UNI
FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Department of Informatics

University of Fribourg
Department of Informatics
Bd. de Pérolles 90
CH-1700 Fribourg
Phone. +41 26 300 84 65
Fax +41 26 300 97 26
http://exascale.info/

University of Fribourg, Dept. of Informatics, CH-1700 Fribourg

Louis Mueller
louis.mueller@students.unibe.ch
Bern
Switzerland

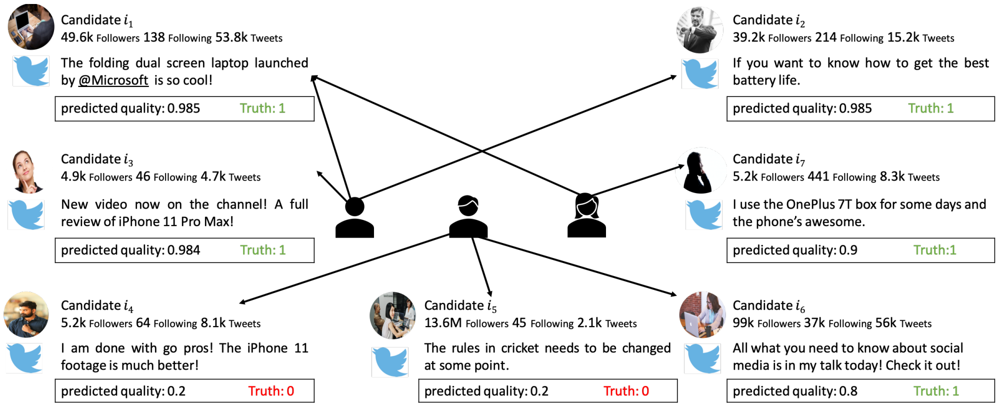
Dr. Mourad Khayati, Ines Arous
Senior researcher
Phone +41 26 300 84 60
Fax +41 26 300 97 26
mkhayati@exascale.info

Fribourg, March 15, 2020

Multiclass Classification of Open-ended Answers:

Work overview:

Open-ended question-answering is an important crowdsourcing task which consists of workers providing answers as free text such as, names of influencers, reviews or image description. OpenCrowd [1] is a recent technique that uses open-ended answers to identify social influencers in Twitter. It combines social features with the reliability of workers to effectively identify influencers (see Figure 1).



Candidate i_1
49.6k Followers 138 Following 53.8k Tweets
The folding dual screen laptop launched by @Microsoft is so cool!
predicted quality: 0.985 Truth: 1

Candidate i_2
39.2k Followers 214 Following 15.2k Tweets
If you want to know how to get the best battery life.
predicted quality: 0.985 Truth: 1

Candidate i_3
4.9k Followers 46 Following 4.7k Tweets
New video now on the channel! A full review of iPhone 11 Pro Max!
predicted quality: 0.984 Truth: 1

Candidate i_7
5.2k Followers 441 Following 8.3k Tweets
I use the OnePlus 7T box for some days and the phone's awesome.
predicted quality: 0.9 Truth:1

Candidate i_4
5.2k Followers 64 Following 8.1k Tweets
I am done with go pros! The iPhone 11 footage is much better!
predicted quality: 0.2 Truth: 0

Candidate i_5
13.6M Followers 45 Following 2.1k Tweets
The rules in cricket needs to be changed at some point.
predicted quality: 0.2 Truth: 0

Candidate i_6
99k Followers 37k Following 56k Tweets
All what you need to know about social media is in my talk today! Check it out!
predicted quality: 0.8 Truth: 1

Figure 1: Example of influencer detection by OpenCrowd in the IT domain.

The main goal of this thesis is to extend the capabilities of OpenCrowd to perform a more fine-grained identification of influencers. For instance, we are interested in classifying real influencers into classes such as established, emerging or a hub. To achieve such a goal, we propose to i) collect social features from candidate influencers, ii) identify the features that can allow to learn the properties of a fine-grained classification and iii) extend OpenCrowd to support such features.

Work tasks:

1. Familiarize yourself with aggregation (truth inference) in crowdsourcing [2].
2. Compute the word embeddings (or use pre-trained embeddings such as Glove [3], word2vec[4] and Bert[5]) to represent the tweets of candidate influencers and classify them into one of the categories (emerging, established, hub or other).
3. Use the collected properties of workers (knowledge in fashion, connectivity to social media and engagement to fashion content on social media) to train a neural network to learn worker reliability priors.
4. Modify the distributions used in OpenCrowd to take into account the multi class classification and derive the corresponding rules. A potential technique to use is Gibbs sampling [6].
5. Evaluate the performance of the implemented techniques against existing answer aggregation methods in fashion and reviews datasets [7].
6. Write a thesis that describes the implemented technique and the result of the experiments.

References:

1. Ines Arous, Jie Yang, Mourad Khayati, and Philippe Cudré-Mauroux. “OpenCrowd: A Human-AI Collaborative Approach for Finding Social Influencers via Open-Ended Answers Aggregation.” In Proceedings of the Web Conference (WWW 2020). Taipei, Taiwan, 2020.[https://exascale.info/assets/pdf/arous2020www.pdf]. Code of OpenCrowd is available here: https://github.com/eXascaleInfolab/opencrowd
2. Zheng, Y., Li, G., Li, Y., Shan, C. and Cheng, R., 2017. Truth inference in crowdsourcing: Is the problem solved?. Proceedings of the VLDB Endowment.
3. Pennington, J., Socher, R. and Manning, C.D., 2014, October. Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)
4. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems
5. Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.
6. Yildirim, I., 2012. Bayesian inference: Gibbs sampling. Technical Note, University of Rochester.
7. Dataset from Amazon: http://snap.stanford.edu/data/

Starting date of thesis: 16.03.2020

Ending date of thesis: TBD

University of Fribourg
Department of Informatics
eXascale Infolab

Mourad Khayati, Ines Arous