

## 一种运用规则类型依赖与复杂短语分析的词语情感极性分类方法

随着 Web 2.0 的兴起, 互联网用户发布的内容量越来越多。这些丰富的自由格式文本具有主观色彩的观点词语并足以影响人们的决策与行动。信息发布者由此对广大用户施加了一定的影响, 例如, 博客内容不仅可以影响人们的购买行为, 也可以左右其对政治的看法以及理财规划等。一般来说, 我们可以通过识别文本所体现的观点倾向来获取人们对于关注主题的态度及其对普通用户产生的影响。本文提出了一种通过挖掘极性规则以及分析词语间复杂关系来识别观点词语情感极性的自动方法。当前情感分析研究, 在准确分析文本内容的语法和语义的条件下, 着眼于通过使用类型依赖语法解析来处理词语间的功能性关系。典型做法是采用启发式方式来识别类型依赖极性模式, 其不足之处在于无法检测到所有规则。我们在本文中提出使用分类序列规则 CSR(Class Sequential Rules)来自动识别各种类型依赖模式, 并将 CSR 与另一种启发式方法进行对比验证。实验结果显示, CSR 的情感极性分类性能有显著的提高, 在测试实例上得到 80% 的 F1 值。此外, 研究还发现词语间的复杂关系足以影响情感极性分类性能, 我们也进一步地探讨了处理词语间这类依赖关系的一些解决方案。