IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 1, July 2023

Emotion Recognition using Deep Learning

Ms. S. Maheshwari, Dr. R. Bhuvana, Ms. S. Sasikala
Assistant Professor, Department of Computer Science
Agurchand Manmull Jain College, Meenambakkam, Chennai, India
maheshwari.s@amjaincollege.edu.in, bhuvana.r@amjaincollege.edu.in, sasikala.s@amjaincollege.edu.in

Abstract*: Speech Emotion Recognition (SER) is critical in Human computer engagement (HCI) because it provides a deeper knowledge of the situation and leads to better engagement. Various machine learning and Deep Learning (DL) methods have been developed over the past decade to improve SER procedures. In this research, we evaluate the features of speech then offer Speech Former++, a comprehensive structure-based framework for paralinguistic speech processing. Following the component relationship in the speech signal, we propose a unit encoder to efficiently simulate intra- and inter-unit information (i.e., frames, phones, and words). We use merging blocks to generate features at different granularities in accordance with the hierarchy connection, which is consistent with the structural structure in the speech signal. Rather than extracting spatiotemporal information from hand-crafted features, we investigate how to represent the temporal patterns of speech emotions using dynamic temporal scales. To that end, we provide Temporal-aware bI- direction Multi-scale Network (TIM-Net), a unique temporal emotional modelling strategy for SER that learns multi-scale contextual affective representations from different time scales. Unweighted Accuracy (UA) of 65.20% and Weighted Accuracy (WA) of 78.29% are accomplished using signal features in low- and high-level descriptions, as well as various deep neural networks and machine learning approaches.*

Keywords: Human computer engagement, Deep Learning, Paralinguistic, Multi-scale Network, Weighted Accuracy

REFERENCES

[1] B. Moore, L. Tyler, and W. Marslen-Wilson, "Introduction. The perception of speech: from sound to meaning," Philosophical transactions of the Royal Society of London. Series B, Biological sciences, vol. 363, no. 1493, pp. 917–921, Mar. 2008.

[2] K. Tokuda, T. Kobayashi, and S. Imai, "Speech parameter generation from HMM using dynamic features," in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 1995, pp. 660–663.

[3] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," IEEE Transactions on Signal Processing, vol. 46, no. 4, pp. 886–902, 1998.

[4] B. Schuller, G. Rigoll, and M. Lang, "Hidden Markov model-based speech emotion recognition," in 2003 International Conference on Multimedia and Expo. ICME '03. Proceedings, vol. 1, 2003, pp. I–401.

[5] J. Cichosz and K. Slot, "Emotion recognition in speech signal using emotion-extracting binary decision trees," Proceedings of affective computing and intelligent interaction, 2007.

[6] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, and M. Huang, "Augmenting end-to- end dialogue systems with commonsense knowledge," in Proceedings of the AAAI Conference on Artificial Intelligence, 2018, pp. 4970–4977.

[7] L. Chen, C. Chang, C. Zhang, H. Luan, J. Luo, G. Guo, X. Yang, and Y. Liu, "L2 learners' emotion production in video dubbing practices," in IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2019, pp. 7430–7434.

[8] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of german emotional speech," in Proceedings of Ninth European Conference on Speech Communication and Technology, 2005.

DOI: 10.48175/IJARSCT-12004

[9] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, “Iemocap: Interactive emotional dyadic motion capture database,” Language Resources and Evaluation, vol. 42, no. 4, p. 335, 2008.

[10] P. Zhan and M. Westphal, “Speaker normalization based on frequency warping,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 1997, pp. 1039– 1042.

[11] Zixuan Peng, Yu Lu, Shengfeng Pan, and Yunfeng Liu, “Efficient speech emotion recognition using multi-scale CNN and attention,” in ICASSP 2021, Toronto, ON, Canada, June 6- 11, 2021. 2021, pp. 3020–3024, IEEE.

[12] Linhui Sun, Sheng Fu, and Fu Wang, “Decision tree SVM model with Fisher feature selection for speech emotion recognition,” EURASIP J. Audio Speech Music. Process., vol. 2019, pp. 2, 2019.

[13] Luefeng Chen, Wanjuan Su, Yu Feng, Min Wu, Jinhua She, and Kaoru Hirota, “Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction,” Inf. Sci., vol. 509, pp. 150–163, 2020.

[14] Jiaxin Ye, Xin-Cheng Wen, Xuan-Ze Wang, Yong Xu, Yan Luo, Chang-Li Wu, Li-Yan Chen, and Kunhong Liu, “GMTCNet: Gated multi-scale temporal convolutional network using emotion causality for speech emotion recognition,” Speech Commun., vol. 145, pp. 21–35, 2022.

[15] J Ancilin and A Milton, “Improved speech emotion recognition with Mel frequency magnitude coefficient,” Applied Acoustics, vol. 179, pp. 108046, 2021.

[16] Arya Aftab, Alireza Morsali, Shahrokh Ghaemmaghami, et al., “LIGHT-SERNET: A lightweight fully convolutional neural network for speech emotion recognition,” in ICASSP 2022, Virtual and Singapore, 23-27 May 2022. 2022, pp. 6912–6916, IEEE.

[17] C. Gorrostieta, R. Lotfian, K. Taylor, R. Brutti, and J. Kane, “Gender de-biasing in speech emotion recognition,” in Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH), Graz, Austria: ISCA, 2019, pp. 2823–2827.

[18] R. Bommasani et al., “On the opportunities and risks of foundation models,” arXiv preprint arXiv:2108.07258, 2021.

[19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in Proceedings of the International Conference on Machine Learning (ICML), Vienna, Austria (virtual), 2020, pp. 1597– 1607.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 2017, pp. 5998–6008.

DOI: 10.48175/IJARSCT-12004