

基于规则的情感本体和词向量的中文情感分类

郑 诚^{1,2}, 潮 旭^{1,2}, 章金平^{1,2}

(1 安徽大学 计算智能与信号处理实验室, 安徽 合肥 230601;

2 安徽大学 计算机学院, 安徽 合肥 230601)

摘 要: 随着电子商务的快速发展, 评论文本的情感倾向研究引起了广大学者的关注. 为了充分利用短文本中的情感本体和语义信息, 提出结合句法规则、情感本体和词向量的中文情感分类方法. 首先利用 Word2vec 训练词向量, 结合句法规则生成短文本向量; 再根据情感特征分布, 创建领域自适应情感词典, 结合句法规则, 得到短文本情感值, 从而构建词向量和情感值相结合的情感模型 VWEQ (Vector with Emotional Ontology). 在酒店评论数据集中, 与已有方法相比, 所提方法在准确率、召回率、F1 值均有明显提升.

关键词: 句法规则; Word2vec; 情感本体; 情感词典

Chinese sentiment classification of emotional ontology and word vector based on rules

ZHENG Cheng^{1,2}, CHAO Xu^{1,2}, ZHANG Jin-ping^{1,2}

(1 Key Laboratory of ICSP Ministry of Education. Anhui University, Hefei 230601, China;

2 School of Computer Science and Technology. Anhui University, Hefei 230601, China)

Abstract: With the rapid development of e-commerce, the research on the emotional tendency of comment texts has attracted the attention of scholars. In order to make full use of the emotional ontology and semantic information in short texts, a Chinese sentiment classification method combining syntactic rules, emotional ontology and word vector is proposed. Word2vec is firstly used to train word vectors and generate short text vectors in conjunction with syntactic rules; Then, according to the distribution of emotional features, a domain adaptive sentiment dictionary is created, and syntactic rules are combined to obtain short text sentiment values, thereby constructing an emotional model called VWEQ (Vector with Emotional Ontology) combining word vectors and emotional values. In the hotel review data set, compared with the existing methods, the proposed method has significantly improved accuracy, recall rate and F1 value.

Key words: syntactic rules; word2vec; emotional ontology; emotional dictionary

作者简介:

郑 诚 男, (1964-), 博士, 副教授. 研究方向为数据挖掘, 智能语义信息检索.

潮 旭 (通讯作者) 男, (1994-), 硕士研究生. 研究方向为数据挖掘, 情感分类, 大数据. E-mail: 1105741346@qq.com.

章金平 女, (1991-), 硕士研究生. 研究方向为数据挖掘, 推荐系统, 大数据.