"I Know Who You Are": Character-Based Features for Conversational Humor Recognition in Chinese

Wenbo Shang[1,*], **Jiangjiang Zhao**[2,*], **Zezhong Wang**[3,4], **Binyang Li**[1,✉], **Fangchun Yang**[2,✉], **Kam-fai Wong**[3,4]
[1]University of International Relations, China
[2]Beijing University of Posts and Telecommunications, China
[3]The Chinese University of Hong Kong, Hong Kong, China
[4] MoE Key Laboratory of High Confidence Software Technologies, China
[1]{wbshang, byli}@uir.edu.cn [2]{zjjbupt, fcyang}@bupt.edu.cn [3]{zzwang, kfwong}@se.cuhk.edu.hk

Abstract

Humor plays an important role in our daily life, as it is an essential and fascinating element in the communication between persons. Therefore, how to recognize punchlines from the dialogue, i.e. conversational humor recognition, has attracted much interest of computational linguistics communities. However, most existing work attempted to understand the conversational humor by analyzing the contextual information of the dialogue, but neglected the character of the interlocutor, such as age, gender, occupation, and so on. For instance, the same utterance could bring out humorous from a serious person, but may be a plain expression from a naive person. To this end, this paper proposes a **C**haracter **F**usion **C**onversational **H**umor **R**ecognition model (CFCHR) to explore character information to recognize conversational humor. CFCHR utilizes a multi-task learning framework that unifies two highly pertinent tasks, i.e., character extraction and punchline identification. Based on deep neural networks, we trained both tasks jointly by sharing weight to extract the common and task-invariant features while each task could still learn its task-specific features. Experiments were conducted on Chinese sitcoms corpus, which consisted of 12,677 utterances from 22 characters. The experimental results demonstrated that CFCHR could achieve 33.08% improvements in terms of F1-score over some strong baselines, and proved the effectiveness of the character information to identify the punchlines.

1 Introduction

Humor recognition is an important task of humor computation, which can not only enable machines to recognize humor, but also lay an important foundation for humor generation. According to the form of humorous text, humor recognition can be generally divided into two types: one-liners humor recognition and conversational humor recognition (CHR). As shown in Table 1, one-liners humor focusing on one single sentence or passage, e.g. joke, while conversational humor is brought out based on the dialogue, which can be widely applied to a variety of scenarios, including chat robots, machine translation, etc. So, to recognize or even understand conversational humor is very significant for both academical and industrial fields (Lin et al., 2016).

One-liners humor	
I failed math so many times at school, I can't even count.	
Conversational humor	
Character	Utterance
志国(Zhiguo) middle age, male	唉，你们不能剩我一人儿啊。 Oh, you can't leave me alone.
圆圆(Yuanyuan) child, female	不剩您一人，还有小芳大妈 You are not alone. Aunt Xiaofang is with you.
志国(Zhiguo)	那还不如剩我一人儿呢 **I'd rather be left alone.**
傅明(Fuming) old man, male	都是你自己惹的麻烦，不剩你剩谁 It's all your own troubles, who else is left without you?

Table 1: Examples of one-liners humor and conversation humor. The sentence in bold is a punchline, while the rest is set-up.

The objective of this paper is to recognize humor from dialogue. Generally speaking, a dialogue is formed by a serial of utterances, and the utterances can be divided into set-up and punchlines. The punchline is the part of the sentences taking the role of laughing, while the rest of utterances are set-ups (Taylor and Mazlack, 2005; Attardo and Raskin, 1991). The conversational humor recognition can be considered as identifying whether an utterance

[*]These authors contributed equally to this work.
✉Corresponding authors

Findings of the Association for Computational Linguistics: EMNLP 2022, pages 2927–2932
December 7-11, 2022

is a punchline, such as **"I'd rather be left alone"** as shown in Table 1.

Different from one-liners humor, the character is one of the most important influencing factors in conversational humor according to the theory of **General Verbal Theory of Humor** (1991). Note that the character refers to the sitcom role in this paper. However, existing studies for conversational humor recognition mainly considered it as a punchline recognition task. Most methods did not distinguish conversational humor from one-liners humor, and focused on the context representation to capture the humorous semantics. To the contrary, character information in the dialogue was neglected, such as, speaking style, gender, age, etc., and these information has some specific features, that are inherently funnier, and more likely to make people laugh. As a result, existing methods performed sub-optimal in conversational humor recognition.

To solve the above issues, this paper attempts to explore character information to facilitate the machines to understand humor in the dialogue. A Character Fusion Conversational Humor Recognition (CFCHR) model is presented by integrating the character information and the contextual information to represent the conversational humorous information.

To capture the contextual information, an utterance embedding is derived from each utterance; to capture the character information, character features and the attributes are learned from utterances and predefined attributes respectively. Then, the utterance embedding and the character embedding are fed into the a multi-task learning framework (Hastie et al., 2009). In this way, CFCHR can capture both of the contextual information and the character information for character extraction and punchline recognition.

We conducted experiments on Chinese sitcoms corpus, which consisted of 12,677 utterances from 22 characters. CFCHR was effective in character extraction task. Compared with some strong baselines, CFCHR could achieve the best performance of punchline recognition, i.e. 51.5%, in terms of F1-score. Moreover, 33.08% improvements were achieved in terms of F1-score over the baseline model, and it proved the effectiveness of the character information in identifying the punchlines.

2 Related Work

Humor recognition is usually formulated as a binary classification problem. Most of the existing work on humor recognition mainly focus on one-liners, and a classifier is usually trained based on the whole texts to predict whether it is humorous or not (Cattle and Ma, 2018; Liu et al., 2018a; Xie et al., 2021; Zhou et al., 2020; Zou and Lu, 2019; Liu et al., 2018b).

In recent years, much research work has a growing interest in conversational humor recognition. Early work on conversational humor recognition attempted to use LSTM, RNN, GRU models (Bertero and Fung, 2016; Ramakrishna et al., 2018) to predict punchlines in dialogues. Due to the lack of corpus for conversational humor recognition, there were some studies (Pamulapati and Mamidi, 2021; Blinov et al., 2019; Pamulapati et al., 2020) later focused on constructing humorous dialogue corpus based on sitcoms. However, none of the above work accounted for the character information, which was a crucial source of information for conversational humor recognition. This paper will explore the character information and integrate multi-faceted information of character into the humorous semantics representation.

3 Task Formulation

Given a dialogue as the input text, and it is consisted of a sequence of M utterances. For an utterance s_i, it is consisted of N words, and denoted by $s_i = \{w_{i,1}, w_{i,2}, \ldots, w_{i,j}, \ldots, w_{i,N}\}$, where $w_{i,j}$ is the j^{th} word in the utterance s_i. Note that each utterance in the dialogue corresponds to an interlocutor with the character features denoted as $Role^V$ and the character attributes denoted as $Role^D$. We detail both of these in Section 4.1. The objective of this paper is to identify if an utterance is a punchline by exploring the character information.

4 CFCHR

In this paper, we propose **C**haracter **F**usion **C**onversational **H**umor **R**ecognition (CFCHR) for punchline recognition from the dialogue. CFCHR consists of three sub-components: character extraction, punchline recognition, and multi-task learning framework. Character extraction module is firstly designed to obtain the character information representation; punchline recognition module is then followed to obtain contextual information representation; and a multi-task learning framework

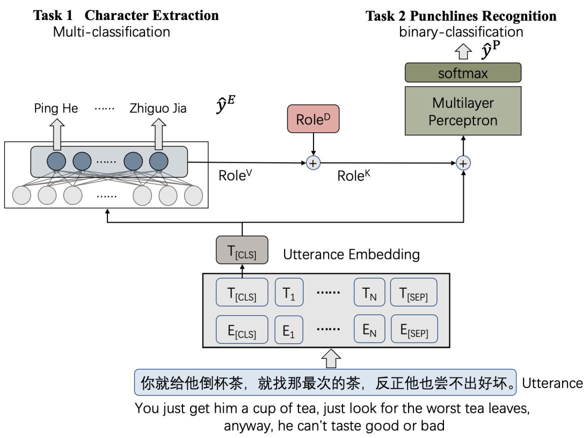

Figure 1: The overall architecture of CFCHR. CFCHR is designed for a multi-task. Task 1 is character extraction to identify each utterance from which character. Task 2 is punchline recognition. Multi-task learning is to introduce the character information in the hidden layer in the process of recognizing punchlines.

is finally employed for character extraction and punchline identification.

Given a dialogue M, input each utterance s_i and its corresponding character information $(Role^V, Role^D)$ into the model. For each utterance, we firstly derive an utterance embedding by the contextual sentence to capture the semantics in the text. To obtain the character features $Role^V$, CFCHR will classify the interlocutor into the predefined category of the character based on each utterance. The character information $Role^K$ can be represented by the character features $Role^V$ together with the character attributes $Role^D$. Then, both of the utterance embedding and the character embedding are fed into the a multi-task learning framework (Hastie et al., 2009). In this way, CFCHR can capture both of the contextual semantics and the character information for character extraction and punchline recognition.

4.1 Character extraction

In the scenario of a dialogue, the character of a interlocutor is essential for creating humor, which contains rich information, including the character attributes, such as gender, etc., and the character features, such as speaking style, personality, etc., which are learned from utterances. Therefore, we represent the character information $Role^K$ by two parts, character features $Role^V$ and character attributes $Role^D$. Hence, we propose to apply utterance embedding and character attributes constructed by human to derive character information representations. In CFCHR, BERT (Devlin et al., 2018) is utlized to derive utterance embedding without loss of generality.

For each utterance s_i, we can obtain the corresponding utterance embedding through the pretrained model, BERT. BERT deploys a multi-layer bidirectional encoder based on transformers with multi-head self-attention (Vaswani et al., 2017), which contains a special token [CLS]. [CLS] can be an embedding to represent the semantics of the whole utterance. Here, the d_H-dimensional utterance embedding is denoted by $T_{[CLS]}$.

Based on $T_{[CLS]}$, we apply multi-layer perceptron with one single hidden layer to identify characters as follows.

$$\hat{y}^E = \underset{e \in \{0,\dots,21\}}{argmax}\ \mathcal{F}(w(T_{[CLS]}) + b)_e \tag{1}$$

$\hat{y}^E$ denotes the prediction results of character classification. e denotes the index of each character class. 22 characters are predefined in total. Since many characters in the sitcom have only appeared once or twice, such as passers-by, couriers and so on, we group the above characters into one category and the remaining 21 characters are the main characters. Thereby, most features of characters like speaking style and personality can be learned in this single layer, which can be extracted from weight. Here, we use $Role^V$ to denote the character features, and use $R^{d_C \times d_H}$ to denote the dimension.

$$Role^V = w, Role^V \in \mathbb{R}^{d_C \times d_H} \tag{2}$$

$\mathcal{F}(\cdot)$ is the activation function, which is set to Relu function (Hastie et al., 2009) in this paper. d_C is the number of character classes, 22. For each $i \in d_C$, $Role_i^V$ represents the i-th category of the characters. Moreover, we can further derive the joint character information $Role_i^K$ to indicate both character features and character attributes by concatenating two different vectors as follows:

$$Role_i^K = [Role_i^V; Role_i^D] \tag{3}$$

$Role_i^K$ denotes the joint character information of the i-th character.

4.2 Punchline recognition

For the task of punchline recognition, it is essential to capture the semantics of the utterance, and we adopt a classic representation(Bertero and Fung, 2016). Since character information is one of the most important factors for creating humor. Thus, we combine the character information embedding $Role_i^K$ and the utterance embedding to capture the overall semantics for punchline recognition as follows.

$$T_{[ALL]i} = Role_i^K \oplus T_{[CLS]i} \quad (4)$$

We deploy MLP (Hastie et al., 2009) to capture the overall semantics representation $T_{[ALL]}$ for recognizing punchlines as follows, .

$$\hat{y}^P = \underset{p\in\{0,1\}}{argmax}\, \mathcal{F}(T_{[ALL]})_p \quad (5)$$

$\hat{y}^P$ represents the prediction results of punchline recognition. p indicates whether it is the punchline.

4.3 Multi-task learning

In our work, CFCHR is designed for character extraction and punchline recognition based on the multi-task model structure (Caruana, 1997). We jointly train both tasks by sharing the weight to extract the common and task-invariant features while each task can still learn its task-specific features. The loss functions of two tasks are calculated separately, and in the same iteration, two gradients are accumulated.

5 Experiment

5.1 Experiment Settings

Dataset To evaluate the performance of our proposed method, experiments were conducted based on a publicly available dataset which was collected from a famous Chinese sitcom 我爱我家*(I Love My Family)*. In this dataset, the scripts of the sitcom was divided into several dialogues depending on the scene and the plot changing. Each dialogue consisted of several utterances, while each utterance was in accordance with a character of the sitcom. For each utterance, it was assigned a label, $punchline$ or $non-punchline$. The statistics of the dataset was shown in Table 2.

Parameter setting During the training processing, we used Adam optimizer (Kingma and Ba, 2015) with an initial learning rate of $1 \times 10-5$, and the batch size was set as 32 utterances. Besides, we also dynamically adjusted the learning rate via a linear function for each iteration of training. For other parameter settings, we followed the standard configuration.

Dataset of **I Love My Family**	
Dialogues	348
Utterances	12,677
Avg. length of dialogues	36.34
Avg. length of utterances	9.99
Punchline ratio (%)	28.76

Table 2: The statistics of dataset.

Model	**P**	**R**	**F1**
$\text{CFCHR}_{\text{Roberta}}$	0.833	0.799	0.775
$\text{CFCHR}_{\text{BERT}}$	**0.877**	**0.824**	**0.784**

Table 3: The results of character extraction task.

Evaluation Metrics We adopted the classic metrics in NLP, precision (P), recall(R), and F1-score(F1) (Ceri et al., 2013; Powers, 2020) to assess the performance.

5.2 Experimental Results

We compared CFCHR with five baseline models, including LSTM+Attention (Lin et al., 2017), Bi-LSTM (Graves and Schmidhuber, 2005), BC-LSTM (Mousa and Schuller, 2017), BERT (Devlin et al., 2018), Roberta (Liu et al., 2019).

Table 3 demonstrated the performance of the character extraction task. From the results, we could find that the representation of the character information could be learned by CFCHR.

Table 4 demonstrated the performance of the punchline recognition task based on the sitcom dataset. It could be seen that after adding character knowledge, the performance of the model was significantly improved. Compared to the baseline models, $\text{CFCHR}_{\text{BERT}}$ achieved the best performance, i.e. 51.5% in terms of F1-score, which has 33.08% improvements over Bert model. $\text{CFCHR}_{\text{Roberta}}$ achieved 30.23% improvements of

Model	**P**	**R**	**F1**
LSTM+attention	0.453	0.258	0.329
Bi-LSTM	-	-	0.326
BC-LSTM	-	-	0.358
Roberta	0.337	0.437	0.381
BERT	0.354	0.428	0.387
$\text{CFCHR}_{\text{Roberta}}$	0.613	**0.458**	0.504
$\text{CFCHR}_{\text{BERT}}$	**0.649**	0.453	**0.515**

Table 4: The results of punchline recognition task. $\text{CFCHR}_{\text{BERT}}$ denotes CFCHR based on BERT, and $\text{CFCHR}_{\text{Roberta}}$ denotes CFCHR based on Roberta.

F1-score over BERT and 32.28% improvements of F1-score over Roberta. It was proved that the character information was able to improve the model's ability of understanding conversational humor. Moreover, the performances in F1-score of BERT model were better than that of Roberta whether considering character information or not.

6 Conclusions

In this paper, we explore character information for conversational humor recognition, and present a multi-task learning framework for character extraction and punchline recognition. Experimental results demonstrated that character information was effective for punchline identification and could achieve the performance of 51.5% in terms of F1-score. Compared with some strong baselines, our proposed model could achieve 33.08% improvement on Chinese sitcom dataset.

7 Limitations

In our work, we explored the character information for punchline recognition, and pre-defined 22 characters on the specific sitcom. However, the characters' personalities are different in different sitcoms, so when transferring to new sitcoms, CFCHR needs to be retrained to learn the new characters' information. For the future work, CFCHR can be improved to learn coarse-grained common features of different clusters of characters.

Acknowledgements

This project was partially supported by National Natural Science Foundation of China (Grant number: 61976066), Beijing Natural Science Foundation (Grant number: 4212031), the Fundamental Research Fund for the Central Universities (Grant numbers:3262021T23), and Research Funds for NSD Construction, University of International Relations (Grant numbers: 2021GA07), Innovation & Technology Commission HKSAR (ITF Project No. PRP-054-21FX).

References

Salvatore Attardo and Victor Raskin. 1991. Script theory revis (it) ed: Joke similarity and joke representation model.

Dario Bertero and Pascale Fung. 2016. A long short-term memory framework for predicting humor in dialogues. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 130–135.

Vladislav Blinov, Valeria Bolotova-Baranova, and Pavel Braslavski. 2019. Large dataset and language model fun-tuning for humor recognition. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 4027–4032.

Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.

Andrew Cattle and Xiaojuan Ma. 2018. Recognizing humour using word associations and humour anchor extraction. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 1849–1858.

Stefano Ceri, Alessandro Bozzon, Marco Brambilla, Emanuele Della Valle, Piero Fraternali, and Silvia Quarteroni. 2013. An introduction to information retrieval. In *Web information retrieval*, pages 3–11. Springer.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6):602–610.

Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Hongfei Lin, Dongyu Zhang, Liang Yang, and Bo Xu. 2016. Computational humor researches and applications. *Journal of Shandong University (Natural Science)*, 51(7):1–10.

Zhouhan Lin, Minwei Feng, Cícero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. 2017. A structured self-attentive sentence embedding. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018a. Exploiting syntactic structures for humor recognition. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, pages 1875–1883.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018b. Modeling sentiment association in discourse for humor recognition. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 586–591.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Amr El-Desoky Mousa and Björn W. Schuller. 2017. Contextual bidirectional long short-term memory recurrent neural networklanguage models: A generative approach to sentiment analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*, pages 1023–1032.

Vaishnavi Pamulapati and Radhika Mamidi. 2021. Developing conversational data and detection of conversational humor in telugu. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 12–19.

Vaishnavi Pamulapati, Gayatri Purigilla, and Radhika Mamidi. 2020. A novel annotation schema for conversational humor: Capturing the cultural nuances in kanyasulkam. In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 34–47.

David M. W. Powers. 2020. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *CoRR*, abs/2010.16061.

Anil Ramakrishna, Timothy Greer, David Atkins, and Shrikanth Narayanan. 2018. Computational modeling of conversational humor in psychotherapy. In *Interspeech*, volume 2018, page 2344.

Julia Taylor and Lawrence Mazlack. 2005. Toward computational recognition of humorous intent. In *Proceedings of Cognitive Science Conference*, pages 2166–2171.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yubo Xie, Junze Li, and Pearl Pu. 2021. Uncertainty and surprisal jointly deliver the punchline: Exploiting incongruity-based features for humor recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 33–39.

Yichao Zhou, Jyun-Yu Jiang, Jieyu Zhao, Kai-Wei Chang, and Wei Wang. 2020. "the boating store had its best sail ever": Pronunciation-attentive contextualized pun recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 813–822.

Yanyan Zou and Wei Lu. 2019. Joint detection and location of english puns. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 2117–2123.