SCIPEDIA Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería RIMNI

Humor detection using deep learning in 10 years: A survey

Chengjuan Ren[1], Ziyu Guo[2], Ping Zhang[3], Yuhan Gao*[4]

1 College of Language Intelligence, Sichuan International Studies University, Chongqing, China

2 Department of Computer Science & Engineering, The Chinese University of Hong Kong, China

3 College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

4 *Corresponding author - KIPP Northeast Denver Middle School, 4635 Walden St. Denver, Colorado 80249, USA

Abstract

Humor is an important part of personal communication. How to create a computational model to recognize humor is still a very challenging task in natural language processing and linguistics. In this survey, we applied some rules to leave 29 articles spanning 10 years (2012 to 2023). The main elements covered by this survey include: recent state-of-the-art detection methods using deep learning from years 2012-2023, (2) summarizing features for humor detection from a linguistic perspective, (3) humor detection datasets, evaluation metrics, data domains and languages, (4) some tricks used in humor detection (e.g. Attention mechanism, multimodal), (5) recognizing open problems and highlight the feasible opportunities for future research directions. To the best of our knowledge, this is the first systematic survey for humor detection using deep learning. The survey can be used to assist novice and prominent researchers to understand the concept of humor, popular method and future research direction and so on.

OPEN ACCESS

Published: 31/01/2024

Accepted: 19/01/2024

Submitted: 13/10/2023

DOI: 10.23967/j.rimni.2024.01.006

Keywords:
Natural language processing
Humor detection
Deep learning
Linguistics

1. Introduction

In the realm of natural language processing, humor detection has emerged as a captivating and challenging research area. Humor, a distinctive and complex aspect of human communication, not only serves as a means of entertainment, but also plays a crucial role in facilitating social interactions and conveying implications in daily communication. Recognizing and comprehending humor in various textual content holds significant implications for various applications, like sentiment analysis.

Since humor is a complex and content-dependent aspect of language use which often involves puns, wordplay, irony, sarcasm, satire, incongruity, cultural interpretations, the intricate nature of humor often poses challenges for humor detection using deep learning. In the humor detection models, the challenges include the availability and quality of specific datasets for training, the domain the languages of the datasets, the appropriateness and effectiveness of metrics that assess model. In order to have a general picture about the rapid advancements in deep learning techniques for humor detection and foster future innovation in the field, this survey aims to present an in-depth analysis of peer-reviewed articles published within the last ten years (2012 to 2023), focusing on the main detection methods employed for humor detection, the results of those metrics, key research challenges as well as opportunities for future investigation. The surveyed articles were selected by first searching the online databases ACM, IEEE Explore, Scopus and Web of Science, Google Scholar for articles which related to humor detection. There were a number of conditions that must be met for an article to be downloaded: (1) using deep learning; (2) humor detection or recognition, and (3) articles of nearly 10 years.

More specifically, we reviewed the 10 years of humor detection using deep learning, and a total number of 29 peer-reviewed articles were included. This survey covers main techniques used for humor detection in these articles, humor analysis from the perspective of linguistics, humor detection datasets, recent state-of-the-art detection methods, as well as potential problems and feasible opportunities for future research directions. This systemic survey can be used to assist novice and prominent researchers to understand the concept of humor, popular methods for humor detection as well as potential revision of the detection methods.

Many papers have analyzed and discussed methods of humor detection. The use of deep learning in humor detection was investigated in this paper [1]. They reviewed the various methods of deep learning used to detect humor. A critical analysis of humor detection methods and future directions were also presented in this paper. However, the work does not explain the theoretical knowledge of humor in linguistics or the problems with humor detection tasks.

The purpose of this work is to investigate computational humor detection through four perspectives: datasets, features and algorithms. Therefore, to detail the computational techniques for humor identification, a review was conducted [2]. However, the description of humor detection methodology is not detailed enough. The methods analyzed in the paper do not focus on the

Correspondence: Chengjuan Ren (99003113@sisu.edu.cn), Ping Zhang (pingzhang@webmail.hzau.edu.cn).

currently popular deep learning methods. Nor is the theoretical knowledge of humor in linguistics addressed in the paper.

2. Humor studies in linguistics

Humor, when analyzed from a linguistic perspective, is a complex and multifaceted phenomenon that is influenced by various cultural, social, and individual factors. Understanding the role of these factors is crucial in comprehending the intricacies of humor in different linguistic contexts. One prominent aspect is the impact of culture on humor comprehension and appreciation. Cultural variation in humor preferences and styles has been widely acknowledged in research [3,4]. Cross-cultural studies have demonstrated that what may be perceived as humorous in one culture could be less amusing or even offensive in another culture due to differences in norms, values, and cultural references [5]. Thus, culture serves as a significant lens through which humor is interpreted, and understanding cultural nuances is vital for effective cross-cultural communication.

In addition to culture, social factors also play a substantial role in humor [6]. Social groups and in-group dynamics influence the type of humor used and the audience it is directed towards. In-group humor refers to humor shared and understood within a specific social group, fostering a sense of belonging and reinforcing group identity. On the other hand, out-group humor often involves jokes or stereotypes targeting individuals or groups outside of one's social circle. Social identity theory highlights how humor can be utilized to establish social boundaries and define group membership through shared laughter [7]. Understanding the social dynamics and context-specific humor is essential for deciphering the intended meaning and impact of humor within a given group or community.

Linguistic mechanisms and devices are fundamental to the creation and comprehension of humor. Various linguistic tools are employed to produce humorous effects, including puns, wordplay, irony, sarcasm, satire, incongruity and parody [8]. Puns, for instance, exploit the multiple meanings or sounds of words to create humorous ambiguity. Irony and sarcasm involve conveying the opposite of the literal meaning to evoke humor or convey a satirical message. Satire and parody, often used in political humor, employ exaggerated or mocking imitations to critique social or political phenomena.

In addition, the role of linguistic context in humor comprehension cannot be understated [9]. Humor is highly dependent on the contextual cues and shared knowledge that enable the listener to interpret and appreciate the intended humor. Contextual information, such as the setting, participants, and preceding discourse, provides crucial cues for decoding the humorous elements in a conversation or text [10]. Understanding the social, cultural, and linguistic context is essential for recognizing the incongruity or deviation from expectations that often underlies humorous situations [11]. Additionally, pragmatic principles, including implicature and presupposition, play a significant role in humor interpretation. Conversational implicatures, based on Grice's Cooperative Principle, involve inferring meanings that go beyond the literal content of the utterance, contributing to the humorous effect.

Beyond linguistic context, the broader cultural and situational context also influences the production and reception of humor. Cultural references and shared knowledge serve as a rich source of humor. Jokes and humorous expressions often draw upon culturally specific knowledge, historical events, or stereotypes that resonate with a particular audience [12]. Understanding the cultural references embedded within humor is essential for accuracy comprehension and appreciation. Furthermore, situational humor capitalizes on the unexpected or peculiar aspects of a given situation to elicit laughter. Contextual triggers, such as timing, setting, and social dynamics, enhance the comedic effect and contribute to the overall humorous experience.

Timing and delivery are critical elements in humor production. The timing of humorous elements, including punchlines or unexpected twists, significantly influences the comedic effect. Effective comedic timing involves creating suspense, building anticipation, and delivering the punchline or humorous element.

3. Problem definitions

We now consider how the problem of automatic humor detection has been delineated. Humor recognition is a challenging problem in natural language because of the differences in the perception of the definition of humor and the different characteristics of different types of humor. The most popular way of formulating humor detection is as a classification task, binary or multi-classification. Given a text, the aim is to predict whether it is humor or not.

4. Approach for humor detection

Mao and Liu descripted on contribution in the 2019 HAHA completion, in which they were offered with an annotated corpus of tweets from the crowd and asked to identify whether tweets are jokes or not, and to predict the funniness score value of tweets [13]. In their network, the pre-trained model was fine-tuned on training data for the HAHA task using BERT. A multi-layer bidirectional transform encoder can help learn deep bidirectional representations in network. The output layer, trained with the mean squared error between the predicted score and the label, was applied to generate the score using float labels to predict the funniness score. At last, they believed that the network was compellable and can be applied to multi-lingual document identification tasks.

A deep convolutional neural network has been deployed in an emotion recognition challenge in which it is successfully able to classify six-second labels as one of those emotions: bored, relaxed, horrified and humorous [14]. This network was used on 14 subjects and obtained a high performance of almost 100% when the test data was randomly selected from the dataset. The authors summarized that although the networks usually perform well for all data from one person, because the electrode placement is consistent, they may ask for extensive adjustment and tuning for good results in another person.

Xie et al. wanted to overcome the failure point how humorous features interact in different modes and how to convert between parts of speech of Chinese words in sentences [15]. Based on interactive attention and text and speech fusion, the authors proposed a multimodal Chinese humor classification algorithm. A distributed word vector was designed to obtain text semantics. Then, feature fusion block was built, combining spectrographic features with Chinese coding to descript language. At last, a mechanism was added to weigh global contextual information to capture the interplay between audio and text characteristics.

Now, the network lacks a mechanism to adapt to the characteristics of each individual, which means that performance is reduced. Kathan et al. intended to solve the issue by applying a new multi-modal recognition approach, where the models are individualized for each of the individuals in the Passau Spontaneous Football Coach Humor (Passau-SFCH) dataset [16]. The first step was the training of a model on

all individuals in the dataset. They then used the data from each individual to fine tune each of the layers of this model. We demonstrate the complementarity of the features by applying a weighted delay fusion technique, which raises the global accuracy to an AUC of 0.9308.

From the perspective of humor-aware language processing models, understanding the use of humor is a computationally challenging task. Godoy have refined contextualized non-verbal representations and developed more humor-aware methods for actual network [17]. This work can be combined with other artificial intelligence techniques to create social robots with personality and the ability to adapt to user and patient conditions.

Dushyant et al. designed two attention-like mechanisms for multi-task detection (humor, offensive, sarcasm and sentiment), namely the inter-task relationship module (iTRM) and the inter-class relationship module (iCRM) [18]. iTRM's primary motivation is to learn how tasks relate to one another and how they support one another. Contrast this with iCRM, which seeks to develop the relationships between the different task classes. The experiment shows that the developed multi-task network leads to better performance compared to single-tasked detection on SemEval 2020.

Xu and Xie suggest that in the humor rating problem, attentional mechanisms attend more to humorous words when claiming the level of humor [19]. By examining the patterns of attentional allocation, they discovered that attentional schemes tend to allocate greater attention to contextual information, instead of just to specific words. They offered a number of methods for investigating attention and proved them to be effective.

Francesco and Saggion studied the detection of irony and humor on Twitter and other social networks by treating it as a taxonomy problem [20]. To train classification procedures, they provided a wealth of features for both text understanding and representation. In domain-spanning semantic classification tests, their model reached and surpasses current state-of-the-art results.

Using BERT sentence embedding, they introduced a new method for recognizing humor in small texts [21]. In approach, BERT is used to embed phrases in a given text, which are then used as input to hidden layer connections in a neural network. In order to forecast the target value, these lines are finally concatenated. They created a new humor detection dataset consisting of 200k formal short texts (100k + and 100k -). The experimental results indicate that our proposal is able to detect humor in short texts with an accuracy of 98.2 per cent and an F1 score of 98.2 percent.

Humor is too abstract to be quantified, and there is no accuracy standard for the evaluation of humor. Wang et al. investigated CNN, RNN, BiLSTM performance in humor detection. To identify the best model, they used the regression and classification methods respectively [22]. The results demonstrated that the use of the CNN network was a superior choice, and in combination with the method of classification, the new model achieved the best performance on 15,000 news headlines.

To predict and detect humor in dialogue, Dario and Fung proposed to compare different supervised machine learning methods from a very famous TV series: "The Big Bang Theory" [23]. Using the canned laughter embedded in the audio track, they built a corpus in which punchlines are annotated. CNN surpasses the other techniques, achieving the best F-score of 68.5% versus CRF's 66.5% and RNN's 52.9%.

Chen and Rau developed a neuronal convolution network that was able to judge whether a phrase was humorous or not, and whether it was Western or Chinese [24]. A total of 463314 English sentences and 111614 Chinese sentences were used to train and test the model. The final model had an accuracy of 96.73% and was optimized through various tests.

It remains very hard to detect the underlying structure of humor, because of its subjective character. Prajapati et al. [25] presented a systematic analysis of different machine learning (logistic regression, decision trees, random forests, passive aggressive classifiers) and deep learning (CNN, LSTM) approaches to classify tweets as humorous or non-humorous. The results show that an increase in the accuracy of humor prediction has been achieved using deep learning in comparison to machine learning.

Li et al. [26] presented HEMOS (Humor-EMOji-Slang-based), a deep learning system for fine-grained Chinese sentiment classification. First, 576 common online slang terms were compiled into a slang dictionary and 109 Weibo emoticons were converted into textual features to create a Chinese emoticon dictionary. In the following phase, the standard sentiment categories of 'positive' and 'negative' are supplemented by the new 'optimistic humorous type' and 'pessimistic humorous type', and two polarity annotations are added. They tested the performance of both lexicons on underlabelled data by applying them to an attention-based bidirectional long short-term memory recurrent neural network (AttBiLSTM).

Ziser et al. [27] presented a deep learning method for the detection of humorous questions in PQA systems. Their framework demonstrates their contribution to humor detection by exploiting two properties of questions - incongruity and subjectivity. They assessed the framework on a realistic dataset and demonstrated an average accuracy of 90.8%, a relative improvement of up to 18.3% over standard methods.

Mahajan and Zaveri [28] used a variety of well-established machine-learning classifiers to address the issue of understanding sentiment based on affective information from text. They have utilized different affective content such as emoticons, writing styles such as punctuation, capitalization, and sentimental words and so on, which inhibit the emotions and feelings of a writer. Using different types of experimental configurations, the presented affect-based humor identification model was validated on the SemEval 2017 HashTagWars dataset and the yelp reviews dataset.

Hasan et al. [29] proposed Humor Knowledge Enriched Transformer (HKT), which captures the meaning of a multi-modal statement by incorporating external context and intelligence. By tracking the ambiguity and sentiment present in the language, the authors incorporated humor-centric external knowledge into the model. The model obtained 77.36% and 79.41% precision in detecting humorous punchlines on UR-FUNNY and MUStaRD datasets.

The study of Yang et al. [30] was aimed at discovering humorous underpinnings, recognizing humorous motifs, and extracting humorous anchoring. The authors initially propose different semantic constructs for humor and develop characteristic sets for each construct, and then apply a computational approach to humor recognition. They then create a quick and powerful method for extracting sentence humor anchors. Experiments carried out on two datasets show that the humor recognizer is capable of discriminating automatically between funny and not funny texts, and that the detected humor anchors agree quite well with human annotations.

To the MuSe 2022 Multimodal Emotion Challenge (MuSe), Xu et

al. [31] first built a predictive model based on Transformer and BiLSTM modules, and then suggest a hybrid fusion method using each modality's predictor results to improve model accuracy. The AUC is 0.8972 for the tested model.

In order to pave the way for understanding the multimodal language used to express humor, Hasan et al. [32] proposed a multifaceted multi-modal dataset named UR-FUNNY. For the natural language processing community, the dataset and related studies provide a framework for multimodal humor detection.

Xiong et al. [33] proposed a humor identification model (MLSN) using popular humor theory and deep learning approaches for humor task. By incorporating the incoherence, phonetic properties and vagueness of a humorous statement as semantic attributes, the model automatically recognizes whether a statement includes a humorous utterance. The results show that the new model has better humor recognition accuracy and can contribute to discourse understanding research on three datasets.

Christ et al. [34] have three datasets for this Multimodal Sentiment Analysis Challenge (MuSe) 2022, the Passau Spontaneous Football Coach Humor (Passau-SFCH) dataset, the Hume-Reaction dataset, the Ulm-Trier Social Stress Test (Ulm-TSST) dataset. For each sub-challenge, a deep learning recurrent neural network with LSTM cells was employed to determine benchmark on the test divisions. They recorded an area under the curve (AUC) of .8480 for MuSe humor; a mean (from 7 classes) Pearson's correlation coefficient (δ) of .2801 for MuSe response, as well as a concordance correlation coefficient (CCC) of .4931 and .4761 for valence and arousal in MuSe stress, respectively.

To aid humor understanding, they elaborated on incongruity and ambiguity and suggested an internal/external attention neural network (IEANN) for humor recognition [35]. To address incongruity and ambiguity in humorous texts, IEANN combined two different types of attentional mechanisms. At the same time, in order to verify the performance and reliability of the model, extensive experiments were conducted on two humor datasets.

Peng-Yu and Von-Wun [36] built and assembled four datasets with different joke type in English and Chinese in order to explore the patterns of humor, to detect humor and even to produce humor. The authors applied a Convolutional Neural Network (CNN) with rich filter size, count and motorway networks to improve network depth. The findings reveal that our model is superior to previous work in detecting different styles of humor, with accuracy, precision and recall benchmarks obtained in both English and Chinese.

Phonetics and ambiguity have been incorporated. Current detection methods suffer from unsuitable feature selection for neural networks. It is shown that the neural network has to learn the phonological structure and ambiguity of confused words in order to represent them. Then, to learn phonetic structure and semantic representation for humor recognition, the authors proposed the Phonetics and Ambiguity Comprehension Gated Attention network (PACGA) [37].

Chen and Zhang [38] recommended a transformer-based hierarchical model for both inherent semantics and cross-modal dependencies of related modalities. First, standard transformers are used to encode the features from each modality. Then, cross-modal transformers are then used to compute the cross-modal interactions from one modality to other modalities. Experimental findings demonstrate that an area under the curve (AUC) of 0.9065 can be reached on the MuSe-Humor test set.

The system submitted for all four subtasks of SemEval-2021 Task-7 is discussed in this paper. Mondal and Sharma [39] proposed two distinct methods of fine-tuning by employing both linear and bi-LSTM layers on the pre-trained BERT model. The output indicates that our model is capable of outperforming baseline models by a wide range.

The main concern of this research area is the development of an exceptional fusion scheme that is able to derive and combine the most important information from different modalities for sentiment analysis [40]. The authors described the bimodal merger network (BBFN), a new, end-to-end network that carries out fusion ('relation enrichment') and separation ('separation difference enrichment') on pair-wise modality descriptions. The two parts were trained simultaneously, simulating the battle between them. To further improve the final output, they used a controlled gating mechanism in the transformer technology.

Deep learning and machine learning were used to support the task of computational humor analysis [41]. The central element of this study was the micro-blogging site Twitter. To detect humor, new classification model was created to perform the task and constructed a funny dataset including 4000 samples.

As you can see from the Table 1, new models were proposed in all the articles surveyed. As this work is investigating humor detection based deep learning. So the proposed models of all the investigated articles are based on deep learning techniques. The articles surveyed came from a variety of domains. There is no pattern for that. However, most of the high-level articles came from challenges related to humor detection. This phenomenon may be correlated with dataset. The labelling of new data is a time-consuming, tedious process. Challenges datasets are publicly available, and it is much easier for authors to obtain. Because of the different datasets and different model, the performance of the models obtained from the articles under investigation cannot be compared.

Table 1. Summary of the information in the surveyed articles

Reference	Domain	Method	Language	Advantage	Future work	Dataset	Performance
[13]	Twitter	BERT-based	English	Major changes from system design to processing flow	Investigation of implementation in other text recognition tasks	HAHA task	Task 1: F1 0.771, Precision 0.724, Recall 0.825, Accuarcy 0.809. Task 2: 2 0.910
[14]	EEG	A new deep convolutional neural network	English	The labelling of the data	Improving generalization for different subjects	GAMEEMO	Accuracy 99.65%
[15]	Psychological activity and psychological state	IPYCA-CB network	Chinese	Using pinyin embedding and spectrogram convolution	Exploring corpus construction and knowledge fusion	Speech and textual information	Non-BERT :Accuracy 0.75, Marco-F1:0.67 On BERT: Accuracy 0.71, Marco-F1 0.63
[16]	None	A new deep convolutional neural network	English, German	Finding a mechanism to adapt to the characteristics of each individual	Extending the method with imbalanced datasets	The Passau Spontaneous Football Coach Humour (Passau-SFCH) dataset	AUC of audio modality 0.773; AUC of video modality 0.925
[17]	Short videoclips of speakers	RAVEN model	English	Developing humor-aware model	Decrease reliance on vast amounts of annotated dataset by using unsupervised learning	CMU-MOSI dataset	Accuracy 87.3%, F1 85.8%

					methods		
[18]	SemEval 2020	Two attention-like mechanisms	English	Understanding the links and similarities between different tasks	Compilation of a large and more balance data set	Released dataset in the Memotion Analysis task @ SemEval 2020	Task C, Precision 27.23%, Recall 27.29%. Task B: Precision 55.82%, Recall 53.84%
[19]	News	A new CNN model	English	Exploring the structure of the distribution of attention	To conduct further research on larger data sets	The collected news headlines	Average weight for each types of words
[20]	Twitter	A new CNN model	English	The features consider occurrence, written/spoken differences, sentiment, vagueness, degree of intensity, synonymousity and structure	Conducting experiments in different scenarios	Data collected by author from twitter	Precision 0.88, Recall 0.88
[11]	Short text (16000 Onet-Liner, Pun of Day, PPT Jokes, and English-Hindi)	A new CNN model	English	Producing embeddings for sentences	Creating high-quality systems for humor detection	Data collected by author from short text	Accuracy 98.2%, F1 98.2%
[21]	News headlines	A new CNN model	English	Exploring the structure of humor sentence	Getting more performance for humor detection	15000 news headlines	Accuracy 0.84
[14]	Jokes and news	A new CNN model	English, Chinese	The model of humor detection can be used for marketing purposes	Developing to detect fake news	Dataset from ZOL and Tsinghua NLP lab 463314 English and 111614 Chinese sentences	Accuracy 96.73%
[23]	Dialogues	A new CNN model	English	Developing model capable in understanding humor in general	Improving the dialog context modeling	Dialogues from TV sitcom: "The Big Bang Theory"	F1-score of 68.5%
[25]	Tweets	A new CNN model	English	Enhanced accuracy of humor detection	Coding mixed data, audio, images, videos etc	Data-set from kaggle	Accuracy 0.865
[26]	Weibo	HEMOS model	Chinese	To analyse the importance of recognising the effects of humour, icons and slang in the social media affective processing task	Create a high-quality sentiment recogniser for a broad range of sentiments in the Chinese language	Data-set from internet	Optimistic humorous: F1-score 66.18%, Recall 60.61%, Precision 72.72%, Pessimistic humorous: Precision 39.29%, Recall 61.11%, F1-score 47.83
[27]	Weibo	A new CNN model	English	Demonstrating contribution of incongruity and subjectivity for humor detection	Exploring more perfromance for humor detection	Product-related humorous questions from Weibo	Accuracy 90.8%
[15]	Weibsite	IEANN	English	To identify and recognise the humour that is implied in the text	Investigate more powerful methods for displaying incongruity and ambiguity properties.	Pun of the Day, 16000 One Liners	Accuracy 89.55%, Precision 90.12%, Recall 92.36%, F1 91.23%
[28]	None	A new CNN model	English	Understanding the meaning of humor	Emoji/emoticons and other new textual modalities used to express emotions must be explored for this task.	emEval 2017 HashTagWars and yelp review dataset	emEval 2017, Accuracy 0.751, yelp review, Accuracy 0.82
[29]	Video utterance	A new CNN model	English	Integrating previous context and external knowledge to understand the gist of a multimodal humorous expression.	Multimodal fusion	UR-FUNNY and MUStaRD datasets	Accuracy: UR-FUNNY 77.36% and MUStaRD 79.4%
[31]	MuSe 2022	A new CNN model	English	Detecting humor and calculate AUC	Improving the performance of the model	MuSe 2022 dataset	AUC 0.8972
[32]	None	A new CNN model	English	Understanding mutimodal laguage used in expressing humor	Rearch of Multimodal (context and punchline) for humor	UR-FUNNY	Accuracy 82.5%
[33]	Websties, twitter, the Kaggle Short Jokes	MLSN	English	Recognising and understanding humorous expressions	Capturing more semantics from sample and expression of more complex lexical items	Pun-of-the-day, 200K-One liners, and SemEval-2021 Task	1. Accuracy 0.994, Precision 0.966, Recall 0.989, F1 0.994 2. Accuracy 0.986, Presion 0.988, Recall 0.984, F1 0.986 3. Accuracy 0.954, Precision 0.945, Recall 0.979, F1 0.963
[34]	MuSe 2022	A new CNN model	English, German	To obtain more performance for challenge	Significant improvements over the reported baseline results may be achieved by more refined methods of combining different modalities and features.	Passau-SFCH, Hume-Reaction and Ulm-TSST dataset	AUC of .8480 for MuSe-Humor; 0.2801 mean (from 7-classes) ð□□□ for Pearson and 0.4761 ð□□□ for MuSe-Stress
[37]	None	PACGA	English	Learn how to use phonetic and semantic representations to recognise humour	Take a deeper dive into how to humour features into a deep learning model	Pun of the Day, 16000 One Liners	Accuracy 88.69%, Presion 88.94%, Recall 92.76%, F1 90.81%
[38]	MuSe 2022	A new CNN model	English, German	To effectively represent both the inner semantics and the cross-modal dependencies of the involved modalities	Perform real-time sentiment analysis on video by embedding advanced pre-trained methods into the system via knowledge distillation	Passau-SFCH dataset	AUC of 0.9065
[39]	SemEval-2021 Task-7	A new CNN model	English	Significant margin of improvement over base model performance	/	SemEval-2021 Task-7 dataset	The first subtask: F1 scores of 0.90 and the third subtask: 0.53. RMSE of 0.57 and 0.58 for the second and fourth subtask
[40]	None	BBFN	English	Extraction and integration of key information from different modalities	Final output further improved	CMU-MOSI, CMUMOSEI, and UR-FUNNY	CMU-MOS, Accuracy 45.0%, F1 84.3%, MAE 0.776; CMU-MOSEI Accuracy 86.2%, F1 86.1%, MAE 0.529, UR-FUNNY, Accuracy 71.68%
[41]	Twitter	A new CNN model	English	To help with the task of calculating and analysing humor	Expanding the dataset for humour recognition,etc.	A new dataset was created	Accuracy 91.2%, Precision 92.5%, Recall 90%
[22]	SemEval-2020 Task 7	A new CNN model	English	To obtain the best framework for humor detection	Developing more understanding for huomor detection	SemEval-2020 Task 7 dataset	Accuracy 0.84
[42]	None	A new CNN model	English, Chinese	To discover the structures of humour, to recognise humour and even to create humour	Comparison with human humour recognition more rigorous, etc.	Pun of the Day, 16000 OneLiners, Short Jokes dataset and PTT jokes	Pun of the Day Accuracy 0.897, Precision 0.899; 16000 One Liners Accuracy 0.894, Precision 0.889; Shor Jokes Accuracy 0.906, Precision 0.902; PTT jokes Accuracy 0.957 Precision 0.927
[30]	None	A new CNN model	English	Recognition humor	Further discovering the characteristics of humour and applying findings to the	Pun of the Day, 16000 One Liners	Pun of the Day: Accuracy 0.854, Precision 0.834, Recall 0.888, F1 0.859 16000 One

					process of humour generation		Liners: Accuracy 0.797, Precision 0.776, Recall 0.836, F1 0.805
[16]	None	A new CNN model	English, Chinese	To find humor and recognize humor	Finding more performance model	Pun of the Day, 16000 One Liners	New model: Accuracy 0.897, Recall 0.903 (16000 One-Liners)

5. Datasets and languages

In the literature we surveyed, there are two types of dataset sources for humor recognition, the first in which is data collected on its own data to the requirements of the task, and the second using public datasets in Table 2. The first case contain 10 papers, which are [15,17,20,21,24,25,26,27,41,42]. Some of this self-collected data comes from Twitter, some from other sites. The remaining papers use the detail of public datasets. The most frequently used public datasets are Passau-SFCH, UR-FUNNY, Pun of the Day and the 16000 One-liner.

Table 2. Summary of dataset from surveyed articles

Datasets	Source	Reference	Explanation
HAHA	HAHA task of Iberian Languages Evaluation Forum 2019	[13]	Training set 24000 tweets and test set 6000 tweets
GAMEEMO	Publicly available EEG database [43]	[14]	14 subjects (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4)
Passau-SFCH	MuSe 2022 challenge	[16], [31], [34], [38]	11 hours of video material from 10 different coaches
Memotion Analysis 1.0	SemEval 2020	[18]	6992 samples, 5 clasesses
News headlines	Social media site Reddit	[19]	287,076 headlines
Humicroedit	https://competitions.codalab.org/competitions/20970	[22]	15,000 news headlines
The Big Bang Theory	https://bigbangtrans.wordpress.com	[23]	35865 utterances
Pun of the Day and the 16000 One-liner	Pun of the Day :punoftheDay.com website the 16,0 0 0 One-liner was constructed by Mihalcea and Strapparava [43]	[30], [33], [35], [37], [42]	2423 humorous sentences; 16000 humorous sentences
HashTagWars dataset and yelp review dataset	HashTagWars: SemEval 2017-Task 6 Yelp review: Challenge dataset	[28]	HashTagWars: Training data consists of 101 files of total 11325 tweets and test data consists of 6 files of total 749 tweets. yelp review dataset: 1.6 million reviews by 366,000 users for 61,000 businesses
UR-FUNNY	UR-FUNNY: from TED talk videos;	[32], [40]	UR-FUNNY:5K humor and 5K non-humor instances;
MUStaRD	MUStaRD: from popular TV shows like Friends, The Big Bang Theory, The golden Girls and Sarcasmaholics [44]	[32]	MUStaRD: 690 video segments
SemEval-2021 Task 7	SemEval-2021	[33], [39]	10,000 texts from Twitter and the Kaggle Short Jokes dataset
Hume-Reaction	MuSe-Reaction sub-challenge	[34]	70 hours of audio and video data, 2222 subjects
Ulm-TSST	The Emotional Stress Sub-Challenge (MuSe-Stress)	[34]	Following the Trier Social Stress Test (TSST); Including four biological signals, EDA, Electrocardiogram, Respiration and heart rate

As can be seen from Figure 1, in total, there are 29 articles. Of these, 22 articles, or 76 percent, used English as data language of humor detection and then English +Chinese. As far as we know, other languages such as Italian, Arabic, Turkish and others are still unexplored, which opens new avenues for researchers to explore these languages as future research by creating new methods and necessary means for these languages such as humor corpus.

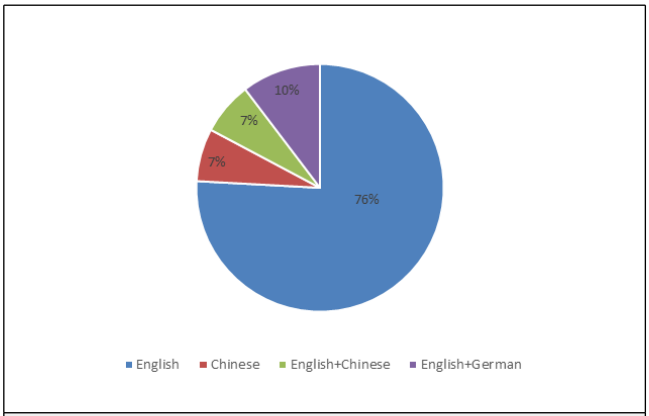

Figure 1. Statistics on the language used in the surveyed articles

6. Trick

6.1 Pre-processing

The first step in a natural language processing task is text (data) preparation or text (data) preprocessing, which aims to break down a document into words so that it can be understood by a computer programmer. The preprocessing step generally consists of: 1) removal of unwanted formatting (e.g., HTML tags), 2) sentence segmentation which splits the document into sentences, 3) word splitting which splits sentences into words, 4) word normalization which converts words into a canonical form, and 5) de-duplication which removes unwanted words. The pipeline outputs lower-case stemmed word sequences, after removing stop words and emoticons (if any) from the data. The specifics in the literature we surveyed are as follows. The stop-words are selected from the NLTK stop-words list [39]. Preprocessing is necessary because neural networks cannot take strings directly as input. Sentences must therefore be converted into numerical representations. The specific steps include word tokenization, lower case, remove stop words, creation of index dictionary, padding [24]. The author use multi-modal fusion technique to extract and integrate semantic information for sentiment analysis. Text, visual, acoustic modality are input stimously in CNN. The input signals in experiments were word-aligned by P2FA to align visual and acoustic signals to the same resolution of text [40]. The author used deep learning-based fine-grained to detect sentiment analysis for social media. The data consist of pictograms and slang. 109 Weibo emojis were converted into textual features and created a Chinese emoji lexicon. In experiment, images and videos were considerated as noise. Chinese slang and emoji lexicon added into Jieba to segment text [26].

6.2 Augmentation

In NLP, the type of data augmentation can be summarized as augmentation in both directions of feature space or data space, which can be divided into four levels, namely feature level, word level, phrase level and document level.

Augmentation methods for feature level in data space: 1) It is the process of adding artificial or natural noise to the original features, for example, changing a car to char. It is also the process of completely random or replacing a character with a neighboring character. 2) Rule modification, rule-based transformations via regular expressions, such transformations are very language dependent. Conversion appearances include misspellings, data changes, injection of entity names and abbreviations, etc.

Data space word level. The focus is on word enhancement operations, using synonym replacement, i.e., replacing words in the original sentence with similar words. Another is embedded replacement enhancement, which searches for words that fit the context of the text as well as possible and do not change the underlying content of the text. To achieve this, words in instances are translated into a potential representation space where words in similar contexts are closer together. The advantage of this data enhancement technique over synonym replacement is that the technique based on distributional assumptions is more general and takes into account the context of the text. Currently the form of more popular is the replacement with language models.

Phrase level in the data space. The structure-based transformation approach can utilise certain features or components of the structure to produce modified text. The other is in numerical analysis, where new data points are constructed from existing points. Generative methods use linguistic generative modelling to create diverse texts, which can contain some new information, and noise, etc.

Data augmentation in feature space is similar to that in data space, and can also incorporate modalities such as noise and interpolation. Of all the articles we surveyed, only one gave an enhancement algorithm. Chen et al. For the first time, mix-up is proposed to be used as a data enhancement for multimodal humor recognition, using linear interpolation to extend multimodal data features [38].

6.3 Attention mechanism

The study of attentional mechanisms has a long history of development, with the main ideas dating back as far as the 1990s of the last centuries. Attention Mechanism is a broad concept that refers to the selective attention and processing of information of different levels of importance by, for example, a human or a machine. Attention mechanisms have different types, roles and applications. In terms of the mechanistic mechanisms by which attention mechanisms work, attention methods are summarized in two broad categories, selective attention mechanisms and self-attentive mechanisms. Of the articles we surveyed, 8 papers dealt with attention mechanisms, as detailed in Table 3.

Table 3. Statistics on the use of attention mechanisms in the surveyed articles

Reference	Attention mechanism	Purpose of using attention mechanism
[18]	iTRM (Inter-Task Relations) and iCRM (Inter-Class Relations) modules	iTRM's primary motivation is to learn about the interrelationship between tasks and how they support one another. Conversely, iCRM develops relationships between different task classes
[26]	Attention-based bi-directional long short-term memory RNN	To improve performance of the model
[35]	Internal and external attention layer	To catch the contradiction and ambiguity in humorous text
[32]	Delta-memory Attention	To capture more word representation
[33]	The new attention Layer	To receive the attention control message of humorous sentences depending on the chosen ambiguity rating of each word
[37]	Gated attention network	To acquire phonetic patterns and semantic mapping for humor identification
[38]	The scaled dot-product attention	To represent both the inherent semantics as well as the cross-modality relations of the different modalities involved
[40]	Gated control mechanism	To improve output of the humor detection

6.4. Multimodal learning

Multimodal learning describes the building of models that allow machines to learn information from multiple modalities and to communicate and transform information from each modality□Multimodal contains a representation of the data information of multiple modalities, which are vectors in a semantic space shared by multiple modalities. A good multimodal representation should have the properties of smoothness, temporal and spatial coherence, sparsity and natural clustering. The multimodal representations corresponding to the different multimodal inputs must reflect the similarity of the information contained in each multimodal input. Multimodal representations can still be generated when some modal data information is missing. According to the multimodal representation it is possible to obtain information about the data of each modality. A total of seven articles in this survey used the multimodal technique in Table 4.

Table 4. Statistics on the use of multimodality in the surveyed articles

Reference	The form of multimodal
[15]	Text and voice
[16]	Image and voice
[18]	Text and image
[31]	Text and image
[32]	Text, audio and video
[38]	Text, video and acoustic
[40]	Text, video and acoustic

7. Evaluation metrics

Among all these studies, the results of deep learning models for humor detection were measured across multiple evaluation methods, guiding researchers in revising their models for optimal performance. The most frequently methods mentioned include F-score, precision, recall, accuracy, while root-mean-squared error was also used for regression tasks.

The F-score, often referred to as the F1-score (F1), is a combined metric that balances precision and recall. It is commonly used in binary and multiclass classification scenarios to balance the trade-off between precision and recall, providing a single performance metric. It is defined as follows:

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (1)$$

Generally speaking, the F-score ranges from 0 to 1, with 1 representing perfect precision and recall, while 0 indicates poor performance.

Precision is a metric commonly employed in binary and multiclass classification tasks to gauge the accuracy of positive predictions. It quantifies the proportion of true positive samples (correctly classified positive instances) out of all predicted positive instances. Formally, precision is defined as follows:

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

where TP represents the number of true positives (correctly classified positive instances) and FP represents the number of false positives (incorrectly classified positive instances). Generally, high precision values signify that the model is adept at making positive predictions and minimizing the number of false positives.

Recall, also known as sensitivity or true positive rate, is a critical metric used in classification tasks to assess the model's capability to correctly identify positive samples from the entire set of actual positive instances. It is calculated as the ratio of true positive predictions to the sum of true positive and false negative predictions (all actual positive samples). A high recall value indicates that the model can effectively capture a large proportion of the actual positive instances, minimizing the number of false negatives.

As one of the most frequently used evaluation method, accuracy

is fundamentally used in classification tasks to measure the overall correctness of the model's predictions. It is calculated as the ratio of the total number of correct predictions to the total number of predictions (both true positives and true negatives):

$$\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / (\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}) \quad (3)$$

While accuracy is widely used and intuitive, it somehow may be misleading in imbalanced datasets.

Root Mean Squared Error (RMSE) is a regression evaluation metric used to measure the difference between predicted values and the actual values in a continuous dataset. It calculates the square root of the average of the squared differences between predicted and actual values.

$$\text{RMSE} = \sqrt{\sum (\text{predicted-actual})^2 / 2} \quad (4)$$

Lower RMSE values indicate better performance, as it means the model's predictions are closer to the ground true values.

8. Future work

Building on the discussion and research presented in the previous sections, we have outlined many directions for future research and identified the unresolved issues. In Figure 2, the most important future areas of research and challenges in humor detection were presented.

Unexplored data domain, dataset problem, language problem, hybrid method, Research of multimodal, improving the performance of humor detection, mining useful feature for humor detection, knowledge distillation of humor detection model.

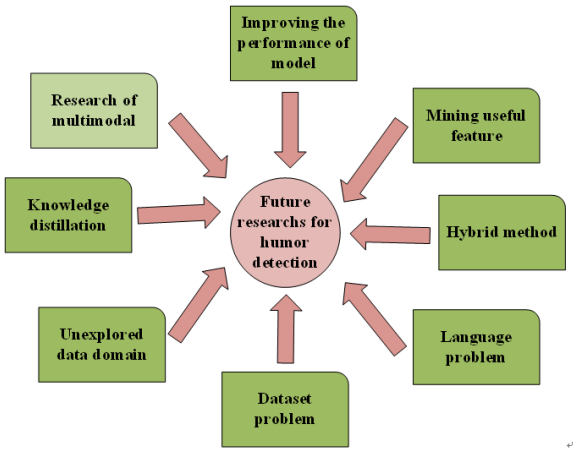

Figure 2. Future research directions for humor detection

As can be seen from Figure 2, there are still many fields in need of more detailed investigation and development. For instance, for humor detection and evaluation, there are few standard datasets available. Those datasets are small. Humor detection is aided by public standards and large datasets. Another problem that emerged from the survey carried out is that almost all of the are in English. Others languages are not yet unexplored in humor detection. One potential future trend is to hybridize several existing methods to overcome the drawbacks of each individual of each individual technique.

An unsolved problem is to explore the implicit and useful feature of humor detection. The use of attentional mechanisms is now a common means of detecting significant features. However, it needs a large amount of data and high computational cost. Perhaps there are more efficient, usable, and less costly methods to be mined to solve feature extraction for humor detection.

In order to enhance the precision of the model, we are used to using sophisticated models (deep network hierarchy, large number of parameters), and even selecting models that are integrated with multiple models, which leads to the need for a large number of computational resources and a huge dataset to support this "big" model. However, when deploying the service, it becomes clear that this "big" model is slow to reason and consumes a lot of memory. So distillation could be another direction to be explored for humor detection.

9. Conclusion

This is the first paper reviewing humor detection to our knowledge. There are a number of significant insights from this survey. Studies on humor detection were carried on almost for English language. However, other languages, such as France and Malay, have not yet been explored and should be taken into consideration for future studies. The survey also shows that the almost works in humor detection based on deep learning use attention mechanism and multimodal technique. With the emergence of large models, models for humor detection have been newly inspired, and their recognition performance has greatly improved. But there's very little mention of pre-processing. In addition, most of the data collected by the authors themselves come from Twitter and did not publish publicly. Use of privately generated data will result in biased performance which would make it difficult to compare new research with the benchmark research. There are many other areas of application that have yet to be explored, such as Weibo, YouTube.

Over the past 10 years, remarkable progress has been made in humor recognition. this article surveyed 29 papers and has addressed some significant topics required of the research landscape in humor detection. A detailed survey of the literatures have been carried out with an emphasis on humor detection, e.g., approach based deep learning, techniques (pre-processing, attention mechanism and multimodal), analysis of dataset, definition of problem, humor studies in linguistics. It also discusses, but is not limited to, some promising future directions at final section of the article.

Acknowledgments

This work is supported by SICHUAN INTERNATIONAL STUDIES UNIVERSITY 2023 Planning Project (sisu202306).

References

[1] Ramakristanaiah C., Namratha P., Ganiya R.K., Reddy M.R. A survey on humor detection methods in communications. 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, 668-674, 2021.

[2] Antony K., Panagiotis A. Computational humor recognition: A Systematic Literature Review. Research Square, 2023.

[3] Attardo S. Humorous texts: A semantic and pragmatic analysis. Series Humor Research, vol. 6, Walter de Gruyter, 2001.

[4] Attardo S. Irony as relevant inappropriateness. Journal of Pragmatics, 32(6):793-826. 2000.

[5] Shahe K. Humour styles, personality, and well-being among Lebanese university students. European Journal of Personality, 18(3):209-219, 2004.

[6] Sonja U. The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. Computers in Human Behavior, 45:1-10, 2015.

[7] Tajfel H., Turner J.C., Austin W.G., Worchel S. An integrative theory of intergroup conflict. Organizational Identity: A reader, 56(65):9780203505984-16, 1979.

[8] Martin R.A., Ford T. The psychology of humor: An integrative approach. Academic Press, 2018.

[9] Giora R. Understanding figurative and literal language: The graded salience hypothesis. Cognitive Linguistics, 8(3):183-206, 1997.

[10] Gibbs R.W. The poetics of mind: Figurative thought, language, and understanding. Cambridge University Press, 1994.

[11] Clark H.H. Using language. Cambridge University Press, 1996.

[12] Purcell D., Brown M.S., Gokmen M. Achmed the dead terrorist and humor in popular geopolitics. GeoJournal 75:373–385, 2010.

[13] Mao J., Liu W. A BERT-based approach for automatic humor detecting and scoring. In IberLEF@ SEPLN, 197-202, 2019.

[14] Heaton J., Givigi S. A deep CNN system for classification of emotions using EEG signals. 2022 IEEE International Systems Conference (SysCon), Montreal, QC, Canada, 1-7, 2022.

[15] Xie J., Tang M., Xiong J. A multimodal Chinese humor classification algorithm based on interactive attention and text and speech fusion. Available at SSRN 4241914, 2022.

[16] Kathan A., Amiriparian S., Christ L., Triantafyllopoulos A., Müller N., König A., Schuller B.W. A personalised approach to audiovisual humour recognition and its individual-level fairness. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge, 29-36, 2022.

[17] Godoy F.C.D. Advancing humor-focused sentiment analysis through improved contextualized. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), 2020.

[18] Chauhan D.S., R D.S., Ekbal A., Bhattacharyya P. All-in-One: A deep attentive multi-task learning framework for humour, sarcasm, offensive, motivation, and sentiment on memes. Proceedings ofthe 1st Conference ofthe Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, 281-290, 2020.

[19] Xu Z., Xie Y. Attention method analysis in sentiment analysis for humor level evaluation. 2022 14th International Conference on Computer Research and Development (ICCRD), 178-185, 2022.

[20] Barbieri F., Saggion H. Automatic detection of irony and humour in twitter. ICCC, 2014.

[21] Annamoradnejad I. ColBERT using BERT sentence embedding for humor detection. arXiv preprint arXiv:2004.12765 1.3, 2021.

[22] Wang C., Xin S., Yi M. Comparative study on deep learning models in humor detection. Proceedings Volume 11933, 2021 International Conference on Neural Networks, Information and Communication Engineering, 2021.

[23] Dario B., Fung P. Deep learning of audio and language features for humor prediction. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 496–501, 2016.

[24] Chen R., Rau P.-L.P. Deep learning model for humor recognition of different cultures. In Cross-Cultural Design. Experience and Product Design Across Cultures, 13th International Conference, CCD 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I 23 Springer International Publishing, 373-389, 2021.

[25] Prajapati P., Jaiswal A., Aastha, Shilpi, Neha, Sachdeva N. Empirical analysis of humor detection using deep learning and machine learning on kaggle corpus. AIR 2022: International Conference on Advancements in Interdisciplinary Research, 300-312, 2022.

[26] Li D., Rzepka R., Ptaszynski M., Araki K. HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. Information Processing & Management, 57(6), 102290, 2020.

[27] Ziser Y., Kravi E., Carmel D. Humor detection in product question answering systems. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 519-528, 2020.

[28] Mahajan R., Zaveri M. Humor identification using affect based content in target text. Journal of Intelligent & Fuzzy Systems, 39(1) 697-708, 2020.

[29] Hasan M.K., Lee S., Rahman W., Zadeh A. Humor knowledge enriched transformer for understanding multimodal humor. In Proceedings of the AAAI Conference on Artificial Intelligence, 12972-12980, 2021.

[30] Yang D., Lavie A., Dyer C., Hovy E. Humor recognition and humor anchor extraction. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2367-2376, 2015.

[31] Xu H., Liu W., Liu J., Li M., Feng Y., Peng Y., Shi Y., Sun X., Wang M. Hybrid multimodal fusion for humor detection. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge, 15-21, 2022.

[32] Hasan M.K., Rahman W., Zadeh A., Zhong J. UR-FUNNY A multimodal language dataset for understanding humor. arXiv preprint arXiv:1904.06618, 2019.

[33] Xiong S., Wang R., Huang X., Chen Z. Multidimensional latent semantic networks for text humor recognition. Sensors (Basel), 22(15), 5509, 2022.

[34] Christ L., Amiriparian S., Baird A., Tzirakis P., Kathan A., Müller N., Stappen L., Meßner E.-M., König A., Cowen A., Cambria E., Schuller B.W. The MuSe 2022 multimodal sentiment analysis challenge. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge, 5-14, 2022.

[35] Fan X., Lin H., Yang L., Diao Y., Shen C., Chu Y., Zou Y. Humor detection via an internal and external neural network. Neurocomputing, 394:105-111, 2020.

[36] Peng-Yu C., Von-Wun S. Humor recognition using deep learning. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 2, 113–117, 2018.

[37] Fan X., Lin H., Yang L., Diao Y., Shen C., Chu Y., Zhang T. Phonetics and ambiguity comprehension gated attention network for humor recognition. Complexity, 2020:1-9, 2020.

[38] Chen C., Zhang P. Integrating cross-modal interactions via latent representation shift for multi-modal humor detection. Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge, 23-28, 2022.

[39] Mondal A., Sharma R. Team KGP at SemEval-2021 Task 7: A deep neural system to detect. Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), 1169–1174, 2021.

[40] Wei H., Hui C., Alexander G., Amir Z. Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis. Proceedings of the 2021 International Conference on Multimodal Interaction, 6–15, 2021.

[41] Arunima J., Monika, Mathur A., Prachi, Sheena Ma. Automatic humour detection in tweets using soft computing paradigms. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), Faridabad, India, 172-176, 2019.

[42] Chen P.-Y., Soo V.-W. Humor recognition using deep learning. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 113-117, 2018.

[43] Burak A.T., Murat G., Ibrahim T. Database for an emotion recognition system based on EEG signals and various computer games–GAMEEMO. Biomedical Signal Processing Control, 60, 101951, 2020.

[44] Santiago C., Devamanyu H., Verónica P.-R., Roger Z., Rada M., Poria S. Towards multimodal sarcasm detection. arXiv preprint arXiv:.01815, 2019.