











communication is one of the fundamental tools for emergency management. Exploring the same CNN model for multilingual short text categorization might help to reduce people anxiety and uncertainty during global emergency.

Currently there are three works in progress. The first work is to collect more data from online social media about a global topic which generate posts in different languages and create a more robust multilingual annotated dataset; the second work is to improve CNN (1+1) performance by model generalization and architecture change, and the last work is to use Recurrent Neural Network (RNN) to process multilingual short text. More contributions are expected as: generate a robust annotated dataset in English, Japanese and Portuguese to be used in other text categorization researches and improve our knowledge about CNN and RNN for multilingual text processing.

## References

- [1] C. Chew and G. Eysenbach, Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak, *PloS one*, 5( 11), page e14118, November 2010.
- [2] C. Corley, D. Cook, A. Mikler and K. Singh, Text and structural data mining of influenza mentions in web and social media, *International Journal of Environmental Research and Public Health*, 7:596-615, February 2010.
- [3] K. Tefrie and K. Sohn, Automated disease outbreak detection and analysis, proceeding of the *Information Science and Applications (ICISA) 2016*, pages 985-993, Springer, Singapore, 2016.
- [4] K. Hu, S. Bianco, S. Edlund and J. Kaufman, The impact of human behavioral changes in 2014 West Africa Ebola outbreak, proceeding of the *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, vol. 9021, pages 75-84, Springer, Cham, March 2015.
- [5] T. Simon, A. Goldberg and B. Adini, Socializing in emergencies—A review of the use of social media in emergency situations, *International Journal of Information Management*, 35:609-619, October 2015.
- [6] F. Sebastiani, Machine learning in automated text categorization, *ACM computing surveys (CSUR)*, 34:1-47, March 2002.
- [7] R. Johnson and T. Zhang, Effective use of word order for text categorization with convolutional neural networks, *arXiv preprint arXiv:1412.1058*, December 2014.
- [8] C. Caragea, A. Silvescu and A. Tapia, Identifying informative messages in disaster events using convolutional neural networks, proceeding of the *International Conference on Information Systems for Crisis Response and Management*, pages 137-147, May 2016.
- [9] Y. Kim, Convolutional neural networks for sentence classification, *arXiv preprint arXiv:1408.5882*, August 2014.
- [10] X. Zhang, J. Zhao and Y. LeCun, Character-level convolutional networks for text classification, proceedings of the *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 649-657, 2015.
- [11] M. Sato, R. Orihara, Y. Sei, Y. Tahara and A. Ohsuga Minato, Japanese text classification by character-level deep ConvNets and transfer learning, proceeding of the *International Conference on Agents and Artificial Intelligence*, pages 175-184, SCITEPRESS, 2017.
- [12] J. Wang, Z. Wang, D. Zhang and J. Yan, Combining knowledge with deep convolutional neural networks for short text classification, proceedings of the *26th International Joint Conference on Artificial Intelligence*, pages 2915-2921. AAAI Press, August 2017.
- [13] Y. Bengio, R. Ducharme, P. Vincent and C. Jauvin, A neural probabilistic language model, *Journal of Machine Learning Research*, 3:1137-1155, February 2003.
- [14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, Distributed representations of words and phrases and their compositionality, proceedings of the *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 3111-3119, 2013.
- [15] X. Zhang and Y. LeCun, Which encoding is the best for text classification in Chinese, English, Japanese and Korean?, *arXiv preprint arXiv:1708.02657*, August 2017