Shizhe Chen

Ph.D. Candidate, Renmin University of China
E-mail: cszhe1@ruc.edu.cn

RESEARCH INTERESTS

Affective Computing, Video Captioning, Deep Learning, Multimodal Machine Learning

EDUCATION

Renmin University of China *Sep 2015 - Present*
Ph.D., Computer Science
- Advisor: Prof. Qin Jin

Renmin University of China *Sep 2011 - Jul 2015*
B.S., Computer Science, Minor in Information Security
- Thesis: Emotion Recognition via Acoustic Signals
- Advisor: Prof. Qin Jin

RESEARCH EXPERIENCE

Multimodal Emotion Recognition

· Key contributor to the emotion recognition challenges including dimensional emotion prediction (AVEC 2015-2017, MediaEval 2016) and categorical emotion recognition (EmotiW 2016-2017, MEC 2016-2017)
· Proposed models to capture temporal and context for emotion recognition
· Explored the performance of multimodal features and proposed the conditional attention model to dynamically fuse multi-modalities

Video Natural Language Description

· Key contributor to the MSR Video to Language Challenge 2016-2017 and the NIST Trecvid Video to Text task. Our team has achieved the best performance in the challenges!
· Proposed the topic-guided caption models to address the topic diversity challenge in video captioning

COMPUTER SKILLS

Python, C, C++, JavaScript, Bash, MATLAB
TensorFlow, Pytorch, HTML, CSS, LaTeX, Markdown

CONFERENCE PUBLICATIONS

[1] **S. Chen**, J. Chen, Q. Jin, and A. Hauptmann. Video captioning with guidance of multimodal latent topics. In *ACM Multimedia*, 2017.

[2] Q. Jin, **S. Chen**, J. Chen, and A. Hauptmann. Knowing yourself: Improving video caption via in-depth recap. In *ACM Multimedia*, 2017.

[3] **S. Chen**, J. Chen, and Q. Jin. Generating video descriptions with topic guidance. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pages 5–13. ACM, 2017.

[4] X. Li, **S. Chen**, and Q. Jin. Facial action units detection with multi-features and-aus fusion. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 860–865. IEEE, 2017.

[5] S. Wang, W. Wang, J. Zhao, **S. Chen**, Q. Jin, S. Zhang, and Y. Qin. Emotion recognition with multimodal features and temporal models. In *ICMI*. ACM, 2017.

[6] **S. Chen**, X. Li, Q. Jin, S. Zhang, and Y. Qin. Video emotion recognition in the wild based on fusion of multimodal features. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 494–500. ACM, 2016.

[7] **S. Chen** and Q. Jin. Multi-modal conditional attention fusion for dimensional emotion prediction. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 571–575. ACM, 2016.

[8] **S. Chen**, Y. Dian, X. Li, X. Lin, Q. Jin, H. Liu, and L. Lu. Emotion recognition in videos via fusing multimodal features. In *Chinese Conference on Pattern Recognition*, pages 632–644. Springer, 2016.

[9] Q. Jin, J. Chen, **S. Chen**, Y. Xiong, and A. Hauptmann. Describing videos using multi-modal fusion. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 1087–1091. ACM, 2016.

[10] Q. Jin, C. Li, **S. Chen**, and H. Wu. Speech emotion recognition with acoustic and lexical features. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4749–4753. IEEE, 2015.

[11] **S. Chen**, Q. Jin, X. Li, G. Yang, and J. Xu. Speech emotion classification using acoustic features. In *Chinese Spoken Language Processing (ISCSLP)*, pages 579–583. IEEE, 2014.

WORKSHOP PUBLICATIONS

[12] **S. Chen**, Q. Jin, S. Wang, and J. Zhao. Multi-task learning for dimensional and continuous emotion recognition. In *ACM Multimedia AVEC workshop*, 2017.

[13] **S. Chen** and Q. Jin. Ruc at mediaeval 2016 emotional impact of movies task: Fusion of multimodal features. In *Proceedings of the MediaEval Workshop*, 2016.

[14] **S. Chen**, Y. Dian, and Q. Jin. Ruc at mediaeval 2016: Predicting media interestingness task. In *Proceedings of the MediaEval Workshop*, 2016.

[15] **S. Chen** and Q. Jin. Multi-modal dimensional emotion recognition using recurrent neural networks. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, pages 49–56. ACM, 2015.

AWARDS AND HONORS

Ranked 1^{st} in ACM Multimedia 2017 Video to Language Grand Challenge	*2017*
Ranked 1^{st} in NIST Trecvid Video to Text Task	*2017*
Ranked 5^{th} in ICMI 2017 EmotiW emotion recognition challenge	*2017*
National Scholarship for Ph.D. Students	*2016*
Ranked 1^{st} in ACM Multimedia 2016 Video to Language Grand Challenge	*2016*
Ranked 2^{nd} in AVEC 2016 Continuous Emotion Recognition Sub-challenge	*2016*
Ranked 2^{nd} in CCPR 2016 Multimodal Emotion Recognition Challenge	*2016*
Ranked 1^{st} in MediaEval 2016 Emotion Impact of Movies Subtask	*2016*
ACM Multimedia Student Travel Grant	*2016*
Second Prize in the Chinese Big Data Contest P2P Fraud Detection Sub-contest	*2015*
Second Prize in IBM Bluemix Cognitive Computation Development Contest	*2015*
Ranked 4^{th} in AVEC 2015 Continuous Emotion Recognition Challenge	*2015*
First Prize in National College Student Information Security Contest	*2014*
Second Prize in the Chinese Big Data Contest Baidu iErmu Sub-contest	*2014*
Meritorious Winner in American Mathematical Contest in Modeling	*2014*
National Second Prize in China Undergraduate Mathematical Contest in Modeling	*2013*
National Scholarship for Undergraduate Students	*2013*