentence	MTL	Product, movie	32k
	Document	Amazon	Product	8k
Emotion	Sentence	NT	News title	1.2k

- **Amazon** [47] This dataset comes from Amazon product reviews of Book, DVD, Kitchen and Electronic domains. It is one of the most widely used MDTC datasets. It was already preprocessed into bag of features.
- **MTL** [28] This dataset is proposed for multi-domain sentiment classification and contains data from 16 domains, 14 of which are Amazon product reviews and 2 from movie reviews.
- **NT** [48] This dataset is extracted from news websites and newspapers and titles of news are labeled with predefined six emotion categories.

3 Fine-grained sentiment analysis

Apart from the overall sentiment orientation of a given text, knowing who or what the opinion is targeted at and what expressions people use to describe their opinions are even more crucial for real word applications. The task of fine-grained sentiment analysis involves a number of above mentioned tasks, which are the identification of opinion targets or aspect terms (ATE), the extraction of opinion expressions and the classification of targeted or aspect based sentiment (ASC). Here, an aspect is an attribute of a target entity, e.g., the keyboard of a laptop. For some dataset, there exists an aspect category classification task (ACC) where a text need to be examined of whether talking about a pre-defined aspect category. Correspondingly, ACC is usually follwed by a aspect category based sentiment classification (ACSC). Figure 4 illustrate an example of aspect based sentiment analysis task.

For the following subsections, we mainly describe related work on opinion expression extraction, opinion target extraction and joint tasks of these problems.

3.1 Opinion target/aspect term extraction

Task definition The task of opinion term extraction studies the extraction of targets or aspect in an opinionated text and serves as the prerequisite for performing target-based and aspect based sentiment classification. For example, in Figure 4, an opinion term extractor need to identify both food and service.

Existing work can be summarized according to their model structures and adopted resources as follows.

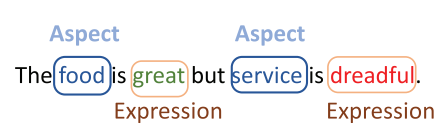

Figure 4 (Color online) An example of fine-grained sentiment analysis.

NN with novel structures Liu et al. [49] first investigated RNN based model with different types of embedding on aspect term extraction and outperforms feature-rich CRF-based models. Li et al. [50] leveraged aspect-aware sentence summarization and coordinate structure and tagging schema constraints on aspect detection to better extract aspect terms. Ma et al. [51] considered aspect term extraction task as a sequence-to-sequence (seq2seq) learning task and incorporate corresponding word representation and position-aware attention during label decoding. Liao et al. [52] studied an unsupervised aspect extraction method which leverage both sentence-level and word-level context. Through investigating the word distribution in sentence-level and word-level context, they find that aspect-words and non-aspect-words do exhibit different context.

NN with domain knowledge Domain-specific knowledge exists both in annotated data and large scale unannotated data. To leverage knowledge of aspect term extraction over existing annotated domains, Shu et al. [53] built a lifelong learning CRF model which continues to improve with experience in its application. Since word embedding trained on domain-specific corpus implicitly push words with similar context as aspect term words into neighboring semantic space, Xu et al. [54] proposed a simple CNN based model that combines both general purpose embeddings and domain-specific embeddings, which achieves superior performance without using any additional supervision.

NN with external resources Apart from domain-specific resources, domain-invariant resources including dependency parsing results, general word embeddings and sentiment lexicons are often used for aspect extraction. Yin et al. [55] encoded dependency path relationship with a RNN and combine that with linear context of each word as feature fed into CRF model for aspect term labeling. He et al. [56] proposed an unsupervised attention-based aspect extraction (ABAE) model where they use word embeddings to discover coherent aspects and apply attention to deemphasize irrelevant words. Karamanolakis et al. [57] explored the task of aspect detection on product review data with a weakly supervised model based on teacher-student schema, which outperforms previous weakly supervised models not effectively leveraging seed words.

Furthermore, some work studies the problem of cross-lingual aspect extraction where alignment of aspect terms in different languages are learned. Wang and Pan [58] developed a transition-based cross-lingual aspect extraction model that leverages invariant configurations in a transition system for transfer. Jebbara and Cimiano [59] leveraged multi-lingual word embeddings to conduct zero-shot cross-lingual opinion target extraction.

3.2 Opinion expression extraction

Task definition The task of opinion expression extraction is to identify the span of text that expresses subject opinion in a sentence or document. Typically, the task is tackled as a sequence labeling problem using BIO tagging schema. For the example in Figure 4, good and dreadful are opinion words that reveal customer's opinions.

Irsoy and Cardie [60] applied a deep bidirectional RNN for opinion expression detection, outperforming shallow RNN and CRF baselines.

3.3 Aspect-based sentiment classification

Instead of given an overall polarity label towards the sentence, aspect-level sentiment analysis [61, 62] tried to infer the sentiment orientation of the whole sentence towards a given target entity, an aspect of that entity (i.e., aspect term) or an aspect category, thus the task is usually considered as a fined-grained classification task in terms of its granularity. We usually refer to it as aspect-based sentiment analysis. An aspect term can be an attribute of entity, e.g, the keyboard of a laptop while an aspect category are pre-defined by users, e.g, atmosphere or service of a restaurant. Take the review in Figure 4 as example, the sentence expresses a positive sentiment towards food while a negative one towards service. Different targets have different opinion context and it is hard to accurately capture the surrounding context for a given sentence. Therefore, sophisticate structures are required to capture the semantic relatedness between a given sentence and each of its target. As most existing work in aspect-based sentiment classification and target-based sentiment classification shares similar structures, here we use "target-" and "aspect-" interchangeably for simplicity.

In the following, we assume that targets or aspects have been given along with the input sentence. Most related work falls into the following categories.

NN for target-aware sentence representation Tang et al. [63] proposed a Target-dependent LSTM (TD-LSTM) and a Target-connection LSTM (TC-LSTM) to model target-aware sentence representation. In TD-LSTM, Tang et al. used forward and backward LSTM towards the position of a target. In TC-LSTM, they added the averaged target representation to each input of TD-LSTM. Tang et al. [64] utilized a deep memory network based model to capture aspect-aware sentence representation with multiple hops of attention process. Their model outperforms LSTM structures in terms of both performance and speed by a large margin. Chen et al. [65] went one step further. They employed a separate RNN instead of a linear transformation to link multiple attention computation layers and leverage the power of such non-linear feature extraction to get better performance on a variety of datasets. Wang et al. [66] employed a attention-based LSTM for aspect-category sentiment classification. They concatenated word embedding and aspect category embedding to form the new input for the network, which took advantage of aspect information and better modeled the inter-dependency between words and the input aspect. Then the hidden layer outputs of BLSTM network were aggregated using attention mechanism with respect to the input target.

Vo and Zhang [67] tackled target-dependent twitter sentiment classification using automatic features including sentiment-specific embeddings, neural pooling operations and sentiment lexicons. Liu and Zhang [68] also split the input text into left and right context with respect to a given target. However, different from Tang et al. [63], they applied attention mechanism instead of simple LSTM to better extract target-aware context from left and right side. Wang and Lu [69] investigated a segmentation-based attention model which leverage the power of CRF in discovering latent opinion words. By modeling such correlation between the given aspect and extracted latent opinion words, their model exhibited great performance gain as well as interpretable opinion expressions for the polarity prediction.

NN with target-sentence co-attention Ma et al. [70] proposed an interactive attention network (IAN) that employed both aspect-to-sentence and sentence-to-aspect attention to capture the most salient feature from both the aspect and the sentence. They took the concatenation of aspect-aware sentence representation and sentence-aware aspect representation for final sentiment classification. More recently, due to the superior capability of pre-trained language model BERT in modeling relationships between sentence pairs, Sun et al. [71] inflated (aspect, sentence) pairs into sentence pairs, which is the same format as BERT's next sentence prediction (NSP) pret-training task. The performance of such transformation outperformed previous attention-based models by a large margin.

NN with structure information. Dong et al. [72] proposed adaptive recursive neural network (AdaRNN) where the sentiment of context towards the target propagates over a dependency tree and the feature vector from the root node were used for final sentiment prediction. He et al. [73] improved previous method by better modeling of target representation and incorporating dependency parse tree into to aspect-to-sentence attention process. Specifically, they used an auto-encoder for ensuring the quality of aspect embeddings and measure the importance of each opinion word in the context by its distance from the aspect in the dependency path. Ruders et al. [74] proposed a hierarchical attention based BLSTM model and demonstrated its effectiveness of modeling the implicit rhetoric structure among sentences in a document.

NN with explicit multi-aspect modeling While previous work deals with aspect terms in a sentence separately, Hazarika et al. [75] and Majumder et al. [76] noticed that it would be helpful to mine the relationship among different aspects in the same sentence. Hazarika et al. first modeled the (aspect, sentence) pair with a structure similar to ATAE of Wang et al. [66] and then fed all the aspect-aware sentence representations to another LSTM by the order of their corresponding aspect in the sentence, which indeed outperformed original ATAE model that did not capture such inter-aspect relationships. Compared with Hazarika et al., Majumder et al. applied an extra memory network based structure on top to strengthen its ability of extracting more abstract aspect-aware sentence representations.

NN with external resources Ma et al. [77] proposed an extension of LSTM, termed Sentic LSTM to integrate commonsense knowledge into recurrent encoder and employ a hierarchical attention mechanism consisting of a target-level attention and a sentence level attention. In their Sentic LSTM, they select a set of knowledge concept candidates to serve as complementary information to the text sequence and encode these candidates into the revised LSTM. As annotating aspect-level data are difficult and expensive, existing aspect-level sentiment classification datasets are usually small, which may hinder the training and application of neural models. Since it is relatively easy to get sentiment polarity label of document and sentences, He et al. [78] exploited document-level data into attention-based LSTM networks through multi-task learning or pre-training feature extractor on document-level sentiment classification data. Chen and Qian [79] proposed TransCap which leverages Capsule Network to transfer sentiment knowledge from document-level labeled data to aspect-level sentiment classification. Similar, Li et al. [80] noticed that even inside aspect-level sentiment classification, it is relative more easy to access aspect category based classification data than aspect term based data, as aspect category is a set of user pre-defined labels while aspect terms vary a lot from each other. To this end, they proposed a multi-granularity alignment network (MGAN) to leverage knowledge from rich-resource aspect category based classification data to low-resource aspect term based classification data through a coarse2fine attention structure.

3.4 Joint tasks

3.4.1 Joint extraction of aspect and opinion term

NN with aspect-opinion interaction Wang et al. [81] proposed a joint model based on recursive neural networks and

CRF to explicitly co-extract aspect terms and opinion words under a unified tagging framework, which prove the effectiveness of exploiting connection between aspect term and opinion expressions. Li and Lam [82] employed two LSTMs with extended memory to extract aspect and opinion terms via memory interaction. They also add another sentimental LSTM to enhance the performance of extraction. Wang et al. [83] offered a coupled multi-layer attention network (CMLA) for joint extraction of aspect terms and opinion words. In each layer, they applied a couple of attention, one for aspect term and the other for opinion word extraction, with tensor operations to exploit relationships between two tasks.

Besides, some work also explores this joint task in cross-domain setting. For instance, Wang and Pan [84] proposed a recursive neural structural correspondence network (RN-SCN) which use syntactic relations as invariant pivot information for domain transfer. Wang and Pan [85] also proposed an interactive memory network, which apply local attention within each domain and global attention between source and target domains, to capture intra-correlation among opinion or aspect words themselves as well as interactions between opinion and aspect words.

3.4.2 Joint extraction of aspect and its sentiment

As this joint task involves both classification and labeling problem, current work usually copes with it using sequence labeling based methods under unified tagging schema.

Joint tagging Luo et al. [86] proposed a novel Dual crOss-sharEd RNN framework (DOER) to solve the joint task of aspect term extraction and aspect-based sentiment analysis, where both tasks are seen as sequence labeling tasks and a cross-shared unit is utilized to explicitly model the interaction between two tasks. Similar to DOER, He et al. [87] proposed an interactive multi-task learning network (IMN) to solve these joint task by introducing an auxiliary classification task at document level. The interactions among different tasks were iteratively passed to each task through a set of shared latent variables. Peng et al. [88] proposed the task of extract aspect sentiment triplet (targeted aspect, polarity, opinion terms). They solved it in a two step manner: (1) first identified the aspect boundary and classify its sentiment using a unified tagging schema, (2) then paired up the tuples with results from first step.

Besides, Li et al. [89] explored the joint task of aspect term extraction and polarity classification under cross-domain settings. Instead of using the common shared syntactic relationship as a bridge for domain transfer, they proposed a Selective Adversarial Learning (SAL) method to dynamically learn an alignment weight for each word.

3.5 Stance detection

Task definition Similar to aspect-level sentiment classification, the goal of stance detection is to determine the attitude of a give text towards a certain entity or claim.

However, different from aspect-level sentiment classification, in stance detection the entity may not appear in the text which makes it even more difficult. Currently, there are many well-established stance detection datasets constructed from social media data, e.g., SemEval2016 [90] from English tweets and NLPCC2016 [91] from Chinese microblogs.

Augenstein et al. [92] proposed a bi-directional conditional encoding method where they use cell output of target BLSTM to initiate the cell state of text BLSTM. By training on the combination of data regarding to different topics, their model significantly outperformed individual classifiers of previous work on tweet dataset. Du et al. [93] built a target-augmented attention layer to force the model concentrate on most salient features relating to target from the give text. Yuan et al. [94] explored the problem of answer stance detection in community question answering platforms where a model needs to decide whether an answer holds a support or against attitude towards the entity or claim mentioned in the question. They employed a recurrent conditional attention to depict the interaction between question and answer iteratively.

3.6 Datasets and evaluation metrics

Here we introduce datasets for the above mentioned fine-grained sentiment analysis tasks, which mostly come from shared tasks on product reviews and social media data. For classification-based tasks, accuracy is adopted as evaluation metric. For extraction-based tasks, F1 score is used for model comparison.

- **SemEval2014 Task4** [61] This dataset introduces annotated data from laptop and restaurant domains. There are two subtasks on laptop domain including aspect term extraction and aspect-based sentiment classification. For restaurant domain, extra tasks of aspect category detection and aspect category polarity classification are provided.
- **SemEval2015 Task12** [62] This dataset is a continuation of SemEval2014 Task4 (Table 3), where a new hotel review corpus is used as out-of-domain ABSA. The restaurant dataset here is usually used by combining with that from SemEval2014.
- **SemEval2016 Task5** [95] This dataset extends the SemEval2014 and SemEval2015 tasks with more domains and languages, where 19 training and 20 testing datasets for 8 languages and 7 domains are provided. Here, 25 were for sentence-level and 14 for text-level ABSA.

Table 3 Statistics of fine-grained sentiment analysis dataset

Name	Domain	Size	Task
SemEval2014 Task 4	Laptop	3.8k	ATE, ASC
SemEval2014 Task 4	Restaurant	3.8k	ATE, ASC, ACC, ACSC
SemEval2015 Task 12	Restaurant	1.3k	ATE, ASC, ACC, ACSC
SemEval2016 Task 5	Product review	70k	Multi-lingual and Multi-domain ASC
Twitter	Twitter	6.9k	ASC
SentiHood	CQA	5.2k	TABSA
SemEval2016 Task6	Twitter	4.8k	Stance Detection
NLPCC2016 Task4	Weibo	5.2k	Stance Detection

- **Twitter** [72] This dataset is collected by using key words (such as "bill gates", "taylor swift", "xbox", "windows 7", "google") to query the Twitter API. It is randomly sampled and balanced.
- **Sentihood** [96] This dataset is extracted from a QA platform and is supposed to be far less constrained compared with data from review specific platforms. This dataset differs from SemEval2014 in that here different aspects of several entities may be discussed in the same unit of text.
- **SemEval2016 Task6** [90] This dataset comes from Twitter platform. The key difference between this dataset and previous datasets is that target of interest may or may not be referred to in the tweet, and it may or may not be the target of opinion. The data contains five different targets ("Atheism", "Climate Change is a Real Concern", "Feminist Movement", "Hillary Clinton", and "Legalization of Abortion").
- **NLPCC2016** [91] This dataset provides a Chinese microblog version of SemEval2016 Task6, where 40000 microblogs of targets of "iPhone SE", "Set off firecrackers in the Spring Festival", "Russia's anti terrorist operations in Syria", "Two child policy", and "Prohibition of motorcycles and restrictions on electric vehicles in Shenzhen" are labeled for Task A. Besides, 2000 unlabeled microblogs for "Genetically modified food" and "The Nuclear Test in DPRK" are used as training data for Task B.

3.7 Discussion

For the past several years, the central research topics in fine-grained sentiment analysis have shifted from single extraction or classification task to jointly solving them together in an end to end fashion, while aspect-based sentiment classification still attracts lots of attention. Methods evolve from modeling the target-dependent sentence representation to leverage external resources like dependency parsing tree and sentiment lexicons. Recent emergence of deep contextualized word embeddings and pre-trained language models have introduced better methodology for utilizing large-scale unlabeled domain corpus. Besides, transfer learning models become more popular for ongoing need of reducing labors in annotation, which includes transfer knowledge from source domain(s) to target domain, from resource-rich tasks to resource-poor tasks and from high-resource languages to low-resource languages.

For future work, we would like to see better utilization of pre-trained languages on specific domains with sentiment-specific regularizations and large-scale structure information for knowledge base when the distribution in training and test data mismatch.

4 Implicit sentiment analysis

The task implicit sentiment analysis arises when no explicit sentiment words exist in a given opinionated text.

There are mainly two types of implicit sentimental texts: one is fact oriented and the other is rhetoric oriented. Fact orient texts with implicit sentiment refer to those texts describing factual information but containing sentiment polarity beneath the surface. Apart from acquiring the overall polarity, knowing what (affective events) causes the feelings and how they are expressed (implicit opinion expressions) are also very important for understanding of human's mental behaviors. Thus, we first introduce several work on implicit sentiment classification and fine-grained mining of this kind of texts. The other type is rhetoric oriented which includes sarcasm detection, irony classification, metaphor understanding, humor detection and etc. In this work, we will focus on sarcasm detection and humor detection.

4.1 Implicit sentiment classification

Task definition Given a text sequence without explicit sentiment words from sentiment lexicons, implicit sentiment classification aims at classify it into positive, neutral or negative. For example, in the hotel review "There is a layer of dust on the table", the author expressed a negative opinion as the table is dirty while it should be clean.

The task is relative difficult due to the lack of explicit opinion words. Some existing work leverages the co-occurrence between implicit and explicit opinion elements. Liao et al. [97] constructed a small factual implicit sentiment corpus and proposed a multilevel semantic fused model that incorporates syntax structure. Wei et al. [98] proposed a multi-polarity attention model which employs orthogonal regularization for attention on sentiment-bearing phrases like "falling down" and "missing the deadline", according to their polarity collected in a sentiment dictionary knowledge.

4.2 Fine-grained implicit sentiment analysis

Task definition Similarly, the task of fine-grained implicit sentiment analysis studies the identification of implicit opinion elements, which involves implicit aspect detection, implicit opinion expression extraction, and aspect-based sentiment classification.

Chen and Chen [99] investigated the problem of implicit aspect term and opinion word extraction and found that 34% of segmented reviews do not contain any opinion words and aspect terms. They extracted implicit aspect and opinion words based on co-occurrence with explicit aspects and opinion words. As no aspect terms exist in such texts, they classified these texts in hotel domain into 10 pre-defined categories and determine corresponding polarity using Bag-of-Words and CNN based models. Moreover, Li et al. [100] attempted to extract verb-expressions that imply negative opinions with Markov network and linguistic features.

Instead of extracting aspects from text in specific hotel domain, affective events requiring word knowledge are also preliminarily studied. An affective event is a desirable or undesirable event that usually implicitly binds with certain polarity, e.g., "going on vacation" is typically considered positive, while "being rushed to the hospital" is often negative. Ding and Riloff [101] explored the problem of extracting such events from personal blogs. They first novelly initiated weights in the event graph using sentiment score from a sophisticated sentiment classifier with rich features. Then they applied a label propagation algorithm to the graph and fetch those event node with sentiment. They [102] further explored the human need behind these implicit affective events.

Discussion Most of current research work on fine-grained implicit sentiment analysis is limited to opinion element extraction in specific domains, e.g., hotel reviews and co-occurrence information are not well-leveraged using simple statistical linking methods. Future work may utilize such co-occurrence information with sophisticated deep contextualized language models. Besides, benchmark datasets for both domain-specific and open-domain scenarios should be proposed to facilitate future research.

4.3 Sarcasm detection

Task definition Given a sentence, the goal is to predict this sentence as sarcastic or non-sarcastic. For example, "I absolutely love to be ignored!" is classified a sarcastic sentence as it describes a conflict feeling of "love" and "ignored".

Methods There are mainly two lines of work conducted on sarcasm detection. One line of methods is to solve it as a common classification task. These methods focus on designing more effective model architectures and incorporating more helpful information like context. Traditional approaches [103] extract hand-crafted features and use machine learning models as their classifiers. Zhang et al. [104] suggested a deep neural network model for extracting local and global features automatically. The other line of methods considers the characteristic of sarcasm detection task and attempts to model the incongruity and contrast which both are obvious clues for detecting sarcasm. Riloff et al. [105] considered sarcasm as contrast between a positive sentiment and a negative situation and the proposed model labels a text as sarcastic if it contains a positive sentiment phrase in close proximity to a negative situation phrase. Tay et al. [106] proposed an intra-attention recurrent network for modeling incongruity.

A lot of researches have been conducted on sarcasm detection. However, there is a long way to enable the models to comprehend the sarcasm in daily conversations like humans since we understand the sarcasm not only depending on text information but also visual and acoustic information such as the facial expressions. Therefore, future work could explore how to detect sarcasm in multimodal contexts.

Datasets Riloff et al. [105] collected 1600 tweets with a sarcasm hashtag (#sarcasm or #sarcastic), and 1600 tweets without these sarcasm hashtags from Twitter. And then they manually labeled these tweets as sarcastic and non-sarcastic. Finally, they obtained 742 sarcastic tweets and 2458 non-sarcastic tweets.

Oraby et al . [107] created a dataset manually which is a subset of Internet Argument Corpus (IAC) from online forums. The dataset consists of 4700 sarcastic posts and 4700 non-sarcastic posts.

Khodak et al. [108] created the Self-Annotated Reddit Corpus (SARC), a large dataset for sarcasm research. The comments ending with the marker "/s" in dataset were labeled automatically as sarcastic without checking. Then 1.3 million out of 533 million comments were labeled as sarcastic.

The statistics of above datasets are listed in Table 4.

Table 4 The statistics of sarcasm detection datasets

Dataset	Sarcastic	Total
Twitter	0.7k	3.2k
IAC	4.7k	9.4k
SARC	1.3M	553M

4.4 Humor detection

Task definition Given a sentence, the goal is to predict it as humorous or non-humorous. Take a humorous example, "The one who invented the door knocker got a No Bell prize". Understanding this example needs external knowledge, which makes this task very challenging.

Methods Humor has been studied for a long time. People are attempting to figure out what would be considered humorous and how to detect humor automatically. Yang et al. [109] investigated the semantic structure behind humor and proposed a method for extracting the humor anchors, the semantic units that enable humor, for better understanding humor. Ahuja et al. [110] focused on predicting the types of the humor and collected many datasets of jokes. Specifically, one of the datasets is topic detection dataset which is used for detecting the topics of the jokes. And they conducted a lot of experiments on these datasets using Logistic Regression and SVM. Chen et al. [111] suggested a model based on CNN and demonstrated the effectiveness of deep learning model for humor detection. Weller et al. [112] used BERT model for detecting humor and gained better results.

Many work focuses on dataset construction, only a few researchers pay attention to models. Future work could propose more effective models considering the characteristic of humor.

Datasets Yang et al. [109] extended two public datasets, the Pun of the Day dataset and the 16000 One-Liner dataset, for researching humor detection. The extended Pun of the Day dataset consists of 2423 humorous and 2403 non-humorous sentences. The extended 16000 One-Liner dataset contains 16000 humorous and 16002 non-humorous sentences.

Chen et al. [111] collected the Short Jokes dataset, which consists of 231657 humorous sentences and 231657 non-humorous sentences. The humorous sentences were collected from an open database on a Kaggle project and the non-humorous were collected from the WMT162 English news crawl.

Zhang et al. [113] created a Chinese dataset for the analysis of humor. This dataset contains 9123 manually annotated jokes. The annotation information includes the place where the jokes occur, the category of humor, and the degree of humor.

The statistics of above datasets are listed in Table 5.

5 Multi-modal sentiment analysis

Sentiment analysis is widely used for mining people's opinions from massive texts on the internet like product reviews. However, more and more online websites allow users to upload videos and images for expressing themselves, which makes the methods only focusing on texts insufficient. Hence, a lot of research works have been proposed for understanding the sentiment in multimodal data. These works not only provide the evidences for that using multimodal data can achieve better performance but also exploit how to make use of multimodal data effectively, which is still a crucial problem to be solved in multimodal sentiment analysis. In this section, we will introduce the related work by task.

There are lots of proposed tasks. We mainly introduces two typical tasks, which are related to sentiment analysis in NLP and have been researched more widely. One is target-oriented multimodal sentiment classification, which can be seen as an extension of aspect-based sentiment classification task. This task aims to identify sentiment polarities over each opinion target in an input sample which contains a sentence and an associated image. The input modalities to be handled are visual (images), and text (words). The other is multimodal sentiment classification, which is similar to sentiment classification in text. The task aims to understand the sentiment expressed by the speakers in videos. The input modalities are visual (video frames), acoustic (audio), and text (spoken words).

5.1 Target-oriented multimodal sentiment classification

Task definition Given a sample c, it contains a sentence S and an associated image I, as well as an target T(a subsequence of words in S). The goal is to predict the sentiment polarity y of the sample c toward the target T. y can be either positive, negative, or neural. An example is shown in Figure 5. The given targets and corresponding sentiment polarities are highlighted.

Methods The framework of methods for target-oriented multimodal sentiment classification is illustrated in Figure 6. First, the encoder components are used for extracting

Table 5 The statistics of humor detection datasets

Dataset	Humorous	Total	Language
Pun of the Day	2423	4826	English
16000 One-Liner	16000	32002	English
Short Jokes	231657	463314	English
Chinese Jokes	9123	9123	Chinese

Figure 5 (Color online) An example of target-oriented multimodal sentiment classification.

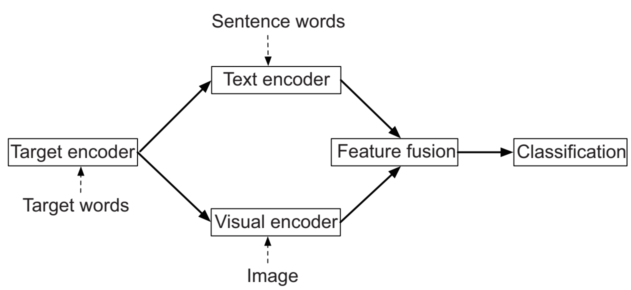

Figure 6 The framework of methods for target-oriented multimodal sentiment classification.

features. Then, the target information is leveraged to refine the visual and textual features. Finally, multimodal features are fused for classification. Yu et al. [114] focused on refining visual features by the given target information and proposed the target-oriented multimodal BERT(TomBERT) motivated by this observation that the associated image in a sample can highlight the focused target and reflect a user's sentiment over the focused target. Specifically, TomBERT first learned target-dependent image features and then fused it with text features using transformer model. However, TomBERT ignored the interaction between the target and the sentence. Xu et al . [115] suggested the multi-interactive memory network (MIMN) which includes two interactive memory networks to supervise the textual and visual information with the given target. Both works demonstrate that effective incorporation of image information can help the models to achieve better performance.

This fresh task has gained attention recently and there are many problems need to be explored. One problem is how to take the advantage of target information. Future works could draw inspiration from related works in aspect-based sentiment analysis to address it.

Datasets Yu et al. [114] built the datasets based on the two publicly available multimodal datasets TWITTER-15 and TWITTER-17 for investigating this task. They annotated the sentiment label towards each target in each tweet. The statistics of the datasets are listed in Table 6.

5.2 Multimodal sentiment classification

Task definition Given an utterance u, it consists of three modalities: text (spoken words), visual (video frames), acoustic (audio). The goal is to predict the sentiment or emotion label y. An utterance example is shown in Figure 7. This utterance conveys negative sentiment by sarcasm.

Methods The framework of methods for multimodal sentiment classification is illustrated in Figure 8. The encoders are first used for extracting multimodal features from raw data. And then the features are fused for classification. Most works focus on how to fuse multimodal features effectively for taking full advantage of multimodal data. Zadeh et al. [116] proposed the tensor fusion network (TFN), which used three-fold Cartesian product to fuse the trimodal utterance-level features. TFN used three subnetworks to learn different modal features and then fused them to obtain the final

Table 6 The statistics of TWITTER-15 and TWITTER-17 datasets

Dataset	Train	Dev	Test
TWITTER-15	3179	1122	1037
TWITTER-17	3562	1176	1234

Figure 7 (Color online) An utterance example which conveys negative sentiment.

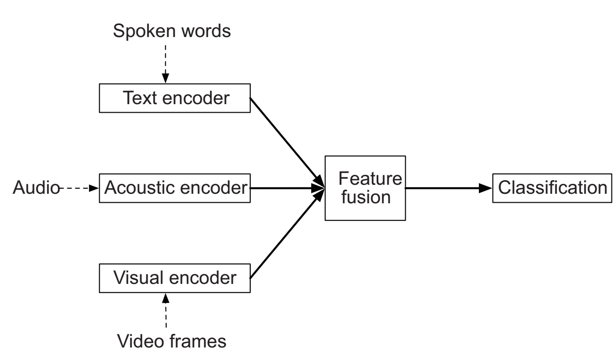

Figure 8 The framework of methods for multimodal sentiment classification.

representation of the utterances.

However, there are two limitations for TFN. One is the exponentially increasing computational complexity, as the outer product operation results in high dimensional representations. To address this problem, Liu et al. [117] suggested the low-rank multimodal fusion (LMF) model, which used low-rank tensors to perform multimodal fusion for reducing computational complexity.

The other is fusing features in utterance-level loses lots of local information which is contained in the word-level features. Chen et al. [118] proposed the gated multimodal embedding LSTM with temporal attention (GME-LSTM(A)) to tackle this problem. GME-LSTM(A) used gate mechanism to refine the multimodal features and fused them in word-level. Zadeh et al. [119] designed the memory fusion network (MFN) which captured the interactions across both different modalities and time. Inspired by the discoveries of cognitive neuroscience, Liang et al. [120] proposed the recurrent multistage fusion network (RMFN) which decomposed the fusion problem into multiple stages. At each recursive stage, RMFN selected a subset of multimodal signals from the signals set in a given timestep and fused with previous fusion representations. Above works all attempt to solve this problem, how to fuse multimodal features effectively, which remains to be explored.

A lot of models have been proposed for fusing the multimodal features effectively. However, the encoder components are not studied enough. Most of existing works use the tools to extract multimodal features and train their own projection layers and fusion layers without finetuning the encoders, which makes the encoders could not extract good enough features for sentiment classification. Therefore, besides the study of fusing multimodal features, future works could pay more attention to encoders.

Datasets ICT YouTube opinion dataset (ICT-Youtube) [121] is a collection of opinion videos collected from YouTube. It consists of 47 videos of people expressing their opinions about a variety of topics. Each video is labeled with sentiment label (negative, neural, positive).

Institute for Creative Technologies' Multi-Modal Movie Opinion (ICT-MMMO) [122] is the dataset of multimodal sentiment classification which contains 370 multimodal review videos. For each video, there is one person expressing his/her opinions. The creators labeled the video utterances with sentiment labels (strongly negative, weakly negative, neural, weakly positive, strongly positive).

CMU multimodal opinion-level sentiment intensity (CMU-MOSI) is created by Zadeh et al. [123] for investigating multimodal sentiment classification. They collected the 93 videos from the YouTube website. The length of the videos varies from 2–5 min. These videos were then split into 2199 short video clips. They labeled these video clips with sentiment labels (highly negative, negative, weakly negative, neutral, weakly positive, positive, highly positive).

CMU multimodal opinion sentiment and emotion intensity (CMU-MOSEI) [124] is the largest dataset of sentiment analysis and emotion recognition recently. it consists of 23453 annotated video utterances from 1000 distinct speakers and 250 topics. Each utterance is annotated with sentiment labels (highly negative, negative, weakly negative, neutral, weakly positive, positive, highly positive) and emotion labels (happiness, sadness, anger, fear, disgust, surprise).

The statistics of above datasets are listed in Table 7. We can see that the size of the created dataset is increasingly large and the annotation information is more and more rich.

6 Sentimental text generation

As emotional intelligence in NLP consist of both the understanding of sentiment in text and the generation of text with sentiment, we discuss various tasks involved in sentimental text generation in this section.

6.1 Affective dialogue system

In dialogue systems, the emotional communication between humans and machines is an important issue. The machine needs to detect the user's emotions to achieve an understanding of the emotions, and the machine also needs to incorporate the appropriate emotions into the response to achieve the expression of emotions. The dialogue system with excellent emotional ability can fully sense the emotional state of the user, realize emotional comfort to the user by replying, and improve the user experience.

There are mainly two tasks. One is Emotion Recognition in Conversation, and the other is Emotional Conversation

Table 7 The statistics of multimodal sentiment classification datasets

Dataset	Utterances	Sentiment Lable	Emotion Lable	Source
ICT-Youtube	47	✓		Youtube
ICT-MMMO	370	✓		Youtube, ExpoTV
CMU-MOSI	2199	✓		Youtube
CMU-MOSEI	23453	✓	✓	Youtube

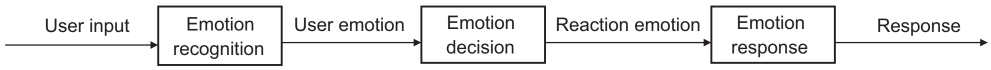

Figure 9 A general architecture of emotional chatbots.

Generation. These two tasks and an extra emotional decision module can be organized into a complete emotional chatbot in a pipeline manner as shown in Figure 9. In this section, we introduce these two tasks in detail.

6.1.1 Emotion recognition in conversation

Task definition Emotion recognition in conversation is a classification task that aims to detect emotions of the utterances in a conversation. The input of the task is a continuous dialogue, and the output is the emotions of all utterances in the dialogue. Figure 10 demonstrates a sample conversation from the IEMOCAP dataset.

Because there are many specific factors in conversation, the emotion detection in utterances is different from that in independent sentences, it requires a comprehensive consideration of the background, context, and speaker information in the dialogue, these are unique challenges of the task.

Emotion Recognition in Conversation can be used in various conversation scenes, such as sentiment analysis of comments in social media, emotion detection of customers in artificial services. In addition, Emotion Recognition in Conversation can also be applied to chatbots to analyze the user's emotions in real time, that is helpful for implementing emotion-driven response generation.

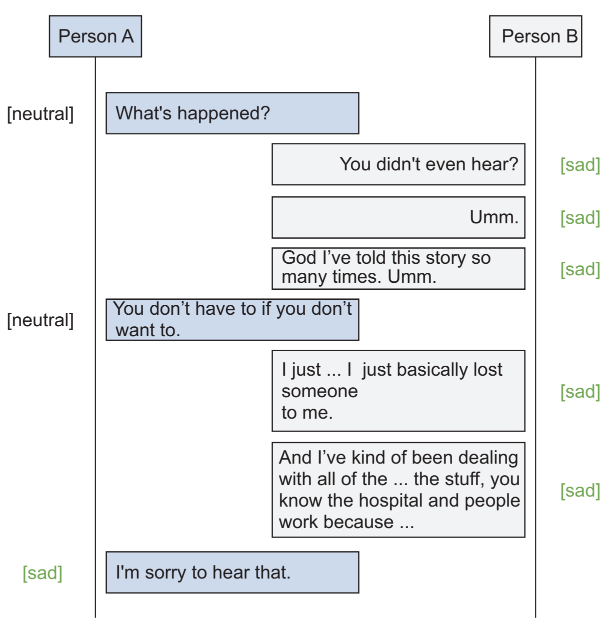

Figure 10 (Color online) An example conversation with annotated labels from the IEMOCAP dataset.

Methods We divide related works into two categories: context-dependent models and speaker-dependent models.

Context-dependent models Dialogue context can provide more information for the emotion detection in utterances. Poria et al. [125] proposed the Contextual LSTM model (C-LSTM), which used LSTM model to capture contextual features. Jiao et al. [126] suggested the hierarchical gated recurrent units (HiGRU), which introduced a word-level GRU and an utterance-level GRU with self-attention and features fusion. Zhong et al. [127] proposed the knowledge-enriched transformer model (KET) to exploit external commonsense knowledge and designed a new graph attention mechanism in the model.

Speaker-dependent models In addition to caring about the context of the conversation, the state of the speakers and the inter-speaker dependency relations also need to be considered. Hazarika et al. [128] proposed the conversational memory network (CMN). CMN used different GRUs for both parties in the conversation to capture contextual features, and then fused the current utterance representation with contextual features using memory networks, that was finally used for sentiment classification. Hazarika et al. [129] proposed the interactive conversational memory network (ICON). ICON is similar to CMN, but the main difference is that ICON introduced a Dynamic Global Influence Module (DGIM) for global contextual features fusion.

Although CMN and ICON considered the influence of different speakers, but which speaker the current utterance belongs to is unknown. Majumder et al. [130] proposed the DialogueRNN to tackle this problem, which set separate states for each speaker and associated states with the speaker's utterance. DialogueRNN assumed that the emotion of an utterance depends on three factors: the speaker, the context given by the preceding utterances, the emotion behind the preceding utterances, and used three GRUs to capture them. Zhang et al. [131] designed the Conversational GCN (ConGCN), which represented each utterance and each speaker as a node and linked the utterances to the speakers by undirected edges. Ghosal et al. [132] proposed the dialogue graph convolutional network (DialogueGCN), which also used a graph convolutional neural network to model the conversation. The graph constructed by DialogueGCN contains only utterance nodes, but the type of edge is determined based on the speaker information that associates the utterances with the speakers indirectly.

This type of method is the mainstream method of emotion recognition in conversation. The general idea is to consider the impact of the user dimension, such as the state of the speakers, the inter-speaker dependency relations, and the emotional continuity. A general architecture is shown in Figure 11.

In summary, all the studies in this task are mainly verified on several datasets whose conversations are temporary. However, the conversations in the real scene are usually continuous, and the speakers also have long-term state. These fators are ignored of existing works and further research is needed.

Datasets We summarize datasets related to Emotion Recognition in Conversation. The details are presented in Table 8.

- IEMOCAP [133] This dataset was collected at SAIL lab at USC. It consists of approximately 12 h of multimodal conversations data. It is annotated into six emotion categories and contains 151 dialogues with a total of 7433 utterances.

- SEMAINE [134] SEMAINE database was collected by Queen's University Belfast. It contains 959 dialogues involving total 150 participants, and each dialogue is annotated in five affective dimensions and 27 associated categories. SEMAINE was used in part by AVEC 2012 [135] whose dataset contains 95 dialogues with a total of 5798 utterances and retains only four affective dimensions. In existing works, AVEC 2012 dataset is widely used instead of SEMAINE database.

- DailyDialog [136] DailyDialog is a manually labelled high-quality multi-turn dialogue dataset. It reflects daily communication way and cover various topics about daily life. It is annotated into seven emotion categories and contains 12218 dialogues with a total of 103607 utterances.

- EmotionLines [137] This dataset contains conversations from Friends TV show transcripts and real chatting logs, each of them contains 1000 dialogues. It is annotated into seven emotion categories and contains 29245 utterances.

- EmoContext [137] EmoContext consists of conversations with only the last utterance labeled. It is annotated into four emotion categories and contains 38421 dialogues with a total of 115263 utterances.

- MELD [138] MELD contains conversations from Friends TV show transcripts, which is a multimodal extension of EmotionLines. It is annotated into seven emotion categories and contains 1433 dialogues with a total of 13708 utterances.

Table 8 Datasets related to emotion recognition in conversation

Dataset	Dialogues	Utterances	Classes
IEMOCAP	151	7433	6
AVEC 2012	95	5798	-
DailyDialog	12218	103607	7
EmotionLines	2000	29245	7
EmoContext	38421	115263	4
MELD	1433	13708	7

6.1.2 Emotional conversation generation

Task definition Emotional conversation generation is a generation task that aims to generate emotional and targeted responses. For emotion in response to be generated, there are generally two different views. One is the emotion of the response needs to be given, the advantage is that the emotion is flexible and controllable, and the challenge is that large-scale emotion annotation dialogue corpus is needed. The other is the emotion of the response is already implicit in the dialogue, the advantage and challenge are the opposite of the other. Table 9 demonstrates an exmaple for this task with given emotion.

Emotional conversation generation is mainly used in chatbots, which allows chatbots to generate emotional responses

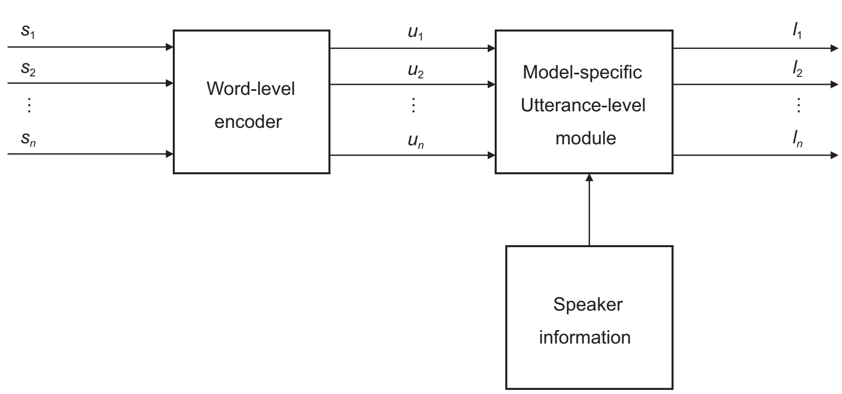

Figure 11 The framework of speaker-dependent models for emotion recognition in conversation.

Table 9 An example for emotional conversation generation with given emotion

Post	Emotion	Response
Sir, this is the police, please open the door immediately.	Neural	Okay, please wait a moment.
	Happy	Here you come finally!
	Fear	No-no-no, no!
	Surprise	Police? Are there any problems?
	Anger	Don't make any noise!

and achieve the expression of emotions.

Methods We divide related works into two categories: conversation models with given emotion and conversation models without given emotion.

Conversation models with given emotion Given the context and emotion information, a response with the specified emotion can be generated. Zhou et al. [139] proposed the Emotional chatting machine (ECM), which is the first work that addresses the emotion factor in large-scale conversation generation. ECM introduced three new mechanisms: emotion category embedding, internal memory, and external memory, that worked together to generate a response with given emotion. ECM adopted perplexity, emotion accuracy and manual evaluation to evaluate the model. Song et al. [140] proposed an emotional dialogue system (EmoDS), which added two modules to seq2seq model: lexicon-based attention mechanism and emotion classification. The lexicon-based attention mechanism was designed to replace the words of the response with their synonyms in an emotion lexicon, and the emotion classification was used to guide the response generation process. EmoDS employed embedding score, BLEU score, distinct, emotion evaluation and human evaluation to evaluate the model.

This type of method is the mainstream method of Emotional conversation generation. The general approach is to add some mechanisms to seq2seq model, which make the generated responses contain emotions. A general architecture is shown in Figure 12. Articles with this type method include: Colombo et al. [141], Huang et al. [142], Zhong et al. [143], Zhou et al. [144] and Lubis et al. [145]. In particular, Zhou et al. [144] used emojis in Twitter data as emotion annotations, which is the only work using emojis in related articles.

Conversation models without given emotion Emotion information is not necessary, this method assumes that the emotion of the response is inherently determined by the context. Asghar et al. [146] proposed the affective neural response generation model, which was based on seq2seq model. The model used three ways to introduce emotional factors: affective word embeddings, affect-based objective functions, and affectively diverse beam search for decoding. This work employed syntactic coherence, naturalness and emotional appropriateness to evaluate dialogue systems.

In summary, all the studies only consider single-turn conversation, but pay no attention to multi-turn conversation. In addition, the study combined emotion recognition in conversation and emotional conversation generation has not appeared yet, which brings difficulties to real applications. In the future, these issues deserve attention.

Datasets We summarize datasets related to emotional conversation generation. The details are presented in Table 10.

- STC [147] STC contains sina weibo corpus without emotion annotation, which are collected by Huawei. This dataset contains 4435959 post-response pairs for training.

- Cornell Movie Dialogs [148] Cornell University collected movie dialogue corpus without emotion annotation and constructed this dataset. This dataset contains 220579 dialogues and 304713 utterances, and involves 9035 characters from 617 movies.

- OpenSubtitles [149] OpenSubtitles consists of movie subtitle corpus collected by Open Subtitles Website without emotion annotation, and the latest version contains 3735070 dialogues. This dataset is usually used after filtering, and millions of dialogues can be left.

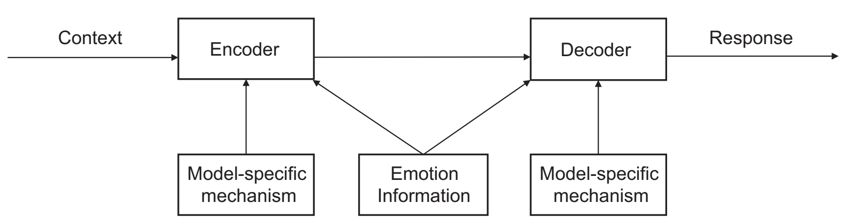

Figure 12 The framework of conversation generation models with given emotion.

Table 10 Datasets related to emotional conversation generation

Dataset	Source	Size	Annotation
STC	Sina Weibo	4.4m	No
Cornell Movie Dialogs	Movie Dialogue	220k	No
OpenSubtitles	Movie Subtitle	3.7m	No
Twitter	Twitter	662k	Yes
DailyDialog	Daily Life	12.2k	Yes
SEMAINE	Human-Recorded	0.9k	Yes

- Twitter [144] This dataset consists of twitter corpus with emoji. It contains 662159 post-response pairs and uses 64 common emojis as labels.
- DailyDialog [136] This dataset is detailed in Sect. 6.1.1.
- SEMAINE [134] This dataset is detailed in Sect. 6.1.1.

6.2 Other related tasks

There are also some other sentimental text generation-related tasks.

Opinion summarization Opinion summarization aims to generate a summary for a given aspect from user opinions, it is useful in situations where there are a large number of user opinions, such as in social media. Opinion summarization was introduced by Hu and Liu [43] in 2004. In that work, they decomposed this task to three subtasks: aspect extraction, sentiment prediction and summary generation, and later studies usually continued this tradition. In the following, we introduce some methods based on neural network for solving this task. Wu et al. [150] proposed two convolutional neural network (CNN) models: cascaded CNN and multitask CNN, which are designed to tackle aspect mapping and sentiment classification. Wang et al. [151] suggested an attention-based neural network model, which gathers information from multiple reviews to construct summaries. Angelidis et al. [152] proposed a neural framework required light supervision, which combines a multi-task training aspect extractor and a multiple instance learning sentiment predictor. Huy et al. [153] used aspect similarity recognition (ASR) to relax the limitation of predefining aspects, which solves the domain adaptation problem of Opinion Summarization. Zhao et al. [154] introduced external domain knowledge and proposed the AspMem method, which is used to identify aspects automatically. In summary, this task is less researched and has not yet received widespread attention.

Text sentiment transfer Text sentiment transfer studies on the problem of change the sentiment of text while preserving the non-sentimental contents. The task is difficult as such parallel corpus is not easily accessible, so researches generally revolves around this issue. Shen et al. [155] proposed a style transfer method, which exploits alignment of latent representations. Li et al. [156] proposed a simple "Delete, Retrieve, Generate" based method which first deletes distinctive sentimental words from the original text, then retrieves candidate phrases with target polarity and employs a neural model to improve fluency of combined text and phrases. Yang et al. [157] replaced the discriminator into a target domain language model in GAN-based unsupervised style transfer systems, which makes the training process more stable. Fu et al. [158] proposed two seq2seq models for style transfer: multi-decoder and style-embedding, which use adversarial networks to learn content and style representations. Prabhumoye et al. [159] leveraged a machine translation model to learn the representations of the input sentences, and then used the adversarial generation method to match the target style. Xu et al. [160] proposed a cycled reinforcement learning (CRL) approach to translate the underlying sentiment of original text. Due to the absence of parallel data, they construct supervision for deleting sentiment words with the help of attention weights from a attention-based sentiment classifier. Then a seq2seq model with two decoders for each sentiment respectively is trained using the deleted sentence and original sentence. In summary, text sentiment transfer is a new task appeared in recent years which attracts many researchers. However, there are still many issues to be resolved, this task will attract more attention in future.

7 Conclusion

In this paper, we give an overview of recent research literature in deep learning based sentiment analysis. We introduce tasks from coarse-grained sentiment classification tasks to fine-grained opinion mining, from surface-level detection to implicit sentiment analysis and from understanding sentiment in text to generating text with sentiment.

While tailor-made neural networks with attention structure and pre-trained language model feature have achieved great success in traditional sentiment classification and extraction tasks, the problems of domain and language shift in fine-grained sentiment analysis have not been extensively studied due the limit amount of annotated data and domains. Though some work makes initial attempt to leverage knowledge from large-scale unlabeled data, linguistic resources and knowledge bases, this work has not figure out what really lies behind this data and what is really needed for specific tasks. Furthermore, benchmark datasets need to be set up to foster those research tasks.

For explainable sentiment analysis beyond opinion expressions and emotion causes, some earlier work proposes to acquire affective events that implicitly carry polarities and classify these events into categories defined by in-depth human needs. In the future, we would like to see automatic detection of those affective events and implicit opinion expression further boosting the performance of general sentiment analysis task.

This work was supported by the National Key R&D Program of China (Grant No. 2018YFB1005103), and the National Natural Science Foundation of China (Grant Nos. 61632011 and 61772153). The first author was supported by China Scholarship Council (CSC) during a visit to the University of Copenhagen.

1 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521: 436–444
2 Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge: MIT Press, 2016
3 Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 4171–4186
4 Cambria E. Affective computing and sentiment analysis. IEEE Intell Syst, 2016, 31: 102–107
5 Kim Y. Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics, 2014. 1746–1751
6 Wang B. Disconnected recurrent neural networks for text categorization. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 2311–2320
7 Lin Z, Feng M, Santos C N d, et al. A structured self-attentive sentence embedding. ArXiv: 1703.03130
8 Wang Y, Sun A, Han J, et al. Sentiment analysis by capsules. In: Proceedings of the 2018 World Wide Web Conference. Lyon: International World Wide Web Conferences Steering Committee, 2018. 1165–1174
9 Teng Z, Vo D T, Zhang Y. Context-sensitive lexicon features for neural sentiment analysis. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austion: Association for Computational Linguistics, 2016. 1629–1638
10 Tay Y, Luu A T, Hui S C, et al. Attentive gated lexicon reader with contrastive contextual co-attention for sentiment classification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 3443–3453
11 Socher R, Perelygin A, Wu J, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: Association for Computational Linguistics, 2013. 1631–1642
12 Tai K S, Socher R, 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 (Volume 1: Long Papers). Beijing: Association for Computational Linguistics, 2015. 1556–1566
13 Zhu X, Sobihani P, Guo H. Long short-term memory over recursive structures. In: Proceedings of the International Conference on Machine Learning. Lille: PMLR, 2015. 1604–1612
14 Qian Q, Huang M, Lei J, et al. Linguistically regularized LSTM for sentiment classification. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 1679–1689
15 Zhang T, Huang M, Zhao L. Learning structured representation for text classification via reinforcement learning. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 6053–6060
16 Tang D, Qin B, 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. Lisbon: Association for Computational Linguistics, 2015. 1422–1432
17 Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: Association for Computational Linguistics, 2016. 1480–1489
18 Tang D, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing: Association for Computational Linguistics, 2015. 1014–1023
19 Chen H, Sun M, Tu C, et al. Neural sentiment classification with user and product attention. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 1650–1659
20 Dou Z Y. Capturing user and product information for document level sentiment analysis with deep memory network. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 521–526
21 Wu Z, Dai X Y, Yin C, et al. Improving review representations with user attention and product attention for sentiment classification. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 5989–5996
22 Yao L, Mao C, Luo Y. Graph convolutional networks for text classification. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 7370–7377
23 Tang D, Wei F, Yang N, et al. 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). Baltimore: Association for Computational Linguistics, 2014. 1555–1565
24 Li Z, Zhang Y, Wei Y, et al. End-to-end adversarial memory network for cross-domain sentiment classification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI Organization, 2017. 2237–2243
25 Yu J, Jiang J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 236–246
26 Li Z, Wei Y, Zhang Y, et al. Hierarchical attention transfer network for cross-domain sentiment classification. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 5852–5859
27 Zhang K, Zhang H, Liu Q, et al. Interactive attention transfer network for cross-domain sentiment classification. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 5773–5780
28 Liu P, Qiu X, Huang X. Adversarial multi-task learning for text clas-

sification. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 1–10
29 Chen X, Cardie C. Multinomial adversarial networks for multi-domain text classification. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 1226–1240
30 Liu Q, Zhang Y, Liu J. Learning domain representation for multi-domain sentiment classification. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 541–550
31 Zheng R, Chen J, Qiu X. Same representation, different attentions: Shareable sentence representation learning from multiple tasks. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm: IJCAI Organization, 2018. 4616–4622
32 Cai Y, Wan X. Multi-domain sentiment classification based on domain-aware embedding and attention. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macco: IJCAI Organization, 2019. 4904–4910
33 Zhou X, Wan X, Xiao J. Cross-lingual sentiment classification with bilingual document representation learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin: Association for Computational Linguistics, 2016. 1403–1412
34 Chen X, Sun Y, Athiwaratkun B, et al. Adversarial deep averaging networks for cross-lingual sentiment classification. Trans Association Comput Linguist, 2018, 6: 557–570
35 Chen X, Awadallah A H, Hassan H, et al. Multi-source cross-lingual model transfer: Learning what to share. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 3098–3112
36 Akhtar M S, Sawant P, Sen S, et al. Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 572–582
37 Yu J, Marujo L, Jiang J, et al. Improving multi-label emotion classification via sentiment classification with dual attention transfer network. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 1097–1102
38 Felbo B, Mislove A, Søgaard A, et al. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 1615–1625
39 Kim Y, Lee H, Jung K. AttnConvnet at SemEval-2018 Task 1: Attention-based convolutional neural networks for multi-label emotion classification. In: Proceedings of The 12th International Workshop on Semantic Evaluation. New Orleans: Association for Computational Linguistics, 2018. 141–145
40 Abdul-Mageed M, Ungar L. EmoNet: Fine-grained emotion detection with gated recurrent neural networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 718–728
41 Wang Z, Lee S, Li S, et al. Emotion detection in code-switching texts via bilingual and sentimental information. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Beijing: Association for Computational Linguistics, 2015. 763–768
42 Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002). Association for Computational Linguistics, 2002. 79–86
43 Hu M, Liu B. Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery
44 Rosenthal S, Farra N, Nakov P. SemEval-2017 task 4: Sentiment analysis in twitter. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). Vancouver: Association for Computational Linguistics, 2017. 502–518
45 Maas A L, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland: Association for Computational Linguistics, 2011. 142–150
46 Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. In: Advances in Neural Information Processing Systems 28. Montreal: Curran Associates, Inc., 2015. 649–657
47 Blitzer J, Dredze M, Pereira F. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Prague: Association for Computational Linguistics, 2007. 440–447
48 Strapparava C, Mihalcea R. SemEval-2007 task 14: Affective text. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). Prague: Association for Computational Linguistics, 2007. 70–74
49 Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 1433–1443
50 Li X, Bing L, Li P, et al. Aspect term extraction with history attention and selective transformation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm: IJCAI Organization, 2018. 4194–4200
51 Ma D, Li S, Wu F, et al. Exploring sequence-to-sequence learning in aspect term extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 3538–3547
52 Liao M, Li J, Zhang H, et al. Coupling global and local context for unsupervised aspect extraction. 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). Hong Kong: Association for Computational Linguistics, 2019. 4579–4589
53 Shu L, Xu H, Liu B. Lifelong learning CRF for supervised aspect extraction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Vancouver: Association for Computational Linguistics, 2017. 148–154
54 Xu H, Liu B, Shu L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne: Association for Computational Linguistics, 2018. 592–598
55 Yin Y, Wei F, Dong L, et al. Unsupervised word and dependency path embeddings for aspect term extraction. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: IJCAI/AAAI Press, 2016. 2979–2985
56 He R, Lee W S, Ng H T, et al. An unsupervised neural attention model for aspect extraction. In: Proceedings of the 55th Annual Meeting of

the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 388–397

57 Karamanolakis G, Hsu D, Gravano L. Leveraging just a few keywords for fine-grained aspect detection through weakly supervised co-training. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 4611–4621

58 Wang W, Pan S J. Transition-based adversarial network for cross-lingual aspect extraction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm: IJCAI Organization, 2018. 4475–4481

59 Jebbara S, Cimiano P. Zero-shot cross-lingual opinion target extraction. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 2486–2495

60 Irsoy O, Cardie C. Opinion mining with deep recurrent neural networks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics, 2014. 720–728

61 Pontiki M, Galanis D, Pavlopoulos J, et al. SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Dublin: Association for Computational Linguistics, 2014. 27–35

62 Pontiki M, Galanis D, Papageorgiou H, et al. Semeval-2015 task 12: Aspect based sentiment analysis. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Denver: Association for Computational Linguistics, 2015. 486–495

63 Tang D, Qin B, Feng X, et al. Effective LSTMs for target-dependent sentiment classification. In: Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 3298–3307

64 Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 214–224

65 Chen P, Sun Z, Bing L, et al. Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 452–461

66 Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 606–615

67 Vo D, Zhang Y. Target-dependent twitter sentiment classification with rich automatic features. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 1347–1353

68 Liu J, Zhang Y. Attention modeling for targeted sentiment. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2. Short Papers. Valencia: Association for Computational Linguistics, 2017. 572–577

69 Wang B, Lu W. Learning latent opinions for aspect-level sentiment classification. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, the 30th innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018. 5537–5544

70 Ma D, Li S, Zhang X, et al. Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI Organization, 2017. 4068–4074

71 Sun C, Huang L, 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, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 380–385

72 Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Baltimore: Association for Computational Linguistics, 2014. 49–54

73 He R, Lee W S, Ng H T, et al. Effective attention modeling for aspect-level sentiment classification. In: Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe: Association for Computational Linguistics, 2018. 1121–1131

74 Ruder S, Ghaffari P, Breslin J G. A hierarchical model of reviews for aspect-based sentiment analysis. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 999–1005

75 Hazarika D, Poria S, Vij P, et al. 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, Volume 2 (Short Papers). New Orleans: Association for Computational Linguistics, 2018. 266–270

76 Majumder N, Poria S, Gelbukh A, et al. IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 3402–3411

77 Ma Y, Peng H, 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. New Orleans: AAAI Press, 2018.

78 He R, Lee W S, Ng H T, et al. Exploiting document knowledge for aspect-level sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne: Association for Computational Linguistics, 2018. 579–585

79 Chen Z, Qian T. Transfer capsule network for aspect level sentiment classification. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 547–556

80 Li Z, Wei Y, Zhang Y, et al. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 4253–4260

81 Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 616–626

82 Li X, Lam W. Deep multi-task learning for aspect term extraction with memory interaction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 2886–2892

83 Wang W, Pan S J, Dahlmeier D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2017. 3316?3322

84 Wang W, Pan S J. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 2171–2181

85 Wang W, Pan S J. Transferable interactive memory network for domain adaptation in fine-grained opinion extraction. In: Proceedings

of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 7192–7199

86 Luo H, Li T, Liu B, et al. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 591–601

87 He R, Lee W S, Ng H T, et al. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 504–515

88 Peng H, Xu L, Bing L, et al. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020. 8600–8607

89 Li Z, Li X, Wei Y, et al. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 4590–4600

90 Mohammad S, Kiritchenko S, Sobhani P, et al. SemEval-2016 task 6: Detecting stance in tweets. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). San Diego: Association for Computational Linguistics, 2016. 31–41

91 Xu R, Zhou Y, Wu D, et al. Overview of NLPCC shared task 4: Stance detection in Chinese microblogs. In: Natural Language Understanding and Intelligent Applications. Kunming: Springer, 2016. 907–916

92 Augenstein I, Rocktäschel T, Vlachos A, et al. Stance detection with bidirectional conditional encoding. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 876–885

93 Du J, Xu R, He Y, et al. Stance classification with target-specific neural attention. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI Organization, 2017. 3988–3994

94 Yuan J, Zhao Y, Xu J, et al. Exploring answer stance detection with recurrent conditional attention. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 7426–7433

95 Pontiki M, Galanis D, Papageorgiou H, et al. Semeval-2016 task 5: Aspect based sentiment analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016). San Diego: Association for Computational Linguistics, 2016. 19–30

96 Saeidi M, Bouchard G, Liakata M, et al. SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In: Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 1546–1556

97 Liao J, Wang S, Li D. Identification of fact-implied implicit sentiment based on multi-level semantic fused representation. Knowledge-Based Syst, 2019, 165: 197–207

98 Wei J, Liao J, Yang Z, et al. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. Neurocomputing, 2020, 383: 165–173

99 Chen H Y, Chen H H. Implicit polarity and implicit aspect recognition in opinion mining. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Berlin: Association for Computational Linguistics, 2016. 20–25

100 Li H, Mukherjee A, Si J, et al. Extracting verb expressions implying negative opinions. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin: AAAI Press, 2015. 2411–2417

101 Ding H, Riloff E. Acquiring knowledge of affective events from blogs using label propagation. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016. 2935–2942

102 Ding H, Riloff E. Human needs categorization of affective events using labeled and unlabeled data. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 1919–1929

103 Ptáček T, Habernal I, Hong J. Sarcasm detection on czech and english twitter. In: Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers. Dublin: Dublin City University and Association for Computational Linguistics, 2014. 213–223

104 Zhang M, Zhang Y, Fu G. Tweet sarcasm detection using deep neural network. In: Proceedings of The 26th International Conference on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 2449–2460

105 Riloff E, Qadir A, Surve P, et al. Sarcasm as contrast between a positive sentiment and negative situation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Seattle: Association for Computational Linguistics, 2013 704–714

106 Tay Y, Luu A T, Hui S C, et al. Reasoning with sarcasm by reading in-between. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 1010–1020

107 Oraby S, Harrison V, Reed L, et al. Creating and characterizing a diverse corpus of sarcasm in dialogue. In: Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Los Angeles: Association for Computational Linguistics, 2016. 31–41

108 Khodak M, Saunshi N, Vodrahalli K. A large self-annotated corpus for sarcasm. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki: European Language Resources Association (ELRA), 2018. 641–646

109 Yang D, Lavie A, Dyer C, et al. Humor recognition and humor anchor extraction. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 2367–2376

110 Ahuja V, Bali T, Singh N. What makes us laugh? Investigations into automatic humor classification. In: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media. New Orleans: Association for Computational Linguistics, 2018. 1–9

111 Chen P Y, Soo V W. Humor recognition using deep learning. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans: Association for Computational Linguistics, 2018. 113–117

112 Weller O, Seppi K. Humor detection: A transformer gets the last laugh. 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), Hong Kong: Association for Computational Linguistics, 2019. 3612–3616

113 Zhang D, Zhang H, Liu X, et al. Telling the whole story: A manually annotated chinese dataset for the analysis of humor in jokes. 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). Hong Kong: Association for Computational Linguistics, 2019. 6403–6408

114 Yu J, Jiang J. Adapting bert for target-oriented multimodal sentiment classification. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macco: IJCAI Organization, 2019. 5408–5414

115 Xu N, Mao W, Chen G. Multi-interactive memory network for aspect based multimodal sentiment analysis. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 371–378
116 Zadeh A, Chen M, Poria S, et al. Tensor fusion network for multimodal sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 1103–1114
117 Liu Z, Shen Y, Lakshminarasimhan V B, et al. Efficient low-rank multimodal fusion with modality-specific factors. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 2247–2256
118 Chen M, Wang S, Liang P P, et al. Multimodal sentiment analysis with word-level fusion and reinforcement learning. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction. Glasgow: Association for Computing Machinery, 2017. 163–171
119 Zadeh A, Liang P P, Mazumder N, et al. Memory fusion network for multi-view sequential learning. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018. 5634–5641
120 Liang P P, Liu Z, Zadeh A B, et al. Multimodal language analysis with recurrent multistage fusion. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 150–161
121 Morency L P, Mihalcea R, Doshi P. Towards multimodal sentiment analysis: Harvesting opinions from the web. In: Proceedings of the 13th international conference on multimodal interfaces. Alicante:Association for Computing Machinery, 2011. 169–176
122 Wollmer M, Weninger F, Knaup T, et al. YouTube movie reviews: Sentiment analysis in an audio-visual context. IEEE Intell Syst, 2013, 28: 46–53
123 Zadeh A. Micro-opinion sentiment intensity analysis and summarization in online videos. In: Proceedings of the ACM on International Conference on Multimodal Interaction. 2015. 587–591
124 Zadeh A B, Liang P P, Poria S, et al. 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 (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 2236–2246
125 Poria S, Cambria E, Hazarika D, et al. Context-dependent sentiment analysis in user-generated videos. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 873–883
126 Jiao W, Yang H, King I, et al. HiGRU: Hierarchical gated recurrent units for utterance-level emotion recognition. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 397–406
127 Zhong P, Wang D, Miao C. Knowledge-enriched transformer for emotion detection in textual conversations. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 165–176
128 Hazarika D, Poria S, Zadeh A, et al. Conversational memory network for emotion recognition in dyadic dialogue videos. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 2122–2132
129 Hazarika D, Poria S, Mihalcea R, et al. ICON: Interactive conversational memory network for multimodal emotion detection. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 2594–2604
130 Majumder N, Poria S, Hazarika D, et al. Dialoguernn: An attentive RNN for emotion detection in conversations. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, and the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019. 6818–6825
131 Zhang D, Wu L, Sun C, et al. Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macco: IJCAI Organization, 2019. 5415–5421
132 Ghosal D, Majumder N, Poria S, et al. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. 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). Hong Kong: Association for Computational Linguistics, 2019. 154–164
133 Busso C, Bulut M, Lee C C, et al. IEMOCAP: Interactive emotional dyadic motion capture database. Lang Resources Evaluation, 2008, 42: 335–359
134 McKeown G, Valstar M F, Cowie R, et al. The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Trans Affective Comput, 2012, 3: 5–17
135 Schuller B W, Valstar M F, Eyben F, et al. AVEC 2012: The continuous audio/visual emotion challenge. In: Proceedings of the International Conference on Multimodal Interaction. Monica: ACM, 2012. 449–456
136 Li Y, Su H, Shen X, et al. DailyDialog: A manually labelled multi-turn dialogue dataset. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Taipei: Asian Federation of Natural Language Processing, 2017. 986–995
137 Hsu C, Chen S, Kuo C, et al. Emotionlines: An emotion corpus of multi-party conversations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. Miyazaki: European Language Resources Association (ELRA), 2018
138 Poria S, Hazarika D, Majumder N, et al. MELD: A multimodal multi-party dataset for emotion recognition in conversations. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 527–536
139 Zhou H, Huang M, Zhang T, et al. Emotional chatting machine: Emotional conversation generation with internal and external memory. 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: AAAI Press, 2018. 730–739
140 Song Z, Zheng X, Liu L, et al. Generating responses with a specific emotion in dialog. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 3685–3695
141 Colombo P, Witon W, Modi A, et al. Affect-driven dialog generation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019. 3734–3743
142 Huang C, Zaïane O, Trabelsi A, et al. Automatic dialogue generation with expressed emotions. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short

Papers). New Orleans: Association for Computational Linguistics, 2018. 49–54

143 Zhong P, Wang D, Miao C. An affect-rich neural conversational model with biased attention and weighted cross-entropy loss. In: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, the Thirty-First Innovative Applications of Artificial Intelligence Conference, the Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019

144 Zhou X, Wang W Y. MojiTalk: Generating emotional responses at scale. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 1128–1137

145 Lubis N, Sakti S, Yoshino K, et al. Eliciting positive emotion through affect-sensitive dialogue response generation: A neural network approach. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, the 30th innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018. 5293–5300

146 Asghar N, Poupart P, Hoey J, et al. Affective neural response generation. In: Advances in Information Retrieval—40th European Conference on IR Research. Grenoble: Springer Lecture Notes in Computer Science, 2018. 154–166

147 Shang L, Lu Z, Li H. Neural responding machine for short-text conversation. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing: Association for Computational Linguistics, 2015. 1577–1586

148 Danescu-Niculescu-Mizil C, Lee L. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In: Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics. Portland: Association for Computational Linguistics, 2011. 76–87

149 Tiedemann J. News from opus-a collection of multilingual parallel corpora with tools and interfaces. In: Proceedings of the International Conference Recent advances in natural language processing. Borovets: Association for Computational Linguistics, 2009. 237–248

150 Wu H, Gu Y, Sun S, et al. Aspect-based opinion summarization with convolutional neural networks. In: Porceedings of the International Joint Conference on Neural Networks. Vancouver: IEEE, 2016. 3157–3163

151 Wang L, Ling W. Neural network-based abstract generation for opinions and arguments. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: Association for Computational Linguistics, 2016. 47–57

152 Angelidis S, Lapata M. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 3675–3686

153 Huy Tien N, Tung Thanh L, Minh Le N. Opinions summarization: Aspect similarity recognition relaxes the constraint of predefined aspects. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). Varna: INCOMA Ltd., 2019. 487–496

154 Zhao C, Chaturvedi S. Weakly-supervised opinion summarization by leveraging external information. In: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020

155 Shen T, Lei T, Barzilay R, et al. Style transfer from non-parallel text by cross-alignment. In: Advances in Neural Information Processing Systems 30. Long Beach: Curran Associates, Inc., 2017. 6830–6841

156 Li J, Jia R, He H, et al. 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). New Orleans: Association for Computational Linguistics, 2018. 1865–1874

157 Yang Z, Hu Z, Dyer C, et al. Unsupervised text style transfer using language models as discriminators. In: Advances in Neural Information Processing Systems 31. Montreal: Curran Associates, Inc., 2018. 7287–7298

158 Fu Z, Tan X, Peng N, et al. Style transfer in text: Exploration and evaluation. In: McIlraith S A, Weinberger K Q, eds. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018. 663–670

159 Prabhumoye S, Tsvetkov Y, Salakhutdinov R, et al. Style transfer through back-translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 866–876

160 Xu J, Sun X, Zeng Q, et al. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 2018. 979–988