doi:10.1631/FITEE.1400405

题目：自底向上事件抽取系统

目的：本文研究自底向上的事件抽取方法。在无需预先人工给定事件类型体系的基础上，实现事件类型体系的自动构建及事件类型识别和事件元素的抽取。

创新点：本文首次提出基于聚类的事件类型自动发现方法。和传统事件抽取技术相比，该方法无需预先定义事件类型，无需先验的领域知识。因此，该方法是对领域移植的一个尝试，尤其适用于知识和资源有限的领域。

方法：该方法依据谓语动词是对领域事件刻画的重要单元的特点，利用依存句法信息抽取领域事件词，利用《知网》(HowNet)对领域事件词进行聚类从而获取不同的事件类型（图 2），随后进行事件元素的抽取。本文提出基于 Bootstrapping 的事件元素抽取框架，该框架核心有三部分：（1）模式获取：该模块负责将事件种子放在互联网上去检索，获得事件实例，并根据事件实例，按照一定的规则生成初始的事件模式（图 3）；（2）模式泛化：初始事件模式由于过于死板，导致遗漏掉很多事件的匹配，因此，本文设计模式泛化方法，将原有的事件模式按照一定规则，进行一定程度上的泛化，使其在保证准确率不变的情况下尽量提高召回率（算法 3）；（3）模式过滤：经泛化后的模式会在一定程度上引入噪声，因此，本文提出一套过滤规则，尽量减少泛化带来的噪声（表 3）。

结论：提出自底向上的事件抽取系统。该系统在公开的 ACE 语料数据集上取得了优于当前最好基线方法的结果。同时在我们手工构造的音乐领域和金融领域数据集上也取得了优秀的实验结果。这表明该方法可以很好地进行领域自适应。

关键词：事件抽取；无监督学习；自底向上