

Paper Reading Note

Surveys

BrightHush

2014 年 11 月 12 日

目录

1 A Review of Relation Extraction	1
1.1 Abstract	1
1.2 Supervised Methods	1

1 A Review of Relation Extraction

1.1 Abstract

现如今的信息抽取、自然语言理解和信息检索都需要理解实体之间的语义关。这篇文章将会比较综合的包括实体关系抽取方法的各个方面，其中重要的监督学习方法和非监督学习方法都会比较充分的进行分析。另外我们会讨论扩展的高阶关系，对监督学习方法和半监督学习方法的评价都是采用常见的评比数据集。最后，本文会介绍实体关系抽取方面重要的应用，问答系统和生物学文本挖掘。

1.2 Supervised Methods

首先我们把实体关系抽取问题当成一个分类问题，为了描述的简洁性，我们限制讨论的实体关系为二元的，多元的关系抽取将会在接下来的部分进行讨论。对于给定的一个句子 $S = w_1, \dots, e_1, \dots, w_j, e_2, \dots$ ，其中 e_1, e_2 是实体，那么函数 $f(.)$ 可以描述如下：

$$f_R(T(S)) = \begin{cases} +1, & \text{If } e_1 \text{ and } e_2 \text{ are related according to relation } R \\ -1, & \text{Otherwise} \end{cases} \quad (1)$$

其中 $T(S)$ 是从句子 S 中提取出来的特征，函数 $f(\cdot)$ 用来判断实体是否存在关系。如果标记好的关系正样本和负样本训练一个分类器，这个分类器可以是Perceptron, Voted Perceptron or Support Vector Machine。这些分类器使用的特征可以通过文本分析得到的POS标记、依存关系，又或者是结构化的一寸句法树。根据分类器输入的本质，实体关系抽取分类方法分成两类（1）基于特征的（2）核方法，下面将会比较细节的讨论这两个方法。

1.2.1 Feature based Methods

1.2.2 Kernel Methods

[1]

参考文献

[1] "Sameer Badaskar" "Nguyen Bach". A review of relation extraction. •.