

Detecting Locations from Twitter Messages

Invited Talk

Diana Inkpen

University of Ottawa
School of Electrical Engineering and Computer Science
800 King Edward Avenue, Ottawa, ON, K1N6N5, Canada

Diana.Inkpen@uottawa.ca

1 Extended Abstract

There is a large amount of information that can be extracted automatically from social media messages. Of particular interest are the topics discussed by the users, the opinions and emotions expressed, and the events and the locations mentioned. This work focuses on machine learning methods for detecting locations from Twitter messages, because the extracted locations can be useful in business, marketing and defence applications (Farzindar and Inkpen, 2015).

There are two types of locations that we are interested in: location entities mentioned in the text of each message and the physical locations of the users. For the first type of locations (task 1), we detected expressions that denote locations and we classified them into names of cities, provinces/states, and countries. We approached the task in a novel way, consisting in two stages. In the first stage, we trained Conditional Random Field models with various sets of features. We collected and annotated our own dataset for training and testing. In the second stage, we resolved cases when more than one place with the same name exists, by applying a set of heuristics (Inkpen et al., 2015).

For the second type of locations (task 2), we put together all the tweets written by a user, in order to predict his/her physical location. Only a few users declare their locations in their Twitter profiles, but this is sufficient to automatically produce training and test data for our classifiers. We experimented with two existing datasets collected from users located in the U.S. We propose a deep learning architecture for the solving the

task, because deep learning was shown to work well for other natural language processing tasks, and because standard classifiers were already tested for the user location task. We designed a model that predicts the U.S. region of the user and his/her U.S. state, and another model that predicts the longitude and latitude of the user's location. We found that stacked denoising auto-encoders are well suited for this task, with results comparable to the state-of-the-art (Liu and Inkpen, 2015).

2 Biography

Diana Inkpen is a Professor at the University of Ottawa, in the School of Electrical Engineering and Computer Science. Her research is in applications of Computational Linguistics and Text Mining. She organized seven international workshops and she was a program co-chair for the AI 2012 conference. She is in the program committees of many conferences and an associate editor of the Computational Intelligence and the Natural Language Engineering journals. She published a book on Natural Language Processing for Social Media (Morgan and Claypool Publishers, Synthesis Lectures on Human Language Technologies), 8 book chapters, more than 25 journal articles and more than 90 conference papers. She received many research grants, including intensive industrial collaborations.

Acknowledgments

I want to thank my collaborators on the two tasks addressed in this work: Ji Rex Liu for his implementation of the proposed methods for both tasks and Atefeh Farzindar for her insight on the first task. I also thank the two annotators of the corpus for task 1, Farzaneh Kazemi and Ji Rex Liu.

This work is funded by the Natural Sciences and Engineering Research Council of Canada.

References

- Atefeh Farzindar and Diana Inkpen. 2015. *Natural Language Processing for Social Media*. Morgan and Claypool Publishers. Synthesis Lectures on Human Language Technologies.
- Diana Inkpen, Ji Liu, Atefeh Farzindar, Farzaneh Kazemi and Diman Ghazi. 2015. Location Detection and Disambiguation from Twitter Messages. In *Proceedings of the 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2015)*, LNCS 9042, Cairo, Egypt, pp. 321-332.
- Ji Liu and Diana Inkpen. 2015. Estimating User Location in Social Media with Stacked Denoising Auto-encoders. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, NAACL 2015*, Denver, Colorado, pp. 201-210.