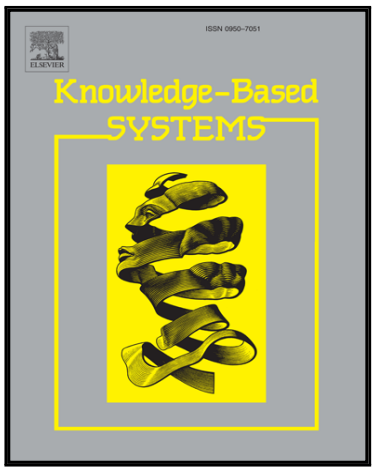

Named entity disambiguation for questions in community question answering

Fang Wang, Wei Wu, Zhoujun Li, Ming Zhou

PII: S0950-7051(17)30144-2
DOI: 10.1016/j.knosys.2017.03.017
Reference: KNOSYS 3866

To appear in: *Knowledge-Based Systems*

Received date: 28 July 2016
Revised date: 22 March 2017
Accepted date: 23 March 2017

Please cite this article as: Fang Wang, Wei Wu, Zhoujun Li, Ming Zhou, Named entity disambiguation for questions in community question answering, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.03.017

Named Entity Disambiguation for Questions in Community Question Answering☆

Fang Wang[a], Wei Wu[b], Zhoujun Li[a,*], Ming Zhou[b]

[a]*State Key Laboratory of Software Development Environment, Beihang University, China*
[b]*Microsoft Research, Beijing, China*

Abstract

Named entity disambiguation (NED) refers to the task of mapping entity mentions in running texts to the correct entries in a specific knowledge base (e.g., Wikipedia). Although there has been a lot of work on NED for long and formal texts like Wikipedia and news, the task is not well studied for questions in community question answering (CQA). The challenges of the task include little context for mentions in questions, lack of ground truth for learning, and language gaps between CQA and knowledge bases. To overcome these problems, we propose a topic modelling approach to NED for questions. Our model performs learning in an unsupervised manner, but can take advantage of weak supervision signals estimated from the metadata of CQA and knowledge bases. The signals can enrich the context of mentions in questions, and bridge the language gaps between CQA and knowledge bases. Besides these advantages, our model simulates people's behavior in CQA and thus is intuitively interpretable. We conduct experiments on both Chinese and English CQA data. The experimental results show that our method can significantly outperform state-of-the-art methods when we apply them to questions in CQA.

Keywords: named entity disambiguation, topic model, unsupervised learning, community question answering

☆This work was done when the first author was an intern in Microsoft Research Asia.

*Corresponding author

Email addresses: fangwang@buaa.edu.cn (Fang Wang), wuwei@microsoft.com (Wei Wu), lizj@buaa.edu.cn (Zhoujun Li), mingzhou@microsoft.com (Ming Zhou)