Danmaku Sentiment and Viewer Activity

Kesong Cao[§*], Zhikang Hao[‡*], Yuting Yan[‡*], Bojun Liu[†*]
Department of Chemistry[†], Department of Computer Sciences[‡], Department of Psychology[§]
University of Wisconsin-Madison
Madison, WI 53706
{kesong.cao, hao45, yan86, bliu293}@wisc.edu

Abstract

Danmaku is a live text comment for videos that allows real-time, interactive communication with viewers. Danmaku sentiment analysis can be a great way to demonstrate the behaviors and psychology of viewers. In this paper, transfer learning is applied to research the danmaku sentiment. The HuggingFace Transformers including ERNIE, XLNet and RoBERTa are trained and finetuned with a labeled dataset. The results show that ERNIE has the highest accuracy in both labeled dataset and partial manually labeled danmaku data. Thus, ERNIE is used to classify the sentiment of the danmaku. An extended investigation on the relationship between sentiment intensity and popularity implies the more intense for the sentiment, the more popular the video.

1 Introduction

Danmaku (English: bullet-screen comments, Japanese: 弾幕, Chinese: 弹幕) is a feature of video websites that allows viewers to send live text comments (named danmaku) which are alongside and overlapping a video. After a viewer hit "send", danmaku are retained at the fixed timestamp in the video, and all future viewers are able to see them overlapping the video, moving from right to left of the screen like bullets. In contrast, "traditional" live comments are displayed alongside but separate from the video and usually in the event of a livestream on sites such as YouTube and Twitch. Figure 1 is an example interface of video plus danmaku. As netizen viewers interact with each other using danmakus in addition to comments and other activities such as "like", "favorite", and subscribe, videos and livestreams with the danmaku feature are shaping a new era of cyberculture[1]. We list some key differences in comment functionalities in Table 1.

Type	Danmaku	Live Comment	Comment
Location on the webpage	overlapping with the video	alongside the video (usually at the left or right)	separate from the video (usually at the bottom)
Length	short	short	short and long
Storage	retained	deleted	retained
Interaction	users can "like"	the livestreamer can read paid "super chats" (SC)	users can "like", reply, share
Anonymity	yes	no	no

Table 1: A comparison of comment functionalities on video websites

Danmaku has much more popularity in East Asian regions compared to other locations, especially in Japan and China. Some examples of danmaku sites include Niconico (https://www.nicovideo.jp/,

[*]Equal contributions.

Preprint. Under review.

Figure 1: User interface of a video with danmaku enabled on Bilibili (in Chinese)

in Japanese) and Bilibili (https://bilibili.com/, in Chinese). Some argue that the varied popularity is because Chinese have much higher Shannon entropy (9.56 bits per character) than other languages such as English (3.9 bits per letter) [2].

The popularity of danmaku not only makes the video or livestreams more entertaining for the audience, but also it can be a great subject for research in many disciplines[3], especially in the sentiment analysis field. Sentiment analysis is an active field of research across many research disciplines such as marketing, social media research, and psychology. For example, in the the marketing field, sentiment analysis offers the abilities to the companies like Twitter and Facebook and individuals that want to monitor their reputation and get timely feedback about their products and actions[4].

Danmaku can be great resource for providing feedback and being applied to sentiment analysis. First, danmaku data can be generated from each video in large quantities, which means in general it can provide a sufficiently large dataset for the analysis. Second, danmaku allows the audience to send comments anonymously and see other' danmaku comments while they are watching videos[3], compared with the 'traditional' live comments which are displayed alongside but separate from the video and usually in the event of a livestream on sites such as YouTube and Twitch. This brand-new comment form can provide audience a stronger sense of participation and makes it easier for the audience to express their true emotions connected with the videos' contents[5, 6]. From an analytical perspective, danmaku data can therefore provide more accurate feedbacks from the audience. Third, perhaps the most important reason, is that danmaku – the user generated, instantaneous, naturally occurring data – offers a more efficient and feasible way to evaluate the audience feedbacks at different timestamps of the video. Intuitively, to summarize and understand the content of a video, a model could perform computer vision tasks and try to do so frame by frame, which is quite challenging as the computed data and information is huge [7]. By contrast, through natural language processing tasks utilizing the timestamp-associated danmaku data as an extra or even alternative modality choice, we can "integrate" viewer sentiments by a time unit, say one second, to gain an accurate and precise viewer sentiment profile of the video. This new form of viewer feedback integration and summarization is only doable with danmuku data, as traditional comments for a video lack the timestamp information: the viewers are expressing their feedback with respect to the overall video content. Hence, timestamp-associated danmaku data enables us to test the following hypotheses with respect to viewer sentiment and activity.

2 Hypotheses

H_1: Sentiments that are more extreme (either positive or negative) from danmaku correlate with higher amounts of viewer activities.

H_2: Danmaku sentiments can better predict view activity amount than sentiments from traditional comments.

3 Related works

There has been so many techniques being wildly used in the field of sentiment analysis corresponding to text mining[8]. For example, many machine learning based methods are extremely powerful in classifying sentiments based on large Users' comments datasets. An improved Naive Bayes Classifier approach was proposed by Kang and Yoo in 2012 for the sentiment analysis of restaurant reviews[9]. Also, some other probabilistic classifiers such as Bayesian Network, Linear Classifiers, Support Vector Machine Classifiers et al. all exhibit great potential in sentiment analysis field[10, 11, 12, 13]. Besides several mentioned probabilistic classifier techniques, more and more NLP (Natural Language Processing) designed deep learning architectures have also been reported to deal with larger and more complex text dataset within recent years. In 2015, Tang et al. introduced several deep learning based approaches for sentiment analyses, such as learning word embedding, sentiment classification, and opinion extraction[14]. Meanwhile, Hassan and Mahmood proved that CNN and RNN models can overcome shortcoming of short text in deep learning models[15]. And recently, Wang et al. analyzed a video interactive dataset - Video-IC using RNN and Transformer techniques[16].

In this paper, we are going to use Transformer NLP models for our sentiment analysis task.

4 Method

Our framework to get the sentiment of each video shown in Figure 2 can be summarized by the following steps: 1) Data preparing, 2) Training model, 3)Evaluation, 4)Prediction, shown in .

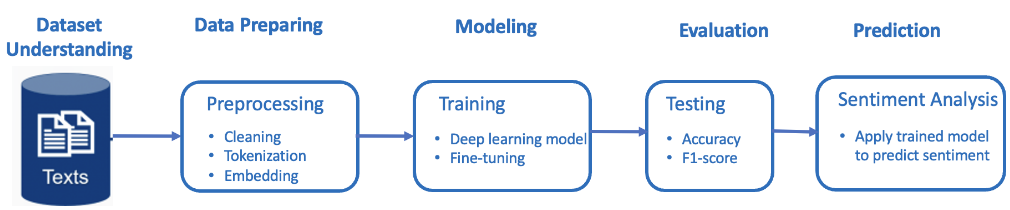

Figure 2: Danmaku sentiment prediction process

4.1 DataSets

Two dataset with Chinese short messages are used in this project. The first one, ChnSentiCorp [17] is a 10,000 Chinese hotel review data set with sentiment labels from HuggingFace library. The dataset would be used to train and evaluate our model. Another is a danmaku dataset [16] from the most popular Bilibili videos, which contains about 4950 video danmaku from all kinds of fields like culture, entertainment, and games. The trained model is applied to danmaku dataset for sentiment analysis.

Data cleaning is one of the key steps in preparing language data. We should remove noise like special characters and stop words to help embedding model easily detect the patterns.

4.2 Word Embedding

Pre-trained word embedding models can reach state-of-art results. The Transformers is a novel architecture developed to solve sequence-to-sequence tasks on TensorFlow platform[18]. Transformers provide hundreds of pretrained models to perform NLP tasks. Both the self-attention and deferentially weighting the significance of each part of the input data are involved in the library. To simplify code and offer convenience for PyTorch users[19], Hugging Face is built as an open-source Transformers community and a Hub for pre-trained models. Users only need to implement the model with a few lines code and fine-tune with an easy way. Hugging Face not only provide the functions in Transformers but also develop tokenizer library, and it's widely used in NLP research.

4.3 Models

In this paper, we applied different models to train our set. The model details are shown in Table 2.

Model name (Hugging Face path)	Parameters
ERNIE 1.0: "nghuyong/ernie-1.0"	Layer:12, Hidden:768, Heads:12, 102M parameters
XLNet: "hfl/chinese-xlnet-base"	Layer:12, Hidden:768, Heads:12, 117M parameters
RoBERTa: "liam168/c2-roberta-base-finetuned-dianping-chinese"	Layer:12, Hidden:768, Heads:12, 110M parameters

Table 2: Training model and parameters,

ERNIE is a word embedding and Bert based embedding developed by Baidu [20]. It is optimized for Sentiment analysis tasks in Chinese characters as well as in English if its mode is set to sentiment analysis. Based on the masking method of BERT, Ernie enhances its knowledge masking strategies to entity-level masking and phrase-level masking. Entities with important information in the sentences are masked in Entity-level strategy. Phrase with several words standing together as a conceptual unit are masked in Phrase-level strategy. It has good performance in multiple Chinese language tasks.

XLNet[21] is a generalized autoregressive pretraining method using the context word to predict the next word in two directions, either forward or backward. The XLNet Chinese[22] is applied in this paper to train our dataset. RoBERTa [23] give a choice of important BERT design choices and training strategies and alternatives making better downstream task performance than BERT.

To fine-tune our models, we use the class Trainer in the Hugging Face Transformers library [24].

5 Experiment

We are using existing datasets and existing models to perform a new type of task: predicting the amount of video viewer activity based on aggregated danmuku sentiment.

5.1 Model comparison

We conduct the experiment on the popular ChnSentiCorp dataset as well as our partially labeled danmaku data using the above chosen models.

We use the ChnSentiCorp dataset to train the models to gauge their performance on whether a relatively formal review is positive or negative. It contains both short opinions with several Chinese characters and long reviews in one sentence. We partition the dataset into training, evaluation, and test sets with 9600, 1200, 1200 records in each. When finetuning the model, we also searched for hyper-parameter space to optimize the model. The model with the best testing accuracy is selected.

Our finetuning results are shown in Table 3. It is noticed that with 3 epochs of training the loss decreases significantly, and validation accuracy approaches or even surpasses 90%. The result, although not as high as those shown in the ERNIE paper, is high enough to become a promising classifier for a binary positive/negative sentiment.

After using the finetuned model to predict the sentiment of Chinese sentence, we manually labeled part of the danmaku data with positive (1), negative (-1), and neutral (0). Neutral are usually the ones containing only some numbers, dates, or meaningless characters. Right now we treat the danmaku being positive when the classified emotion is excitement, and being negative when neutral or meaningless. The reason for this choice lay on the fact that although some of the phrase may

Model name	Accuracy on ChnSentiCorp	Accuracy on labeled danmaku data
ERNIE	94%	85%
XLNet	82%	73%
RoBERTa	83%	73%

Table 3: Finetuned Model Performance

be negative in daily use, they are not indicating the user have negative sentiment. However, there are some sentence which must be interpreted with the context of video. For example, satires is strongly correlated with the video content, or even previous danmaku; it is hard to retrieve the order of danmaku sent by different user. Currently, we manually remove any danmaku data with sign of satire either to the video or other danmaku to avoid any influence on the model weights.

Since the pretrained model itself may contain bias or weights, we firstly uses the raw model to predict the dataset and see if the distribution of prediction matches the distribution of dataset.

The number of neutral reviews are taking up 65% of our danmaku dataset while only about 10% the prediction of our model are positive, resulting accuracy is about 6%. The result shows that the initial embedding has no benefit in predicting the sentiment of the danmaku and the prediction prefers identifying danmaku as neutral.

In the future, we may want to find out what kind of danmaku are difficult to predict and add more kinds of emotion to our model to further classify the sentiment of the danmaku (or more directly, the sending user). Our study has already identified there is a correlation between the strength of the sentiment and the number of viewer. What kind of sentiment contributes to the number of viewer or a function of the strength of each sentiment to the number of viewer.

5.2 Empirical statistics

We then chose the ERNIE model to look into our empirical hypotheses. After aggragating danmakus by seconds, we predict a binary positive (1) or negative (-1) label for each second of danmaku in a video, for all 4950 videos in the VideoIC dataset. We then sum up the 1 and -1 labels for each video and all this the "net positive sentiment" (NetPos) of a video. We can then derive the proportion of the net posivie sentiment, NetPosProp, of a video by dividing NetPos using the total seconds of the video. Finally, we link this NetPosProp variable with the viewer activity data for the videos which we scraped from Bilibili ourselves.

After filtering out invalid or deleted videos and rows with errors, we get a dataframe of viewer activity and NetPosProp info for $N = 4503$ videos. Viewer activity includes "view","danmaku","reply","favorite","coin","share","like". "view" records how many users have clicked the video. "danmaku" and "reply" refer to time-instant comments in the video and comments below the video, respectively. And "favorite","coin","share","like" all represent the positive feedback from the users, such as how many times the video has been favorited, shared and liked. And "coin" specifically indicates the number of donations received by the video.

We first looked at the descriptive statistics of our focus variables, as shown in Figure 3. We can see that the viewer activity variables are all significantly positively correlated with each other, which indicates that these variables as scales for viewer interaction have a high reliability score. However, there are still qualitative differences in these variables, as each of them measure different kinds of user feedbacks for one video, as discussed above.

The critical variable NetPosProp, which is an output from sentiment analysis and an input for viewer activity prediction, looks like a Gaussian distribution ($M = 0.21, SD = 0.27$), as shown in Figure 4. Although we don't know the exact accuracy for the sentiment labels that NetPosProp is derived from, this Gaussian shape is a good signal that confirms our computed NetPosProp is reasonable.

We choose the viewer activity variable "coin" for our following inferential analysis. Users spend their virtual non-monetary, relatively scarce coins on a video to indicate their highest level of appreciation besides monetary donations. Therefore, it is an excellent measurement of positive viewer feedback. We divide the number of coins a video gets by its total number of clicks to get the CoinRatio.

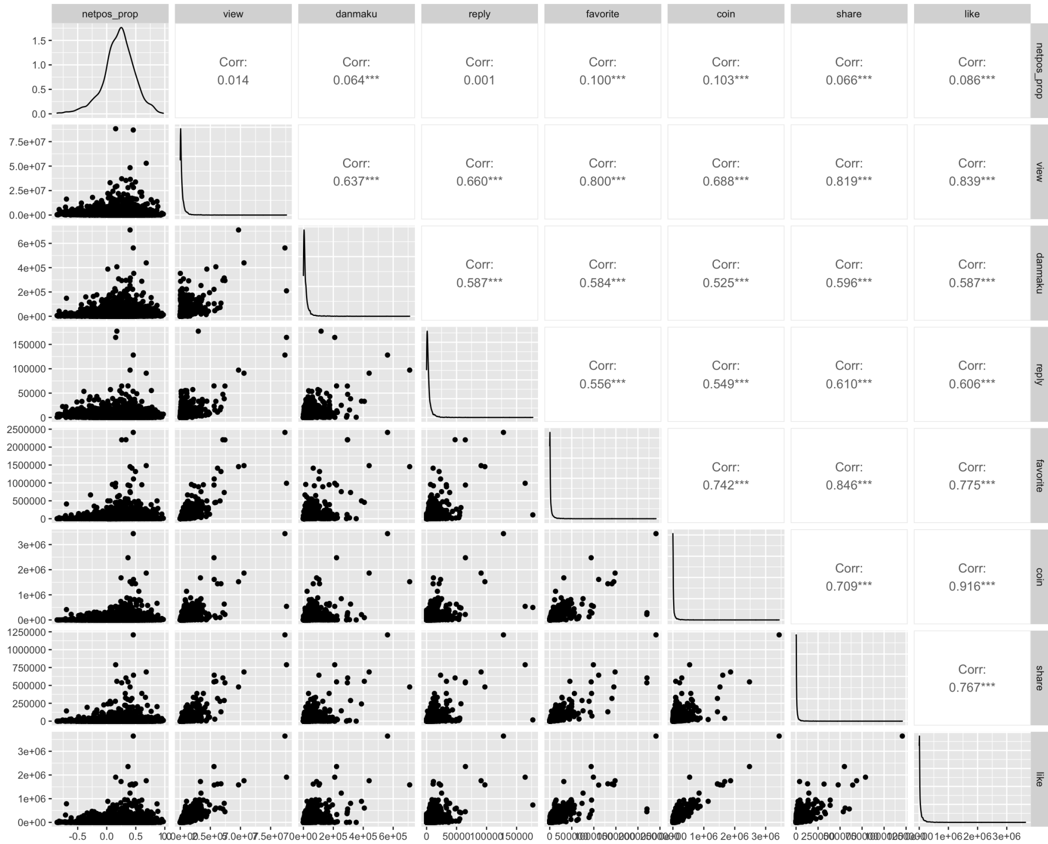

Figure 3: Bivariate statistics of focus variables

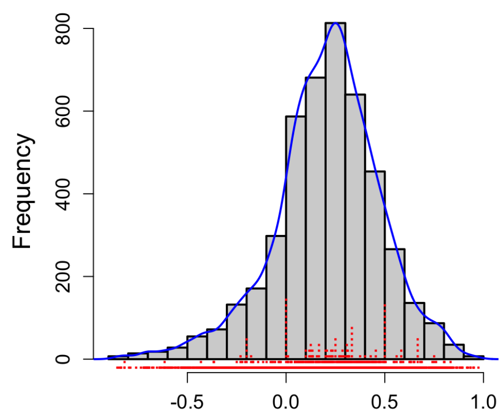

Figure 4: Univariate statistics of NetPosProp

We then performed a simple linear regression of CoinRatio by NetPosProp as shown in Figure 5. The effect of NetPosProp on CoinRatio ($b_1 = 0.023$) is significant ($se = 0.0014, F(1, 4499) = 270.2, p < .001, \eta_p^2 = .06$). This indicates that more positive sentiment over the video correlates with greater positive view feedback, which is expected and partly confirms with our hypothesis H_1, that sentiments that are more extreme (either positive or negative) from danmuku correlate with higher amounts of viewer activities.

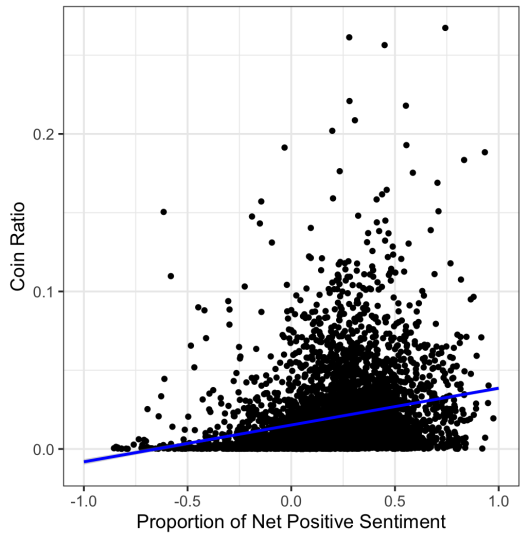

Figure 5: Linear regression of CoinRatio by NetPosProp

6 Discussion

Due to time constraints, we were unable to continue with our experiments and analyses.

In this paper, we attempt to find the relationship between sentiment intensity of danmaku and the hotness of the video. We employ the transfer learning in our research to classify the sentiment of the danmaku in the video. We uses am existing dataset on reviews to pre-train the model and further finetune the model with manually labeled data. The result shows that the reviews in relatively formal sentence can help training the classifier for danmaku data and we can apply the BERT-like model directly to danmaku. Further analysis on danmaku and video indicates a strong relationship between sentiment intensity and popularity. We can then build a regression model to predict the number of viewer using the sentiment intensity.

Currently the linear regression doesn't fully depict the correlation between NetPosProp and CoinRatio. Again, because we were short on time, we didn't get the chance to explore other regression models. We are still very excited that we went through this far, and this project leaves us a great list of intriguing future work directions.

7 Future Work

7.1 Data augmentation

In the future, we may want to filter out certain types of danmaku data, such as purely marking one's presence. Such danmaku may contribute to identify the hotness of the video by numbers, but whether these meaningless danmaku without specific sentiment is a powerful enough indicator of hotness and future popularity is still unknown. If there is more time, we may want to train a model to identify such sentence and research the relationship between them.

7.2 Semi- or self-supervised learning

The biggest problem in our research is deficit dataset and labeling. One possible approach is to manually label the remaining data but is very time consuming. Another approach is to create a script or build a model to label these data. Semi-supervised learning might be suitable for such task where the sentence or phrase in danmaku can be aggregated into multiple base phrase. (KNN center) If we can find a function to calculated the distance of phrase, we can use limited dataset containing multiple kind of sentiment to infer the label of the whole dataset.

7.3 New models

Currently we are using existing Transformer models to perform our danmaku sentiment analysis. These models are doing a fairly good job, bu We want to use the full value of the longitudinal timestamp information in the danmaku dataset. Therefore we should consider modifying model structure and design a new model that can capture the longitudinal information in our dataset.

7.4 More datasets

For us to look into our second hypothesis H_2 that danmaku sentiments can better predict view activity amount than sentiments from traditional comments, we would need additional datasets such as Twitch.tv Chat Log Data **??** and Trending YouTube Video Statistics (https://www.kaggle.com/datasnaek/youtube-new).

8 Acknowledgement

We greatly appreciate the feedback provided by Prof. Fred Sala and the TAs, Changho Shi and Harit Vishwakarma, for our project proposal. We are also thankful for the 4 extra days being granted by the deadline extension. Thank you for a great semester!

References

[1] Moegirlpedia users. danmuku - moegirlpedia. in chinese.

[2] John D. Cook. Chinese character frequency and entropy, oct 2019.

[3] Qingchun Bai, Qinmin Vivian Hu, Linhui Ge, and Liang He. Stories that big danmaku data can tell as a new media. *IEEE Access*, 7:53509–53519, 2019.

[4] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[5] Qingyuan LIN and Angela Lu. Why people participate in sending danmuku? a perspective from herding effect. 2021.

[6] Xixian Peng, Yuxiang Chris Zhao, and Hock-Hai Teo. Understanding young people's use of danmaku websites: the effect of perceived coolness and subcultural identi. *PACIS*, 252, 2016.

[7] Shan Sun, Feng Wang, and Liang He. Movie summarization using bullet screen comments. *Multimedia Tools and Applications*, 77(7):9093–9110, 2018.

[8] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

[9] Hanhoon Kang, Seong Joon Yoo, and Dongil Han. Senti-lexicon and improved naïve bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5):6000–6010, 2012.

[10] Huaxia Rui, Yizao Liu, and Andrew Whinston. Whose and what chatter matters? the effect of tweets on movie sales. *Decision support systems*, 55(4):863–870, 2013.

[11] Yung-Ming Li and Tsung-Ying Li. Deriving market intelligence from microblogs. *Decision Support Systems*, 55(1):206–217, 2013.

[12] Chien Chin Chen and You-De Tseng. Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4):755–768, 2011.

[13] Jonathan Ortigosa-Hernández, Juan Diego Rodríguez, Leandro Alzate, Manuel Lucania, Inaki Inza, and Jose A Lozano. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. *Neurocomputing*, 92:98–115, 2012.

[14] Duyu Tang, Bing Qin, and Ting Liu. Deep learning for sentiment analysis: successful approaches and future challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(6):292–303, 2015.

[15] Abdalraouf Hassan and Ausif Mahmood. Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)*, pages 705–710. IEEE, 2017.

[16] Weiying Wang, Jieting Chen, and Qin Jin. Videoic: A video interactive comments dataset and multimodal multitask learning for comments generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 2599–2607, 2020.

[17] Songbo Tan and Jin Zhang. An empirical study of sentiment analysis for chinese documents. *Expert Systems with applications*, 34(4):2622–2629, 2008.

[18] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow, Large-scale machine learning on heterogeneous systems, 11 2015.

[19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[20] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*, 2019.

[21] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

[22] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, Online, November 2020. Association for Computational Linguistics.

[23] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.

[24] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. Huggingface's transformers: State-of-the-art natural language processing. *CoRR*, abs/1910.03771, 2019.