Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231410

Research on Automatic Recognition of Rhetorical Questions' Types in Modern Chinese Texts

Gaoxiang LI[1], Yongquan LI
College of Chinese Language and Culture, Jinan University, Guangzhou, 510610, China

Abstract. Effective social dialogue patterns must understand rhetorical means of expression in a broad sense. Rhetorical questioning is a commonly used rhetorical expression that typically employs an interrogative form to convey a negative function. As the analysis progresses, we find that there is a continuum from interrogative to rhetorical question, with the degree of questioning decreasing and the degree of negation increasing. The rhetorical question allows for a response, and the form of the response reflects the addressee's understanding of the degree of negation. The article manually marked 1002 rhetorical questions by experts and divided them into four categories: Awakening, Questioning, Discovering, and Constantly Changing. By comparing CNN, RNN, Transformer, FastText, Baidu *PaddlePaddle* model, and CNN with FastText composite model, it was found that CNN alone can achieve an accuracy of 53.12%, with good prediction performance, and performs well in various types of rhetorical questions. Therefore, CNN can be applied in natural language processing of rhetorical questions in fields such as sentiment analysis, translation, and automatic writing.

Keywords. Chinese; rhetorical question; text classification; negative degree

1. Introduction

Effective social dialogue patterns must understand rhetorical means of expression in a broad sense. Rhetorical questioning is a commonly used rhetorical expression that typically employs an interrogative form to convey a negative function. It exhibits significant commonalities in both Chinese and foreign languages. Rhetorical Question ("fanwen ju") has three characteristics that distinguish them from genuine interrogative sentences: firstly, they use interrogative form but are not intended to seek information [1-4]; secondly, they make an assertion of opposite polarity, thereby negating the original statement [1,5-12]; and thirdly, they do not require a response [1,6,13]. However, as the analysis progresses, we find that there is a continuum from interrogative to rhetorical question, with the degree of questioning decreasing and the degree of negation increasing. The rhetorical question allows for a response, and the form of the response reflects the addressee's understanding of the degree of negation.

In the area of natural language processing (NLP), sentiment analysis refers to the utilization of natural language processing, text mining, and computational linguistics to identify and extract individuals' opinions, emotions, and perspectives on issues,

[1] Corresponding author: Gaoxiang LI, College of Chinese Language and Culture, Jinan University, Guangzhou, 510610, China; e-mail: dauphin94@126.com

subsequently categorizing them based on their polarity [17-20]. There are four primary approaches to sentiment analysis: sentiment lexicon and rule-based approaches, conventional machine learning-based approaches, deep learning-based approaches, and hybrid approaches.

In emotion recognition and analysis, identifying and understanding rhetorical structures in sentences can more accurately capture the speaker's emotional tendencies. For example, a rhetorical question may convey strong emotions such as anger, sarcasm, ridicule, or sadness. By analyzing the language features, intonation, and context of rhetorical questions, it is possible to delve deeper into the emotional state of the speaker, providing richer information for emotional recognition and analysis.

The article categorizes the differences in the degree of negation of rhetorical questions into four categories through expert manual classification. Machine learning algorithms such as Transformer are used to model the dataset, and then the optimal model is found to apply to sentiment analysis systems to deepen NLP's research on rhetorical questions.

2. Introduction to Relevent Terms

2.1. Modern Chinese

Modern Chinese refers to the language form mainly used in contemporary China, which is a tool for daily communication and expression of ideas among the Han people. Its language features include the following: syllables are the basic units of composition, with tones; Chinese characters form a writing system, and a character can have different meanings; a rich vocabulary that is compatible across ancient and modern times; flexible grammar and variable word order; emphasis on context and conciseness; and the presence of standard Mandarin to promote national language uniformity.

Modern Chinese can be understood in both a broad sense and a narrow sense, with the narrow sense specifically referring to Mandarin. Mandarin serves as the common language for the Han Chinese people and is the official national language of the country, excluding other Chinese dialects [2][15]. Our focus is specifically on rhetorical questions in Mandarin.

2.2. Rhetorical Question

Rhetorical Question is often translated as "fanwen ju" in Chinese. Rhetorical questions have been a hot topic of linguistic research in China and abroad, focusing on the syntactic, semantic, pragmatic, and rhythmic properties of rhetorical questions and how to distinguish them from "real" questions [5-7,15,24-30]. Subsequently, the researcher shifted their focus towards the discourse function of rhetorical questions [13,31-34]. The most important feature of the discourse function of rhetorical questions is negativity.

A rhetorical question is a rhetorical expression that signifies negation. Lv [11] points out that rhetorical questions and interrogatives share a common surface form but differ in their function; rhetorical questions serve as a form of negation. Scholars such as Huang and Liao [2] and Zhu [3] also argue that rhetorical questions, despite their interrogative form, do not convey an actual inquiry. They do not require an answer and, in fact, their form and intended meaning are completely opposite. Rhetorical questions serve as a means of expressing discursive negation. Shao [15] asserts that rhetorical questions serve as a manifestation of the speaker's inner dissatisfaction. Liu [12] puts forward that rhetorical questions convey four negative evaluative stances, that is reminding, opposing, unexpected, and reprimanding. Furthermore, rhetorical questions have negative functions such as challenging [35], retorting [31], and complaining [36].

As delineated in the Introduction, research on sentiment analysis has primarily concentrated on social media [37-39], online reviews [41-42], and business evaluations [43]. The expression of emotions in language is intricately complex, often employing rhetorical means to subtly convey the underlying emotional value. Rhetorical questioning is one commonly employed technique, yet it has received relatively less attention in the research of language computing models.

Rhetorical questions serve the function of negation, and the consideration of negation is essential in sentiment analysis. Negation can be expressed through the use of negative words such as "bu" (no) and "mei" (not), as well as through various other linguistic forms. For example, the sentence "Ni bushi jide ma?" which means that the listener forgot something, serves as a way to express negation towards the listener's behavior. However, identifying such negation can pose a significant challenge for computers [19][22]. The current research on negation sentiment analysis includes negation detection, negation scope detection [44], and negation computation. In negation identification, there is a predominant focus on explicit negation, while there is a scarcity of research on implicit negation recognition.

Negation sentiment analysis methods encompass approaches based on sentiment lexicons [40,45], traditional machine learning techniques [46], deep learning methodologies [47], and hybrid methods [48].

3. Experimental Design and Result Analysis

3.1. Corpus Data Sources and Classification

The data for this study is derived from two sources. The first source consists of collected Mandarin daily spoken dialogues, totaling 40 hours, which include audio and video recordings. The transcriptions were conducted based on the transcription system developed by Gail Jefferson [49] and tailored to the specific needs of Mandarin corpus transcription. The second source is the DMC corpus from the Ocean University of China, which comprises telephone conversations between friends or relatives. All participants in the conversations are native Mandarin speakers. All data were collected with the informed consent of the participants, and privacy information has been anonymized.

The sentence pattern "bushi … ma?" is one of the most common forms of rhetorical questions in Chinese and exhibits typical characteristics of a rhetorical question. However, it has a non-negative usage in certain contexts, such as interrogative, affiliation. Ranganath [50] demonstrates that a given question can be interpreted both as an information-seeking question and as a rhetorical question from the perspective of natural language processing. It is evident that there exists a continuum from interrogative to rhetorical questions, and the functional realization of a rhetorical question is influenced by the position within the conversational turn-taking and its response, which are contextual factors. This article investigates using "bushi … ma?" as an example.

We have found that "bushi ... ma?" serves functions such as negating, providing contextual background, and drawing conversational inferences. We categorize these functions as Questioning, Awakening, Discovering and Constantly Changing respectively, with a decreasing degree of negation. The variation in functional realization is related to the position within the conversational turn-taking and the type of response received. After manual annotation by experts, we obtained 1002 text fragments from conversations. In the following text, the four types of awakening, questioning, discovering, and Constantly Changing are referred to as types A, B, C, and D. Please refer to the table 1 below for details:

Table 1. 1002 sentence classification details.

ID	Type	Quantity	Proportion
A	Awakening	530	52.89%
B	Questioning	221	22.06%
C	Discovering	145	14.47%
D	Constantly Changing	106	10.58%
Sum		1002	100%

3.2. Analysis of Experimental Process

The code in this article mainly comes from the code hosting platform-https://github.com/649453932/Chinese-Text-Classification-Pytorch. We have made some modifications to make the codes of this project suitable for the research on the classification of rhetorical questions in the article. Tested the performance of Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Transformer, and FastText models in the classification of rhetorical questions.

3.2.1. All Data Experiment

In this experimental section, we trained the model with all data and tested its performance on the entire dataset. The experimental results are shown in table 2.

Table 2. Classification and prediction performance of all datasets.

Model	Type	Precision	Recall	F1-Score
CNN	A	0.9638	0.9963	0.9797
	B	0.9773	0.9729	0.9751
	C	0.9927	0.9379	0.9645
	D	0.9794	0.8962	0.936
RNN	A	0.5308	1	0.6935
	B	0	0	0
	C	0	0	0
	D	0	0	0
Transformer	A	0.5863	0.9925	0.7371
	B	0.8974	0.1584	0.2692
	C	0.8519	0.1586	0.2674
	D	0.8889	0.3019	0.4507
FastText	A	0.9148	0.985	0.9486
	B	0.9052	0.9502	0.9272
	C	0.9836	0.8276	0.8989
	D	0.961	0.6981	0.8087

For the convenience of observing the data, we drew figure 1 and 2.

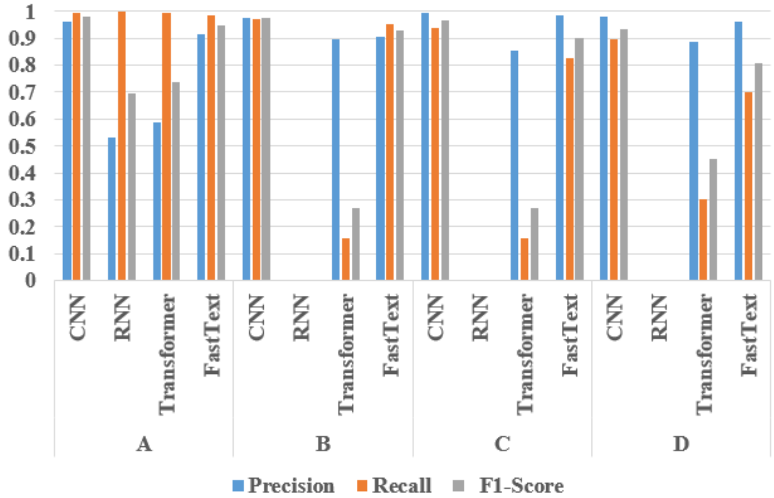

Figure 1. Performance of all datasets.

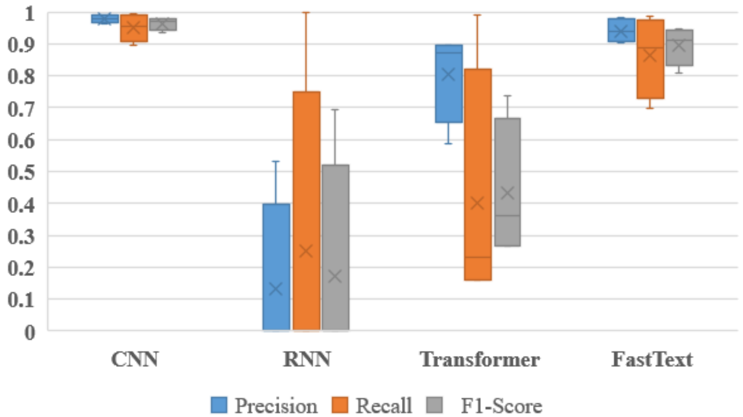

Figure 2. Performance of all datasets classified by models.

These data are measurements of the precision, recall, and F1-Score of different models in different categories. It can be seen that different models have significant differences in performance across different categories (A, B, C, D). According to overall accuracy, the order is CNN 0.9722 > FastText 0.9245 > Transformer 0.6163 > RNN 0.5308. Some models perform well in certain categories while performing poorly in other categories. Transformer has a significant change in F1-Score across different categories, while CNN has a relatively small change in F1-Score across different categories. This indicates that Transformer is more sensitive to data changes in different categories, while CNN has better adaptability to category changes. There may be imbalanced categories in the data, such as the significantly lower recall rate in categories B, C, and D in Transformer compared to other categories. This is because the sample size of these categories is relatively small, making it difficult for the model to correctly recognize these categories. RNN has zero accuracy, recall, and F1-Score in categories B, C, and D. This indicates that the model can't obtain any correct predictions in these categories. In a word, CNN performs best.

3.2.2. 70% Data Experiment

In this section, we adopted a stratified sampling method, randomly extracting 70% of the data for modeling, and using the remaining 30% of the data for performance evaluation. The performance of each model is shown in table 3.

Table 3. Classification and prediction performance of 70% dataset.

Model	Type	Precision	Recall	F1-Score
CNN	A	0.619	0.8357	0.7112
	B	0.4412	0.2542	0.3226
	C	0.125	0.0606	0.0816
	D	0.1176	0.0833	0.0976
RNN	A	0.5469	1	0.7071
	B	0	0	0
	C	0	0	0
	D	0	0	0
Transformer	A	0.556	0.9929	0.7128
	B	0.4	0.0339	0.0625
	C	0	0	0
	D	0	0	0
FastText	A	0.5437	0.6214	0.58
	B	0.2278	0.3051	0.2609
	C	0.5	0.0303	0.0571
	D	0.0667	0.0417	0.0513

For the convenience of observing the data, we also drew figure 3 and 4.

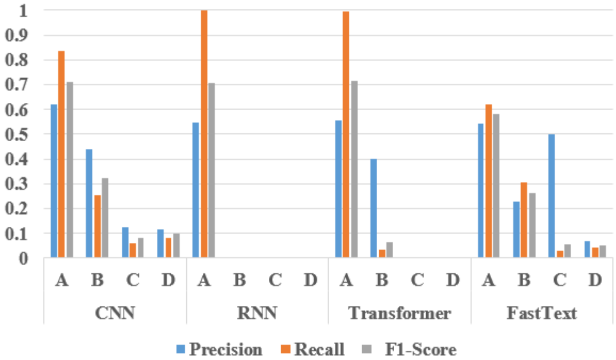

Figure 3. Performance of 70% datasets.

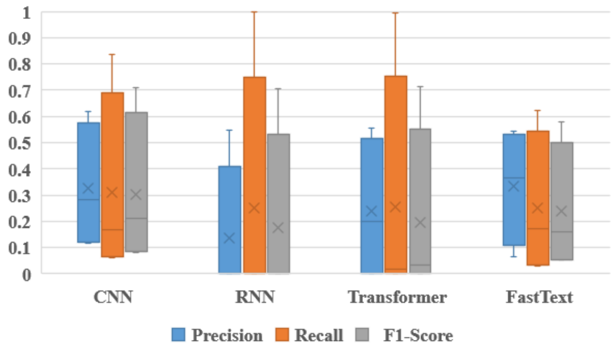

Figure 4. Performance of 70% datasets classified by models.

In terms of accuracy, the model performance is ranked as follows: Transformer 0.5508 > CNN 0.5312 > RNN 0.5469 > FastText 0.4180. But the Transformer predicted completely wrong in C and D, which means it can't predict the C and D types. Combining the overall performance with the predicted performance of each type, CNN can be regarded as the most efficient model. However, CNN does not perform as well as FastText in predicting type C. Therefore, in order to improve the performance of model prediction, CNN can be combined with FastText to predict categories, that is, CNN can be used to predict types A, B, and D, while FastText is used to determine types C.

3.2.3. Comparison

And put the same 70% dataset into the EasyDL (https://ai.baidu.com/easydl/) of Baidu PaddlePaddle. The platform conducts model training using the Big Model ERNIE (https://wenxin.baidu.com/). Finally, an accuracy rate of 55.7%, a recall rate of 25.0%, and a F1-Score performance of 17.9% were achieved. But it can only predict text of type A, and can't recognize any of B, C, or D. Therefore, in terms of the predictable types of the model, its performance is not as good as CNN.

To see how CNN and FastText combine to predict performance, we trained and tested the performance again on a given 70% dataset. The final accuracy of the composite model is 52.73%. This indicates that the composite model did not significantly improve accuracy, but rather decreased predictive performance as the model became more complex.

4. Conclusions

Effective social dialogue patterns must understand rhetorical means of expression in a broad sense. Rhetorical questioning is a commonly used rhetorical expression that typically employs interrogative form to convey a negative function. As the analysis progresses, we find that there is a continuum from question to rhetorical question, with the degree of questioning decreasing and the degree of negation increasing. The rhetorical question allows for a response, and the form of the response reflects the addressee's understanding of the degree of negation.

The article manually marked 1002 rhetorical questions by experts and divided them into four categories: Awakening, Questioning, Discovering and Constantly Changing. By comparing CNN, RNN, Transformer, FastText, Baidu *PaddlePaddle* model, and CNN with FastText composite model, it was found that CNN alone can achieve an accuracy of 53.12%, with good prediction performance, and performs well in various types of rhetorical questions. Therefore, CNN can be applied in NLP of rhetorical questions in fields such as sentiment analysis, translation, and automatic writing.

Acknowledgments

Thank you to the anonymous reviewers for their assistance in the writing process of this article. This article can't be separated from the careful guidance of my respected teacher Y. P. Zeng. Thank you for their selfless help and guidance.

References

[1] Yin S 2009 *Xiandai Hanyu Fanwenju Yanjiu (Research on rhetorical questions in Mandarin)* (Harbin: Heilongjiang University Press)
[2] Huang B and Liao X 2017 *Xiandai Hanyu (Modern Chinese)* (Beijing: Higher Education Press)
[3] Zhu D 1982 *Yufa Jiangyi (The Lecture notes of grammar)* (Beijing: The Commercial Press)
[4] Brown P and Levinson S C 1978 Universals in language usage: Politeness phenomena *Questions and politeness: strategies in social interaction* (Cambridge University Press) pp 56–311
[5] Freed A F 1994 The form and function of questions in informal dyadic conversation *Journal of Pragmatics* **21** 621–44
[6] Han C 2002 Interpreting interrogatives as rhetorical questions *Lingua* **112** 201–29
[7] Heritage J 2002 The limits of questioning: negative interrogatives and hostile question content *Journal of Pragmatics* **34** 1427–46
[8] Koshik I 2002 A conversation analytic study of yes/no questions which convey reversed polarity assertions *Journal of Pragmatics* **34** 1851–77
[9] Ljiljana P 1993 Negative polarity: Entailment and binding | SpringerLink *Linguistics and Philosophy* 149–80
[10] Sadock J M 1974 *Toward a linguistic theory of speech acts* (New York: Academic Press)
[11] Lv S 1982 *Zhongguo Wenfa Yaolue (Chinese Grammar Summary)* (Beijing: The Commercial Press)
[12] Liu Y 2010 *On Negative Rhetorical Questions and Wh-word Rhetorical Questions in Mandarin Conversations* Doctoral dissertation (Shanghai, China: Fudan University)
[13] Frank J 1990 You call that a rhetorical question?: Forms and functions of rhetorical questions in conversation *Journal of Pragmatics* **14** 723–38
[14] Richard A H 1975 The Meaning of Questions *Language* **51** 1–31
[15] Shao J 2014 *Xiandai Hanyu Yiwenju Yanjiu (Research on Interrogative Sentences in Mandarin)* (Beijing: The Commercial Press)
[16] Lu X and Ni B 2021 *Python3 Yuliaoku Jishu yu Yingyong (Corpus Techniques and Applications in Python3)* (Xiamen: Xiamen University Press)
[17] Sharif W, Samsudin N A, Deris M M and Naseem R 2016 Effect of negation in sentiment analysis *2016 Sixth International Conference on Innovative Computing Technology (INTECH)* 2016 Sixth International Conference on Innovative Computing Technology (INTECH) (Dublin, Ireland: IEEE) pp 718–23
[18] Asmi A and Ishaya T 2012 Negation Identification and Calculation in Sentiment Analysis The Second International Conference on Advances in Information Mining and Management pp 1–7
[19] G V and RM C 2012 Sentiment Analysis and Opinion Mining: A Survey *IJARCSSE* **2** 281–92
[20] Chen L, Guan Z, He J and Peng J 2017 A Survey on Sentiment Classification *Journal of Computer Research and Development* **54** 1150–70
[21] Zhong J, Liu W, Wang S and Yang H 2021 Review of Methods and Applications of Text Sentiment Analysis *Data Analysis and Knowledge Discovery* **5** 1–13
[22] Hussein D M E-D M 2018 A survey on sentiment analysis challenges *Journal of King Saud University - Engineering Sciences* **30** 330–8
[23] Lu W and Wang Y 2012 Review of Chinese text sentiment analysis *Application Research of Computers* **29** 2014–7
[24] Zahner K, Xu M, Chen Y, Dehé N and Braun B 2020 The prosodic marking of rhetorical questions in Standard Chinese *Speech Prosody 2020* Speech Prosody 2020 (ISCA) pp 389–93
[25] Caponigro I and Sprouse J 2007 Rhetorical Quesitons as Questions Proceedings of Sinn und Bedeutung 11 (Barcelona) pp 121–33
[26] Han C-H 1998 Deriving the Interpretation of Rhetorical Questions proceedings of WCCFL 16 pp 1–17
[27] Guo J 1997 Fanwenju de Yuyi Yuyong Tedian(The semantic and pragmatic characteristics of rhetorical questions) *Studies of the Chinese Language* 111–21
[28] Chang Y and Lan C 2008 Fanwenju Fouding Yuyong Gongneng Lun (The negation pragmatics of rhetorical questions) *SOCIAL SCIENTIST* 151-154+157
[29] Stainton R J and Ilie C 1996 What Else can I Tell You? A Pragmatic Study of English Rhetorical Questions as Discursive and Argumentative Acts *Language* **72** 429
[30] Liu Y and Tao H 2011 Indexing Evaluative Stances with Negative Rhetorical Interrogatives in Mandarin Conversation *Studies of the Chinese Language* 110-120+191
[31] Schaffer D 2005 Can rhetorical questions function as retorts?*1Is the Pope Catholic? *Journal of Pragmatics* **37** 433–60
[32] Raymond W G 2000 Irony in talk among friends *Metaphor and Symbol* **15** 5–27
[33] Roberts R M and Kreuz R J 1994 Why Do People Use Figurative Language? *Psychol Sci* **5** 159–63

[34] Petty R E, Cacioppo J T and Heesacker M 1981 Effects of rhetorical questions on persuasion: A cognitive response analysis. *Journal of Personality and Social Psychology* **40** 432–40
[35] Cerović M 2016 When suspects ask questions: Rhetorical questions as a challenging device *Journal of Pragmatics* **105** 18–38
[36] Monzoni C M 2009 Direct complaints in (Italian) calls to the ambulance: The use of negatively framed questions *Journal of Pragmatics* **41** 2465–78
[37] Awwalu J 2019 Hybrid N-gram model using Naïve Bayes for classification of political sentiments on Twitter *Neural Computing and Applications* 9207–20
[38] Kiritchenko S, Zhu X and Mohammad S M 2014 Sentiment Analysis of Short Informal Texts *Journal of Artificial Intelligence Research* **50** 723–62
[39] Ren Z, Peng Z, Lan Y, Zhang Q, Xia Y and Cui Y 2019 Emotional Tendency Prediction of Emergencies Based on the Portraits of Weibo Users *Journal of Intelligence* **38** 126–33
[40] Wu J and Lu K 2019 Chinese Weibo Sentiment Analysis Based on Multiple Sentiment Lexicons and Rule sets *Computer Applications and Software* **36** 93–9
[41] Lei M and Zhu M 2016 Applications of sentiment analysis in movie recommendation system *Computer Engineering and Applications* **52** 59-63+107
[42] He X, Yang W, Silamu W, Yang B, Yin Y and Li Y 2019 Sentiment analysis of tourist reviews combined with syntactic rules and CNN *Computer Engineering and Design* **40** 3306–12
[43] Griffith J 2020 Emotions in the Stock Market *Journal of Behavioral Finance* **21** 42–56
[44] Reitan J, Reitan J, Gambäck B and Bungum L 2015 Negation Scope Detection for Twitter Sentiment Analysis *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (Lisboa, Portugal: Association for Computational Linguistics) pp 99–108
[45] Dong L, Zhao F and Zhang X 2014 Analysing Propensity of Product Reviews Based on Domain Ontology and Sentiment Lexicon *Computer Applications and Software* **31** 104-108+194
[46] Councill I G, McDonald R and Velikovich L 2010 What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis Proceedings of the Workshop on Negation and Speculation in Natural Language Processing pp 51–9
[47] Singh P K and Paul S 2021 Deep Learning Approach for Negation Handling in Sentiment Analysis *IEEE Access* **9** 102579–92
[48] Zhang L, Tan Y, Zhu L and Dong W 2019 Analyzing the Features of Negative Sentiment Microblog *Information Studies:Theory & Application* **42** 132-137+170
[49] Jefferson G 2004 Glossary of Transcript Symbols with an Introduction *Conversation Analysis: Studies from the First Generation* (JohnBenjamins) pp 13–31
[50] Ranganath S, Hu X, Tang J, Wang S and Liu H 2018 Understanding and Identifying Rhetorical Questions in Social Media *ACM Trans. Intell. Syst. Technol.* **9** 1–22