

# 基于词向量的中文情感分析方法比较与应用探讨

马胜位

江南机电设计研究所 贵州 贵阳 550009

**【摘要】**：中文情感分析作为自然语言处理中的重要任务，广泛应用于舆情监测、产品评价等领域。基于词向量的情感分析方法通过将词语映射到高维空间，提取语义特征，有助于提高情感分类的准确性。本文从不同的词向量模型出发，比较了它们在中文情感分析中的应用效果。通过实验分析，发现不同模型在文本特征表示上的差异直接影响情感分析的效果。同时，探讨了基于词向量的情感分析方法在实际应用中的优势和挑战，提出了未来改进的方向。

**【关键词】**：词向量；中文情感分析；机器学习；应用比较；文本分类

DOI:10.12417/3041-0630.26.05.040

情感分析是自然语言处理领域中的一项核心任务，尤其在中文文本中，由于其独特的语法结构和多义性，给情感分析带来了较大的挑战。近年来，词向量技术的兴起为情感分析提供了新的思路。通过将词语映射到向量空间，词向量不仅能够有效捕捉词语的语义信息，还能克服传统方法中无法处理复杂语义的局限。尽管词向量方法在情感分析中取得了一定成果，但不同词向量模型在中文情感分析任务中的应用效果存在较大差异。深入探讨不同模型的表现及其在实际场景中的应用，将为中文情感分析技术的进一步发展提供有价值的参考。

## 1 中文情感分析中的词向量模型

中文情感分析的核心在于如何有效地从文本中提取出准确的情感信息。传统的情感分析方法主要依赖于人工设计的特征，如情感词典或规则基方法，但这些方法在处理复杂语义时往往力不从心。近年来，词向量模型作为一种基于大规模语料学习的技术，逐渐成为中文情感分析的重要工具。词向量通过将词语映射到一个高维空间，能够保留词语间的语义关系，使得词语的语义信息得以更为准确地捕捉。常见的词向量模型如 Word2Vec、GloVe 和 FastText，都通过训练大量的语料库，学习词汇之间的上下文关系，进而生成具有语义连贯性的词向量。

在中文情感分析中，由于中文的语法结构与英文有所不同，词向量模型的训练与应用需要特殊处理<sup>[1]</sup>。中文文本的分词问题常常成为情感分析的障碍，因为中文缺乏空格来区分词汇，这就要求在进行词向量建模时，必须对中文文本进行准确的分词处理。为了适应中文的特点，许多研究者在传统词向量模型的基础上进行改进，提出了适合中文的词向量训练方法，并且结合领域知识优化模型的表现。

值得注意的是，尽管词向量能够捕捉到词汇的语义特征，但在情感分析的应用中，单纯依赖词向量的效果并不总是理想。中文情感分析的难点不仅在于对词语情感的准确判定，还在于上下文的处理。同一词语在不同情境下可能会表达不同的

情感。为了更好地解决这些问题，一些基于词向量的情感分析模型采用了深度学习技术，通过构建卷积神经网络（CNN）或循环神经网络（RNN）模型，进一步增强了情感分析模型的表达能力和准确性。这些方法通过有效地结合上下文信息，提高了对复杂情感表达的理解能力，尤其在面对具有多义性的词汇时，能够提供更加精准的情感分类结果。中文情感分析中的词向量模型不仅要准确地表达单词的语义，还需要在情感分类任务中有效地处理上下文信息，才能在实际应用中取得良好的效果。

## 2 词向量模型在情感分类中的比较与实验分析

不同的词向量模型在中文情感分类中展现出不同的优劣，主要体现在词语表示能力、上下文处理以及训练效率等方面。Word2Vec 作为最早的字向量模型之一，通过连续词袋（CBOW）和跳字模型（Skip-gram）有效地学习到词语的语义信息，并且能够捕捉到词与词之间的相似性。这种模型忽视了上下文的细微差异，仅能通过局部语境建立词汇间的联系，对于情感分析任务中多义词的处理并不总是理想。针对这种局限性，后续的 GloVe 模型提出了通过全局共现矩阵来增强词向量的语义信息，并且能够在计算效率上进行优化，使得大规模语料训练成为可能。虽然 GloVe 提高了词与词之间的全局语义关联性，但在处理中文情感分析时仍然会遇到词性多样性及情感倾向性的模糊判断。

FastText 则在 Word2Vec 的基础上进行了进一步扩展，主要通过将词分解为子词进行学习，从而提升了对未登录词或低频词的表示能力。这一特点使得 FastText 在中文情感分类中展现出较好的表现，特别是在处理新词和词根时，能够更准确地捕捉到情感信息<sup>[2]</sup>。FastText 虽然在处理稀疏数据时具有优势，但其训练过程相对较慢，且在较大语料库中存在计算开销较大的问题。

基于深度学习的 BERT 模型通过预训练和微调的方式，极大地提升了情感分类的准确性。BERT 采用了双向编码器，使

得每个词的表示能够同时考虑左右上下文的信息，从而提供了比传统词向量更为丰富的上下文信息。这种上下文感知的模型在处理复杂的情感分类任务时表现出色，尤其在面对语境变化较大的中文情感分析中，能够更精确地判断情感倾向。BERT模型的训练成本较高，需要大量的计算资源，这使得在资源有限的情况下其应用受到一定限制。通过对不同词向量模型比较实验，结果表明，尽管每种模型在某些特定场景下都有其优势，但综合考虑准确率、效率以及对中文特性的适应能力，基于深度学习的模型在情感分类中的表现最为出色。

### 3 中文情感分析的应用与优化策略

中文情感分析在实际应用中面临许多挑战，尤其是在处理复杂语境、表达多样性和情感极性时。社交媒体、电子商务评论和在线论坛等平台都生成了大量的中文文本数据，这些数据具有情感色彩，且往往是非结构化和多样化的。针对这些特点，如何有效地提取情感信息并进行分析，已经成为提高用户体验和商业决策的重要手段。在这种背景下，基于词向量的情感分析方法逐渐成为主流，但其在实际应用中仍需进一步优化。为了提高中文情感分析的准确性，改进词向量模型的训练和应用成为了重要的优化方向。一个关键的策略是结合领域特定的情感词典与词向量模型。通过引入行业或平台特有的情感词典，可以帮助模型更好地识别与特定领域相关的情感表达，从而提高情感分析的准确性。细化分词策略也能有效提升情感分析效果。在中文情感分析中，由于词语的多义性和模糊性，传统的分词方法往往难以做到精确识别。通过深度学习模型与领域适

配的分词方法结合，可以更好地捕捉到文本中的情感信息。

除了分词和情感词典，结合上下文信息也对情感分析的准确性有显著影响。由于中文中有许多情感表达可能依赖于前后语境，采用基于上下文的深度学习模型显得尤为重要<sup>[3]</sup>。比如BERT等预训练语言模型能够在全局上下文中理解词汇的情感倾向，从而实现更加准确的情感分类。在这个基础上，通过微调技术，可以使模型在特定领域的数据集上进行训练，进而提高情感分析的效果。数据增强也是提升中文情感分析应用效果的一个策略。在实际应用中，由于标注数据的匮乏或不平衡，模型的训练效果往往受到限制。通过数据增强技术，如回译、同义词替换等方法，可以扩展训练数据集，增加模型的泛化能力，从而提升其在各种场景下的表现。通过这些优化策略，中文情感分析不仅能够准确识别情感倾向，还能够为实际业务应用提供有力的数据支持，推动情感分析技术在更多领域的落地与发展。

### 4 结语

中文情感分析作为自然语言处理的重要任务，在不断发展的技术支持下，得到了越来越广泛的应用。词向量模型的引入极大提升了情感分析的准确性，但在实际应用中仍面临着挑战。针对中文文本的复杂性，结合领域知识、优化分词策略和利用深度学习模型等优化策略，能够有效提升情感分析的效果。未来，随着技术的不断进步，中文情感分析的应用将更加成熟，为各行各业的智能化发展提供更加精准的数据支持。

### 参考文献：

- [1] 黄伟, 戴滕好. 基于历时语料库和词向量的汉语形容词功能演变研究[J]. 当代语言学, 2026, 28(01): 80-97.
- [2] 任菊香, 刘忠宝. 融合多粒度语义特征的中文情感分析方法[J]. 华东师范大学学报(自然科学版), 2023, (06): 95-107.
- [3] 王桂江, 黄润才, 黄勃. 基于 K-BERT 和残差循环单元的中文情感分析[J]. 数据与计算发展前沿, 2023, 5(04): 127-138.