SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Cognition-Driven Multimodal Personality Classification

Xiaoya Gao, Jingjing Wang*, Shoushan Li, Min Zhang & Guodong Zhou

[1]*School of Computer Science and Technology, Soochow University, Suzhou* 215000*, China*

Appendix A Additional Evaluation Metrics

The detail macro-precision (P.) and macro-recall (R.) of various kinds of approaches are shown in Table A1 and Table A2 for better analyzing our proposed MASN approach. Here, P. and R. are the average of the precision/recall scores for all categories. From Table A1 and Table A2, we can see that the performances of the macro-precision and macro-recall for various approaches are similar. This is reasonable since the categories are balanced. In addition, our MASN approach could still perform better than all the baselines, which further justifies the effectiveness of our approach.

Table A1 Complementary evaluation metrics (i.e., macro-precision and macro-recall) for the approaches in Table 1.

	Approaches	**5 Binary Classification Tasks**										**All Traits**					
		Romantic		*Calm*		*Scornful*		*Gloomy*		*Aggressive*		Top-1		Top-5		Top-10	
		P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.
Unimodality (Text)	Char-RNN [8]	69.2	68.9	62.6	62.2	60.3	60.2	58.7	58.6	61.1	61.0	5.9	6.5	28.9	20.1	44.3	30.2
	HS-LSTM [26]	70.5	69.8	64.0	63.3	62.9	62.3	62.0	61.4	63.4	62.9	8.0	6.9	29.5	21.0	45.7	31.6
	BERT [6]	79.4	79.3	74.1	73.3	75.6	75.5	72.2	72.0	76.8	76.8	8.8	9.6	34.0	29.4	47.4	41.4
	MASN (Text)	80.3	80.2	76.7	74.4	76.5	76.5	73.7	73.2	79.2	78.4	9.8	10.9	33.6	29.7	46.9	41.6
Unimodality (Image)	SIFT+SVM [7]	54.8	54.7	52.2	52.2	43.8	43.9	48.7	48.8	46.5	46.6	0.1	0.5	2.7	1.9	9.2	4.2
	VGG [43]	58.3	57.9	56.6	55.6	56.6	56.1	59.9	57.3	52.6	52.6	1.5	1.3	1.9	2.2	4.2	4.4
	ResNet [7]	61.7	60.4	60.4	56.7	57.3	57.1	62.2	59.2	55.5	55.3	2.0	1.9	1.9	2.3	4.6	4.5
	MASN (Image)	63.6	62.5	61.4	58.9	59.8	59.2	64.4	61.0	62.4	58.6	1.0	2.4	2.5	2.9	4.8	5.7
Multimodality (Text+Image)	DAN [45]	73.6	73.6	64.6	64.5	63.9	63.8	63.5	62.8	67.6	66.4	7.8	8.5	33.9	24.4	47.0	35.6
	FMN [18]	72.1	71.8	64.2	63.3	69.2	65.4	56.9	56.1	61.0	60.4	5.9	6.2	30.5	20.3	44.8	30.8
	CoATT [46]	70.1	69.8	65.6	65.6	70.6	69.4	61.9	61.6	64.1	63.8	7.0	8.1	29.2	23.9	41.8	34.0
	UDMF [19]	71.8	71.7	65.7	65.6	66.4	66.3	61.0	61.0	65.0	64.7	7.2	8.4	31.4	24.7	44.1	35.3
	COMMA [28]	74.2	73.4	66.7	66.7	71.7	67.3	64.7	64.6	66.5	65.5	7.4	8.2	26.2	24.2	38.1	34.8
	BERT+ResNet	78.3	78.3	75.5	72.8	75.2	75.0	68.5	68.3	77.7	76.7	9.1	10.7	**36.2**	29.3	**49.0**	41.5
	ViLT [47]	79.3	79.2	74.5	74.4	78.6	78.6	72.0	72.0	79.6	79.3	11.0	11.1	36.8	29.9	49.1	41.7
	MASN (Random)	76.5	76.4	68.9	68.9	73.8	73.5	69.8	69.5	72.4	71.6	7.8	9.0	30.1	25.4	42.5	36.1
	MASN	**83.1**	**83.0**	**78.4**	**76.7**	**79.6**	**79.6**	**74.8**	**74.4**	**82.0**	**81.2**	**11.5**	**12.1**	34.3	**31.3**	47.5	**44.4**

* Corresponding author (email: djingwang@suda.edu.cn)

Table A2 Complementary evaluation metrics (i.e., macro-precision and macro-recall) for the approaches in Table 2.

Approaches	5 Binary Classification Tasks										All Traits					
	Romantic		*Calm*		*Scornful*		*Gloomy*		*Aggressive*		Top-1		Top-5		Top-10	
	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.
MASN	**83.1**	**83.0**	**78.4**	**76.7**	**79.6**	**79.6**	**74.8**	**74.4**	**82.0**	**81.2**	**11.5**	**12.1**	34.3	**31.3**	47.5	**44.4**
w/o state-sharing	80.4	79.3	73.0	70.0	77.6	74.5	72.6	72.0	75.9	75.9	10.2	11.5	**34.5**	30.8	47.3	42.8
w/o reward-sharing	78.6	78.3	75.9	72.2	77.0	75.7	72.9	72.6	77.7	77.6	10.4	11.2	34.2	30.0	47.5	42.5
w/o opinion-word selection	80.2	80.1	74.2	71.1	76.6	76.5	71.6	70.7	79.9	79.3	11.4	12.0	34.0	31.1	47.3	43.4
w/o image-region selection	80.3	80.2	76.1	73.3	75.6	75.5	70.1	69.5	78.8	78.5	10.4	11.8	34.3	30.7	47.1	42.6
using objects as image-regions	81.2	81.1	75.8	73.3	76.8	76.5	71.0	70.3	79.3	79.3	10.3	11.4	33.0	30.5	46.2	42.7
using soft-attention as selectors	81.1	80.2	74.2	71.1	72.9	72.4	72.1	72.0	80.6	80.2	10.8	11.2	**34.5**	29.7	**48.1**	42.1

Appendix B Additional Performance Comparison on The Normalized Dataset

To investigate the effect of the fuzzy boundary issue among similar traits, we normalize the PERSONALITY-CAPTIONS data by grouping some similar traits (e.g., *miserable* and *gloomy*). With this operation process, the number of traits is reduced from 215 to 100. On this basis, we investigate our MASN approach on this new dataset and the relevant results are shown in Table B1 and B2. From Table B1, we can see that the Top1 (Acc.) performance of our MASN approach is significantly improved from 12.1% in Table 2 to 20.8% in Table B1. This indicates that normalizing the dataset can alleviate the fuzzy boundary issue and thus boost the performance of CMPC. In addition, our MASN approach could still perform better than almost all baselines, which further justifies the effectiveness of our CMPC approach.

Table B1 Performance comparison of various approaches to CMPC on the normalized dataset.

	Approaches	Top-1				Top-5				Top-10			
		P.	R.	F1	Acc.	P.	R.	F1	Acc.	P.	R.	F1	Acc.
Unimodality (Text)	Char-RNN [8]	9.9	8.8	7.6	14.5	51.4	25.9	27.7	37.6	70.3	38.5	42.7	52.2
	HS-LSTM [26]	15.9	10.5	10.4	15.4	49.4	27.8	30.4	38.8	66.6	40.2	44.7	53.9
	BERT [6]	16.9	13.3	12.6	18.5	53.8	**37.5**	38.2	46.3	70.6	**52.5**	**55.5**	**62.9**
	MASN (Text)	19.9	13.7	13.5	20.4	60.5	34.5	38.1	46.0	73.7	47.5	53.2	60.0
Unimodality (Image)	SIFT+SVM [44]	0.4	0.8	0.3	0.8	8.6	4.3	2.9	4.1	28.8	9.2	7.3	8.9
	VGG [43]	0.3	1.0	0.8	3.1	4.0	4.9	4.1	17.0	9.1	10.0	9.1	30.0
	ResNet [7]	0.4	1.1	0.9	3.4	5.1	5.0	4.3	17.2	10.1	10.0	9.2	30.1
	MASN (Image)	2.3	1.3	1.2	4.1	7.2	5.5	5.0	17.9	11.5	10.5	9.7	31.4
Multimodality (Text+Image)	DAN [45]	15.9	11.3	11.0	16.3	53.7	28.8	31.4	39.8	69.7	40.7	45.5	53.9
	FMN [18]	11.2	8.9	8.0	14.7	53.9	24.7	26.7	37.2	72.3	37.4	41.9	52.1
	CoATT [46]	13.1	10.3	9.5	16.3	56.2	27.6	30.3	40.0	73.9	40.1	45.2	54.2
	UDMF [19]	16.9	10.9	10.8	16.8	56.6	29.1	32.1	41.1	71.9	41.8	47.2	55.6
	COMMA [28]	12.5	8.3	7.9	12.3	48.7	20.6	22.8	31.2	66.9	30.0	34.3	43.5
	BERT+ResNet	16.0	12.2	11.3	18.9	59.8	32.4	35.5	44.8	**77.6**	46.4	52.0	60.0
	ViLT [47]	19.3	14.6	14.1	19.5	59.0	35.0	37.7	45.7	73.1	48.5	53.3	60.3
	MASN (Random)	15.9	12.6	12.5	17.4	47.9	30.5	32.8	40.7	66.0	43.6	48.1	55.1
	MASN	**20.1**	**15.0**	**14.6**	**20.8**	**61.5**	35.7	**39.0**	**46.6**	74.3	49.1	54.5	61.0

Table B2 Ablation study of our proposed MASN approach on the normalized dataset.

Approaches	Top-1				Top-5				Top-10			
	P.	R.	F1	Acc.	P.	R.	F1	Acc.	P.	R.	F1	Acc.
MASN	**20.1**	**15.0**	**14.6**	**20.8**	61.5	**35.7**	**39.0**	**46.6**	74.3	49.1	**54.5**	**61.0**
w/o state-sharing	12.0	13.0	12.5	19.4	62.8	33.5	36.7	45.2	**76.1**	47.1	52.6	60.0
w/o reward-sharing	17.3	14.5	13.8	20.0	59.1	35.4	38.2	46.2	71.9	49.0	53.9	60.8
w/o opinion-word selection	19.7	14.6	14.1	20.2	57.3	35.5	38.2	45.9	72.3	**49.5**	54.3	**61.0**
w/o image-region selection	16.5	12.0	10.9	18.4	**62.9**	32.2	35.1	44.3	75.9	45.7	51.0	59.0
using objects as image-regions	17.0	13.8	13.1	19.7	59.3	35.2	38.1	46.2	73.9	48.8	53.8	60.7
using soft-attention as selectors	16.9	13.7	12.8	19.6	58.5	34.7	37.3	45.5	74.9	48.3	53.1	60.4

References

1 Lewis R Goldberg. An alternative "description of personality" : The big-five factor structure. Journal of personality and social psychology, 1990, 59(6):1216

2 Esteban Andrés Ríssola, Seyed Ali Bahrainian, Fabio Crestani. Personality recognition in conversations using capsule neural networks. In: Proceedings of Web Intelligence, Thessaloniki, 2019. 180-187

3 Yunan Li, Jun Wan, Qiguang Miao, et al. CR-Net: A deep classification-regression network for multimodal apparent personality analysis. International Journal of Computer Vision, 2020, 128:2763–2780

4 Charles S, Carver, Michael F Scheier. Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. Psychological bulletin, 1982, 92(1):111

5 Jingjing Wang, Jie Li, Shoushan Li, et al. Aspect sentiment classification with both word-level and clause-level attention networks. In: Proceedings of International Joint Conference on Artificial Intelligence, Stockholm, 2018. 4439–4445

6 Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. BERT: Pre-training of deep bidirec-tional transformers for language understanding. In: Proceedings of North American Chapter of the Association for Computational Linguistics, Minneapolis, 2019. 4171–4186

7 Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al. Deep residual learning for image recognition. In: Proceedings of Computer Vision and Pattern Recognition, Las Vegas, 2016. 770–778

8 Fei Liu, Scott Nowson, Julien Perez. A language-independent and compositional model for personality trait recognition from short texts. In: Proceedings of European Chapter of the Association for Computational Linguistics, Valencia, 2017. 754–764

9 Kosuke Yamada, Ryohei Sasano, Koichi Takeda. Incorporating textual information on user behavior for personality prediction. In: Proceedings of Association for Computational Linguistics, Florence, 2019. 177–182

10 Pierre-Hadrien Arnoux, Anbang Xu, Neil Boyette, et al. 25 Tweets to know you: A new model to predict personality with social media. In: Proceedings of International Conference on Web and Social Media, Montréal, 2017. 472-475

11 Xiangguo Sun, Bo Liu, Jiuxin Cao, et al. Who am I? personality detection based on deep learning for texts. In: Proceedings of International Conference on Communications, Kansas City, 2018. 1-6

12 Barbara Barbosa Claudino da Silva, Ivandré Paraboni. Personality recognition from facebook text. In: Proceedings of the Portuguese Language, Canela, 2018. 107-114

13 Pizzolli, Daniele, Strapparava, et al. Personality traits recognition in literary texts. In: Proceedings of Storytelling Workshop, 2019. 107-111

14 Leqi Liu, Daniel Preotiuc-Pietro, Zahra Riahi Samani,et al. Analyzing personality through social media profile picture choice. In: Proceedings of International Conference on Web and Social Media, Cologne, 2016. 211–220

15 Bruce Ferwerda, Marko Tkalcic. Predicting users' personality from instagram pictures: Using visual and/or content features?. In: Proceedings of User Modeling, Adaptation and Personalization, Singapore, 2018. 157-161

16 Noura Al Moubayed, Yolanda Vazquez-Alvarez, Alex McKay, et al. Face-based automatic personality perception. In: Proceedings of ACM-MM, Orlando, 2014. 1153–1156

17 Jia Xu, Weijian Tian, Yangyu Fan, et al. Personality trait prediction based on 2.5D face feature model. In: Proceedings of Cloud Computing and Security, Haikou, 2018. 611-623

18 Onno Kampman, Elham J. Barezi, Dario Bertero, et al. Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In: Proceedings of Association for Computational Linguistics, Melbourne, 2018. 606–611

19 Golnoosh Farnadi, Jie Tang, Martine De Cock, et al. User profiling through deep multimodal fusion. In: Proceedings of Web Search and Data Mining, Marina Del Rey, 2018. 171–179

20 Tao Lei, Regina Barzilay, Tommi S. Jaakkola. Rationalizing neural predictions. In: Proceedings of Empirical Methods in Natural Language Processing, Austin, 2016. 107-117

21 Hongyu Guo. Generating text with deep reinforcement learning. 2015. ArXiv:1510.09202

22 Qiuyuan Huang, Zhe Gan, Asli Çelikyilmaz, et al. Hierarchically structured reinforcement learning for topically coherent visual story generation. In: Proceedings of the Association for the Advance of Artificial Intelligence, Honolulu, 2019. 8465-8472

23 Jiwei Li, Will Monroe, Alan Ritter, et al. Deep reinforcement learning for dialogue generation. In: Proceedings of Empirical Methods in Natural Language Processing, Austin, 2016. 1192–1202

24 Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, et al. A hierarchical framework for relation extraction with reinforcement learning. In: Proceedings of the Association for the Advance of Artificial Intelligence, Honolulu, 2019. 7072–7079

25 Heng Wang, Shuangyin Li, Rong Pan, et al. Incorporating graph attention mechanism into knowledge graph reasoning based on deep reinforcement learning. In: Proceedings of Empirical Methods in Natural Language Processing, Hong Kong, 2019. 2623-2631

26 Tianyang Zhang, Minlie Huang, Li Zhao. Learning structured representation for text classification via reinforcement learning. In: Proceedings of the Association for the Advance of Artificial Intelligence, New Orleans, 2018. 6053–6060

27 Jun Feng, Heng Li, Minlie Huang, et al. Learning to collaborate: multi-scenario ranking via multi-agent reinforcement learning. In: Proceedings of World Wide Web, Lyon, 2018. 1939–1948

28 Tao Gui, Liang Zhu, Qi Zhang, et al. Cooperative multimodal approach to depression detection in twitter. In: Proceedings of the Association for the Advance of Artificial Intelligence, Honolulu, 2019. 110–117

29 Michael L. Littman. Markov games as a framework for multi-agent reinforcement learning. In: Proceedings of International Conference of Machine Learning, New Brunswick, 1994. 157–163

30 Richard S. Sutton, David A. McAllester, Satinder P. Singh, et al. Policy gradient methods for reinforcement learning with function approximation. In: Proceedings of Neural Information Processing Systems, Denver, 1999. 1057–1063

31 Yonghui Wu, Mike Schuster, Zhifeng Chen, et al. Google's neural machine translation system: bridging the gap between human and machine translation. 2016. ArXiv:1609.08144

32 Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need. In: Proceedings of Neural Information Processing Systems, Long Beach, 2017. 6000–6010

33 Tao Shen, Tianyi Zhou, Guodong Long, et al. Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling. In: Proceedings of International Joint Conference on Artificial Intelligence, Stockholm, 2018. 4345–4352

34 Jiasen Lu, Caiming Xiong, Devi Parikh, et al. Knowing when to look: Adaptive attention via a visual sentinel for image captioning. In: Proceedings of Computer Vision and Pattern Recognition, Honolulu, 2017. 3242–3250

35 Shaoqing Ren, Kaiming He, Ross B. Girshick, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proceedings of Neural Information Processing Systems, Montreal, 2015. 91-99

36 Richard S. Sutton, Andrew G. Barto. Reinforcement learning: An introduction. IEEE Trans, Neural Networks, 1998, 9(5):1054–1054

37 Serena Yeung, Vignesh Ramanathan, Olga Russakovsky, et al. Learning to learn from noisy web videos. In: Proceedings of Computer Vision and Pattern Recognition, Honolulu, 2017. 7455–7463
38 Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning, 1992, 8:229–256
39 Kurt Shuster, Samuel Humeau, Hexiang Hu, et al. Engaging image captioning via personality. In: Proceedings of Computer Vision and Pattern Recognition, Long Beach, 2019. 12516–12526
40 Xavier Glorot, Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of Artificial Intelligence and Statistics, Chia Laguna Resort, 2010. 249-256
41 Diederik P. Kingma, Jimmy Ba. Adam: A method for stochastic optimization. In: Proceedings of International Conference on Learning Representations, San Diego, 2015
42 Yiming Yang, Xin Liu. A re-examination of text categorization methods. In: Proceedings of SIGIR, Berkeley, 1999. 42–49
43 Karen Simonyan, Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In: Proceedings of International Conference on Learning Representations, San Diego, 2015
44 Murat Olgun, Ahmet Okan Onarcan, Kemal Özkan, et al. Wheat grain classification by using dense SIFT features with SVM classifier. Comput. Electron. Agric., 2016, 122:185–190
45 Hyeonseob Nam, Jung-Woo Ha, Jeonghee Kim. Dual attention networks for multimodal reasoning and matching. In: Proceedings of Computer Vision and Pattern Recognition, Honolulu, 2017. 2156–2164
46 Qi Zhang, Jinlan Fu, Xiaoyu Liu, et al. Adaptive co-attention network for named entity recognition in tweets. In: Proceedings of the Association for the Advance of Artificial Intelligence, New Orleans, 2018. 5674–5681
47 Wonjae Kim, Bokyung Son, Ildoo Kim. ViLT: Vision-and-Language Transformer Without Convolution or Region Supervision. 2021. ArXiv:2102.03334
48 Yangtao Zheng, Di Huang, Songtao Liu, et al. Cross-domain object detection through coarse-to-fine feature adaptation. In: Proceedings of Computer Vision and Pattern Recognition, Seattle, 2020. 13763-13772
49 Fei Yu, Jiji Tang, Weichong Yin, et al. ERNIE-ViL: Knowledge enhanced vision-language representations through scene graph. 2020. ArXiv:2006.16934
50 Di Qi, Lin Su, Jia Song, et al. ImageBERT: Cross-modal pre-training with large-scale weak-supervised image-text data. 2020. ArXiv:2001.07966