Digital Communications and Networks 8 (2022) 745–762

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

A survey on deep learning for textual emotion analysis in social networks

Sancheng Peng [a], Lihong Cao [b,*], Yongmei Zhou [c], Zhouhao Ouyang [d], Aimin Yang [e], Xinguang Li [a], Weijia Jia [f], Shui Yu [g]

[a] Laboratory of Language Engineering and Computing, Guangdong University of Foreign Studies, Guangzhou, 510006, China
[b] School of English Education, Guangdong University of Foreign Studies, Guangzhou, 510006, China
[c] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, 510006, China
[d] School of Computing, University of Leeds, Wood-house Lane, Leeds, West Yorkshire, LS2 9JT, United Kingdom
[e] School of Computer Science and Intelligence Education, Lingnan Normal University, Zhanjiang 524048, China
[f] BNU-UIC Institute of Artificial Intelligence and Future Networks, Beijing Normal University (BNU Zhuhai), Zhuhai, 519087, China
[g] School of Computer Science, University of Technology Sydney, Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Text
Emotion analysis
Deep learning
Sentiment analysis
Pre-training

ABSTRACT

Textual Emotion Analysis (TEA) aims to extract and analyze user emotional states in texts. Various Deep Learning (DL) methods have developed rapidly, and they have proven to be successful in many fields such as audio, image, and natural language processing. This trend has drawn increasing researchers away from traditional machine learning to DL for their scientific research. In this paper, we provide an overview of TEA based on DL methods. After introducing a background for emotion analysis that includes defining emotion, emotion classification methods, and application domains of emotion analysis, we summarize DL technology, and the word/sentence representation learning method. We then categorize existing TEA methods based on text structures and linguistic types: text-oriented monolingual methods, text conversations-oriented monolingual methods, text-oriented cross-linguistic methods, and emoji-oriented cross-linguistic methods. We close by discussing emotion analysis challenges and future research trends. We hope that our survey will assist readers in understanding the relationship between TEA and DL methods while also improving TEA development.

1. Introduction

Textual Emotion Analysis (TEA) is the task of extracting and analyzing user emotional states in texts. TEA not only acts as a standalone tool for information extraction but also plays an important role for various Natural Language Processing (NLP) applications, including e-commerce [1], public opinion analysis [2], big search [3], information prediction [4], personalized recommendation [5], healthcare [6], and online teaching [7].

The idiom's "seven emotions and six desires" are joy, love, anger, sadness, fear, evil, and desire. Among them, only a few are positive emotions and the rest are negative, indicating that people are naturally more sensitive to negative emotions. During real-world life, negative emotions are also easier to propagate than positive emotions. The 2011 Annual Report of China's Internet Public Opinion Index [8,9] stated, "In 2011, there are more than 80% negative events for the total number of topics. The negative events on the Microblog and Tianya forum accounted for 75.6% and 95.8%, respectively, which are higher than those of other social media."

With the rapid development of social networks [10–13], people have changed from general users to network information producers. According to the 38th statistical report on the development of China's Internet published by China Internet Information Center [14], the number of Internet users in China has reached 710 million, and the Internet penetration rate reached 51.7% in June 2016. Among them, 656 million were mobile Internet users, 242 million used Microblog, and more than 100 million had daily blogs. Negative emotions are the most prevalent among this massive number of short text messaging.

Emotion analysis [15] aims to automatically extract user emotional states from their social network text activity (e.g., blogs, tweets). Early research focused on either a positive/negative bipartition or a positive/negative/neutral tripartition of emotion analysis [16,17]. However, such partitioning ignores subtle user emotion changes and their psychological states, preventing a full expression of people's complex

* Corresponding author.
E-mail addresses: psc346@aliyun.com (S. Peng), 201610130@oamail.gdufs.edu.cn (L. Cao), yongmeizhou@163.com (Y. Zhou), tal-darim@foxmail.com (Z. Ouyang), amyang18@qq.com (A. Yang), lxg@gdufs.edu.cn (X. Li), jiawj@uic.edu.cn (W. Jia), shui.yu@uts.edu.au (S. Yu).

https://doi.org/10.1016/j.dcan.2021.10.003
Received 27 December 2020; Received in revised form 22 September 2021; Accepted 8 October 2021
Available online 14 October 2021

inner emotional world. This gave rise to bipartition-oriented emotion analysis being named "sentiment analysis" [18,19], while a more encompassing emotion analysis was dubbed "fine-grained sentiment analysis".

In 2012, Deep Learning (DL) methods [20,21] were introduced to NLP after they achieved successful object recognition via ImageNet [22]. DL methods improved statistical learning results in many fields. At present, a neural network-based NLP framework has achieved new levels of quality and become the dominating technology for NLP tasks, such as sentiment analysis, machine translation, and question answering systems.

Popular DL methods are used to model emotion analysis, including Deep Averaging Networks (DANs) [23], Denoising Autoencoders (DAEs) [24], Convolutional Neural Networks (CNNs) [25], Recurrent Neural Networks (RNNs) [26], Long Short-term Memory (LSTM) networks [27], Bi-directional Long Short-term Memory (Bi-LSTM) networks [28], Gated Recurrent Units (GRUs) [29], attention [30],and Multi-head Attention (MHA) [31,32]. In general, researchers may integrate multiple methods into a specific model (rather than just one) to improve emotion analysis performance.

To our knowledge, this is the first paper to provide a comprehensive survey on TEA and DL combinations. We collected papers from various sources such as ACL Web Anthology, AAAI, IEEE, ACM, Elsevier, and Springer. We confined our collection to papers published from January 2015 to November 2020 to gauge recent DL popularity. We used keywords (such as "deep learning", "emotion", "pre-training", "embeddings", "natural language processing", "short text", and "sentiment") to retrieve 210 relevant papers. Removing duplicate papers and those unrelated to TEA reduced the number to 130. We then conducted a manual review to obtain 70 of the most relevant papers from the remainder. The main contributions of our survey are summarized as follows.

1. We provide a comparative study covering recent DL-based TEA published between January 2015 and November 2020 by analyzing the basic characteristics of typical TEA methods.
2. We first attempt to provide a comprehensive review of the related TEA methods.
3. We provide a detailed overview of different definitions and classification models of emotion.
4. We provide a detailed overview of the related TEA applications.
5. We provide a detailed description and comparative analysis of related pre-training TEA methods.
6. We provide a detailed description and comparison analysis of existing DL-based TEA methods.

The remainder of this paper are organized as follows: In Section 2, we provide an overview of the term "emotion"; we then survey emotion analysis applications in Section 3. In Section 4, we provide an overview of DL methodology before surveying pre-training methods in Section 5. In Section 6, we discuss emotion analysis methods based on DL and then present emotion analysis challenges in Section 7. In Section 8, we discuss the future trends before concluding this paper in Section 9.

2. Emotion overview

Due to the variability and sensitivity of human emotions, people have different understandings of emotion, causing the term to be classified differently by different fields. At present, there is no unified standard for defining and classifying emotion academically; however, researchers have performed in-depth studies of emotion classification, presenting multiple definitions and classification models.

2.1. Defining emotion

In this subsection, we will explain the meaning of emotion. The basic understanding of emotion is summarized as follows.

Definition 1. Emotion [33] is defined by the Merriam-Webster Dictionary as "A conscious mental reaction (e.g., anger or fear) subjectively experienced as strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body".

Definition 2. Emotion [34] denotes people's attitude experience and corresponding behavioral responses to objective things.

Definition 3. Emotions [35] are "generated states in humans that reflect evaluative judgments of the environment, the self and other social agents".

Definition 4. Emotion [36] is the feeling or reaction that people have due to a certain event. "Happy", "sad", "angry", "fear" are a few examples of emotions that can be expressed. "Emotions" and "sentiments" are often considered interchangeable, yet the latter represent emotional polarity or general emotional states (i.e., positive, negative, or neutral).

Definition 5. Emotion [37] is a mental state that arises spontaneously rather than through a conscious effort; it is also often accompanied by physiological changes.

Definition 6. Emotion [38] is often defined as an individual's mental state associated with thoughts, feelings, and behavior.

2.2. Classifying emotion

In this subsection, we will explain the classification of emotion. The basic understanding of emotion types is surveyed as follows.

Ekman [39] categorized emotions into six basic types: anger, disgust, fear, happiness, sadness, and surprise.

Parrott [40] proposed an emotion classification model based on a tree structure with six kinds of emotions: love, joy, surprise, anger, sadness, and fear.

Plutchik [41] visualized a wheel-shaped emotion classifier (based on Ekman's model [39] with two additional categories) with four bipolar sets: joy and sadness; anger and fear; trust and disgust; and surprise and anticipation.

Lin Chuanding [42], a modern Chinese psychologist, divided emotions into 18 Shuowen-based categories: joy, quiet, caress, worry, fright, pity, fear, grief, shame, sorrow, anger, vexation, reverence, hatred, arrogance, greed, jealousy, and shame.

These existing emotional category approaches focus on modeling emotions based on distinct emotion classes or labels. These models assume discrete emotion categories exist.

3. Emotion analysis application

Emotion analysis has been widely studied in psychology, neuroscience, and behavioral science, as emotions are an important element of human nature. Such analysis plays an important role in many application fields, including e-commerce, public opinion analysis, big search, information prediction (e.g., financial prediction, presidential election prediction), personalized recommendation, healthcare (e.g., depression screening), and online teaching.

3.1. E-commerce

With the development of mobile Internet, online shopping is increasingly welcomed by users, and users usually publish personal comments on products purchased via Taobao, Jingdong, Amazon, and other e-commerce platforms. Using this source of product comments [1], real-time emotional analysis is conducted to obtain useful emotional and behavioral consumer features, enabling the prediction of trend changes in consumer preferences. Such information would help a majority of consumers deeply understand the quality of goods, pre-sale and after-sales services, logistic services, and other related information,

guiding them through their future purchases. Manufacturers would also benefit from first-hand consumer feedback, timely product shortage warnings, and improved product quality and design. Sellers would benefit from knowing consumer psychological states as they relate to available commodities and related services. Sellers who can capture consumer psychology can make timely sales, purchases, and marketing decisions, allowing them to reach market dominance.

3.2. Public opinion analysis

Network public opinion [2] refers to different views of popular social issues expressed on the Internet. This public form of social opinion carries appreciable influence and allows tendentious opinions of issues affecting the reality through the Internet. Public opinion analysis is used to objectively reflect the state of public opinion by collecting and sorting out people's attitudes as well as discovering relevant opinion tendencies. Many irrational emotions (such as negative feelings towards the rich, official, powerful, or market) are expressed and strengthened by means of Internet violence or entertainment, driving people towards more extreme emotional reaction. Irrational netizen emotion causes national and societal security risks. Thus, relevant national management departments require knowledge of network public opinion trends to guide that opinion properly and in a timely fashion. However, when such information is obtained through various channels, its complexity prevents manual processing. This shortfall makes developing accurate and effective emotion analysis systems significant as well as the automatic processing of network public opinion information necessary to maintain national security and social stability.

3.3. Big search

With network space expansion, network application mode development, and the arrival of the big data era, the Internet has become ubiquitous and given rise to big search technology [3]. The big search is becoming a strong tool and catalyst for network development. Big search, the next generation search engine for cyberspace, is becoming an urgent need. Compared with traditional search, the big search can understand user search intentions on a semantic level while also perceiving user needs according to their spatio-temporal location, emotional state, and historical preferences. Big search can also remove false data and protect user privacy. In addition, big search solutions can provide intelligent answers to users, making retrieval technology based on user emotion analysis an important research task for big search.

3.4. Information prediction

As the Internet has developed, many people rely on it for information and communication sharing, particularly for social network interactions (e.g., Microblog, Wechat, stock, and futures forums). Emotion analysis technology can be used to analyze the impact of social networks on user lives and predict developing trends through commentary, news articles, and other content. The main application of information prediction includes the following three aspects.

3.4.1. Financial prediction

Many financial investors are turning to their networks for financial information and investment opinion exchanges. This makes professional forums on subjects such as stocks and futures rich with financial data and investor sentiment information (i.e., important factors affecting investor behavior and psychology). Behavioral finance dictates that the psychology and behavior of irrational stock investors affect stock market situations, causing stock prices to deviate from their correct value. Thus, we can predict stock market volatility by reviewing investor emotion and behavior information drawn from a stock forum.

3.4.2. Election prediction

Emotion analysis plays an increasingly important role in the prediction of democratic elections. Paul et al. [4] presented a Compass framework that used the 2016 U S. presidential election as an example to analyze election-related crowd emotion. They built a spatial-temporal sentiment map through Compass for the election and used that map to match election results to an extent. Their study showed that any political event could be described by its popularity in negative and positive senses. In addition, Ceron et al. [43] used emotion analysis to calculate Twitter support rates for political leadership candidates in the 2011 Italian parliamentary election and the 2012 French presidential election.

3.4.3. Other prediction

Emotion analysis can also be used to predict public opinion regarding various policy events (such as personal income tax adjustment, medical insurance reform, and retirement delays) as well as provide support for national policy formulation. In addition, emotion analysis can be applied to natural disaster prediction and judgment, including epidemics [44] and earthquakes. With the application of information prediction, emotion analysis technology has received greater attention. With emotion analysis technology to analyze Internet news, blogs, and other information sources, developing event trends can be predicted accurately.

3.5. Personalized recommendation

The emergence of personalized recommendation systems [45–47] has provided users with a tool to address information overload issues. However, traditional recommendation technology only considers overall user scores while ignoring emotional information contained in user comments. Such commentary usually contains subjective user views, preferences, and emotions regarding certain attributes of things, reflecting user emotional tendencies for those attributes. Mining and exploiting user comments to the greatest extent can produce more accurate personalized recommendations, and help solve the problems of cold starts, data sparsity, and low recommendation accuracy.

3.6. Healthcare

According to text recorded from social networks on psychological counseling, the emotional state information of patients suspected to be depressed was analyzed and screened using emotion analysis technology [6]. Yang et al. [48] used the disease analysis interview corpus from the University of Southern California to analyze whether patients with poor mental health (e.g., those diagnosed with anxiety or PTSD) also suffered from depression. To lighten the psychological burden of interviewees, interviews were conducted by human-controlled robots. Collected data also included recorded texts and PHQ-8 questionnaires to determine possible depression conditions.

3.7. Online teaching

With the popular application of Massive Open Online Courses (MOOCs) [49,50], a large number of online courses and reviews have been generated. Most reviews allow students to express their emotions and opinions. Tucker et al. [7] found a positive correlation between student emotional tendencies from their forum-based reviews and their learning performance on the MOOC platform. Thus, using emotion analysis technology to analyze comment information on the MOOC platform allowed the authors to obtain course-related emotional information. Such information can help teachers find problems in curriculum arrangement, knowledge systems and teaching methods, enabling timely teaching plan and method optimization to further improve teaching quality and student learning efficiency.

4. DL methodology

4.1. DL overview

The core task of a DL method is feature learning. In essence, it is a method of learning complex feature representation based on original feature input through multi-layer nonlinear processing. If combined with specific domain tasks, DL can construct new classifiers or generating tools through the feature representation of automatic learning and realize domain-oriented classification or other tasks. The specific steps of the algorithm for a DL model are listed as follows [51,52]:

Step 1: Construct a learning network with the random initialization, set the total number of network training layers n, initialize unlabeled data as the input set of network training, and initialize training network layer $i = 1$.
Step 2: Based on the input set, an unsupervised learning algorithm is used to pre-train the learning network of the current layer.
Step 3: The training results of each layer are used as input for the next layer, constructing the input set once again.
Step 4: If i is less than n, then $i = i + 1$, and return to Step 2; otherwise, proceed to Step 5.
Step 5: The supervised learning method is used to adjust network parameters of all layers, forcing any errors to meet practical requirements.
Step 6: Complete classifier construction (such as neural network classifiers) or complete deep generation model construction (such as a Deep Neural Network (DNN)).

4.2. DL-related methods

In this subsection, we will provide an overview of related DL methods. The basic understanding of related methods is summarized as follows.

4.2.1. DAN

DANs are constructed by stacking nonlinear layers over traditional neural bag-of-words models. For each document, a DAN takes the arithmetic mean of the word vectors as input and passes it through one or more feed-forward layers until there is a softmax for classification. The framework is shown in Fig. 1.

According to text classification, a DAN needs to map an input sequence of n tokens to one of k labels, and it also requires the following three steps to function:

Step 1: Take the vector average of the embedding associated with n input sequences of tokens n:

$$av = \sum_{i=1}^{n} c_i n \tag{1}$$

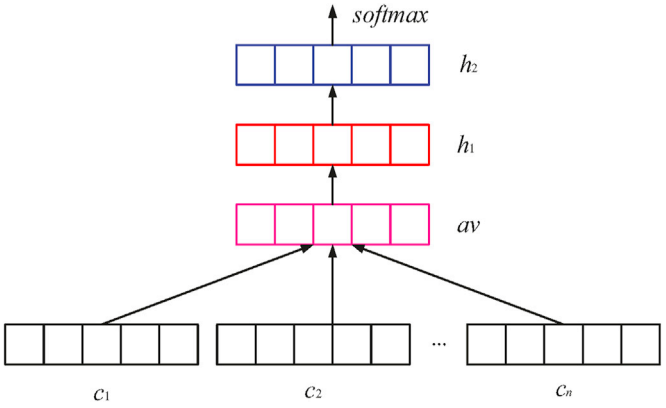

Fig. 1. A DAN with two feed-forward layers.

where c_i denotes the word embedding sequence.

Step 2: Pass the average through one or more feed-forward layers. If there is one layer, the average word embedding is described as follows:

$$h_1 = f(w_1 \cdot av + b_1) \tag{2}$$

where b denotes the offset term and w denotes the $k \times d$ weight matrix. If there are more layers, the average word embedding is described as follows:

$$h_i = f(w_i \cdot h_{i-1} + b_i) \tag{3}$$

Step 3: Conduct (linear) classification on the representation of the final layer:

$$\mathrm{softmax}(q) = \frac{\exp q}{\sum_{j=1}^{k} \exp q_j} \tag{4}$$

4.2.2. DAE

A DAE is an unsupervised learning algorithm that acts as an autoencoder modification. It can form a DL network with multiple stacked layers. A denoising autoencoder (as shown in Fig. 2) consists of an encoder, a hidden layer, and a decoder.

Encoder $f(x)$ is used to reduce the dimensionality of high-dimensional input. Input x is added noise to obtain a destroy version x, which is input into $f(x)$. Implicit coding result y is then obtained through linear transformation and activation function. Decoder $g(y)$ is used to obtain a reconstructed vector z. The specific computing for y and z is described as follows:

$$y = f(x) = S_f(Wx + b_y) \tag{5}$$

$$z = g(y) = S_g(W^T y + b_z) \tag{6}$$

$$S_f = \mathrm{sigmoid}(x) = 1(1 + e^{-x}) \tag{7}$$

$$S_g = \mathrm{sigmoid}(y) = 1(1 + e^{-y}) \tag{8}$$

where S_f denotes an activation function of the >nonlinear transformation, S_g denotes an activation function of the decoder, b denotes the offset term, and W denotes the weight matrix.

4.2.3. CNN

A CNN is a kind of feed-forward neural network, mainly consisting of an input layer, a convolution layer, a pooling layer, a full connection layer, and an output layer. The basic framework is shown in Fig. 3.

In the field of NLP, the application steps of a CNN mainly include: taking a vectorized sentence matrix as the input, convoluting the original input with multiple convolution kernels through the convolution layer, and obtaining multiple feature representations of the original input. Extracted features are then sent through the pooling layer as input and sampled to obtain more abstract features. Finally, through the full connection layer, the corresponding classification function is used to classify results to complete the corresponding task.

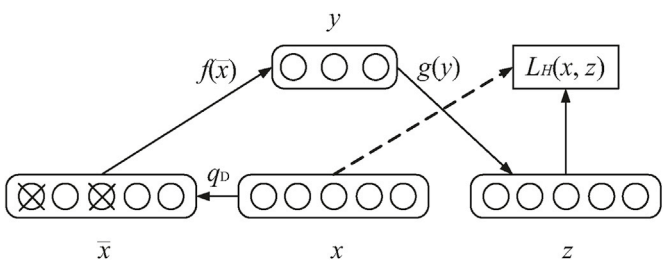

Fig. 2. The framework of DAE.

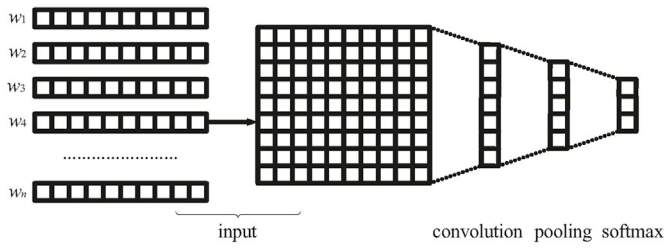

Fig. 3. The framework of CNN.

Input layer: It is responsible for the vectorization of input data. For a length n of a given sentence, the matrix of input layer can be expressed as:

$$E \in \mathbb{R}^{n \times m} \tag{9}$$

where m denotes the dimensionality of a word vector.

Convolution layer: It uses different convolution kernels to perform the convolution operation for the input matrix, extract local features from the input, and obtain feature maps of the convolution kernel. The specific representation is described as follows:

$$X_j^l = f\left(\sum_{i=1}^{M} X_i^{l-1} * K_{ij}^l + b_j^l\right) \tag{10}$$

where X_j^l denotes the output of convolution operation between the input and the j-th convolution kernel, X_i^{l-1} denotes the output of the previous layer (i.e., the i-th local receiving domain of the current convolution layer l), K_{ij}^l denotes the j-th convolution kernel of l, b_j^l denotes the offset term of the current convolution kernel, and M denotes all local received domains that the convolution kernel needs to traverse.

Pooling layer: It uses the corresponding sampling function to sample characteristics of the matrix generated by the convolution operation, and extracts more important features to reduce matrix dimensionality. In addition, it simplifies the calculation process while preventing key features from being discarded. Common pooling algorithms include max pooling and average pooling.

Full connection layer: It classifies input, obtains the classification results, and is responsible for taking those results to the output layer.

4.2.4. RNN

An RNN is a kind of feed-forward neural network with a ring structure and a specific memory function. Its input includes current input samples as well as information obtained in the previous time so that information can be cycled in the network at any time. The framework of an RNN is shown in Fig. 4, in which x denotes the input layer, O denotes the output layer, H denotes the hidden layer, and u, V, and w denote the weights of the above each respective layer. The output of the hidden layer is described by:

$$h_t = \sigma(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{11}$$

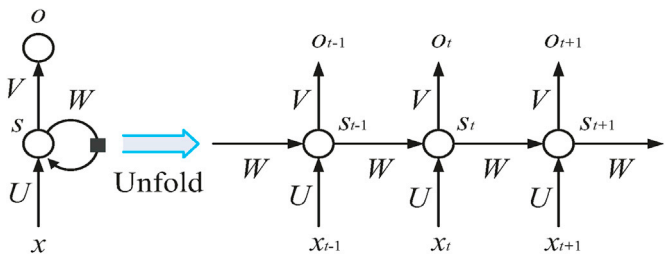

Fig. 4. The framework of RNN.

The output of the out layer is described by:

$$o_{t+1} = \sigma(W_{hy}h_t + b_y) \tag{12}$$

$$y_t = \text{softmax}(o_t) \tag{13}$$

where y_t denotes the output in time t, W_{hh}, W_{xh}, and W_{hy} denote the weight matrix, b_h and b_y denote the bias term, h_t denotes the hidden state in t, and x_t denotes the input data.

4.2.5. LSTM

An LSTM network is a kind of RNN that can learn the relationship of long-term dependency, characterize the information of a time sequence, and effectively solve the problem of gradient vanishing or gradient exploding faced during RNN training. It was first proposed by Hochreiter and Schmidhuber in 1997 [27]. After that, many researchers have optimized and improved upon it, causing rapid development and leading to its wide use among various NLP aspects.

Each unit of an LSTM network consists of four components: a memory cell, input gate, output gate, and forget gate. Memory cells are connected circularly with each other. Three nonlinear gate cells can be used to adjust the information of memory cell input and output flows. The framework of an LSTM network is shown in Fig. 5.

The forward computing process of an LSTM network is described as follows.

Input gate:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{14}$$

Forget gate:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{15}$$

Memory cell:

$$c_t = f_t \bullet c_{t-1} + i_t \bullet \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{16}$$

Output gate:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{17}$$

Result output:

$$h_t = o_t \bullet \tanh(c_t) \tag{18}$$

where x_t denotes the input vector (such as a word vector) at time t; f, i, and o denote the activation vectors of the forget gate, input gate, and output gate, respectively; c denotes the memory unit vector, h denotes the output vector the LSTM unit, W_i, U_i, W_f, U_f, W_c, U_c, W_o, and U_o denote the weight matrix; b_i, b_f, b_c, and b_o denote the bias vector; σ and tanh denote the activation function.

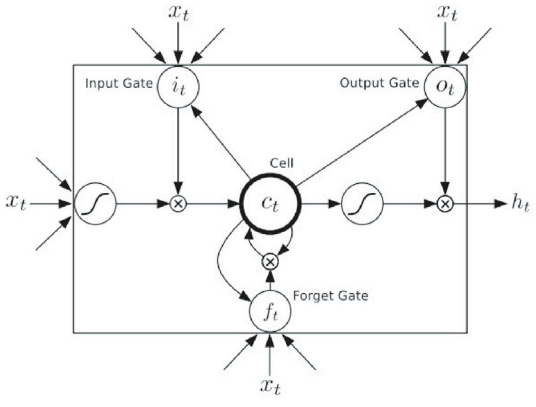

Fig. 5. The framework of LSTM.

4.2.6. Bi-LSTM

A Bi-LSTM is an improved LSTM model. One-directional LSTM uses previous information to deduce subsequent information, requiring information processing and preventing it from accessing future context or integrating context information, which affects system prediction performance. A Bi-LSTM uses two LSTM networks to train together and start their respective sequences from opposite ends while being connected back to the same output layer. Thus, it can integrate the past and future information of each point. A Bi-LSTM includes forward and backward calculations. The horizontal direction represents the bi-directional flow of the time sequence, and the vertical direction represents the one-directional flow from the input layer to the hidden layer and on to the output layer. The network structure of a Bi-LSTM network is shown in Fig. 6.

The forward calculation of hidden vector h is described as follows:

$$h_t = \mathrm{LSTM}(x_t, h_{t-1}) \tag{19}$$

The backward calculation of hidden vector h is described as follows:

$$h_t = \mathrm{LSTM}(x_t, h_{t-1}) \tag{20}$$

The output is described as follows:

$$y_t = g(W_{hy}h_t + W_{hy}h_t + b_y) \tag{21}$$

where x_t denotes the input data, y_t denotes the output at time t, W_{hy} and W_{hy} denote the weight matrix, and b_y denotes the bias term.

4.2.7. GRU

A GRU is an LSTM variant. It is well known that LSTM can overcome the problem of gradient vanishing or gradient exploding when dealing with the relationship of long-distance dependence, and it can keep the dependency relationship of long-distance and short-distance for temporal data. A GRU retains these advantages while offering a simpler network structure. Compared with an LSTM network's three-gate structure, a GRU has only two gates: an update gate and a reset gate. The network structure of a GRU is shown in Fig. 7.

The forward calculation of a GRU is described as follows:

Update gate:

$$u_t = \sigma(W_u x_t + V_u h_{t-1} + b_u) \tag{22}$$

Reset gate:

$$r_t = \sigma(W_r x_t + V_r h_{t-1} + b_r) \tag{23}$$

Memory cell:

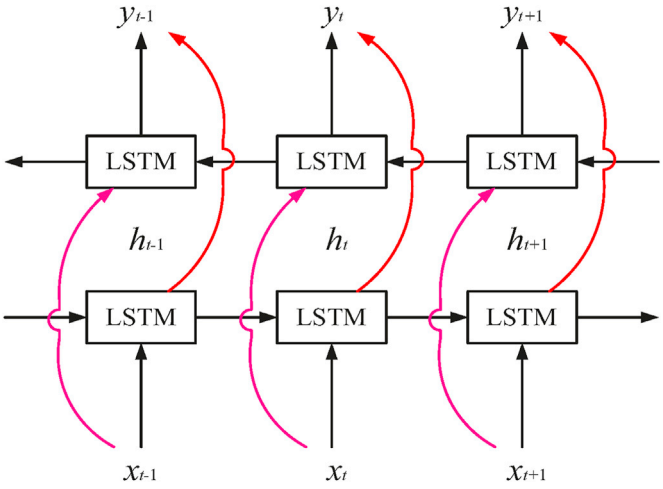

Fig. 6. The framework of Bi-LSTM.

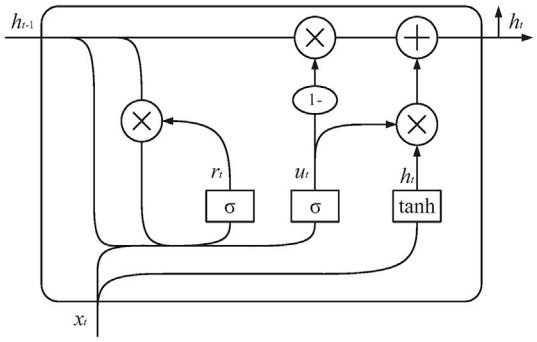

Fig. 7. The framework of GRU.

$$\tilde{h}_t = \tanh(W_h x_t + V_h(r_t * h_{t-1}) + b_r) \tag{24}$$

Output:

$$h_t = u_t h_{t-1} + (1 - u_t)\tilde{h}_t \tag{25}$$

where W_u, U_r, W_h, V_u, V_r, and V_h denote the weight matrix, b_i, b_f, b_c, and b_o denote the bias vector, and tanh denotes the activation function.

4.2.8. Attention

An attention mechanism imitates human visual processing (i.e., it aligns internal experience with external feelings to increase the observation precision of certain areas). For example, when browsing a picture, people first scan the global image quickly to obtain a target area (i.e., attention point) that requires focus. More attention is then devoted to that point to obtain more detailed information while other useless information is suppressed. The specific framework is shown in Fig. 8.

The calculation process is mainly divided into the following three steps.

Step 1: The similarity between the query and each key is calculated to obtain the weight. Common similarity functions include the dot product, concatenating, and perceptron. The related descriptions are described as follows:

$$f(Q,K) = \begin{cases} Q^T K, \text{dot} \\ Q^T W_a K, \text{general} \\ W_a[Q;K], \text{concat} \\ v_a^t \tanh(W_a Q + U_a K), \text{perceptron} \end{cases} \tag{26}$$

Step 2: Generally, a softmax function (see Equation (27)) is used to normalize these weights,

$$a_i = \mathrm{softmax}(f(Q,K)) = \frac{\exp(f(Q,K_i))}{\sum_{j=1}^{n}\exp f(Q,K_j)} \tag{27}$$

Step 3: The final attention is obtained by weighting and summing the

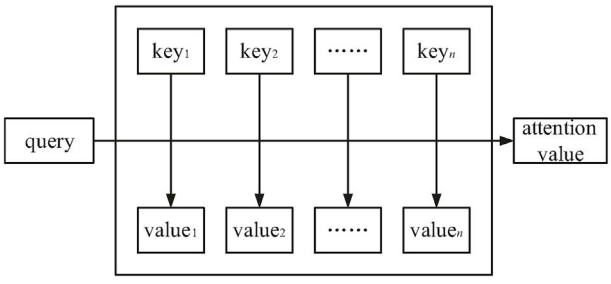

Fig. 8. The framework of attention.

weight and corresponding key value, which is described as follows:

$$\text{attention}(Q,K,V)=\sum_i a_i V_i \tag{28}$$

At present, NLP research tends to use identical keys and values.

4.2.9. MHA

MHA is an attention mechanism variant that uses multiple queries to extract multiple groups of different information in parallel from input information for concatenating. The multiple attention mechanism is shown in Fig. 9.

First, a linear transformation is made for the query, key, and value. Then, they are input into the Scaled Dot product Attention (SDA) mechanism, and the same operation is repeated h times. The input of each time is the linear transformation of the original input. The SDA framework is shown in Fig. 10.

An SDA is an attention mechanism of similarity calculation using a point product, a series of queries, a series of keys with dimensions d_k, and a series of values with dimensions d_v. The calculation process is described as follows:

$$\text{attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{29}$$

"Multi-head" denotes that a head is calculated each time, parameter W of the linear transformation for Q, K, and V is different each time, and the results of the h times SDA mechanism are concatenated. They are then conducted in a linear transformation once more to obtain a value, which is the MHA result. The calculation is described as follows:

$$\text{multihead}(Q,K,V)=\text{concat}(head_1,head_2,\ldots,head_h)W^o \tag{30}$$

$$head_i=\text{attention}(QW_i^Q,KW_i^K,VW_i^V) \tag{31}$$

where $W_i^Q\in R^{d_k d}$, $W_i^K\in R^{d_k d}$, $W_i^V\in R^{d_v d}$, and $W^o\in R^{hd_v d}$ denote a single attention function with d-dimensional keys, values, and queries.

5. Pre-training method overview

The emergence and development of pre-trained methods have brought NLP into a new era. The related methods have been divided into word-oriented representation learning and sentence-oriented representation learning.

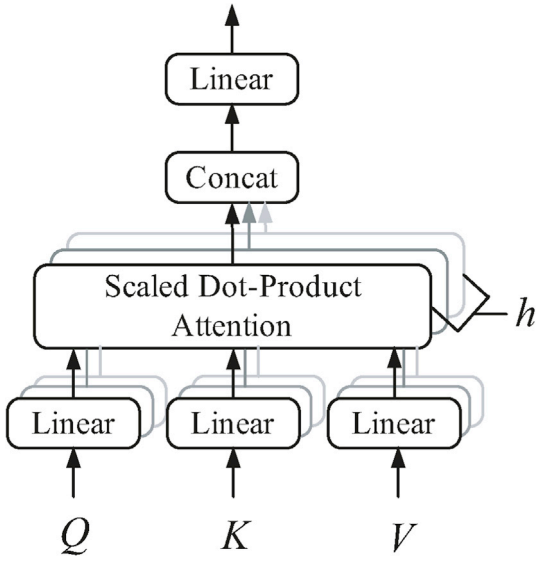

Fig. 9. The framework of MHA.

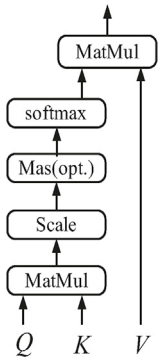

Fig. 10. The framework of SDA.

5.1. Word-oriented representation learning

In this subsection, we will provide an overview of word-oriented representation learning. The basic understanding for related approaches are summarized as follows.

5.1.1. Word2vec

The word2vec technique [53] consists of two models: continuous bag-of-words (CBOW) and continuous skip-gram. The CBOW model uses the average/sum of context words as input to predict a current word. The skip-gram model uses a current word as input to predict each contextual word. Word2vec has fewer dimensions than the previous embedding methods, making it faster, more versatile, and able to be used by various NLP tasks. Although it has strong generality, it cannot be dynamically optimized for specific tasks or solve the problem of polysemy.

5.1.2. Bandwidth expansions (BWEs)

BWEs [54] is an unsupervised neural model for learning bilingual semantic embedding of words across Chinese and English. It is a bilingual word embedding method that applies initialization and optimization constraints while using machine translation alignments. However, it also cannot solve the problem of polysemy.

5.1.3. GloVe

GloVe [55] is a word representation tool based on count base and global corpus statistics. It calculates global co-occurrence statistics first using a fixed-size context window and then minimizes its least-squares objective function using the stochastic gradient descent, which essentially factorizes the log co-occurrence matrix. It can support parallelization and has appreciable speed, but also requires more memory resources than word2vec.

5.1.4. BiDRL

A Bilingual Document Representation Learning (BiDRL) method [56] is used for cross-lingual sentiment classification. It can learn vector representations for both words and documents in bilingual texts.

5.1.5. Emoji2Vec

Emoji2Vec [57] is a method for learning emoji representations that contains a complete set of Unicode emoji representations. It can be a pre-trained embedding for all Unicode emoji methods and can be used to pre-train a large text corpus. It can provide emoji embedding, but does not capture the context-dependent definitions of emoji (e.g., sarcasm, appropriation via other cultural phenomena).

5.1.6. Sentiment-specific word embedding (SSWE)

SSWE [58] can encode both positive/negative sentiment and syntactic contextual information in a vector space. It has the effectiveness of incorporating sentiment labels in word-level information for sentiment-related tasks compared with other word embedding methods.

However, it only focuses on binary labels, weakening its generalization ability on other affect tasks.

5.1.7. FastText

The fastText library [59] can handle Out-of-Vocabulary (OOV) words by predicting their word vectors based on learned character n-grams embedding. While it requires little training time, without sharing parameters, it has poor generalization for large output spaces.

5.1.8. Context2vec

The context2vec model [60] is a learning contextual representation method for predicting a single word from both left and right contexts, based on Bi-LSTM. It can learn generic context embedding of wide sentential contexts and can encode the context around a pivot word.

5.1.9. REF

REF [61] is a word vector refinement model that refines existing semantically oriented word vectors used sentiment lexicons. It can be applied to existing pre-trained word vectors (e.g., word2vec and GloVe). Both semantic and sentiment word vectors can be obtained with this model.

5.1.10. ELMo

ELMo [62] is a deep contextualized word representation method based on Bi-LSTM that can solve the problem of polysemy. It can pre-train a large text corpus and model polysemy, but it also requires long training times and cannot solve long-distance dependency.

5.1.11. KUBWE

KUBWE [63] is a word embedding algorithm that builds a symmetric co-occurrence matrix from the corpus, then calculates an adjusted form of the pointwise mutual information matrix to remove insignificant and uninformative co-occurrences. It uses a spherical representation for the latent space in which points are located on the surface of a hypersphere.

5.1.12. Maximum a posteriori (MAP)

MAP [64] is a probabilistic word embedding model based on MAP estimation. It is a generalized word embedding model that considers a wide range of parametrized GloVe variants and incorporates priors on those parameters. In the model, word vectors are learned by finding the parameters that maximize a posterior probability.

5.1.13. CoVe

CoVe [65] is a contextualized representations method that uses LSTM. It can train context-based word vectors for machine translation. Due to the high training complexity and high decoding delay of LSTM, this model's training time is excessive.

5.1.14. Emo2Vec

Emo2Vec [66] is a multi-task learning method that encodes emotional semantics into vectors by using a CNN. It is trained by six different emotion-related tasks and can encode emotional semantics into real-valued, fixed-sized word vectors.

5.1.15. ULMFit

ULMFit [67] is a transfer learning method that can be applied to any NLP task. It consists of two pre-trained models: a forward model trained from left to right and a backward model trained from right to left.

5.1.16. NTUA SLP

NTUA-SLP [68] is a word embedding method based on word2vec. It consists of a two-layer Bi-LSTM network with a deep self-attention mechanism. It can overcome the problem of OOV words.

5.1.17. SVD NS

SVD NS [69] is a word embedding method in the context of NLP. It not only learns word-context co-occurrences, but also learns the abundance of unobserved or insignificant co-occurrences, improving word distributions in latent embedded space.

5.1.18. cw2vec

The cw2vec [70] method is used for learning Chinese word embedding. It can use stroke n-grams to capture semantic and morphological level information of Chinese words. However, it only learns Chinese word embedding.

5.1.19. MNLM

This unsupervised Multilingual Neural Language Model (MNLM) [71] is used for word embedding. It can jointly learn word embedding of different languages in the same space and generate multilingual embedding without any parallel data or pre-training. However, it cannot exploit character and subword information.

The specific comparison of existing word representation learning methods is listed in Table 1.

5.2. Sentence-oriented representation learning

In this subsection, we will provide an overview of sentence-oriented representation learning. The basic understanding of related approaches are summarized as follows.

5.2.1. Paragraph vector

This unsupervised algorithm learns fixed-length semantic representations from variable text lengths [72]. It is a strong alternative sentence embedding model and has been widely applied to learning representations for sequential data.

5.2.2. Skip-Thoughts

Skip-Thoughts [73] is a method used to train a sentence encoder by predicting preceding and following sentences using a current sentence. It follows the same idea as the skip-gram model of the word2vec embedding method. It can predict the probability of a sentence appearing in a given context through the current sentence, but its model training speed is slow.

5.2.3. DeepMoji

DeepMoji [74] is a model for detecting sentiment, emotion, and sarcasm by using an attention mechanism and a two-layer Bi-LSTM. However, the model still faces the problem of long-distance dependency.

5.2.4. BERT

BERT [75] is a language representation model for bidirectional encoder representations through transformers. It consists of two steps: pre-training and fine-tuning. It is designed to pre-train deep bidirectional representations from the unlabeled text by jointly conditioning both left and right contexts in all layers. It can obtain context sensitive bidirectional feature representation. However, there is an inconsistency between its pre-training process and generation process, which leads to poor effect on the generation task. It also consumes more computing resources than other existing models.

5.2.5. InferSent

InferSent [76] is a universal representation method for learning sentence embedding, based on a Bi-LSTM architecture with max pooling. It is the first attempt to use the Stanford natural language inference to build sentence encoders. However, the problem of long-distance dependency exists for this model.

5.2.6. CCTSenEmb

CCTSenEmb [77] is an unsupervised method for discovering hidden associations between sentences and integrating discriminative topics into the learning process. It can leverage latent associations between

Table 1
Comparison of existing word representation learning methods.

No.	Name	Method	Tasktype	Language	Year	Affiliation
1	word2vec [53]	feed-forward neural network, logistic regression	unsupervised	multi-language	2013	Google
2	BWEs [54]	neural network	unsupervised	Chinese and English	2013	Stanford University
3	Glove [55]	weighted least squares regression	unsupervised	multi-language	2014	Stanford University
4	BiDRL [56]	logistics regression	semi-supervised	cross-language	2016	Peking University
5	Emoji2Vec [57]	logistic regression	supervised	English	2016	Princeton University and University College London
6	SSWE [58]	feed-forward neural network	supervised	English	2016	Harbin Institute of Technology et al.
7	fastText [59]	probability statistics	supervised	multi-language	2016	Facebook
8	context2vec [60]	Bi-LSTM	unsupervised	multi-language	2016	Bar-Ilan University
9	REF [61]	nearest neighbor ranking	NA	English	2017	Yuan Ze University et al.
10	ELMo [62]	Bi-LSTM, CNN	semi-supervised	multi-language	2018	Allen Institute for Artificial Intelligence et al.
11	KUBWE [63]	kernel-based	unsupervised	English	2019	Dalhousie University
12	MAP [64]	weighted least squares regression	unsupervised	English	2019	University of Kent et al.
13	CoVe [65]	LSTM	supervised	English and German	2017	NA
14	Emo2Vec [66]	CNN	supervised	English	2018	Hong Kong University of Science and Technology
15	ULMFit [67]	LSTM	supervised, semi-supervised	multi-language	2018	University of San Francisco et al.
16	NTUA-SLP [68]	Bi-LSTM, attention	unsupervised	multi-language	2018	National Technical University of Athens et al.
17	SVD-NS [69]	singular value decomposition	unsupervised	English	2018	Dalhousie University
18	cw2vec [70]	stroke n-grams	unsupervised	Chinese	2018	Ant Financial Services Group et al.
19	MNLM [71]	LSTM	unsupervised	multi-language	2019	Nara Institute of Science and Technology et al.

sentences by directly predicting a sentence given the semantic information of a neighboring sentence.

5.2.7. CAMSE

CAMSE [78] is a multi-scale sentence embedding method for encoding sentences into an embedding tensor, based on contextual self-attention and multi-scale techniques. It is a supervised learning framework sentence embedding to answer medical questions. However, this model also suffers from long-distance dependency.

5.2.8. OPAI GPT

OPAI GPT [79] is a semi-supervised method for language understanding tasks that uses a combination of unsupervised pre-training and supervised fine-tuning. It is a two-stage training procedure that starts by using a language modeling objective on unlabeled data to learn the initial parameters of a neural network model. It then adapts these parameters to a target task using the corresponding supervised objective. It is a unidirectional auto-regressive language model and cannot obtain context-sensitive feature representation.

5.2.9. FastSent

FastSent [80] is a model for obtaining sentence embedding that can predict words in context sentences based on a current sentence. Its disadvantage is that it loses sentence sequencing information.

5.2.10. ERNIE

ERNIE [81] uses a multi-layer transformer as a basic encoder to capture contextual information. It is a method for learning language representation enhanced by knowledge masking strategies, including basic-level masking, entity-level masking, and phrase-level masking.

5.2.11. GenSen

GenSen [82] is a sentence representation method that combines the benefits of many sentence-representation learning models into a multi-task framework. It is a large-scale reusable sentence representation model obtained by combining a set of training objectives with the level of diversity studied here (e.g., Skip-Thoughts), natural language inference, machine translation, and constituency parsing.

5.2.12. Universal Sentence Encoder (USE)

A USE [83] provides sentence-level embedding in English. It can achieve the best performance by using sentence-level and word-level transfers.

5.2.13. Sent2Vec

Sent2Vec [84] is a simple unsupervised model for learning universal sentence embedding by using word vectors along with *n*-gram embedding. It can be used to train distributed representations of sentences.

5.2.14. DisSent

DisSent [85] uses a Bi-LSTM sentence encoder to yield high-quality sentence embedding, using global max pooling to construct the encoding for each sentence. It can serve as a supervised fine-tuning dataset for large models (e.g., BERT).

The specific comparison of existing sentence representation learning methods is listed in Table 2.

6. DL methods for TEA

TEA can characterize the emotional attitudes of people from a multi-dimensional view. Existing TEA methods based on DL are divided into four categories according to their text structures and linguistic types: text-oriented monolingual methods, text conversation-oriented monolingual methods, text-oriented cross-linguistic methods, and emoji-oriented cross-linguistic methods.

6.1. Text-oriented monolingual emotion analysis models

DL methods have been proven effective for many NLP tasks, including sentiment and emotion analysis. The following are emotion analysis models for a single language based on DL methods.

Abdul-Mageed and Ungar [86] proposed a fine-grained emotion detection method using Gated Recurrent Neural Networks (GRNNs). Tafreshi and Diab [87] proposed a joint multi-task learning model using a GRNN and trained it with a multigenre emotion corpus to predict emotions for four types of genres. Kulshreshtha et al. [88] proposed a neural architecture, Linguistic-featured Emoji-based Partial Combination of

Table 2
Comparison of existing sentence representation learning methods.

No.	Name	Method	Tasktype	Language	Year	Affiliation
1	Paragraph Vector [72]	log-bilinear	unsupervised	English	2014	Google
2	Skip-Thoughts [73]	RNN, GRU	unsupervised	English	2015	University of Toronto et al.
3	DeepMoji [74]	Bi-LSTM, attention	supervised	English	2017	Massachusetts Institute of Technology et al.
4	BERT [75]	Transformer	unsupervised	multi-language	2018	Google
5	InferSent [76]	Bi-LSTM, max pooling	supervised	English	2017	Facebook et al.
6	CCTSenEmb [77]	Gaussian	unsupervised	English	2019	Beijing Institute of Technology
7	CAMSE [78]	Self-attention, Bi-LSTM	supervised	English	2019	Tsinghua University
8	OPAI GPT [79]	Transformer	semi-supervised	multi-language	2018	OpenAI
9	FastSent [80]	log-bilinear	unsupervised	multi-language	2016	University of Cambridge et al.
10	ERNIE [81]	Transformer	unsupervised	multi-language	2019	Baidu
11	GenSen [82]	GRU	supervised	multi-language	2018	Microsoft Research Montreal
12	USE [83]	Transformer, DAN	unsupervised, supervised	English	2018	Google
13	Sent2Vec [84]	Optimization theory, n-grams	unsupervised	English	2018	Iprova SA, Switzerland
14	DisSent [85]	Bi-LSTM	supervised	English	2019	Stanford University

Deep Neural Networks (LE-PC-DNNs), for emotion intensity detection based on a CNN. LE-PC-DNNs can combine CNN layers with fully connected layers in a non-sequential or parallel fashion to improve system performance.

Mohammadi et al. [89] proposed a neural feature extraction method for contextual emotion detection. The model utilized an attention-based RNN and conducted experiments with Glove and ELMo embeddings, alongside Part-of-speech (POS) tags as input, LSTM and GRU as recurrent units, and a neural or a Support Vector Machine (SVM) classifier. Li et al. [90] presented a method for emotion classification of short text based on the skip-gram model and LSTM. Rathnayaka et al. [91] presented an approach for implicit emotion detection, called Sentylic, based on bidirectional GRUs and a capsule network.

Akhtar et al. [92] proposed a stacked ensemble method to predict emotion and sentiment intensity, designing three DL models based on CNN, LSTM, and GRU, respectively, and one classical supervised model based on Support Vector Regression (SVR). Batbaatar et al. [93] proposed a neural network architecture, called Semantic-emotion Neural Network (SENN), which can use both semantic/syntactic and emotion information by adopting pre-trained word representations. There are two sub-networks for SENN: the first uses Bi-LSTM to capture contextual information and focuses on semantic relationship, while the second uses a CNN to extract emotion features. Zhang et al. [94] proposed a multi-task CNN for TEA based on emotion distribution learning.

Khanpour and Caragea [95] proposed a method for emotion detection in online health communities, called ConvLexLSTM. It combined the output of a CNN with lexicon-based features, then fed everything into an LSTM network to produce the final output via the softmax mechanism. Yang et al. [96] proposed an interpretable neural network model for the relevant emotion ranking, using a multi-layer feed-forward neural network. Kratzwald et al. [97] proposed a text-based emotion recognition approach using an RNN, named sent2affect, which was a tailored form of transfer learning for affective computing. Yang et al. [98] proposed a framework called Interpretable Relevant Emotion Ranking with Event-driven Attention (IRER-EA), based on RNNs and the attention mechanism.

The specific comparison of existing text-oriented monolingual emotion analysis models is listed in Table 3.

6.2. Text conversation-oriented monolingual emotion analysis models

There are numerous emotions in textual conversations. As people use text messaging applications (such as Wechat, Facebook) and conversation agents (such as Amazon Alexa) to communicate more frequently than ever, context emotion detection in the text is becoming more important to emotion analysis. If we can effectively detect the emotion in a conversation, it has great commercial value (e.g., online customer service of an e-commerce platform).

Ghosal et al. [99] presented the Dialogue Graph Convolutional Network (DialogueGCN) for emotion recognition in conversation based on the Bi-GRU. DialogueGCN consists of three integral components, sequential context encoder, speaker-level context encoder, and emotion classifier. Zhong et al. [100] proposed a Knowledge-Enriched Transformer (KET) framework for emotion detection in textual conversations. They used the hierarchical self-attention to interpret contextual utterances and used a context-aware graph attention mechanism to leverage the external commonsense knowledge. Zhang et al. [101] proposed a Graph-based Convolutional neural Network towards Conversations, namely ConGCN, to model both context-level and speaker-level dependence for emotion detection.

Majumder et al. [102] presented a neural architecture, called DialogueRNN, which is based on the RNN to detect emotion in a conversation, where the textual feature of each utterance is extracted by the CNN. Ishiwatari et al. [103] proposed a relational position encodings method based on Relational Graph ATtention networks (RGAT) to recognize human emotions in textual conversation. Zhang et al. [104] proposed a Knowledge Aware Incremental Transformer with Multi-task Learning (KAITML) to conduct emotion classification. In KAITML, a dual-level graph attention mechanism was designed to leverage commonsense knowledge, which augments the semantic information of the utterance; an incremental transformer was used to encode multi-turn contextual utterances. In addition, multi-task learning was used to improve the performance of emotion recognition.

Jiao et al. [105] proposed a Hierarchical Gated Recurrent Unit (HiGRU) framework with two Bi-GRUs. The lower-level Bi-GRU was used to learn the individual utterance embedding and the upper-level Bi-GRU was used to learn the contextual utterance embedding. Li et al. [106] proposed a fully data-driven Interactive Double States Emotion Cell Model (IDS-ECM) for textual dialogue emotion prediction. In the model, the Bi-LSTM and attention mechanism were used to extract the emotion features. Li et al. [107] proposed a transformer-based context- and speaker-sensitive model for emotion detection in conversations, namely HiTrans, which consists of two hierarchical transformers. One was used to generate local utterance representations using BERT, and another was used to obtain the global context of the conversation.

Ghosal et al. [108] proposed a method for emotion detection in conversations, named COSMIC, which modeled various aspects of commonsense knowledge by considering mental states, events, actions, and cause-effect relations. Lu et al. [109] proposed an iterative emotion interaction network for emotion recognition in conversations. The network consists of the utterance encoder, the emotion interaction-based context encoder, and the iterative improvement mechanism. Li et al. [110] proposed a Hierarchical Transformer (HiTransformer) framework to address utterance-level emotion recognition in dialogue systems. It used a lower-level transformer to model word-level input, an upper-level transformer to capture the contexts of utterance-level embeddings, and BERT to obtain better individual utterance embeddings. Mundra et al. [111] proposed an Emotion Detection approach using Neural Networks

Table 3
Comparison of existing monolingual emotion analysis models.

No.	Name	DL method	Pre-training	Year	Dataset	Accuracy
1	Abdul-Mageed's [86]	GRNN	NA	2017	Twitter	0.8758
2	Tafreshi's [87]	GRU	fastText, word2vec, Glove	2018	TweetEN, BLG + HLN, MOV	TweetEN: 0.781, BLG + HLN: 0.836, MOV: 0.91
3	LE-PC-DNN [88]	CNN	DeepMoji, word2vec	2018	EmoInt-2017	0.791
4	Mohammadi's [89]	GRU, LSTM, SVM, attention	Glove, ELMo	2019	SemEval 2019 (EmoContext)	0.7303
5	Li's [90]	LSTM	word2vec	2017	WeChat	0.2512
6	Sentylic [91]	Bi-GRU, capsule networks	word2vec	2018	WASSA 2018	0.692
7	Akhtar's [92]	CNN, LSTM, GRU	GloVe and word2vec	2020	EmoInt-2017, SemEval-2017	0.748
8	SENN [93]	CNN, Bi-LSTM	word2vec, GloVe, and FastText	2019	Dailydialogs, CrowdFlower, TEC, Tales-Emotions, ISEAR, EmoInt, Electoral-Tweets, Grounded-Emotions, Emotion-Cause, SSEC	0.848, 0.511, 0.613, 0.746, 0.910, 0.563, 0.593, 0.988 and 0.708
9	Zhang's [94]	CNN	word2vec	2018	SemEval-2007	0.4141
10	ConvLexLSTM [95]	CNN, LSTM	word2vec	2018	Cancer Survivors' Network	Joy: 0.932, Sad: 0.923
11	Yang's [96]	multi-layer feed-forward neural network	NA	2018	Sina Social News, Ren-CECps corpus, SemEval 2007	News: 0.7108, Blogs: 0.6187, SemEval: 0.7081
12	sent2affect [97]	RNN	GloVe	2018	SemEval 2007, SemEval 2018	SemEval 2007: 0.584, SemEval 2018: 0.586
13	IRER-EA [98]	RNN, attention	Glove	2019	SemEval 2007, Ren-CECps corpus, Sina Social News	News: 0.7379, Blogs: 0.6304, SemEval: 0.7538

driven by Emotion Vectors (ED-NNEV) to predict the emotion category of each turn in a conversation, based on the CNN.

The specific comparison of existing text conversation-oriented monolingual emotion analysis models is listed in Table 4.

In addition, SemEval-2019 Task 3 [112] introduced a task to detect contextual emotion (e.g., happiness, sadness, anger) in conversational text. Its purpose was to invite research interest to the area of emotion detection in textual conversation.

Agrawal and Suri [113] proposed the Neural and Lexical Combiner (NELEC) model that combined lexical and neural features for emotion classification. Basile [114] designed different architectures (such as the three-input, two-output, Universal Sentence Encoder (USE), and Bidirectional Encoder Representations from Transformers (BERT) models) based on DL for emotion classification. Huang et al. [115] proposed an ensemble approach for emotion detection comprised of two DL models, the Hierarchical LSTMs for Contextual Emotion Detection model and the BERT model.

Winata et al. [116] used the hierarchical attention for dialogue emotion classification based on logistic regression and XGBoost. Bae et al. [117] proposed a method to detect emotion using a Bi-LSTM encoder for higher-level representation. Liang et al. [118] proposed hierarchical ensemble classification of contextual emotion using three sets of CNN-based neural network models trained for four-emotion classification, Angry-Happy-Sad classification, and Others-or-not classification respectively.

Xiao [119] designed a set of transfer learning methods using pre-trained language models (ULMFiT, OpenAI GPT, and BERT). He also trained a DL model from scratch using pre-trained word embedding and Bi-LSTM architecture with the attention mechanism. The conducted experimental result reveals that ULMFiT can perform best due to its fine-tuning technique. Li et al. [120] proposed a multi-step ensemble neural network for emotion analysis in the text. They used four DL models (LSTM, GRU, CapsuleNet, and Self-Attention) and obtained eight different models by combining two different word embedding models. They then used Dropout to support improved model convergence. Finally, at each model output, the four predicted probability categories

Table 4
Comparison of existing text conversations-oriented monolingual emotion analysis models.

No.	Name	DL method	Pre-training	Year	Dataset	macro-F1/weighted-F1
1	DialogueGCN [99]	GRU, GCN	Golve	2019	IEMOCAP	0.6418
2	KET [100]	MHA	Golve	2019	EC, DailyDialog, MELD, EmoryNLP, IEMOCAP	0.7413, 0.5337, 0.5818, 0.3439, 0.5956
3	ConGCN [101]	graph convolutional network	GloVe	2019	MELD	0.574
4	DialogueRNN [102]	RNN, CNN, attention	NA	2019	IEMOCAP	0.6275
5	RGAT [103]	relational graph attention networks	BERT	2020	DailyDialog, MELD, EmoryNLP, IEMOCAP	0.5431, 0.6091, 0.3442, 0.6522
6	KAITML [104]	graph attention mechanism, incremental transformer	GloVe	2020	EC, DailyDialog, MELD, EmoryNLP, IEMOCAP	0.7539, 0.5471, 0.5897, 0.3559, 0.6143
7	HiGRU [105]	Bi-GRU	word2vec	2019	Friends, EmotionPush, IEMOCAP	0.744, 0.771, 0.821
8	IDS-ECM [106]	Bi-LSTM	DeepMoji	2020	DailyDialog, EC	0.3885, 0.3623
9	HiTrans [107]	relational graph attention networks	BERT	2020	MELD, EmoryNLP, IEMOCAP	0.6194, 0.3675, 0.6450
10	COSMIC [108]	Bi-GRU	RoBERTa	2020	DailyDialog, MELD, EmoryNLP, IEMOCAP	0.5105, 0.6521, 0.3811, 0.6528
11	Lu's [109]	Bi-GRU	GloVe	2020	MELD, IEMOCAP	0.6072, 0.6437
12	HiTransformer [110]	Bi-LSTM, MHA	BERT	2020	Friends, EmoryPush, EmoryNLP	0.6788, 0.6543, 0.3304
13	ED-NNEV [111]	CNN	word2vec	2017	Chats data of phone	0.7438

were obtained. Ragheb et al. [121] presented a model to detect textual conversational emotion. They used deep transfer learning, self-attention mechanisms, and turn-based conversational modeling to classify emotion.

Lee et al. [122] proposed a multi-view turn-by-turn model. In this model, the vectors were generated from each utterance using two encoders: a word-level Bi-GRU encoder and a character-level CNN encoder. The model could predict emotion with the contextual information, which was grasped by combining the vectors. Ma et al. [123] proposed a DL architecture that combined the Bi-LSTM and the attention mechanism to extract emotional information from an utterance. Ge et al. [124] proposed an attentional LSTM-CNN model for dialogue emotion classification. They used a combination of CNNs and long-short term neural networks to capture both local and long-distance contextual information in conversations. In addition, they applied the attention mechanism to recognize and attend to important words within conversations. They also used ensemble strategies by combing the variants of the proposed model with different pre-trained word embedding via weighted voting.

The specific comparison of existing text conversation-oriented monolingual emotion analysis models from SemEval-2019 Task 3 is listed in Table 5.

6.3. Text-oriented cross-linguistic emotion analysis models

In this subsection, we will provide a survey on the text-oriented cross-linguistic emotion analysis model. The basic understanding of related approaches is summarized as follows.

Wang et al. [125] proposed a Bilingual Attention Network (BAN) model based on LSTM and the attention mechanism. BAN can aggregate monolingual and bilingual informative words to form vectors from document representations; it can also integrate attention vectors to conduct emotion prediction. Zhou et al. [126] proposed an attention-based cross-lingual sentiment classification model that learns the distributed semantics of documents in both source and target languages. In each language, they used LSTM to model documents and introduced a hierarchical attention mechanism for the model. Chen et al. [127] presented an Adversarial DAN (ADAN) for cross-lingual sentiment classification. ADAN could transfer knowledge learned from labeled English data to Chinese and Arabic, where little or no annotated data existed.

Zhou et al. [128] proposed a Bilingual Sentiment Word Embeddings (BSWE) method based on DL technology, for English-Chinese cross-language sentiment classification. BSWE could use DAE to learn bilingual embeddings for Cross-language Sentiment Classification (CLSC). Feng and Wan [129] proposed a Cross-Language In-domain Sentiment Analysis (CLIDSA) model based on LSTM. It was an end-to-end method that leveraged unlabeled data in multiple languages and multiple domains. Barnes et al. [130] proposed a Bilingual Sentiment Embeddings (BLSE) model that used a two-layer feed-forward averaging network to predict text sentiment. Ahmad et al. [131] built a DL model for emotion detection of the Hindi language. They used a CNN, Bi-LSTM, cross-lingual embeddings, and different transfer learning strategies for their purpose.

The specific comparison of existing text-oriented cross-linguistic emotion analysis models is listed in Table 6.

6.4. Emoji-oriented cross-linguistic emotion analysis model

Emojis are defined by the Oxford Dictionary [132] as "A small digital image or icon used to express an idea or emotion". To enhance the visual effect and meaning of a short text, emojis are becoming one of the indispensable components in any instant messaging platform or social media service. Due to emojis' increasing importance in emotion analysis of social networks, the SemEval-2018 task 2: Emoji Prediction in English and Spanish [133] was introduced in 2018. The aim was to attract greater NLP attention. The basic understanding of related methods is summarized as follows.

Coltekin and Rama [134] designed a supervised system consisting of an SVM classifier with bag-of-n-grams features. Baziotis et al. [135] proposed an architecture to predict emojis using Bi-LSTM and a context-aware attention mechanism. Beaulieu and Owusu [136] proposed a method to predict English and Spanish emojis using a bag-of-words model and a linear SVM. Coster et al. [137] built a linear

Table 5
Comparison of existing text conversations-oriented monolingual emotion analysis models from SemEval-2019 Task 3.

No.	System	Ranking	DL method	Pre-training	macro-F1	Country
1	NELEC [113]	3	GRU, LSTM, attention	Emoji2Vec, GloVe	0.7765	the U.S.
2	SymantoResearch [114]	4	Bi-LSTM, attention	BERT	0.7731	Germany
3	ANA [115]	5	multi-head self-attention, LSTM	BERT, GloVe, ELMO, DeepMoji	0.7709	Canada
4	CAiRE HKUST [116]	6	LSTM, hierarchical attention	BERT, GloVe, ELMO, DeepMoji	0.7677	China
5	SNU IDS [117]	7	Bi-LSTM, multi-dimensional attention	Word2Vec, ELMO, Emoji2Vec	0.7661	Korea
6	THU-HCSI [118]	8	Bi-LSTM, LSTM, CNN, attention	word2vec, NTUA-SLP	0.7616	China
7	Figure Eight [119]	9	Bi-LSTM, attention	ULMFit, BERT, NTUASLP, DeepMoji, OpenAI- GPT	0.7608	the U.S.
8	YUN-HPCC [120]	10	Bi-LSTM, GRU, Capsule-Net, attention	ELMO, GloVe	0.7588	China
9	LIRMM-Advanse [121]	11	Bi-LSTM, AWD-LSTM, attention	ULMFit	0.7582	France
10	MILAB [122]	12	CNN, Bi-GRU	GloVe	0.7581	Korea
11	PKUSE [123]	14	Bi-LSTM, attention	GloVe	0.7557	China
12	THU NGN [124]	15	CNN, LSTM, attention	GloVe, word2vec, ekphrasis	0.7542	China

Table 6
Comparison of existing text-oriented cross-linguistic emotion analysis models.

No.	Name	DL method	Pre-training	Year	Dataset	Accuracy
1	BAN [125]	LSTM, attention	Skip-gram	2016	Weibo	0.672
2	Zhou's [126]	LSTM, Bi-LSTM, attention	NA	2016	NLPCC 2013	0.824
3	ADAN [127]	DAN	BWE	2018	Yelp reviews, Chinese hotel reviews	0.4249, 0.5454
4	BSWE-CLSC [128]	denoising autoencoder	BSWE	2015	NLPCC 2013	0.8068
5	CLIDSA [129]	LSTM	Unsupervised CLCA	2019	Amazon review	0.8483
6	BLSE [130]	DAN	word2vec	2018	OpeNER English and Spanish datasets, MultiBooked Catalan and Basque	ES: 0.803, CA: 0.85, EU: 73.5
7	Ahmad's [131]	CNN, Bi-LSTM	fastText, alignment matrices	2020	SemEval-2018, Emo-Crowd-EN, Hindi review	Emo-Dis-HI: 0.477, Emo-SemEval-EN: 0.863

SVM model to predict emoji in Spanish tweets using the SKLearn SGDClassifier.

Jin and Pedersen [138] built a classifier for Spanish emoji prediction using naive Bayes, logistic regression, and random forests. Basile and Lino [139] presented an approach to predict Spanish emoji based on the SVM model. Liu [140] presented a model for English emoji prediction using a gradient boosting regression tree method. Lu et al. [141] proposed a method to address Twitter emoji prediction based on Bi-LSTM and the attention mechanism.

The specific comparison of existing emoji-oriented cross-linguistic emotion analysis models is listed in Table 7.

7. Challenges of emotion analysis

Due to the increasing development of social networks and DL technology, unprecedented challenges to emotion analysis have been posed. Though many researchers have proposed potential solutions for some of the discussed issues, there are still many other open issues requiring further exploration and deep study [142]. In this section, we summarize the challenges of emotion analysis and point out the future trends in this field.

7.1. Emotion description

At present, there is no unified definition for emotion and no unified standard to classify emotions effectively and scientifically, which may affect emotional feature extraction performance involving texts. However, because of the three unique components of human emotion (physiological arousal, subjective experience, and external expression), different fields possess different understandings. For example, social psychology, developmental psychology, and neuroscience deem it impossible for researchers to have the same understanding of emotion [142]. Thus, there is difficultly in determining a unified standard to accurately characterize human emotions.

7.2. Data imbalance

Emotion classification has made great progress in NLP. However, most existing works assume there are as many positive samples as negative samples, while positive and negative samples are often distributed unevenly in practice. Emotion analysis is a more fine-grained classification based on sentiment analysis, yet most of that work also assumes balanced sample sizes for each emotion category, which is not consistent with reality [143]. Thus, when methods suitable for balanced classification are used to deal with unbalanced data, analysis results often fail to achieve their intended effects, directly affecting the performance of emotion classification.

7.3. Language imbalance

Most of the existing emotion analysis methods are aimed at English texts. Some recent methods have focused on Chinese texts, but these methods are based on emotion dictionary or semantic knowledge base that rely on external resources of specific languages [144]. It is difficult to transfer English-based emotion analysis methods to other languages (e.g., Japanese and French). In addition, training and test sets of non-English emotional analysis are relatively scarce, particularly uncommon language resources, which has a serious negative impact on the research and performance of non-English language emotion analysis methods.

7.4. Domain relevance

Descriptive words and phrases, such as “a long time”, can express different emotions depending on their domain. For instance, food and beverage reviews often express negative emotion in relation to long waiting times, while smartphone reviews express positive emotion in relation to long battery standby times [145]. Thus, the domain relevance of words must be considered by emotion analysis. Cross-domain emotion analysis presents numerous pressing problems for resolution.

7.5. Understanding short texts

Social networks limit the length of their commentary, making short text (with its sparseness, non-standard use of words, and massive data) common instead of traditional long text. At the same time, insufficient contextual semantic information, single-word polysemy, and multi-word synonymy make topic information extraction difficult to perform accurately, affecting final emotion analysis performance. Thus, understanding short texts is a very challenging task in NLP.

7.6. Emotion cause extraction (ECE)

ECE aims to identify important potential causes or stimuli for observed emotions during in-depth emotion analysis [146]. However, most existing works focus on annotating emotions before causing extraction, which greatly limits the latter's application in real-world scenarios, and ignores mutual indications between two emotions. In addition, due to the inherent subtlety and ambiguity of emotional expression, ECE has become a very challenging task.

7.7. DL model training

The repetitious process of DL adjusts model parameters; during DL model training, model training speed presents the biggest problem due to slow convergence speeds and long training times. Thus, model training efficiency deserves consideration. To improve the convergence speed of the model, we need to reduce the number of iterations and consistent training times, while to improve the training speed, we need to reduce the number of training times, so as to reduce the chance of trying different super parameters. Both improvements will affect model accuracy. In addition, training model performance relates to training dataset size, with larger datasets producing better results [147]. However, larger datasets increase training times and can require larger amounts of computing resources (e.g., GPUs).

8. Future research trends

As Internet +, AI +, 5G, and other opportunities arise, many new

Table 7
Comparison of existing emoji-oriented cross-linguistic emotion analysis models.

No.	System	DL method	Pre-training	EnglishRanking	Spanish Ranking	Country
1	Tubingen-Oslo [134]	SVM	NA	1	1	Germany, Norway
2	NTUA-SLP [135]	Bi-LSTM, attention	word2vec	2	NA	Greece
3	EmoNLP [136]	gradient boosting regression tree	NA	4	NA	NULL
4	ECNU [137]	Bi-LSTM, attention	POS embedding	5	7	China
5	UMDuluth-CS8761 [138]	SVM	NA	6	3	the U.S.
6	Hatching Chick [139]	SVM, gradient descent optimization	NA	29	2	Holland
7	TAJJEB [140]	SVM	POS embedding	8	4	Malta, Spain
8	Duluth UROP [141]	naive Bayes, logistic regression, random forests	NA	18	5	the U.S.

applications (e.g., multi-language, multi-modal, cross-domain, big data) have emerged and provided fresh opportunities for emotion analysis development. As emotion analysis plays an important role in grasping public sentiment trends quickly, predicting public opinions of relevant development trends, and satisfying daily human needs, emotion analysis will change qualitatively by integrating different media, forms, scales, and domains of emotional information. In addition, the rapid development of social network analysis and DL technology offers many new research directions for emotion analysis. Some researchers are gradually changing the research focus of emotion analysis, from single language, single media, a single domain, and small-scale data samples to multi-language, multi-modal, cross-domain, and big data [148,149]. According to existing technology development trends, future emotion analysis research will include the following aspects.

8.1. Multi-language emotional analysis

Due to increasing cultural exchanges, multi-language network information affects and merges with itself. Existing work has focused on a single language and corpus resources collected for a single-language emotion analysis model cannot be applied to multi-language emotion analysis. In addition, corpus resources for the emotion analysis of different languages are also unbalanced, making their application to multi-language environments difficult.

8.2. Multi-modal emotional analysis

While traditional emotion analysis focuses on single forms of media, multi-modal information (e.g., audio, video, and image) [150,151] can often express emotional effect than text with greater description and vividness. In addition, as the main carrier of emotional information expression, voice can accurately reflect current user emotions. Thus, allowing for the combined study of various social media big data types (e.g., image, audio, text, video), improved application prospects will become available for researching multi-model user emotion analysis.

8.3. Cross-domain emotional analysis

The main idea of the cross-domain emotion analysis method is that emotions present in current comment information can be identified accurately and quickly, provided said information contains words expressing various emotions from different domains. However, traditional emotion analysis methods often ignore domain dependency characteristics of emotional words and may even deliberately choose domain-independent features (such as emoji). With an increasing demand for practical application and the emergence of emotional corpus resources in different domains, cross-domain emotion analysis will draw greater research attention and focus.

8.4. Emotion analysis based on social network analysis

With the rapid development of social networks, a large amount of user interaction data has been generated. These data reflect not only static user characteristics (e.g., number of friends, activity, frequency of surfing the Internet), but also dynamic user characteristics (e.g., thoughts, social relations, social influence). By analysing social networks, we can understand how different individuals and social groups express their emotions and how group emotional tendencies relate to popular events. Therefore, research on emotion analysis technology based on social network analysis can better describe public opinion trends while providing technical support for applications involving big search, public opinion analysis, personalized recommendation, etc.

8.5. Emotion analysis based on big data analysis

With the increasing scale of social networks, massive data are produced every day. Mining these data can produce substantially valuable information for products and services, but a significant portion of that data is stored in an unstructured form after being collected by crawlers. When emotion analysis is carried out on text data, traditional methods of probabilistic latent semantic analysis will have difficulty meeting the needs set by large-scale data training. This will drive method proposals for emotion analysis based on big data.

8.6. In-depth emotion analysis

The purpose of extracting emotional cause is to recognize the potential cause or stimulus of an observed emotion. Existing methods of emotion analysis focus on the shallow tasks of emotion recognition and classification. However, emotion cause identification requires in-depth emotion analysis that focuses on emotional keywords in text to identify causes automatically. Although current mainstream methods are based on linguistic rules and statistics, the wide application of DL will continue to attract increasing attention in ECE research.

8.7. Automatic recognition of negative emotions in short texts

At present, there are a large number of comments on Wechat, Twitter, Taobao, and other social networks. Much of this massive amount of short text information contains negative emotion, making the automatic identification of negative emotion from such information an urgent need along with mastering the intelligent extraction of negative emotion features. Such needs will play important roles in national cyberspace security, driving greater numbers of scholars to study automatic negative emotion recognition in short texts.

8.8. Negative emotion evolution analysis

Due to expression methods, popular events drive negative emotions with the highest concentration of these emotions being expressed on social networks. Such networks (e.g., Microblog) often have propagation characteristics of "weak information and strong emotion", which has caused negative emotions to be propagated widely across Microblog [142]. Once information with negative emotions is released, it may propagate by means of nuclear fission. If it is amplified by users with great social influence (e.g., opinion leaders) [152,153], it will influence and possibly become public opinion. Thus, characterizing the internal evolution of negative emotions for popular events will become a notable focus of NLP research and cyberspace security.

9. Conclusions

In this survey, our purpose was to review existing studies on DL-based TEA solutions and provide a comprehensive understanding for new researchers. We introduced the background of TEA, a brief on sentiment analysis approaches, the current state-of-the-art challenges, and future research trends. We began by introducing preliminaries, such as emotion definition and classification, a summary of emotion analysis applications, basic DL methods, and pre-training methods. We then reviewed the literature based on various DL methods and compared these studies according to our understanding. Finally, we presented readers with challenges and future directions for emotion analysis. We hope that this survey can provide a good reference for designing DL-based emotion analysis models with improved performance.

Declaration of competing interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China under Grant Nos. 61876205 and 61877013, the Ministry of Education of Humanities and Social Science project under Grant Nos. 19YJAZH128 and 20YJAZH118, the Science and Technology Plan Project of Guangzhou under Grant No. 201804010433, and the Bidding Project of Laboratory of Language Engineering and Computing under Grant No. LEC2017ZBKT001.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dcan.2021.10.003.

References

[1] S. Bharti, B. Vachha, R. Pradhan, K. Babu, S. Jena, Sarcastic sentiment detection in tweets streamed in real time: a big data approach, Digit.Commun.Network 2 (2016) 108–121.
[2] Y. Hao, Q. Zheng, Y. Chen, C. Yan, Recognition of abnormal behavior based on data of public opinion on the web, J. Comput. Res. Dev. 53 (3) (2016) 611–620.
[3] B. Fang, Y. Jia, A. Li, L. Yin, Research progress and trend of cyberspace big search, J. Commun. 36 (12) (2015) 1–8.
[4] D. Paul, F. Li, M.K. Teja, X. Yu, R. Frost, Compass: spatio temporal sentiment analysis of US Election what Twitter says!, in: Proceedings of the 23rd ACM International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1585–1594. Halifax, Canada.
[5] L. Zhang, C. Xu, Y. Gao, Y. Han, X. Du, Z. Tian, Improved dota2 lineup recommendation model based on a bidirectional lstm, Tsinghua Sci. Technol. 25 (6) (2020) 712–720.
[6] M.D. Choudhury, S. Counts, E.J. Horvitz, A. Hoff, Characterizing and predicting postpartum depression from shared Facebook data, in: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore, MD, USA, 2014, pp. 626–638.
[7] C. Tucker, B. Pursel, A. Divinsky, Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes, The ASEE Computers in Education (CoED) Journal 5 (4) (2014) 84.
[8] Development Report of China New Media£Propagation of Microblog, Forum and Other We-Media Increase the Threat on Modern Society, China Reading Newspaper, November 28 2012. URL, http://epaper.gmw.cn/zhdsb/html/2012-11/28/nw.D110000zhdsb_20121128_3-18.htm (Accessed 15 July 2020).
[9] H. Zhu, X. Shan, J. Hu, China Internet public opinion analysis report (full text), July 2012), http://yuqing.people.com.cn/n/2012/0727/c209170-18615551.html, 2011 (Accessed 15 July 2020).
[10] S. Peng, G. Wang, Y. Zhou, C. Wan, C. Wang, S. Yu, J. Niu, An immunization framework for social networks through big data based influence modeling, IEEE Trans. Dependable Secure Comput. 16 (6) (2019) 984–995.
[11] Z. Zhang, X. Li, C. Gan, Identifying influential nodes in social networks via community structure and influence distribution difference, Digit.Commun.Network 7 (1) (2021) 131–139.
[12] D. Camachoa, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, E. Cambria, The four dimensions of social network analysis: an overview of research methods, applications, and software tools, Inf. Fusion 63 (2020) 88–120.
[13] S. Peng, Y. Zhou, L. Cao, S. Yu, J. Niu, W. Jia, Influence analysis in social networks: a survey, J. Netw. Comput. Appl. 106 (2018) 17–32.
[14] The 38th Statistical Report on the Development of china's Internet, August 2016. URL, http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwt- jbg/20-1608/t20160803_54392.htm (Accessed 15 July 2020).
[15] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective computing: from unimodal analysis to multimodal fusion, Inf. Fusion 37 (2017) 98–125.
[16] R. Li, Z. Lin, H. Lin, W. Wang, D. Meng, Text emotion analysis: a survey, J. Comput. Res. Dev. 55 (1) (2018) 30–52.
[17] M. Bouazizi, T. Ohtsuki, Multi-class sentiment analysis on twitter: classification performance and challenges, Big Data Mining and Analytics 2 (3) (2019) 181–194.
[18] M. Usama, B. Ahmad, E. Song, M.S. Hossain, M. Alrashoud, G. Muhammad, Attention-based sentiment analysis using convolutional and recurrent neural network, Future Generat. Comput. Syst. 113 (2020) 571–578.
[19] E. Cambria, S. Poria, A. Gelbukh, M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intell. Syst. 32 (6) (2017) 74–80.
[20] M. Zhou, N. Duan, S. Liu, H.-Y. Shum, Progress in neural NLP: modeling, learning, and reasoning, Engineering (2020), https://doi.org/10.1016/j.eng.2019.12.014 (Accessed 15 July 2020).
[21] T. S. amd M.A. Chishti, Deep learning for the internet of things: potential benefits and use-cases, Digit.Commun.Network 7 (4) (2021) 526-542. https://doi.org/10.1016/j.dcan.2020.12.002..
[22] J. Deng, W. Dong, R. Socher, L. Li, K. Li, F. Li, ImageNet: a large-scale hierarchical image database, in: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248–255.
[23] M. Iyyer, V. Manjunatha, J. Boyd-Graber, H. Daumé III, Deep unordered composition rivals syntactic methods for text classification, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 2015, pp. 1681–1691.
[24] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol, Extracting and composing robustfeatures with denoising autoencoders, in: Proceedings of the 25th International Conference on Machine Learning, USA, 2008, pp. 1096–1103.
[25] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Browse J.Mag. 86 (11) (1998) 2278–2324.
[26] C.L. Giles, G.M. Kuhn, R.J. Williams, Dynamic recurrent neural networks: theory and applications, IEEE Trans. Neural Network. 5 (1994) 153–156.
[27] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[28] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681.
[29] K. Cho, B. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014, pp. 1724–1734.
[30] L. Itti, C. Koch, Computational modelling of visual attention, Nat. Rev. Neurosci. 2 (3) (2001) 194–203.
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017, pp. 5999–6009.
[32] Y. Lv, F. Wei, L. Cao, S. Peng, J. Niu, S. Yu, C. Wang, Aspect-level sentiment analysis using context and aspect memory network, Neurocomputing 428 (2021) 195–205.
[33] Dictionary by merriam-webster, emotion, URL, https://www.merriam-webster.com/dictionary/e-motion (Accessed 15 July 2020).
[34] X. Huang, Introduction to Psychology, Peoples Education Press, OBeijing, 1991.
[35] E. Hudlicka, Guidelines for designing computational models of emotions, Int. J. Synth. Emot. (IJSE) 2 (1) (2011) 26–79.
[36] M. Munezero, C. Montero, E. Sutinen, J. Pajunen, Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text, IEEE Trans.Affect.Computing 5 (2) (2014) 101–111.
[37] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.
[38] S. Poria, N. Majumder, R. Mihalcea, E. Hovy, Emotion recognition in conversation: research challenges, datasets, and recent advances, IEEE Access 7 (2019) 100943–100953.
[39] P. Ekman, An argument for basic emotions, Cognit. Emot. 6 (3/4) (1992) 169–200.
[40] W. Parrott, Emotions in Social Psychology: Essential Readings, Psychology Press, Oxford, UK, 2001.
[41] R. Plutchik, The nature of emotions, Phil. Stud. 89 (4) (2001) 393–409.
[42] C. Lin, Emotional problems in socialist psychology, Science of Social Psychology 21 (83) (2006) 37–62.
[43] A. Ceron, L. Curini, S.M. Iacus, G. Porro, Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to Italy and France, New Media Soc. 16 (2) (2014) 340–358.
[44] B. Alkouz, Z. Aghbari, J. Abawajy, Tweetluenza: predicting flu trends from twitter data, Big Data Mining and Analytics 2 (4) (2019) 273–287.
[45] J. Zhang, Y. Wang, Z. Yuan, Q. Jin, Personalized real-time movie recommendation system: practical prototype and evaluation, Tsinghua Sci. Technol. 25 (2) (2020) 180–191.
[46] P. Zhang, X. Huang, L. Zhang, Information mining and similarity computation for semi-/unstructured sentences from the social data, Digit.Commun.Network 7 (4) (2021) 518-525, https://doi.org/10.1016/j.dcan.2020.08.001.
[47] H. Chen, C. Yin, W.R.R. Li, Z. Xiong, B. David, Enhanced learning resource recommendation based on online learning style model, Tsinghua Sci. Technol. 25 (3) (2020) 348–356.
[48] C. Yang, X. Lai, Z. Hu, Y. Liu, P. Shen, Depression tendency screening use text based emotional analysis technique, J. Phys. Conf. 1237 (2019) 1–10.
[49] Z. Xie, Modelling the dropout patterns of mooc learners, Tsinghua Sci. Technol. 25 (3) (2020) 313–324.
[50] J. Liao, J. Tang, X. Zhao, Course drop-out prediction on mooc platform via clustering and tensor completion, Tsinghua Sci. Technol. 24 (4) (2019) 412–422.
[51] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, in: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, Florida, USA, 2009, pp. 448–455.
[52] X. Xi, G. Zhou, A survey on deep learning for natural language processing, Acta Autom. Sin. 42 (10) (2016) 1445–1465.
[53] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, in: Proceedings of the 1st International Conference on Learning Representations (ICLR 2013), Scottsdale, Arizona, USA, 2013.
[54] W.Y. Zou, R. Socher, D. Cer, C.D. Manning, Bilingual word embeddings for phrase-based machine translation, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, 2013, pp. 1393–1398.
[55] J. Pennington, R. Socher, C. Manning, GloVe: global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, 2014, pp. 1532–1543. Doha, Qatar.

[56] X. Zhou, X. Wan, J. Xiao, Cross-lingual sentiment classification with bilingual document representation learning, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ume 1, Long Papers, Berlin, Germany, 2016, pp. 1403–1412.
[57] B. Eisner, T. Rocktaschel, I. Augenstein, M. Bosnjak, S. Riedel, emoji2vec: learning emoji representations from their description, in: Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media, Austin, TX, 2016, pp. 48–54.
[58] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, M. Zhou, Sentiment embeddings with applications to sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (2) (2016) 496–509.
[59] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 2017, pp. 427–431.
[60] O. Melamud, J. Goldberger, I. Dagan, context2vec: learning generic context embedding with bidirectional LSTM, in: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016), Berlin, Germany, 2016, pp. 51–61.
[61] L. Yu, J. Wang, K. Lai, X. Zhang, Refining word embeddings for sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017), Copenhagen, Denmark, 2017, pp. 534–539.
[62] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018, pp. 2227–2237. New Orleans, Louisiana, USA.
[63] B.H. Soleimani, S. Matwin, Fast PMI-based word embedding with efficient use of unobserved patterns, in: Proceeding of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019), 2019, pp. 7031–7038.
[64] S. Jamee, Z. Fu, B. Shi, W. Lam, S. Schockaert, Word embedding as maximum a posteriori estimation, in: Proceeding of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019), 2019, pp. 6562–6569.
[65] B. McCann, J. Bradbury, C. Xiong, R. Socher, Learned in translation: contextualized word vectors, in: Proceedings of the Thirty-First Conference on Neural Information Processing Systems (NIPS 2017), Long Beach CA, USA, 2017, pp. 1–12.
[66] P. Xu, A. Madotto, C. Wu, J. Park, P. Fung, Emo2vec: learning generalized emotion representation by multitask training, in: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Brussels, Belgium, 2018, pp. 292–298.
[67] J. Howard, S. Ruder, Universal language model fine-tuning for text classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ume 1, Long Papers), Melbourne, Australia, 2019, pp. 328–339.
[68] C. Baziotis, N. Athanasiou, A. Chronopoulou, A. Kolovou, G. Paraskevopoulos, N. Ellinas, S. Narayanan, A. Potamianos, NTUA-SLP at SemEval-2018 task 1: predicting affective content in tweets with deep attentive RNNs and transfer learning, in: Proceedings of the 12th International Workshop on Semantic Evaluation, Louisiana, 2018, pp. 245–255. New Orleans.
[69] B.H. Soleimani, S. Matwin, Spectral word embedding with negative sampling, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Lousiana, USA, 2018, pp. 5481–5487.
[70] S. Cao, W. Lu, J. Zhou, X. Li, cw2vec: learning Chinese word embeddings with stroke n-gram information, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Lousiana, USA, 2018, pp. 5053–5061.
[71] T. Wada, T. Iwata, Y. Matsumoto, Unsupervised multilingual word embedding with limited resources using neural language models, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, pp. 3113–3124.
[72] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014, pp. 1188–1196.
[73] R. Kiros, Y. Zhu, R.R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, S. Fidler, Skip-thought vectors, Adv. Neural Inf. Process. Syst. (2015) 3294–3302.
[74] B. Felbo, A. Mislove, A. Sogaard, I. Rahwan, S. Lehmann, 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, Denmark, 2017, pp. 1616–1626.
[75] J. Devlin, M. Chang, K. Lee, K. Toutanova, 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, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[76] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, A. Bordes, Supervised learning of universal sentence representations from natural language inference data, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 670–680.
[77] Y. Gao, Y. Xu, H. Huang, Q. Liu, L. Wei, L. Liu, Jointly learning topics in sentence embedding for document summarization, IEEE Trans. Knowl. Data Eng. 32 (4) (2020) 688–699.
[78] Y. Hao, X. Liu, J. Wu, P. Lv, Exploiting sentence embedding for medical question answering, in: Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, 2019, pp. 938–945.
[79] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, Improving language understanding with unsupervised learning,technical report, URL, https://cdn.openai.com/research-covers/langu-age-unsupervised/language_understanding_paper.pdf, 2018 (Accessed 15 July 2020).
[80] F. Hill, K. Cho, A. Korhonen, Learning distributed representations of sentences from unlabelled data, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, California, San Diego, 2016, pp. 1367–1377.
[81] Y. Sun, S. Wang, Y. Li, S. Feng, H. Tian, H. Wu, H. Wang, ERNIE 2.0: a continual pre-training framework for language understanding, in: Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020), 2020, pp. 8968–8975.
[82] S. Subramanian, A. Trischler, Y. Bengio, C.J. Pal, Learning general purpose distributed sentence representations via large scale multi-task learning, in: Proceedings of the Sixth International Conference on Learning Representations, Canada, Vancouver, 2018, pp. 1–16.
[83] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R.S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y. Sung, B. Strope, R. Kurzweil, Universal sentence encoder for English, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (System Demonstrations), Brussels, Belgium, 2018, pp. 169–174.
[84] M. Pagliardini, P. Gupta, M. Jaggi, Unsupervised learning of sentence embeddings using compositional n-gram features, in: Proceedings of NAACL-HLT 2018, Louisiana, New Orleans, 2019, pp. 528–540.
[85] A. Nie, E.D. Bennett, N.D. Goodman, DisSent: learning sentence representations from explicit discourse relations, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, pp. 4497–4510.
[86] M. Abdul-Mageed, L. Ungar, Emonet: fine-grained emotion detection with gated recurrent neural network, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 718–728. Vancouver, Canada.
[87] S. Tafreshi, M. Diab, Emotion detection and classification in a multigenre corpus with joint multi-task deep learning, in: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 2905–2913.
[88] D. Kulshreshtha, P. Goel, A. Singh, How emotional are you? Neural architectures for emotion intensity prediction in microblogs, in: Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 2914–2926.
[89] E. Mohammadi, H. Amini, L. Kosseim, Neural feature extraction for contextual emotion detection, in: Proceedings of Recent Advances in Natural Language Processing, Varna, Bulgaria, 2019, pp. 785–794.
[90] P. Li, J. Li, F. Sun, P. Wang, Short text emotion analysis based on recurrent neural network, in: Proceedings of the 6th International Conference on Information Engineering, Dalian Liaoning, China, 2017, pp. 1–5.
[91] P. Rathnayaka, S. Abeysinghe, C. Samarajeewa, I. Manchanayake, M. Walpola, Sentylic at IEST 2018: gated recurrent neural network and capsule network based approach for Iimplicit emotion detection, in: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Brussels, Belgium, 2018, pp. 254–259.
[92] M. Akhtar, A. Ekbal, E. Cambria, How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble, IEEE Comput. Intell. Mag. (2020) 64–75.
[93] E. Batbaatar, M. Li, K. Ryu, Semantic-emotion neural network for emotion recognition from text, IEEE Access 7 (2019) 111866–111878.
[94] Y. Zhang, J. Fu, D. She, Y. Zhang, S. Wang, J. Yang, Text emotion distribution learning via multi-task convolutional neural network, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018, pp. 4595–4601.
[95] H. Khanpour, C. Caragea, Fine-grained emotion detection in health-related online posts, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018, pp. 1160–1166.
[96] Y. Yang, D. Zhou, Y. He, An interpretable neural network with topical information for relevant emotion ranking, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018, pp. 3423–3432.
[97] B. Kratzwald, S. Ilic, M. Kraus, S. Feuerriegel, H. Prendinger, Deep learning for affective computing: text-based emotion recognition in decision support, Decis. Support Syst. 115 (2018) 24–35.
[98] Y.Y. an D. Zhou, Y. He, M. Zhang, Interpretable relevant emotion ranking with event-driven attention, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 2019, pp. 177–187.
[99] D. Ghosal, N. Majumder, S. Poria, N. Chhaya, A. Gelbukh, 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), 2019, pp. 154–164. Hong Kong, China.
[100] P. Zhong, D. Wang, C. Miao, Knowledge-enriched transformer for emotion detection in textual conversations, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP), 2019, pp. 165–176. Hong Kong, China.
[101] D. Zhang, L. Wu, C. Sun, S. Li, Q. Zhu, G. Zhou, 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, IJCAI), 2019, pp. 5415–5421.

[102] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, E. Cambria, Dialoguernn: an attentive rnn for emotion detection in conversations, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 6818–6825.
[103] T. Ishiwatari, Y. Yasuda, T. Miyazaki, J. Goto, Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP), 2020, pp. 7360–7370.
[104] D. Zhang, X. Chen, S. Xu, B. Xu, Knowledge aware emotion recognition in textual conversations via multi-task incremental transformer, in: Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 2020, pp. 4429–4440.
[105] W. Jiao, H. Yang, I. King, M.R. Lyu, 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, ume 1, (Long and Short Papers), Minneapolis, Minnesota, 2019, pp. 397–406.
[106] D. Li, Y. Li, S. Wang, Interactive double states emotion cell model for textual dialogue emotion prediction, Knowl. Base Syst. 189 (2020) 1–11.
[107] J. Li, D. Ji, F. Li, M. Zhang, Y. Liu, HiTrans: a transformer-based context- and speaker-sensitive model for emotion detection in conversations, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 4190–4200.
[108] D. Ghosal, N. Majumder, A. Gelbukh, R. Mihalcea, S. Poria, COSMIC: commonsense knowledge for emotion identification in conversations, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP), 2020, pp. 2470–2481.
[109] X. Lu, Y. Zhao, Y. Wu, Y. Tian, H. Chen, B. Qin, An iterative emotion interaction network for emotion recognition in conversations, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 4078–4088.
[110] Q. Li, C. Wu, K. Zheng, Z. Wang, Hierarchical transformer network for utterance-level emotion recognition, Appl. Sci. 10 (13) (2020) 4447.
[111] S. Mundra, A. Sen, M. Sinha, S. Mannarswamy, S. Dandapat, S. Roy, Fine-grained emotion detection in contact center chat utterances, Pacific-Asia Conference on Knowledge Discovery and Data Mining (2017) 337–349.
[112] A. Chatterjee, K.N. Narahari, M. Joshi, P. Agrawal, Semeval-2019 task 3: emocontext: contextual emotion detection in text, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 39–48.
[113] P. Agrawal, A. Suri, NELEC at SemEval-2019 task 3: think twice before going deep, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 266–271.
[114] A. Basile, M. Franco-Salvador, N. Pawar, S. Stajner, M.C. Rios, Y. Benajiba, SymantoResearch at SemEval-2019 task 3: combined neural models for emotion classification in human-chatbot conversations, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 330–334.
[115] C. Huang, A. Trabelsi, O.R. Zaiane, ANA at SemEval-2019 task 3: contextual emotion detection in conversations through hierarchical LSTMs and BERT, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 49–53.
[116] G.I. Winata, A. Madotto, Z. Lin, J. Shin, Y. Xu, P. Xu, P. Fung, CAiRE HKUST at SemEval-2019 task 3: hierarchical attention for dialogue emotion classification, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 142–147.
[117] S. Bae, J. Choi, S. Lee, SNU IDS at SemEval-2019 task 3: addressing training-test class distribution mismatch in conversational classification, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 312–317.
[118] X. Liang, Y. Ma, M. Xu, THU-HCSI at SemEval-2019 task 3: hierarchical ensemble classification of contextual emotion in conversation, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 345–349.
[119] J. Xiao, Figure Eight at SemEval-2019 task 3: ensemble of transfer learning methods for contextual emotion detection, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 220–224.
[120] D. Li, J. Wang, X. Zhang, YUN-HPCC at SemEval-2019 task 3: multi-step ensemble neural network for sentiment analysis in textual conversation, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 360–364.
[121] W. Ragheb, J. Aze, S. Bringay, M. Servajean, LIRMM-Advanse at SemEval-2019 task 3: attentive conversation modeling for emotion detection and classification, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 251–255.
[122] Y. Lee, Y. Kim, K. Jung, MILAB at SemEval-2019 task 3: multi-view turn-by-turn model for context-aware sentiment analysis, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 256–260.
[123] L. Ma, L. Zhang, W. Ye, W. Hu, PKUSE at SemEval-2019 task 3: emotion detection with emotion-oriented neural attention network, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 287–291.
[124] S. Ge, T. Qi, C. Wu, Y. Huang, THU NGN at SemEval-2019 task 3: dialog emotion classification using attentional LSTM-CNN, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, 2019, pp. 340–344.
[125] Z. Wang, Y. Zhang, S. Lee, S. Li, G. Zhou, A bilingual attention network for code-switched emotion prediction, in: Proceedings of the 26th International Conference on Computational Linguistics, Osaka, Japan, 2016, pp. 1624–1634.
[126] X. Zhou, X. Wan, J. Xiao, Attention-based LSTM network for cross-lingual sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 247–256. Austin, Texas.
[127] X. Chen, Y. Sun, B. Athiwaratkun, C. Cardie, K. Weinberger, Adversarial deep averaging networks for cross-lingual sentiment classification, Trans.Assoc.Comput. Ling. 6 (2018) 557–570.
[128] H. Zhou, L. Chen, F. Shi, D. Huang, Learning bilingual sentimentword embeddings for cross-language 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, Beijing, China, 2015, pp. 430–440.
[129] Y. Feng, X. Wan, Towards a unified end-to-end approach for fully unsupervised cross-lingual sentiment analysis, in: Proceedings of the 23rd Conference on Computational Natural Language Learning, Hong Kong, China, 2019, pp. 1035–1044.
[130] J. Barnes, R. Klinger, S. Walde, Bilingual sentiment embeddings: joint projection of sentiment across languages, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), Melbourne, Australia, 2018, pp. 2483–2493.
[131] Z. Ahmad, R. Jindal, A. Ekbal, P. Bhattachharyya, Borrow from rich cousin: transfer learning for emotion detection using cross lingual embedding, Expert Syst. Appl. 139 (2020) 1–12, 112851.
[132] Oxford English and Spanish dictionary, emoji, URL, https://www.lexico.com/definition/emoji (Accessed 15 July 2020).
[133] F. Barbieri, J. Camacho-Collados, F. Ronzano, L. Espinosa-Anke, M. Ballesteros, V. Basile, V. Patti, H. Saggion, Semeval 2018 task 2: multilingual emoji prediction, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 24–33.
[134] C. Coltekin, T. Rama, Tubingen-Oslo at SemEval-2018 task 2: SVMs perform better than RNNs at emoji prediction, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 32–36.
[135] C. Baziotis, A. Nikolaos, A. Kolovou, G. Paraskevopoulos, N. Ellinas, A. Potamianos, NTUA-SLP at SemEval-2018 task 2: predicting Emojis using RNNs with context-aware attention, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana. 2018, pp. 438–444.
[136] J. Beaulieu, D.A. Owusu, UMDuluth-CS8761 at SemEval-2018 task 2: emojis: Too many choices?, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 397–401.
[137] J. Coster, R.G. van Dalen, N.A.J. Stierman, Hatching chick at SemEval-2018 task 2: multilingual emoji prediction, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 442–445. New Orleans, Louisiana.
[138] S. Jin, T. Pedersen, Duluth UROP at SemEval-2018 task 2: multilingual emoji prediction with ensemble learning and oversampling, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 479–482.
[139] A. Basile, K.W. Lino, TAJJEB at SemEval-2018 task 2: traditional approaches just do the job with emoji prediction, in: Proceedings of the 12th International Workshop on Semantic Evaluation, 2018, pp. 467–473. New Orleans, Louisiana.
[140] M. Liu, EmoNLP at SemEval-2018 task 2: English emoji prediction with gradient boosting regression tree method and bidirectional lstm, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 387–391.
[141] X. Lu, X. Mao, M. Lan, Y. Wu, ECNU at SemEval-2018 task 2: leverage traditional nlp features and neural networks methods to address twitter emoji prediction task, in: Proceedings of the 12th International Workshop on Semantic Evaluation, New Orleans, Louisiana, 2018, pp. 430–434.
[142] L. Cao, S. Peng, P. Yin, Y. Zhou, A. Yang, X. Li, A survey of emotion analysis in text based on deep learning, in: Proceedings of the IEEE 8th International Conference on Smart City and Informatization (iSCI 2020), Guangzhou, China, 2020, pp. 81–88.
[143] R. Xu, T. Chen, Y. Xia, Q. Lu, B. Liu, X. Wang, Word embedding composition for data imbalances in sentiment and emotion classification, Cognitive Computation 7 (2015) 226–240.
[144] S.F. Yilmaz, E.B. Kaynak, A. Koc, H. Dibeklioglu, S.S. Kozat, Multi-label sentiment analysis on 100 languages with dynamic weighting for label imbalance, IEEE .Trans. Neural Network.Learn Syst (2021), https://doi.org/10.1109/TNNLS.2021.3094304.
[145] Z. Cao, Y. Zhou, A. Yang, S. Peng, Deep transfer learning mechanism for fine-grained cross-domain sentiment classification, Connect. Sci. 33 (4) (2021) 911–928.
[146] R. Xia, M. Zhang, Z. Ding, Rthn a rnn-transformer hierarchical network for emotion cause extraction, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19), 2019, pp. 5285–5291.
[147] S. Peng, L. Cao, Y. Zhou, J. Xie, P. Yin, J. Mo, Challenges and trends of android malware detection in the era of deep learning, in: Proceedings of the IEEE 8th International Conference on Smart City and Informatization (iSCI 2020), Baltimore, MD, USA, 2020, pp. 37–43.
[148] S. Peng, G. Wang, D. Xie, Social influence analysis in social networking big data: opportunities and challenges, IEEE Network 31 (1) (2017) 11–17.

[149] M. Mahmud, J. Huang, S. Salloum, T. Emara, K. Sadatdiynov, A survey of data partitioning and sampling methods to support big data analysis, Big Data Mining and Analytics 3 (2) (2020) 85–101.
[150] B. Liu, S. Tang, X. Sun, Q. Chen, J. Cao, J. Luo, S. Zhao, Context-aware social media user sentiment analysis, Tsinghua Sci. Technol. 25 (4) (2020) 528–541.
[151] W. Peng, X. Hong, G. Zhao, Adaptive modality distillation for separable multimodal sentiment analysis, IEEE Intell. Syst. 36 (3) (2021) 82–89, https://doi.org/10.1109/MIS.2021.3057757.
[152] S. Peng, A. Yang, L. Cao, S. Yu, D. Xie, Social influence modeling using information theory in mobile social networks, Inf. Sci. 379 (2017) 147–159.
[153] J. Wu, N. Wang, Approximating special social influence maximization problems, Tsinghua Sci. Technol. 25 (6) (2020) 703–711.