



基于Tri-training半监督学习的中文组织机构名识别

<http://www.firstlight.cn> 2010-01-01

针对中文组织机构名识别中的标注语料匮乏问题，提出了一种基于协同训练机制的组织机构名识别方法。该算法利用Tri-training学习方式将基于条件随机场的分类器、基于支持向量机的分类器和基于记忆学习方法的分类器组合成一个分类体系，并依据最优效用选择策略进行新加入样本的选择。在大规模真实语料上与co-training方法进行了比较实验，实验结果表明，此方法能有效利用大量未标注语料提高算法的泛化能力。

[存档文本](#)