Le-kui Zhou, Si-liang Tang, Jun Xiao, Fei Wu, Yue-ting Zhuang, 2017. Disambiguating named entities with deep supervised learning via crowd labels. *Frontiers of Information Technology & Electronic Engineering*, **18**(1):97-106. http://dx.doi.org/10.1631/FITEE.1601835

Disambiguating named entities with deep supervised learning via crowd labels

Key words: Named entity disambiguation; Crowdsourcing; Deep learning

Corresponding author: Fei Wu
E-mail: wufei@zju.edu.cn
ORCID: http://orcid.org/0000-0003-2139-8807

Motivation

- Linking the references of entities in online documents to entity-based knowledge is helpful for machines to understand semantics in unstructured documents, and named entity disambiguation (NED) is one of the most important tasks for solving this problem.
- Crowdsourcing has become a common way to employ knowledge from the crowd.
- There are existing works that tackles NED tasks with crowdsourcing technique (Demartini, 2012), however, they simply consider crowd workers as a convenient classifier.
- We are interested in how the different interpretation of the problem between human and machine can be reflected in features, and whether the features learned from the crowd are able to enhance machine-based features.

Main idea

- Inspired by the ideas that the feature space of human workers may be reconstructed given crowd labels, and the relation between feature spaces of human and machine may be learned via a deep network, we propose an approach called ‘deep supervised learning via crowd labels (DeCLs)’.

- The proposed approach can be used to enhance existing NED methods by extending the their feature space.

Method

- A crowd model is devised to handle sparse and noisy crowd labels by considering each worker's ability, and elicit 'crowd features' of entities and mentions from them.
- A dynamic convolutional neural network (DCNN) is employed to learn a mapping from hand-crafted features to the crowd features.
- The learned DCNN is employed to obtain 'deep crowd features' to enhance the performance NED classification tasks.

Analysis on Real Data

- Looking into the dataset we use, we found evidences that human and machines are good on different questions, which supports our idea of their different feature space.

Table 2 Two samples from the ZenCrowd dataset

Mention detected	User ID / Algorithms	Candidate entities	Mention detected	User ID / Algorithms	Candidate entities
··· trains will share a track between New York Avenue and **Rhode Island Avenue** stations ···	33	**RIA (station)**[a]	··· they planned to work closely with the **United States**, Europe and India to plan ···	7	These United States (rock band)[c]
	26	**RIA (station)**[a]		59	Electoral College (institution)[d]
	47	RIA (avenue)[b]		39	**United States (state)**[e]
	tf-idf + CosDistance	RIA (avenue)[b]		tf-idf + CosDistance	**United States (state)**[e]

In the first columns, there are pieces of sentences in the news articles, in which the mention is marked in bold font. The second columns indicate whether the label is produced by a user or an algorithm, and user IDs or the algorithm descriptions are presented correspondingly. The third columns specify the entity that users or algorithms link the mention to, with correct links marked in bold font. Note that users are allowed to reject all candidate entities in this dataset, and such cases are not shown in this table due to the limited space

[a] http://dbpedia.org/page/Rhode_Island_Avenue_%E2%80%93_Brentwood_%28WMATA_station%29

[b] http://dbpedia.org/page/Rhode_Island_Avenue_%28Washington,_D.C.%29

[c] http://dbpedia.org/page/These_United_States

[d] http://dbpedia.org/page/Electoral_College_%28United_States%29

[e] http://dbpedia.org/page/United_States

Major results

- Experiments show that conventional NED algorithms benefits even from very compact ‘deep crowd features’, as long as there are enough crowd labels.

Table 3 Accuracy comparisons on NED tasks in terms of two dimensions of deep crowd features

Feature	Accuracy						
	Deep crowd feature dimension: 5			Deep crowd feature dimension: 300			
	9:1	8:2	7:3	9:1	8:2	7:3	5:5
tf-idf + DCF	78%	71%	69%	77%	71%	69%	67%
tf-idf	74%	70%	69%	74%	70%	69%	69%

Accuracies of NED tasks are given by applying AdaBoost on combined features (tf-idf + deep crowd features (DCF)) and tf-idf features respectively, in the settings of different ratios (9:1, 8:2, 7:3, and 5:5) of training data size to testing data size. tf-idf dimensions are the same, i.e., 18 478

Conclusions

- Our methods can effectively improve conventional vector space NED algorithms.
- The benefit of our method is obvious, even with compact ‘deep crowd feature’ space.
- Such benefit becomes more significant as there are more crowd labels involved.
- Our method is a promising way to make full use of crowdsourcing labels in NED tasks.