doi:10.1631/FITEE.1601846

题目：基于依存关系和多义词分析的句法词嵌入

概要：现有大多数词嵌入学习模型存在以下问题：（1）基于词袋上下文的模型完全忽略句子的句法结构关系；（2）每个词使用单个嵌入向量使多义词共享一个嵌入向量；（3）词嵌入往往趋向于句子上下文共性。为解决这些问题，提出一种基于依存关系和多义词分析的句法词嵌入（syntactic word embedding, SWE）。该算法主要处理：（1）基于主题模型，提出一个多义词识别算法；（2）采用符号“+”和“−”表示依存关系方向；（3）删除停用词及其依存关系；（4）引入“skip”依存关系表示依存关系之间的间接关系；（5）将基于依存关系的上下文输入到 Word2Vec 模型中训练语言模型。实验结果表明，SWE 模型在词相似度评测任务中表现出优异性能。基于依存关系句法上下文捕获词语的语义和句法特征，使词语表现出较少的上下文主题相似性和更多的句法和语义相似性。综上，包含更多信息的 SWE 模型性能优于单一的词嵌入学习模型。

关键词：基于依存关系的上下文；多义词表示；表示学习；句法词向量