

doi:10.1631/FITEE.1601835

题目：基于众包标签数据深度学习的命名实体消歧算法

概要：命名实体消歧主要研究如何将存在歧义的文本描述映射到其对应知识库（例如 Wikipedia）中的实体。

本文结合群体智能（即群体用户生成的标签）和深度学习（即数据驱动学习），提出了在命名实体消歧过程中生成区别度更高的特征方法。具体来说，通过设计一个众包模型，学习文本描述或实体所对应“众包特征”，然后利用“众包特征”对动态卷积神经网络(Dynamic convolutional neural network, DCNN)进行优化，最后用优化得到的 DCNN 来提取“深度众包特征”，以此来解决传统命名实体消歧算法中单独依赖手工设计特征的不足。本文所提出方法巧妙将群体认知（由众包标签反映）结合到命名实体消歧深度学习框架中。实验分析表明，当有足够多众包标签时，所提出方法优于传统手工设计特征。

关键词：命名实体消歧；众包计算；深度学习