Supporting Comedy Writers: Predicting Audience's Response from Sketch Comedy and Crosstalk Scripts

Maolin Li
Independent Research*
maolin.li.cs@gmail.com

Abstract

Sketch comedy and crosstalk are two popular types of comedy. They can relieve people's stress and thus benefit their mental health, especially when performances and scripts are high-quality. However, writing a script is time-consuming and its quality is difficult to achieve. In order to minimise the time and effort needed for producing an excellent script, we explore ways of predicting the audience's response from the comedy scripts. For this task, we present a corpus of annotated scripts from popular television entertainment programmes in recent years. Annotations include a) text classification labels, indicating which actor's lines made the studio audience laugh; b) information extraction labels, i.e. the text spans that made the audience laughed immediately after the performers said them. The corpus will also be useful for dialogue systems and discourse analysis, since our annotations are based on entire scripts. In addition, we evaluate different baseline algorithms. Experimental results demonstrate that BERT models can achieve the best predictions among all the baseline methods. Furthermore, we conduct an error analysis and investigate predictions across scripts with different styles.[1]

1 Introduction

Comedy plays a major role in people's lives in that it relieves stress and anxiety (Williams et al., 2005; Sarıtaş et al., 2019). There are two popular types of comedy: sketch comedy and crosstalk. A sketch comedy usually presents a short story and is performed by multiple comedians in various short scenes; while in a crosstalk performance, which is similar to a talk show, there are usually two performers telling humorous stories behind a desk. Although these two types of comedy are different, both of them are performed based on scripts. A script breaks down a story into pieces along with the details that describe which performer should take what action or say which lines at a specific point (Blake, 2014). Therefore, the quality of the script is critical and it directly influences whether the audience enjoys the performance.

However, it is difficult for script writers to ensure a high-quality comedy script and be productive. Firstly, writers have to assess if audiences will react as expected, in particular laughing at specific points. It is necessary to rehearse multiple times to continuously improve the script, which is time-consuming and can be costly. Secondly, to develop laughter triggers, writers need to identify the potential points from the script where there are possibilities for performers to use funny body moves, tone or tell amusing stories to make the audience laugh. Thirdly, the more times a script is publicly performed, the less laughter it can bring, since the audience have become too familiar with it. Thus, it is essential for comedy writers to explore new laughter triggers constantly.

Since natural language processing (NLP) has been widely and successfully applied to a number of fields (Carrera-Ruvalcaba et al., 2019; Rao and McMahan, 2019), we investigate how NLP methods can support comedy writers to produce high-quality scripts more efficiently. This paper specifies this challenge as a new task, i.e. the prediction of the audience's response to sketch comedy and crosstalk scripts. To address this challenge, we explore the use of two different NLP methodologies: 1) Text Classification: we predict whether or

* The research was conducted during non-working time. The idea of this research was inspired by a discussion with my friend about an entertainment TV programme in which the comedians mentioned the difficulties of producing a high-quality script.

[1]The corpus and source code can be freely downloaded from https://github.com/creatememo/supporting-comedy-writers

Proceedings of the First Workshop on Computational Approaches to Discourse, pages 42–52
Online, November 20, 2020. ©2020 Association for Computational Linguistics
https://doi.org/10.18653/v1/P17

Label	Actor's Line	Source
1	宋小宝：我的人生格言是，在哪里跌倒，就在哪里睡一觉。 Xiaobao SONG: My life motto is to have a sleep where you've fallen.	碰瓷（*An Incident-Faking Extortionist*） Joyful Comedians (Season 1), 2015
1	张小斐：我可能是洗的你藏私房钱的这条裤子。 Xiaofei ZHANG: The trousers I wash might be the ones you hide your secret purse.	幸福牛家村（*Happy Niu Families' Village*） JSTV Chinese New Year Gala, 2019
0	沈腾：大妈，你好好回忆一下，真的没有撞你。 Teng SHEN: Please recall exactly what happened. I really did not hit you.	扶不扶（*Help Her Up or Not*） CCTV Chinese New Year Gala, 2014

Table 1: Text classification annotation examples taken from different comedies in our corpus. In the *Label* column, ***1*** and ***0*** indicate whether or not this line makes audiences laugh respectively; In the ***Action Line*** column, we present the performer's names and their lines; The ***Source*** column indicates the title of the comedy and the venue where it is performed.

Actor's Line with Annotations	Source
贾玲：这种装修风格显得你家**特别的大**。（客厅几乎是空的） Ling JIA: With this decoration style, your house seems to be ***incredibly big***. (The living room is almost empty)	懒汉相亲（Idler's Blind Date） Ace VS Ace (Season 4), 2019
岳云鹏：不能，不退票**是我们的服务宗旨**。 Yunpeng YUE: No way! Our policy ***is no refund***.	非一般的爱情（*Unusual Love*） Joyful Comedians (Season 2), 2016
贾冰：恩。（贾冰**乖乖地闭上了双眼**） Bing JIA: Okay. (He duly ***closes his eyes***)	贾总的演讲（*Manager JIA's Presentation*） Legend of Laughter (Season 1), 2017

Table 2: Information extraction annotation examples taken from different comedies in our corpus. In the first column, we highlight the **text spans** that trigger laughs from audiences. Note that, we also collected the performer's moves (e.g., "***duly closes his eyes***" in the third example).

not an actor's lines[2] can make audiences laugh. In other words, we formulate the task of predicting as a binary text classification problem. 2) Information Extraction: we predict the text spans from an actor's lines indicating the specific words that trigger an audience's laughter.

Contributions Firstly, we introduce a Chinese corpus of annotated comedy scripts collected from popular TV entertainment programmes. Our annotations include both text classification and information extraction labels. Tables 1 and 2 present annotation examples. The corpus can be used to build an intelligent system to benefit the script writing for comedy writers. It may also be useful for dialogue system research and discourse analysis. Secondly, we evaluate a number of NLP methods and the results demonstrate that BERT models (Devlin et al., 2019) are able to achieve the best prediction performance among all methods. We also further conduct an error analysis which may be useful for further improving the performance. Lastly, we experimentally show that our corpus can also be used to predict laughter triggers for scripts which have very different styles compared to training data.

2 Related Work

Our work is closely related to humour detection, which has been widely studied for many years in natural language processing. Mihalcea and Strapparava (2006); Yang et al. (2015); Chen and Soo (2018); Blinov et al. (2019) investigated if a text fragment is a one-liner.[3] Zhang and Liu (2014); Ortega-Bueno et al. (2018); Chiruzzo et al. (2019) explored the humour classification task on tweets. Castro et al. (2018) collected humour values and funniness scores of Spanish tweets by using crowdsourcing. Chiruzzo et al. (2019) proposed a regression task that predicts the humour score for a tweet. Li et al. (2020) collected Chinese Internet slang expressions and combined them with a humor detecting method to analyse the sentiment of Weibo[4] posts. It should be noted that the examples in all of the corpora used or constructed in the above-mentioned studies are independent of each other. Since our corpus is based on entire scripts, the annotated lines and text spans might also benefit the researchers who are interested in modelling long-context-aware algorithms to understand humour. Apart from the studies on short text fragments, Bertero (2019) and Hasan et al. (2019) created corpora from television (TV) sitcoms such as *The Big Bang Theory*[5] and *TED talks*[6] respectively. Their goal is to predict whether or not a sequence of texts will trigger immediate laughter. Yang et al. (2015); Zhang et al. (2019) extracted the key words such as *sing*, *sign language* and *pretty handy* from jokes, which are similar to our information extraction annotations.

[2] The lines are from the dialogue of a comedy performance. Each line consists of an actor's name and the sentences this actor speaks in performance.

[3] A one-liner is a joke that is delivered in a single line which only contains a few words.

[4] Weibo is a Chinese micro-blogging website similar to Twitter: https://www.weibo.com/

[5] https://the-big-bang-theory.com/

[6] https://www.ted.com/talks

3 Corpus

3.1 Data Collection

Source Selection In order to ensure the high-quality of scripts, we carefully selected thirty performances (the total duration is approximately 473 minutes), including both sketch comedies and crosstalks, of which the leading roles are famous Chinese comedians. These performances were played on well-known Chinese TV entertainment programmes such as Chinese New Year Gala and *Ace VS Ace*[7]. Since there were many people in the audience present for the recording of these performances, the annotators can judge whether the audience laughed based on the performance videos. Please refer to the appendix for the full list of performances which gives details of their titles, leading comedians and sources. Lastly, we manually typed up actors' lines for each performance and completed thirty scripts. Although there may be differences between our scripts and the real scripts used by comedians in terms of format or content, we assume that our scripts contain the key information about the real scripts, i.e., the actors' lines. Therefore the corpus can be useful for the development of intelligence-assistant comedy script writing systems.

Diversity We also took the comedy style into consideration. In order to ensure the diversity and its balance: a) The performances were selected from three main different types of sources[8] as shown in Table 3, including the topic descriptions of selected performances. It can be observed that the corpus has a wide range of topics. b) As a preliminary study, we selected six popular Chinese comedians who have various and distinctive styles, and we chose five representative performances of each comedian.

Corpus Statistic Table 4 illustrates the statistics and Figure 1 shows the laughter rates of each script. The highest line-level and character-level rates are 45.39% and 13.12%, while the lowest rates are 16.03% and 3.49%. We note that the character-level laughter rates vary in different scripts. This may be due to density of laughter triggers of a line or the topic of the script.

Source	Topics
Chinese New Year Galas	- Love stories and blind dates between old people; - Reflecting social phenomena to call for a better society (e.g. avoid judging people by their appearances, do not spoil children, care more about lonely seniors, the woman builds a good relationship with her mother-in-law, spend more time with children, be wary of scams); - Funny family stories during spring festival;
Reality Shows	- Stories happened in ancient times; - Stories about young people (e.g. encounter ex-boyfriends or ex-girlfriends, relationships between best friends, blind dates); - Reflecting social phenomena to call for a better society (e.g. give seats to vulnerable people);
Comedy Competition Shows	- Love stories; - Hot topics (e.g. support the COVID-19 frontline fighters); - Funny stories that happened among friends and in families; - Reflecting social phenomena to call for a better society (e.g. be wary of scams, care more about orphans in orphanage);

Table 3: Topics of the selected comedies.

Statistics	Value
# of Comedy Scripts	30
Year Range	2014—2020
Total Duration	473.44 mins
Average Duration	15.78 mins
# of Actors' Lines	6087
Laughter Rate (Line-Level)	28.62%
# of Characters	120451
Laughter Rate (Character-Level)	8.16%

Table 4: Corpus statistics. ***# of Actors' Lines*** and ***Characters*** correspond to the total number of lines and characters in our corpus respectively. ***Laughter Rate*** is the rate of lines/characters that trigger laughter.

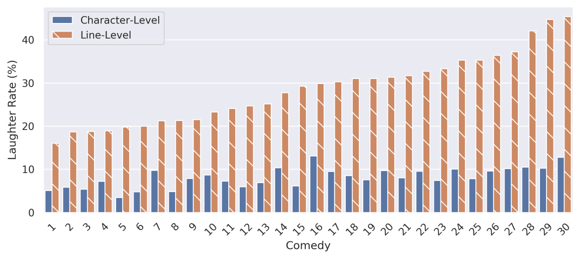

Figure 1: Each script's laughter rate in our corpus.

3.2 Annotation

The annotation was completed on *Doccano* platform (Nakayama et al., 2018) and the annotators are two native Chinese speakers. The annotations were produced based on the studio audiences' responses as observed in the videos, and are not based on the annotators' responses.

Annotation Instruction Annotating text classification labels is easy; annotators are requested to simply assigned label *1* to the lines that make audiences laugh, and *0* to the others. With regard to

[7] https://es.wikipedia.org/wiki/Ace_vs_Ace

[8] The three sources are: **Chinese New Year Galas**—the annual televised Chinese New Year celebrations which are the most viewed TV shows in China. The shows consist of various performances including sketch comedies and crosstalks; **Reality Shows**—the programmes that show the unscripted actions of participants such as playing games and talking. We selected the shows in which comedians were involved; **Comedy Competition Shows**—the programmes where different comedians present their comedy performances to a studio audience and the winners are selected based on the audience's votes.

the information extraction annotations, annotators are requested to identify text spans which are usually phrases. The span consists of the words that immediately made the audience laugh after the comedians said them. For example, as indicated in Table 2, the span *incredibly big* was annotated. In this case, only annotating *big* would be considered as an incorrect annotation, because the comedian was using *incredibly* to strongly emphasise *big* which was her first impression of a man's house in a blind date. Only annotating *incredibly* would also be incorrect, because the main reason why the audience laughed was because the comedian said the house looked *big*.[9]

Annotation Process The annotation process was as follows: Firstly, the annotators conducted discussions about the conflicting annotations after several attempts to annotate the same three scripts. Secondly, once agreements about how to solve the conflicts had been reached, they started to annotate their assigned scripts. Afterwards, since information extraction annotation is more complex than that of classification annotations, we measured its quality by computing three types of inter-annotator agreement. We asked the annotators to annotate the same six scripts having different styles and then calculated the Overall Percent Agreement (OPA), Fleiss's kappa (Fleiss, 1971) and Randolph's kappa (Randolph, 2005). We found that the agreement rates were high (OPA 98.09%, Fleiss's Kappa 0.85, Randolph's Kappa 0.96). This is due to the fact that the discussions about solving conflicts were in-depth and the laughter triggers were usually clear in the lines.

4 Baselines and Results Discussion

In order to understand how well the machine learning methods work on our corpus, we evaluate the performances of a number of models on 5-fold cross-validation random splits of the scripts in our corpus and report the average results in this section.[10] All the BERT models were pre-trained by using a mixture of large Chinese corpora.[11] Please refer to the appendix for the results of each fold, statistics of splits, computing infrastructure, each model's running time, parameter details and hyper-parameter settings.

Model	P	R	F	Acc.
CNN (Kim, 2014)	42.53	64.14	51.07	66.29
RCNN (Lai et al., 2015)	41.21	**68.52**	50.89	63.54
BiLSTM (Liu et al., 2016)	41.17	57.13	47.66	65.69
+ Attention (Zhou et al., 2016)	39.97	59.91	47.44	63.94
FastText (Joulin et al., 2017)	40.61	66.26	50.12	63.72
DPCNN (Johnson and Zhang, 2017)	42.46	63.25	50.76	66.32
Transformer (Vaswani et al., 2017)	42.60	64.71	51.24	66.18
BERT-tiny (Jiao et al., 2019)	47.56	53.38	48.91	66.21
BERT-small (Turc et al., 2019)	47.29	56.21	51.21	70.78
BERT-base (Devlin et al., 2019)	**47.60**	56.64	**51.61**	**70.94**

Table 5: Text classification performance. ***P***, ***R***, ***F*** and ***Acc.*** are ***Precision***, ***Recall***, ***F1-score*** and ***Accuracy*** respectively.

Model	P	R	F
HMM (Rabiner and Juang, 1986)	22.19	7.43	11.04
CRF (Lafferty et al., 2001)	28.56	6.11	10.01
BiLSTM (Huang et al., 2015)	**31.21**	1.64	3.09
BiLSTM-CRF (Lample et al., 2016)	30.33	9.81	14.48
BERT-tiny (Jiao et al., 2019)	26.26	19.89	22.57
BERT-small (Turc et al., 2019)	28.82	17.52	21.56
BERT-base (Devlin et al., 2019)	30.15	**21.47**	**24.59**

Table 6: Information extraction performance. Relaxed metrics are used. The exact-match metric is over-strict because the length of text spans in this corpus is much longer than general named entities. The computation of these metrics can be found in (Nguyen et al., 2017).

Baselines Tables 5 and 6 respectively present the results of text classification and information extraction. BERT-base has the best F1-scores among all the methods. We also note that the classification recall of RCNN (Lai et al., 2015) is much higher than other methods. Therefore, we suggest using this model if users prefer a classifier with a high recall. In addition, we observe that the scores are not high, especially for the information extraction task. The reason may be if the audience laughter highly depends on the conversation contexts which were not considered by baselines. Therefore, taking a longer conversation context of a line into consideration is a worthy research direction.

Prediction Errors Tables 7 and 8 present examples incorrectly predicted by the BERT-base. The first 3 examples describe how the model failed to predict the laughter triggers, while the last 3 examples show false positive predictions. Incorporating the context information of lines may further reduce these errors.

Cross-Style Performance We further investigate the performance of predicting laughter triggers on scripts with a totally different style compared to

[9]The house is actually small. Since there is almost no furniture in the house, the comedian said it looked big.

[10]Model implementations were adapted from `https://github.com/649453932/Chinese-Text-Classification-Pytorch`, `https://github.com/luopeixiang/named_entity_recognition` and Zhao et al. (2019).

[11]More details are listed in `https://github.com/dbiir/UER-py/wiki/Modelzoo`.

G	P	Actor's Line	Translation
1	0	(雷电、下雨)	(Thunder, Rain)
1	0	岳云鹏: 今天别开生面，我给大家说一段单口相声。	Yunpeng YUE: Today, to start with something new, I will present a monologue comic talk.
1	0	(马丽出场摔倒)	(Li MA falls down right after she comes on stage)
0	1	快递哥: (电话) 啊，开了。	Courier: (Phone) Yeah, the door opens.
0	1	张小斐: 哎呀，快快快，哎呀，这怎么弄呀，焦头烂额呀，快，擦擦。	Xiaofei ZHANG: Oops! Think quick quick quick. Oh no! What can I do? I'm in a dilemma. Clean it quickly.
0	1	何欢: 爷爷，我错了，我以后再也不撒谎了，我错了爷爷。	Huan HE: It's my fault, grandpa. I'll never lie again. My mistake, grandpa.

Table 7: Examples of incorrect classifications taken from different comedies in our corpus. ***G*** and ***P*** are ***Gold*** and ***Predicted*** labels, respectively.

Actor's Line	Translation
贾冰：这事我不敢保证**我现在还认不[认字啊]**。	Bing JIA: I can't guarantee ***if I am still [able to read this] now***.
岳云鹏：你**什么时候来的啊** 。	Yuepeng YUE: Huh? ***When did you come***?
沈腾：走肯定是不赶趟了，**我得跑了** 。	Teng SHEN: Walking is not fast enough, ***I have to run to escape this***.
快递哥：[散打那个] 。	Courier: [The one who is good at free combat].
张小斐：哎呀，这是善意的谎言，可问题是，老师也不会[演戏呀]！	Xiaofei ZHANG: Well, this is a white lie. But the problem is, I don't know how [to pretend]!
何欢：不是，你[找谁呀]？	Huan HE: Eh? [Who are] you [looking for]?

Table 8: Examples of incomplete and incorrect extractions taken from different comedies in our corpus. Characters in **bold** and located in [ ] are the gold annotations and predicted results respectively.

Model	P	R	F	Acc.
CNN (Kim, 2014)	42.75	65.69	51.63	65.04
RCNN (Lai et al., 2015)	42.32	**71.26**	**52.32**	62.79
BiLSTM (Liu et al., 2016)	41.35	60.76	48.81	63.71
+ Attention (Zhou et al., 2016)	43.20	53.76	47.57	66.55
FastText (Joulin et al., 2017)	40.86	67.43	50.70	62.62
DPCNN (Johnson and Zhang, 2017)	41.26	66.82	50.42	62.44
Transformer (Vaswani et al., 2017)	41.57	70.15	51.92	63.06
BERT-tiny (Jiao et al., 2019)	43.01	56.76	48.69	66.23
BERT-small (Turc et al., 2019)	44.95	55.59	49.09	67.38
BERT-base (Devlin et al., 2019)	**47.28**	58.13	51.72	**69.39**

Table 9: Cross-style text classification performance.

Model	P	R	F
HMM (Rabiner and Juang, 1986)	19.79	7.52	10.68
CRF (Lafferty et al., 2001)	25.29	5.03	8.35
BiLSTM (Huang et al., 2015)	29.05	2.90	5.23
BiLSTM-CRF (Lample et al., 2016)	**30.72**	8.77	12.56
BERT-tiny (Jiao et al., 2019)	26.30	19.20	22.14
BERT-small (Turc et al., 2019)	24.93	27.31	25.12
BERT-base (Devlin et al., 2019)	24.64	**31.65**	**26.51**

Table 10: Cross-style information extraction performance.

the styles in the training data. Firstly, since the six comedians in the corpus have distinctive comedy styles, we split the entire corpus into a 6-fold cross-validation manner. The comedies in each fold are performed by the same leading comedian. Secondly, we train baseline models on five of the folds and evaluate the performance on the remaining fold. Tables 9 and 10 present the average results and the full results are available in the appendix. The results demonstrate that the laughter triggers can be detected even though the styles in the training data are very different compared to the testing data.

5 Conclusion and Future Work

We study the prediction of laughter triggers from comedy scripts by using text classification and information extraction methods. Firstly, we introduce a corpus including high-quality and annotated sketch comedy and crosstalk scripts. Secondly, we evaluate a number of baselines and find that BERT models achieve the best performance. We note that the information extraction performance was very low, indicating that this task is particularly challenging. We also conduct an error analysis of incorrect predictions. The errors suggest the incorporation of rich context information may further improve the performance. Therefore, it is worth investigating a model which can take such information into consideration. Furthermore, it is also worth extending the corpus to a multimodal one by aligning scripts to corresponding audios or videos, because certain intonations or scenes can also make audiences laugh. The multimodal corpus can also benefit the creation of silent comedy. Enriching the corpus by including scripts in other languages may also be a potential direction. Lastly, the encouraging cross-style prediction performance shows the usefulness of our corpus for predicting new scripts with different styles.

Moreover, it is also interesting to explore human performances by asking annotators to make predictions based purely on the scripts of unwatched comedies, and to investigate if the script writers find the model predictions insightful.

We hope this study will benefit script writing by inspiring the community to develop intelligent systems for comedy writers and other artists in the field. The corpus might also be useful for researchers who are working on related or similar tasks, such as discourse analysis and humorous response generation for dialogue systems.

Acknowledgements

We would like to express our sincere appreciation to all comedians and backstage teams for their hard work to make audiences happy. We also sincerely appreciate the valuable comments of all anonymous reviewers. My deepest thanks should also be given to my best friend who inspired me to conduct such research and whose encouragement was important for me to be able to complete the work.

References

Dario Bertero. 2019. *Conversational humor recognition and generation through deep learning*. PhD dissertation, Hong Kong University of Science and Technology.

Marc Blake. 2014. *How to be a Comedy Writer*, volume 1. Andrews UK Limited.

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, Florence, Italy. Association for Computational Linguistics.

Ernesto Carrera-Ruvalcaba, Johnson Ekedum, Austin Hancock, and Ben Brock. 2019. Leveraging natural language processing applications and microblogging platform for increased transparency in crisis areas. *SMU Data Science Review*, 2(1):6.

Santiago Castro, Luis Chiruzzo, Aiala Rosá, Diego Garat, and Guillermo Moncecchi. 2018. A crowd-annotated Spanish corpus for humor analysis. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 7–11, Melbourne, Australia. Association for Computational Linguistics.

Peng-Yu Chen and Von-Wun Soo. 2018. Humor recognition using deep learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 113–117, New Orleans, Louisiana. Association for Computational Linguistics.

Luis Chiruzzo, S Castro, Mathias Etcheverry, Diego Garat, Juan José Prada, and Aiala Rosá. 2019. Overview of haha at iberlef 2019: Humor analysis based on human annotation. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019). CEUR Workshop Proceedings, CEUR-WS*, Bilbao, Spain.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychometrika*, 76(5):378.

Md Kamrul Hasan, Wasifur Rahman, AmirAli Bagher Zadeh, Jianyuan Zhong, Md Iftekhar Tanveer, Louis-Philippe Morency, and Mohammed (Ehsan) Hoque. 2019. UR-FUNNY: A multimodal language dataset for understanding humor. 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)*, pages 2046–2056, Hong Kong, China. Association for Computational Linguistics.

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang, and Qun Liu. 2019. Tinybert: Distilling bert for natural language understanding. *arXiv preprint arXiv:1909.10351*.

Rie Johnson and Tong Zhang. 2017. Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 562–570, Vancouver, Canada. Association for Computational Linguistics.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, page 2267–2273. AAAI Press.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.

Da Li, Rafal Rzepka, Michal Ptaszynski, and Kenji Araki. 2020. Hemos: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Information Processing & Management*, 57(6):102290.

Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. 2016. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 2873–2879. AAAI Press.

Rada Mihalcea and Carlo Strapparava. 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence*, 22(2):126–142.

Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. Software available from https://github.com/doccano/doccano.

An Thanh Nguyen, Byron Wallace, Junyi Jessy Li, Ani Nenkova, and Matthew Lease. 2017. Aggregating and predicting sequence labels from crowd annotations. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 299–309, Vancouver, Canada. Association for Computational Linguistics.

Reynier Ortega-Bueno, Carlos E Muniz-Cuza, José E Medina Pagola, and Paolo Rosso. 2018. Uo upv: Deep linguistic humor detection in spanish social media. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)*, pages 203–213, Seville, Spain.

Lawrence Rabiner and B Juang. 1986. An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16.

JJ Randolph. 2005. Free-marginal multirater kappa: an alternative to fleiss' fixed-marginal multirater kappa. *Joensuu University Learning and Instruction Symposium*.

Delip Rao and Brian McMahan. 2019. *Natural language processing with PyTorch: build intelligent language applications using deep learning.* " O'Reilly Media, Inc.".

Serdar Sarıtaş, Hasan Genç, Şerafettin Okutan, Ramazan İnci, Ahmet Özdemir, and Gülnaz Kizilkaya. 2019. The effect of comedy films on postoperative pain and anxiety in surgical oncology patients. *Complementary Medicine Research*, 26(4):231–239.

Iulia Turc, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Well-read students learn better: The impact of student initialization on knowledge distillation. *arXiv preprint arXiv:1908.08962*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Sandy M Williams, Pauline K Arnold, and Jennifer N Mills. 2005. Coping with stress: a survey of murdoch university veterinary students. *Journal of Veterinary Medical Education*, 32(2):201–212.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2367–2376, Lisbon, Portugal. Association for Computational Linguistics.

Dongyu Zhang, Heting Zhang, Xikai Liu, Hongfei Lin, and Feng Xia. 2019. Telling the whole story: A manually annotated Chinese dataset for the analysis of humor in jokes. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6402–6407, Hong Kong, China. Association for Computational Linguistics.

Renxian Zhang and Naishi Liu. 2014. Recognizing humor on twitter. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, page 889–898, New York, NY, USA. Association for Computing Machinery.

Zhe Zhao, Hui Chen, Jinbin Zhang, Xin Zhao, Tao Liu, Wei Lu, Xi Chen, Haotang Deng, Qi Ju, and Xiaoyong Du. 2019. UER: An open-source toolkit for pre-training models. 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): System Demonstrations*, pages 241–246, Hong Kong, China. Association for Computational Linguistics.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany. Association for Computational Linguistics.

A Appendices

A.1 Computing Resources

Table 11 describes the details of the computing resources used for all of our experiments. These resources are freely available from *Paperspace*[12].

A.2 Model Details

Below we present **model hyper-parameter values**[13] and the **average running time of one epoch**.

[12]`https://www.paperspace.com/`

[13]The size of model's trainable parameters can be found in original papers.

Name	Description
CPU	8 vCPUs, 30GB RAM
GPU	NVIDIA Quadro P5000 Graphics Card, 16GB
Processor	Intel Xeon
Clock Speed	2.60 GHz

Table 11: Details of computing resources.

A.2.1 Text Classification Models

CNN (Kim, 2014) Dropout = 0.5, Number of Epochs = 20, Batch Size = 128, Learning Rate = 0.001, Number of Filters = 256, Filter Sizes = 2,3,4, Average Running Time = 5s.

RCNN (Lai et al., 2015) Dropout = 1.0, Number of Epochs = 10, Batch Size = 128, Learning Rate = 0.001, Hidden Size = 256, Number of Layers = 1, Average Running Time = 5s.

BiLSTM (Liu et al., 2016) Dropout = 0.5, Number of Epochs = 10, Batch Size = 128, Learning Rate = 0.001, Hidden Size = 128, Number of Layers = 2, Average Running Time = 5.4s.

+ Attention (Zhou et al., 2016) Dropout = 0.5, Number of Epochs = 10, Batch Size = 128, Learning Rate = 0.001, Hidden Size = 128 and 64 respectively, Number of Layers = 2, Average Running Time = 5.74s.

FastText (Joulin et al., 2017) Dropout = 0.5, Number of Epochs = 20, Batch Size = 128, Learning Rate = 0.001, Hidden Size = 256, Average Running Time = 22.5s.

DPCNN (Johnson and Zhang, 2017) Dropout = 0.5, Number of Epochs = 20, Batch Size = 128, Learning Rate = 0.001, Number of Filter = 250, Average Running Time = 5s.

Transformer (Vaswani et al., 2017) Dropout = 0.5, Number of Epochs = 20, Batch Size = 128, Learning Rate = 0.0005, Number of Head = 5, Number of Encoder = 2, Average Running Time = 6.528s.

BERT-tiny (Jiao et al., 2019) Dropout = 0.1, Number of Epoch = 20, Batch Size = 64, Learning Rate = 0.00002, Size of Embedding = 384, Feedforward Size = 1536, Hidden Size = 384, Number of Head = 6, Number of Layer = 3, Average Running Time = 20.63s.

BERT-small (Turc et al., 2019) Dropout = 0.5, Number of Epoch = 20, Batch Size = 64, Learning Rate = 0.00002, Size of Embedding = 512, Feedforward Size = 2048, Hidden Size = 512, Number of Head = 8, Number of Layer = 6, Average Running Time = 36.65.

BERT-base (Devlin et al., 2019) Dropout = 0.1, Number of Epoch = 10, Batch Size = 64, Learning Rate = 0.00002, Size of Embedding = 768, Feedforward Size = 3072, Hidden Size = 768, Number of Head = 12, Number of Layer = 12, Average Running Time = 99s.

A.2.2 Information Extraction Models

HMM (Rabiner and Juang, 1986) Uniform Distribution for Initialisation, Average Total Running Time = 8.47s.

CRF (Lafferty et al., 2001) LBFGS algorithm, c1 = 0.1, c2 = 0.1, Max Iteration = 100, Average Total Running Time = 11.72s.

BiLSTM (Huang et al., 2015) Number of Epoch = 30, Batch Size = 64, Learning Rate = 0.001, Size of Embedding = 128, Hidden Size = 128, Average Running Time = 8.43s.

BiLSTM-CRF (Lample et al., 2016) Number of Epoch = 30, Batch Size = 64, Learning Rate = 0.001, Size of Embedding = 128, Hidden Size = 128, Average Running Time = 9.35s.

BERT Models We use the same hyper-parameter settings as used in the text classification models with the exception of Batch Size = 16. The average running time of BERT-tiny, BERT-small and BERT-base for information extraction are 34.00s, 55.73s, and 120s respectively.

A.3 Details of Baseline Experiments

Table 12 shows the statistics of each fold.

Fold	Text Classification		Information Extraction	
	# of Lines	Total	# of Characters	Total
0	277/685	962	1666/20010	21676
1	290/854	1144	1703/19354	21057
2	285/789	1074	1666/19609	21275
3	358/909	1267	2251/23535	25786
4	459/1181	1640	2632/28025	30657

Table 12: Statistics of each fold in the baseline experiments. The number before slash indicates how many actor's lines or characters that make the audience laugh. The number after slash indicates the number of lines or characters without causing audiences laugh.

	Model	Fold-0	Fold-1	Fold-2	Fold-3	Fold-4	Average
P	CNN	42.64	40.43	42.61	42.27	44.68	42.53
	RCNN	43.47	38.30	43.93	35.44	44.93	41.21
	BiLSTM	43.56	37.80	40.19	42.20	42.12	41.17
	+Attention	39.53	38.93	41.75	37.62	42.01	39.97
	FastText	39.37	44.64	36.96	41.12	40.98	40.61
	DPCNN	42.44	41.67	41.73	41.86	44.61	42.46
	Transformer	**46.67**	38.17	42.43	42.08	43.66	42.60
	BERT-tiny	42.26	42.05	44.62	**54.43**	**54.43**	47.56
	BERT-small	44.74	46.13	**46.02**	52.35	47.22	47.29
	BERT-base	45.65	**47.73**	42.66	53.03	48.91	**47.60**
R	CNN	**70.04**	64.02	60.70	66.48	59.48	64.14
	RCNN	66.06	**70.08**	59.65	**86.03**	60.78	**68.52**
	BiLSTM	51.26	65.15	58.95	56.70	53.59	57.13
	+Attention	55.23	65.91	58.60	75.14	44.66	59.91
	FastText	68.23	56.82	**71.58**	63.41	**71.24**	66.26
	DPCNN	67.87	62.50	61.05	65.36	59.48	63.25
	Transformer	60.65	69.70	64.91	62.29	66.01	64.71
	BERT-tiny	62.09	41.03	59.65	49.72	54.43	53.38
	BERT-small	59.93	47.24	62.81	59.22	51.85	56.21
	BERT-base	62.45	50.69	55.09	61.17	53.81	56.64
F	CNN	53.01	49.56	50.07	51.68	51.03	51.07
	RCNN	52.44	49.53	50.60	50.20	51.67	50.89
	BiLSTM	47.10	47.84	47.80	48.39	47.17	47.66
	+Attention	46.08	48.95	48.76	50.14	43.29	47.44
	FastText	49.93	**50.00**	48.75	49.89	52.03	50.12
	DPCNN	52.22	**50.00**	49.57	51.04	50.98	50.76
	Transformer	**52.75**	49.33	51.32	50.23	**52.56**	51.24
	BERT-tiny	50.29	41.54	51.05	51.97	49.72	48.91
	BERT-small	51.23	46.68	**53.12**	55.57	49.43	51.21
	BERT-base	52.74	49.16	48.09	**56.81**	51.24	**51.61**
A	CNN	64.24	66.41	67.88	64.88	68.05	66.29
	RCNN	65.49	63.18	69.09	51.78	68.17	63.54
	BiLSTM	66.84	63.38	65.83	65.82	66.40	65.65
	+Attention	62.79	64.55	67.32	57.77	67.26	63.94
	FastText	60.60	70.70	60.06	64.01	63.23	63.72
	DPCNN	64.24	67.77	67.04	64.56	67.99	66.32
	Transformer	**68.71**	63.09	67.32	65.11	66.65	66.18
	BERT-tiny	64.66	70.72	69.65	74.03	51.97	66.21
	BERT-small	67.15	72.64	**70.58**	73.24	70.30	70.78
	BERT-base	67.78	**73.43**	68.44	**73.72**	**71.34**	**70.94**

Table 13: Each fold's text classification baseline experiments and their overall average performance. ***P***, ***R***, ***F*** and ***A*** are ***Precision***, ***Recall***, ***F1-score*** and ***Accuracy*** respectively.

Tables 13 and 14 describe the detailed performance of text classification and information extraction baseline experiments.

	Model	Fold-0	Fold-1	Fold-2	Fold-3	Fold-4	Average
P	HMM	18.26	21.15	26.37	19.46	25.70	22.19
	CRF	24.62	**30.86**	**34.17**	23.33	29.83	28.56
	BILSTM	26.25	37.14	26.00	**37.31**	29.35	**31.21**
	BiLSTM-CRF	30.50	24.43	32.46	29.71	**34.56**	30.33
	BERT-tiny	27.64	23.98	28.89	26.73	24.08	26.26
	BERT-small	31.71	24.12	30.83	28.04	29.38	28.82
	BERT-base	**36.23**	28.16	26.56	30.88	28.93	30.15
R	HMM	7.19	9.64	7.38	5.93	7.02	7.43
	CRF	6.49	5.49	6.32	6.47	5.76	6.11
	BILSTM	2.24	1.23	0.55	2.47	1.70	1.64
	BiLSTM-CRF	16.76	8.00	9.25	6.30	8.73	9.81
	BERT-tiny	17.37	19.54	20.96	**22.46**	19.10	19.89
	BERT-small	16.90	19.22	16.26	21.35	13.86	17.52
	BERT-base	**21.62**	**21.19**	**27.17**	21.02	**16.35**	**21.47**
F	HMM	10.32	13.24	11.53	9.09	11.03	11.04
	CRF	10.27	9.33	10.67	10.13	9.65	10.01
	BILSTM	4.13	2.37	1.08	4.63	3.22	3.09
	BiLSTM-CRF	21.63	12.05	14.40	10.39	13.94	14.48
	BERT-tiny	21.33	21.53	24.30	24.41	**21.30**	22.57
	BERT-small	22.05	21.40	21.29	24.24	18.83	21.56
	BERT-base	**26.01**	**24.18**	**26.86**	**25.02**	20.89	**24.59**

Table 14: Each fold's information extraction baseline experiments and their overall average performance.

A.4 Details of Cross-Style Experiments

Table 15 shows the statistics of the scripts performed by specific leading comedians. Tables 16 and 17 present the prediction results.

Comedian	Text Classification		Information Extraction	
	# of Lines	Total	# of Characters	Total
Xiaobao SONG	315/815	1150	2001/20366	22367
Yuepeng YUE	436/1547	1983	2362/25579	27941
Ling JIA	195/501	696	1135/14414	15549
Xiaofei ZHANG	190/495	685	1056/15946	17002
Teng SHEN	166/350	516	1071/13143	14214
Bing JIA	367/690	1057	2293/21085	23378

Table 15: Statistics of the scripts performed by specific leading comedians.

A.5 Full List of Selected Comedy Performances

Tables 18 and 19 show the full list of performances in our corpus with details.

	Model	0	1	2	3	4	5	Average
P	CNN	41.99	40.00	41.85	39.67	45.70	47.29	42.75
	BiLSTM	38.17	42.18	39.35	36.26	45.26	46.90	41.35
	+ Attention	42.70	38.16	**45.06**	38.49	48.02	46.78	43.20
	RCNN	**46.93**	44.01	36.36	37.23	46.26	43.10	42.32
	FastText	41.88	37.99	36.64	37.81	45.24	45.57	40.86
	DPCNN	42.01	42.21	36.03	35.25	45.33	46.71	41.26
	Transformer	41.97	36.32	36.94	41.30	46.50	46.41	41.57
	BERT-tiny	40.73	37.41	42.38	40.81	45.30	51.40	43.01
	BERT-small	39.08	39.67	42.97	48.40	43.22	56.38	44.95
	BERT-base	42.50	**44.40**	42.73	**48.68**	**48.05**	**57.31**	**47.28**
R	CNN	61.59	51.38	69.74	66.48	**80.61**	64.31	65.69
	BiLSTM	51.75	47.02	62.56	70.33	75.15	57.77	60.76
	+ Attention	36.19	55.05	53.85	53.30	58.79	65.40	53.76
	RCNN	55.87	51.38	82.05	**75.27**	78.79	**84.20**	**71.26**
	FastText	62.22	55.50	73.85	66.48	**80.61**	65.94	67.43
	DPCNN	60.95	47.25	75.38	74.18	79.39	63.76	66.82
	Transformer	62.22	**62.39**	**84.10**	66.48	**80.61**	65.12	70.15
	BERT-tiny	**67.62**	47.71	58.46	47.89	63.86	55.04	56.76
	BERT-small	61.90	44.50	56.41	47.89	71.08	51.77	55.59
	BERT-base	64.76	47.25	49.74	58.42	74.10	54.50	58.13
F	CNN	49.94	44.98	**52.31**	49.69	58.33	54.50	51.63
	BiLSTM	43.94	44.47	48.32	47.85	56.49	51.77	48.81
	+ Attention	39.18	45.07	49.07	44.70	52.86	54.55	47.57
	RCNN	51.01	**47.41**	50.39	49.82	58.30	**57.01**	**52.32**
	FastText	50.06	45.11	48.98	48.21	57.95	53.90	50.70
	DPCNN	49.74	44.59	48.76	47.79	57.71	53.92	50.42
	Transformer	50.13	45.91	51.33	50.95	**58.98**	54.20	51.92
	BERT-tiny	50.84	41.94	49.14	44.07	53.00	53.16	48.69
	BERT-small	47.91	41.95	48.78	48.15	53.76	53.98	49.09
	BERT-base	**51.32**	45.78	45.97	**53.11**	58.29	55.87	51.72
A	CNN	66.17	72.37	64.37	61.72	62.89	62.72	65.04
	BiLSTM	63.83	74.18	62.50	56.41	62.70	62.63	63.71
	+ Attention	69.22	70.50	**68.68**	62.50	66.21	62.16	66.55
	RCNN	**70.61**	74.94	54.74	56.87	63.67	55.91	62.79
	FastText	66.00	70.30	56.90	59.38	62.30	60.83	62.62
	DPCNN	66.26	74.18	55.60	53.91	62.50	62.16	62.44
	Transformer	66.09	67.68	55.32	63.59	63.87	61.78	63.06
	BERT-tiny	64.17	70.95	66.09	66.28	63.57	66.32	66.23
	BERT-small	63.13	72.92	66.81	**71.39**	60.66	69.35	67.38
	BERT-base	66.35	**75.39**	67.24	**71.39**	**65.89**	**70.10**	**69.39**

Table 16: Performance of text classification in predicting the scripts performed by specific leading comedians (***0***: Xiaobao SONG, ***1***: Yuepeng YUE, ***2***: Ling JIA, ***3***: Xiaofei ZHANG, ***4***: Teng SHEN, ***5***: Bing JIA).

	Model	0	1	2	3	4	5	Average
P	HMM	21.67	19.47	22.03	16.69	14.61	24.25	19.79
	CRF	22.72	20.54	**39.79**	21.43	18.28	28.96	25.29
	BILSTM	**26.32**	21.48	31.70	24.53	**32.39**	37.89	29.05
	BiLSTM-CRF	**26.32**	**23.35**	33.08	**31.86**	25.71	**44.02**	**30.72**
	BERT-tiny	24.43	21.87	29.61	26.37	27.19	28.32	26.30
	BERT-small	23.93	20.19	26.01	23.85	22.80	32.81	24.93
	BERT-base	20.48	19.10	23.74	24.41	26.58	33.51	24.64
R	HMM	6.91	6.26	6.96	9.50	8.79	6.68	7.52
	CRF	5.80	4.87	7.47	4.41	3.13	4.48	5.03
	BILSTM	4.40	2.17	2.40	2.20	3.71	2.49	2.90
	BiLSTM-CRF	7.64	14.06	17.52	7.30	3.20	2.90	8.77
	BERT-tiny	16.88	19.89	22.26	19.15	19.06	**17.98**	19.20
	BERT-small	22.63	31.31	29.31	**32.78**	**32.42**	15.39	27.31
	BERT-base	**39.61**	**36.43**	**35.25**	32.01	28.64	17.95	**31.65**
F	HMM	10.48	9.48	10.58	12.11	10.97	10.47	10.68
	CRF	9.24	7.87	12.59	7.31	5.34	7.76	8.35
	BILSTM	7.60	3.94	4.46	4.04	6.65	4.67	5.23
	BiLSTM-CRF	11.87	17.55	22.90	11.88	5.70	5.44	12.56
	BERT-tiny	19.97	20.83	25.42	22.19	22.41	22.00	22.14
	BERT-small	23.26	24.55	27.56	27.61	26.77	20.95	25.12
	BERT-base	**27.00**	**25.06**	**28.37**	**27.70**	**27.57**	**23.37**	**26.51**

Table 17: Performance of information extraction in predicting the scripts performed by specific leading comedians.

Performance Title	Translation	Comedians	Source	Translation
扶不扶	Help Her Up or Not	Teng SHEN etc.	央视春晚	CCTV Chinese New Year Gala
碰瓷	An Incident-Faking Extortionist	Xiaobao SONG etc.	欢乐喜剧人(第一季)	Joyful Comedians (Season 1)
一念天堂	Heaven or Hell?	Teng SHEN etc.	欢乐喜剧人(第一季)	Joyful Comedians (Season 1)
纯闺蜜	We Are Pure Besties	Teng SHEN etc.	欢乐喜剧人(第一季)	Joyful Comedians (Season 1)
以貌取人	Judge By Appearances	Xiaobao SONG etc.	辽宁卫视春晚	LNTV Chinese New Year Gala
闺蜜小时代之怀孕	Story of My Bestie - Pregnancy	Xiaofei ZHANG etc.	喜剧班的春天(第一季)	Comedy Class of Spring (Season 1)
非一般的爱情	Unusual Love	Yunpeng YUE and Yue Sun	欢乐喜剧人(第二季)	Joyful Comedians (Season 2)
选妃记	Select Imperial Concubine	Xiaobao SONG etc.	王牌对王牌(第一季)	Ace VS Ace (Season 1)
暴走街区	Violent Teenagers	Xiaofei ZHANG etc.	欢乐喜剧人(第三季)	Joyful Comedians (Season 3)
前男友前女友	Ex-boyfriend and Ex-girlfriend	Xiaofei ZHANG etc.	喜剧班的春天(第二季)	Comedy Class of Spring (Season 2)
落跑姐妹之还债	Escapted Sisters - Repay a Debt	Ling JIA etc.	喜剧班的春天(第二季)	Comedy Class of Spring (Season 2)
公交故事之让座	Give Up Seats on a Bus	Ling JIA etc.	喜剧班的春天(第二季)	Comedy Class of Spring (Season 2)
贾总的演讲	Manager JIA's Presentation	Bing JIA etc.	笑声传奇(第一季)	Lengend of Laughter (Season 1)
一碗元宵	A Bowl of Yuanxiao	Xiaofei ZHANG etc.	湖南卫视元宵喜乐会	Mango TV Lantern Festival Party
非诚来扰	Blind Date with Me If You Are Not Sincere	Xiaobao SONG etc.	辽宁卫视春晚	LNTV Chinese New Year Gala
学车	Learn Driving	Bing JIA etc.	央视春晚	CCTV Chinese New Year Gala
爱回家	Love Back Home	Bing JIA etc.	东方卫视春晚	Tomato TV Chinese New Year Gala
幸福牛家村	Happy Niu Families' Village	Xiaofei ZHANG etc.	江苏卫视春晚	JSTV Chinese New Year Gala
关于爱情	Something About Love	Yunpeng YUE and Yue Sun	辽宁卫视春晚	LNTV Chinese New Year Gala
懒汉相亲	Idler's Blind Date	Ling JIA etc.	王牌对王牌(第4季)	Ace VS Ace (Season 4)
啼笑皆非	Not Know Whether to Laugh or Cry	Ling JIA etc.	央视春晚	CCTV Chinese New Year Gala
占位子	Grab The Best Seat In The Classroom for My Children	Teng SHEN etc.	央视春晚	CCTV Chinese New Year Gala
你膨胀了	Arrogant You	Yunpeng YUE and Yue Sun	东方卫视春晚	Tomato TV Chinese New Year Gala
乌龙快递	Express Delivery for COVID-19 Frontline Fighters	Bing JIA etc.	欢乐喜剧人(第六季)	Joyful Comedians (Season 6)
父与子	Father and Son	Bing JIA etc.	辽宁卫视春晚	LNTV Chinese New Year Gala
猜谜语	Guess Riddles	Yunpeng YUE and Yue Sun	辽宁卫视春晚	LNTV Chinese New Year Gala
想说爱你不容易	Not Easy to Say Love You	Xiaobao SONG etc.	山东卫视春晚	SDTV-1 Chinese New Year Gala
生活趣谈	Funny Stories in Life	Yunpeng YUE and Yue Sun	央视春晚	CCTV Chinese New Year Gala
婆婆妈妈	Husband's Mother	Ling JIA etc.	央视春晚	CCTV Chinese New Year Gala
走过场	Go Through The Motions	Teng SHEN etc.	央视春晚	CCTV Chinese New Year Gala

Table 18: Full list of the selected comedy performances with their titles, leading comedians and source.

Performance Title	Translation	Year	Duration (mins)	Number of Lines	Laughter Rate (Line-Level)	Number of Characters	Laugher Rate (Character-Level)
扶不扶	Help Her Up or Not	2014	14.75	114	44.74%	3411	10.29%
碰瓷	An Incident-Faking Extortionist	2015	16.07	202	23.27%	3902	8.69%
一念天堂	Heaven or Hell?	2015	10.92	55	32.73%	1518	9.55%
纯闺蜜	We Are Pure Besties	2015	11.3	85	35.29%	2468	7.86%
以貌取人	Judge By Appearances	2015	12.87	220	31.36%	3919	9.72%
闺蜜小时代之怀孕	Story of My Bestie - Pregnancy	2015	8.5	81	33.33%	1894	7.44%
非一般的爱情	Unusual Love	2016	41.68	317	27.76%	4836	10.36%
选妃记	Select Imperial Concubine	2016	12.62	141	21.28%	2522	9.79%
暴走街区	Violent Teenagers	2017	12.67	190	20%	4768	4.8%
前男友前女友	Ex-boyfriend and Ex-girlfriend	2017	10.43	106	19.81%	2723	3.49%
落跑姐妹之还债	Escapted Sisters - Repay a Debt	2017	12.27	135	25.19%	2894	6.91%
公交故事之让座	Give Up Seats on a Bus	2017	9.28	95	42.11%	1795	10.53%
贾总的演讲	Manager JIA's Presentation	2017	23.5	228	36.4%	5319	9.64%
一碗元宵	A Bowl of Yuanxiao	2018	17.58	177	31.07%	4465	7.57%
非诚来扰	Blind Date with Me If You Are Not Sincere	2018	31.3	357	35.29%	7854	10.06%
学车	Learn Driving	2018	14.83	165	30.3%	3306	9.5%
爱回家	Love Back Home	2019	15.31	152	45.39%	3738	12.79%
幸福牛家村	Happy Niu Families' Village	2019	12.18	131	31.7%	3152	8.03%
关于爱情	Something About Love	2019	21.83	590	21.53%	8044	7.91%
懒汉相亲	Idler's Blind Date	2019	8.77	101	18.81%	1946	5.4%
啼笑皆非	Not Know Whether to Laugh or Cry	2019	17.72	178	24.72%	4726	5.99%
占位子	Grab The Best Seat In The Classroom for My Children	2019	14	140	29.29%	3786	6.15%
你膨胀了	Arrogant You	2020	12.65	241	29.88%	3797	13.12%
乌龙快递	Express Delivery for COVID-19 Frontline Fighters	2020	15.43	195	24.1%	4672	7.3%
父与子	Father and Son	2020	23.33	317	37.22%	6343	10.2%
猜谜语	Guess Riddles	2020	19.4	523	18.93%	7142	7.24%
想说爱你不容易	Not Easy to Say Love You	2020	12.53	230	18.7%	4170	5.85%
生活趣谈	Funny Stories in Life	2020	11.5	312	16.03%	4122	5.09%
婆婆妈妈	Husband's Mother	2020	16.42	187	31.02%	4188	8.55%
走过场	Go Through The Motions	2020	11.8	122	21.31%	3031	4.88%

Table 19: Full list of the selected comedy performances with their titles, years, duration, number of lines/characters and laughter rate at line/character-level.