

doi:10.1631/FITEE.1601476

题目：基于分布式表示语义组合的查询子主题挖掘

概要：推断查询意图对于信息检索具有重要意义。查询子主题挖掘旨在找到可能的子主题，用于表示给定查询的潜在意图。由于查询较短，子主题挖掘具有挑战性。学习词或句子分布式表示推动和影响了很多领域的发展。然而，没有清晰的结论表明该分布式表示是否有助于应对查询子主题挖掘面临的挑战。提出并比较利用分布式表示的语义组合进行查询子主题挖掘。采用两种分布式表示策略：能学习任意长度文本分布式表示的段落向量（paragraph vector）以及词向量的语义组合。探索了语义组合策略和数据类型对查询表示的影响。在国家信息学研究所信息获取研究试验平台和社区（National Institute of Informatics Testbeds and Community for Information Access Research, NTCIR）提供的公开数据集上的实验结果表明，与传统语义表示相比，分布式语义表示能获得更优查询子主题挖掘性能。文中做了更多深入探讨。

关键词：查询子主题挖掘；查询意图；分布式表示；语义组合