

Jointly Identifying Rhetoric and Implicit Emotions via Multi-Task Learning

Xin Chen¹, Zhen Hai², Deyu Li^{1,3}, Suge Wang^{1,3*} and Dian Wang¹

1. School of Computer and Information Technology, Shanxi University, Taiyuan, China

2. R&D Center Machine Intelligence Technology Alibaba DAMO Academy, Singapore

3. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan, China

cx_simple@163.com, haiz0001@e.ntu.edu.sg

{lidy, wsg}@sxu.edu.cn, sxu_wdd@163.com

Abstract

Rhetorical implicit emotion identification is one of important and challenging tasks in natural language processing. We observe that each rhetoric may express certain evidence of semantic and syntactic patterns. Then, we design a gate mechanism based classification module to capture respective rhetorical representation and identify each rhetoric. Moreover, sentences carved with rhetoric tends to express emotions in subtle ways. We thus propose a new multi-task learning framework that can encode the categorical correlation between tasks to improve the performance of rhetoric and emotion identification problem. Experimental results validate the benefit of the proposed model over state-of-the-art baselines for rhetoric and emotion identification tasks. In addition, a new Chinese rhetorical implicit emotion dataset was constructed and will be released in this work.

1 Introduction

Rhetoric is formed by decorating the semantics and adjusting the structure of sentences in text (Kennedy, 2009; Wang, 2013). The sentences that express implicit emotions via rhetorics are commonly used in literary works and spiritual user reviews. They tend to express semantics at a high cognitive level and convey emotions in an obscure way. The example sentence below actually expresses the emotion of “love” and meanwhile consists of two rhetoric forms of parallelism and simile.

宽容是一块块木板，可以架起人与人之间的桥梁；宽容是一根根丝带，可以系住人与人之前的友谊；宽容是一团团火焰，可以融化人与人之间的隔阂。(Tolerance is a piece of board, which can build a bridge between people; Tolerance is a ribbon, which can tie the friendship

between people; Tolerance is a flame, which can melt the barriers between people.)

Clearly, one can understand and predict the emotion (“love”) and rhetoric (“simile” and “parallelism”) of the sentence by capturing the semantic meanings of content words (“tolerance”, “board”, “bridge”, “ribbon”, “friendship”) as well as the syntactic structure and pattern of parallelism clauses in the sentence (“Tolerance is ..., which can ...”). Semantics and syntax are pivotal cornerstones for rhetoric, and are attached to different rhetoric. Moreover, positive emotions are more likely to be expressed by rhetorical categories such as simile, parallelism and so on, while negative emotions are more commonly conveyed by the rhetorical categories such as rhetorical questions, irony and so on.

Sentences carved with rhetoric can convey emotions in an implicit way, without explicitly using emotional words, which pose new challenges for textual understanding and rhetorical emotion analysis in natural language processing.

Various efforts have been made to cope with emotion identification (Delbrouck et al., 2020; Chen et al., 2018; Klinger et al., 2018) and rhetoric detection (Liu et al., 2018; Wen et al., 2019; Mao et al., 2019). However, constructing heuristic rules and syntactical patterns for rhetoric identification is time-consuming and labor-intensive, and it may perhaps result in poor generalization to develop a specific detection model for each rhetoric category (Liu et al., 2018; Mao et al., 2019). To our knowledge, the correlation between rhetoric and emotion identification tasks has not been exploited in previous work.

In this work, we aim to cope with rhetoric and emotion analysis. Following previous studies, we formulate each of the rhetoric and emotion identification as a multi-label classification problem.

*Corresponding author: Suge Wang.

Then, we propose a new multi-task learning model called REI-MUL, which can jointly leverage the task correlation to address the rhetoric and emotion identification in a unified framework.

In particular, a pre-trained language model (Devlin et al., 2019) and a tree-structured long short-term memory networks (Tree-LSTMs) (Tai et al., 2015) are used to encode the semantic and syntactic representations of sentences, respectively. Taking the representations as input, two correlation modules are introduced to learn the categorical correlations between rhetoric and emotion tasks. One key benefit of REI-MUL is that it exploits the correlation between the two tasks, and may allow better learning of model parameters from real-life natural language data.

2 Methodology

2.1 Overview

Formally, let C be a labeled set of sentences, $C = \{(S_1, Y_1^r, Y_1^e), (S_2, Y_2^r, Y_2^e), \dots, (S_L, Y_L^r, Y_L^e)\}$, where S_l is an input sentence, while $Y_l^r = \{y_{l_1}^r, y_{l_2}^r, \dots, y_{l_R}^r\}$ and $Y_l^e = \{y_{l_1}^e, y_{l_2}^e, \dots, y_{l_E}^e\}$ denote the rhetoric label set (with R categories) and the emotion label set (with E categories) of the sentence S_l , respectively. Note that $y_{l_i}^r, y_{l_j}^e \in \{0, 1\}$, $y_{l_i}^r = 1$ and $y_{l_j}^e = 1$ mean the S_l is labeled as the i -th rhetoric and j -th emotion categories.

Based on the labeled set C , we develop a new multi-task learning model (REI-MUL), as shown in Figure 1, which can exploit task correlation to deal with rhetoric identification and emotion recognition in a unified framework.

2.2 Semantic Representation

As one of the state-of-the-art language representation models, the BERT model (Devlin et al., 2019) is employed to learn the semantic representations of sentences. Specifically, the contextualized representation of $[CLS]$ token is used as the encoding of whole sequence, which is denoted by $sr^{sem} \in \mathbb{R}^{d_1}$. For simplicity, we omit the subscript of input sentence “ l ” for subsequent sections.

2.3 Syntactic Representation

A well-known Tree-LSTMs (Tai et al., 2015) is exploited to encode syntactic representations of sentences. Word sense is a critical factor for structural representation. We thus represent the meaning of a word via connecting the *comprehensive* embedding

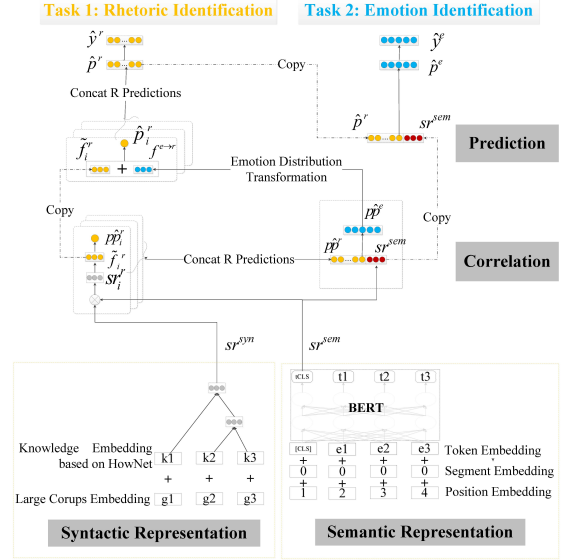


Figure 1: The architecture of the proposed multi-task learning model for rhetoric and emotion identification (REI-MUL). REI-MUL consists of four parts.

(Song et al., 2018) and the *sememes knowledge* embedding from SE-WRL Model (Niu et al., 2017).

Taking the word representations as input, the hidden state of root node in Tree-LSTMs model is taken as syntactic representation of each input sentence, as shown in Eqn (1).

$$sr^{syn} = Tree - LSTMs(X, X^{syn}) \quad (1)$$

where $X = \{x_1, x_2, \dots, x_N\}$, x_i means the input embedding of each word w_i , X^{syn} is the dependency parsing of S by using Stanford Parser, and $sr^{syn} \in \mathbb{R}^{d_2}$.

2.4 Correlation Layer

Both rhetorical and emotional modules of correlation layer are designed to obtain the label distributions of sentences, which can help to explicitly encode the semantic and syntactic correlations between rhetoric and emotion identification tasks.

2.4.1 Rhetoric Distribution

A gate mechanism (Cho et al., 2014) based classification model is proposed for each type of rhetoric, which assigns different weights to semantic and syntactic representations. Taking the i -th rhetoric identification as an example, the classifier can derive three types of representations, i.e., sentence representation, feature representation, and rhetoric distribution via Eqn (2)-(7).

$$r_i = \sigma(W_i^r * [sr^{syn}, sr^{sem}]) \quad (2)$$

<https://nlp.stanford.edu/software/lex-parser.html> (“version 4.0.0, Chinese”)

$$z_i = \sigma(W_i^z * [sr^{syn}, sr^{sem}]) \quad (3)$$

$$\tilde{sr}_i = \tanh(W_i^{\tilde{sr}} * [r_i \odot sr^{syn}, sr^{sem}]) \quad (4)$$

$$sr_i^r = (1 - z_i) * sr^{syn} + z_i * \tilde{sr}_i \quad (5)$$

where r_i , z_i and $sr_i^r \in \mathbb{R}^{d_2}$ denote the reset gate, update gate and sentence representation, respectively. σ and \tanh denote the sigmoid and hyperbolic tangent activation functions, and \odot is elementwise multiplication.

$$\tilde{f}_i^r = sr_i^r W_i^{fr} \quad (6)$$

where \tilde{f}_i^r is the feature representation for i -th rhetoric identification, and $\tilde{f}_i^r \in \mathbb{R}^{d_3}$.

$$pp_i^r = \tilde{f}_i^r W_i^{pr} \quad (7)$$

where pp_i^r refers to rhetoric distribution for i -th rhetoric identification, and $pp_i^r \in \mathbb{R}^1$.

Next, the respective R rhetoric distributions are concatenated as one comprehensive rhetoric distribution via Eqn (8), which would be used to compute the emotion distribution.

$$p\hat{p}^r = [pp_1^r || pp_2^r || \dots || pp_R^r] \quad (8)$$

where $p\hat{p}^r \in \mathbb{R}^R$.

2.4.2 Emotion Distribution

The emotional features are constructed by concatenating the semantic representation and rhetorical distribution of S . Taking the features as input, a classification model can be adopted to compute emotional distribution via Eqn (10), which would be next used to improve the rhetorical prediction task.

$$sr^e = [sr^{sem} || p\hat{p}^r] \quad (9)$$

$$p\hat{p}^e = sr^e W^{pe} \quad (10)$$

where $p\hat{p}^e$ denotes emotional distribution of input sentence S , and $p\hat{p}^e \in \mathbb{R}^E$.

2.5 Prediction Layer

2.5.1 Rhetoric Prediction Layer

An emotional distribution is incorporated into original rhetorical features, which is extracted from sentence representations. Then a sigmoid classifier is designed to take the correlated representation of a sentence as input, and predicts rhetorical probability scores. Formally, Eqn (11)-(13) calculate the transformed features based on emotional distribution ($f^{e \rightarrow r}$), new correlated features (f_i^r) and predicted probability (\hat{p}_i^r), respectively, for i -th rhetoric identification task.

$$f^{e \rightarrow r} = p\hat{p}^e W^{e \rightarrow r} \quad (11)$$

$$f_i^r = \tilde{f}_i^r + f^{e \rightarrow r} \quad (12)$$

$$\hat{p}_i^r = \sigma(f_i^r W_i^{pr}) \quad (13)$$

where $W^{e \rightarrow r}$ represents the feature transformation matrix of emotional distribution.

Next, the R predicted scores are concatenated as one rhetorical predicted probability distribution $\hat{p}^r \in \mathbb{R}^R$.

2.5.2 Emotion Prediction Layer

Similarly, a new correlated representation of input sentence is computed based on the predicted rhetorical distribution and semantic representation of the sentence via Eqn (14). And then we predict probability scores of emotion by a sigmoid classifier via Eqn(15).

$$sr^e = [sr^{sem} || \hat{p}^r] \quad (14)$$

$$\hat{p}^e = \sigma(sr^e W^{pe}) \quad (15)$$

where $\hat{p}^e \in \mathbb{R}^E$ denotes the emotional prediction probabilities of input sentence.

3 Experiments

3.1 Data

To our knowledge, there is a labeled dataset for emotion analysis of metaphorical expressions, which was constructed from a cognitive process (Zhang et al., 2018). However, there are no datasets available for joint rhetoric and emotion identification problem from the perspective of linguistics. We constructed the first dataset in Chinese, where each sentence may contain multiple rhetoric/emotion labels. Table 1 shows the detailed statistics of the dataset.

Category Name	Train	Val	Test	All
similes/metaphors	5,121	650	652	6,432
parallelism	1,903	238	234	2,375
personification	1,755	223	226	2,204
rhetorical questions	1,582	180	183	1,945
irony	649	82	82	813
Total	11,010	1,373	1,377	13,760
joy	133	22	22	177
love	6,329	784	808	7,921
anger	109	15	10	134
sadness	724	91	78	893
fear	97	12	13	122
disgust	2,042	256	239	2,537
surprise	58	5	14	77
Total	9,492	1,185	1,184	11,861
Sentence Number	9,456	1,182	1,182	11,820

Table 1: The statistics of the annotated dataset.

Each sentence of the dataset was collected from

	joy	love	anger	sadness	fear	disgust	surprise
similes/metaphors	83	5,446	25	406	88	376	23
parallelism	51	2,034	2	181	8	125	2
personification	29	1,876	17	149	24	71	7
rhetorical questions	21	272	87	232	9	1,282	43
irony	0	4	6	10	0	790	3

Table 2: The correlation between rhetorical and emotional classes on the dataset.

various sources such as literary works, textbooks, microblog, and websites. Three graduate students were hired to annotate the rhetorical and emotional labels of each sentence, where the inter-rater kappa coefficients are 0.848, 0.692, 0.757 for rhetoric annotation task and 0.458, 0.512, and 0.556 for emotion task, respectively. We randomly selected 80% of the annotated dataset as training data, while the rest 20% was equally divided as testing (test) and development (dev) data. All the models were evaluated on the same data split.

Table 2 shows the correlation statistics between rhetorical and emotional categories on the dataset. For example, the emotion “love” of a sentence is more likely to be expressed by rhetorical categories such as simile, parallelism and personification, while the emotion “disgust” is more commonly conveyed by the rhetorical categories such as rhetorical questions and irony.

3.2 Comparison Systems

We compared the proposed models with the following well-established baselines and ablation models.

- **CNN-Adversarial-MUL**: Liu et al. (2017) proposed an adversarial multi-task learning framework for text classification, alleviating the shared and private latent feature spaces from interfering with each other.
- **BERT-MUL**: Liu et al. (2019, 2020) presented a multi-task deep neural network for learning representations across multiple natural language understanding tasks.
- **RI-Single, EI-Single**: the proposed model is simplified a single task model for rhetoric or emotion identification.
- *w/o RheFusing, w/o EmoFusing*: the proposed model without predicted rhetorical or emotional distribution.
- *w/o Gate*: the proposed model without a gate mechanism.

- *w/o Tree*: the proposed model without syntactic representation.
- *w/o ComprehensiveEmb, w/o KnowledgeEmb*: the proposed model without *comprehensive* or *sememes knowledge* embedding in Tree-LSTMs.

3.3 Experimental Setting

The version “bert-base-chinese” was employed for the BERT module in the framework. The dimensionality of word embeddings in Tree-LSTMs was set 200. We carefully tuned and specified the values of the dimensionalities of Tree-LSTMs hidden state and feature representation of rhetoric identification as 64. We employ a heuristic threshold method to specify and predict rhetorical and emotional labels for each sentence, where different threshold scores were evaluated by the grid search method and selected as 0.88 and 0.73 on validation sets, respectively.

The joint one-versus-all cross-entropy loss was used to train the proposed multi-task learning. We applied the Adamax optimizer (Kingma and Ba, 2014) with learning rate as 0.00005, the batch size was set as 6, and the training epoch was set as 15. In addition, we applied dropout (Srivastava et al., 2014) on per layer of BERT model and prediction layers, and the value was 0.1.

3.4 Experimental Results

We compare the proposed method with state-of-the-art multi-task learning methods. Table 3 shows the results of rhetoric and emotion identification.

In particular, the proposed REI-MUL model achieves the best performance in terms of F1 for both tasks. Generally, the disadvantage of CNN-Adversarial-MUL may lie in the fact that the sentence representation by CNN module is not as good as the semantic representation of the BERT module of the proposed model. Compared to BERT-MUL, the proposed REI-MUL attains better results. This may suggest that the learning from associations

Model	Rhetoric(%)			Emotion(%)		
	P	R	F1	P	R	F1
CNN-Adversarial-MUL	89.56	87.80	88.67	87.16	86.57	86.86
BERT-MUL	92.11	93.32	92.71	88.27	87.67	87.97
REI-MUL	94.88	92.81	93.83	91.05	86.74	88.84

Table 3: The rhetoric and emotion identification results (P: Precision, R: Recall, F1: F1-score). Our proposed method is significantly better than the baselines given p-value 0.05 for the emotion identification.

Model	Rhetoric(%)			Emotion(%)		
	P	R	F1	P	R	F1
REI-MUL	94.88	92.81	93.83	91.05	86.74	88.84
RI-Single/EI-Single	92.84	93.79	93.31	89.74	87.76	88.74
<i>w/o RheFusing</i>	94.80	92.59	93.68	90.90	86.06	88.42
<i>w/o EmoFusing</i>	91.40	94.12	92.74	91.63	85.98	88.71
<i>w/o Gate</i>	93.77	92.88	93.32	88.63	88.26	88.45
<i>w/o Tree</i>	94.55	91.94	93.23	91.89	85.14	88.38
<i>w/o ComprehensiveEmb</i>	93.88	92.52	93.20	90.69	86.40	88.49
<i>w/o KnowledgeEmb</i>	92.66	93.54	93.10	88.95	88.34	88.64

Table 4: The ablation study results of rhetoric and emotion identification via the proposed model.

between rhetoric and emotion categories is really beneficial for rhetoric and emotion tasks. In addition, the improved results of rhetoric detection show the independent rhetorical classifiers for each rhetoric detection are effective.

3.5 Ablation Study

We designed the ablation study to evaluate the impact of different parts of the proposed REI-MUL model for rhetoric and emotion identification. Specifically, we removed each of the following modules from our proposed model, i.e., correlation between two tasks, independent rhetorical classifier based on gating mechanism, word embedding of Tree-LSTMs. Table 4 shows the comparison results of REI-MUL and reduced models on annotated dataset. Clearly, these results manifest remarkable capability of each part of the proposed model.

Table 5 shows the detailed identification results for each rhetorical category. Experimental results show that the proposed method performs well for minority categories, such as parallelism and rhetorical questions. In addition, the results for personification and irony are not as good as the rest, which may lie in their higher dependence on human cognition and pose a challenge for the identification.

Category	P	R	F1
similes/metaphors	95.50	97.55	96.51
parallelism	94.21	97.44	95.80
personification	91.44	75.66	82.81
rhetorical questions	96.24	97.81	97.02
irony	96.97	78.05	86.49
All (macro)	94.87	89.30	91.72

Table 5: The identification results via the proposed model for each rhetorical category (%).

4 Conclusion

In this paper, we have proposed a new multi-task and multi-label learning model, which can exploit the task correlation to jointly address the rhetoric and emotion identification in a unified framework. The experimental results show the benefit of the proposed model over the state-of-the-art baselines.

Acknowledgments

The authors would like to thank all anonymous reviewers for their valuable comments and suggestions which have significantly improved the quality of this work. The work is supported by the National Key Research and Development Project of China (2018YFB1005103) and the National Natural Science Foundation of China (61632011,62076158).

References

- Ying Chen, Wenjun Hou, Xiyao Cheng, and Shoushan Li. 2018. Joint learning for emotion classification and emotion cause detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 646–651.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734.
- Jean-Benoit Delbrouck, Noé Tits, Mathilde Brousmiche, and Stéphane Dupont. 2020. A transformer-based joint-encoding for emotion recognition and sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1–7.
- 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*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- George A Kennedy. 2009. *A new history of classical rhetoric*. Princeton University Press.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Roman Klinger, Orphee De Clercq, Saif Mohammad, and Alexandra Balahur. 2018. Iest: Wassa-2018 implicit emotions shared task. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 31–42.
- Lizhen Liu, Xiao Hu, Wei Song, Ruiji Fu, Ting Liu, and Guoping Hu. 2018. Neural multitask learning for simile recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1543–1553.
- Pengfei Liu, Xipeng Qiu, and Xuan-Jing Huang. 2017. Adversarial multi-task learning for text classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1–10.
- Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496.
- Xiaodong Liu, Yu Wang, Jianshu Ji, Hao Cheng, Xueyun Zhu, Emmanuel Awa, Pengcheng He, Weizhu Chen, Hoifung Poon, Guihong Cao, and Jianfeng Gao. 2020. The microsoft toolkit of multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 118–126.
- Rui Mao, Chenghua Lin, and Frank Guerin. 2019. End-to-end sequential metaphor identification inspired by linguistic theories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3888–3898.
- Yilin Niu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Improved word representation learning with sememes. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 2049–2058.
- Yan Song, Shuming Shi, Jing Li, and Haisong Zhang. 2018. Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 2, pages 175–180.
- Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.
- K S Tai, R Socher, and C D Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, volume 1, pages 1556–1566.
- Yang Wang. 2013. *Chinese Rhetoric*. Shanghai Jiao Tong University Press (in Chinese).
- Zhi Wen, Yang Li, Suge Wang, Jian Liao, and Xin Chen. 2019. Feature enhanced CNN for rhetorical questions identification. *Journal of Chinese information processing (in Chinese)*, 33(1):68–76.
- Dongyu Zhang, Hongfei Lin, Liang Yang, Shaowu Zhang, and Bo Xu. 2018. Construction of a chinese corpus for the analysis of the emotionality of metaphorical expressions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 144–150.