



开放科学  
(资源服务)  
标识码  
(OSID)

# 基于多跳动态记忆网络和情感词典的情感分析模型

沈金金 陈荔

上海理工大学 上海 200093

**摘要:** [目的/意义] 文本情感分析是自然语言处理的一大重要分支。论文结合了深度学习模型在特征提取方面的优势以及情感词典对网络词情感识别的敏感性,提出了一种将动态记忆网络和情感词典方法相结合的网络文本情感分类模型。[方法/过程] 在传统的动态记忆网络中设计情感问题向量,利用基于注意力机制的多跳结构识别并提取句子情感特征,同时构建扩充网络情感词的情感词典并将情感分值与多跳记忆网络所得情感分类结果进行线性加权,减少了对于情感词库的完备性、判断规则质量以及事先标注语料的依赖,提升了训练效率。[结果/结论] 在 WEIBO\_SENTI\_100K 和 NLPCC2013 两个数据集上进行实验,证明与单个模型相比,论文模型在两个数据集上都获得了更好的分类效果。

**关键词:** 动态记忆网络; 情感词典; 注意力机制; 情感分析

**中图分类号:** G35

## Emotion Analysis Model Based on Dynamic Memory Network and Emotion Dictionary

SHEN Jinjin CHEN Li

University of Shanghai for Science & Technology, Shanghai 200093, China

**Abstract:** [Objective/Significance] Text sentiment analysis is an important branch of natural language processing. Combining the advantages of deep learning model in feature extraction and the sensitivity of sentiment dictionary to identify the emotions of emerging online words, this paper proposes an online text emotion classification model combining dynamic memory network and sentiment dictionary. [Methods/Process] Emotion problem vectors are designed in the traditional dynamic memory network, and the multi-hop structure based on attention mechanism is used to identify and extract sentence emotion features. Meanwhile, the sentiment dictionary with extended network emotion words is constructed, and the emotion score is weighted with the emotion classification results obtained from multi-hop memory network. It reduces the dependence on the completeness of sentiment lexicon, the quality of judgment rules and the pre-annotation corpus, and improves the training efficiency. [Results/

**基金项目** 国家自然科学基金项目“产业互联”智造“供需网的结构、演化及其动力学研究”(71871144)。

**作者简介** 沈金金(1998-), 硕士研究生, 主要研究方向为自然语言处理, 机器学习, E-mail: li540766180@163.com; 陈荔(1967-), 博士, 副教授, 主要研究方向为机器学习, 数据分析, 计算机软件及计算机应用。

**引用格式** 沈金金, 陈荔. 基于多跳动态记忆网络和情感词典的情感分析模型[J]. 情报工程, 2022, 8(2): 3-18.

Conclusions] Experiments on two datasets, WEIBO\_SENTI\_100K and NLPCC2013, prove that the proposed model achieves better classification effect on both datasets compared with a single model.

**Keywords:** Dynamic memory network; sentiment dictionary; attention mechanism; sentiment analysis

## 引言

随着传统信息传播方式的改变以及互联网用户数量的爆炸式增长, 研究人员愈发关注网络数据所蕴含的巨大价值。其中, 从网民的观点和评论中挖掘用户的主观性情感倾向对于政府的管控、社会发展和商业决策等多个领域有着重大的指导意义。因此为了高效地从巨量互联网文本资源中挖掘出用户的情感倾向, 情感分析作为自然语言处理的一大重要分支成为了近年来学界研究的热门话题。它具有许多应用场景, 比如产品在线评论数据<sup>[1-3]</sup>、用户体验感知<sup>[4]</sup>, 以及相关个性化推荐<sup>[5]</sup>等。

在早期比较传统的方法是基于情感词典的文本情感分类, 是指根据所建立的词典和语法规则计算情感分值最后判断情感倾向。比较成熟的中文情感词典有知网 HowNet、NTUSD<sup>[6]</sup>等, 英语情感词典有 sentiwordnet<sup>[7]</sup>等。许多学者在这些词典的基础上做了大量的研究。Ebert 等<sup>[8]</sup>将词典词向量和语义词向量合并取得了分类准确率的提高。陈国兰<sup>[9]</sup>根据微博文本的特点基于开源词典作出了改进, 提出了微博文本情感计算方法。虽然基于情感词典的方法应用范围比较广且易于操作和理解, 但是它过于依赖所创建的情感词典, 特别在社交文本的情感倾向分析方面, 网络新词层出不穷, 需要对情感词典不断地扩充才能保证情感分类的准确率, 词典的构建需要耗费大量时间和资源且迁移性

差。因此越来越多的研究者基于机器学习和深度学习对情感分类任务进行了研究, 并证明了这两种方法在情感分析任务上可以取得更高的准确率<sup>[10]</sup>。Ahmad 等<sup>[11]</sup>介绍了常见的基于机器学习的情感分类算法, 例如 SGD、随机森林、SVM 和最大熵等。Singh 等<sup>[12]</sup>对朴素贝叶斯、J48、BFTree 和 OneR 四种用于情感分析的机器学习分类器进行了比较, 其中朴素贝叶斯的学习速率比较快而 OneR 在分类准确率上更胜一筹。深度学习模型可以挖掘隐含的文本特征, 解决数据稀疏性问题从而获得更好的分类效果。Kim 等<sup>[13]</sup>将卷积神经网络用于句子级情感分类任务, 并在多个基准数据集上获得了很好的结果。Zhu 等<sup>[14]</sup>将双向长短期记忆网络和多列卷积神经网络相结合捕捉了更高层次的语义特征。然而基于深度学习的方法属于有监督学习, 其训练效果过于依赖训练文本的质量而导致情感分类不准确。

综上所述, 三种不同的情感分类算法均有自己的优缺点, 因此论文尝试结合基于情感词典和深度学习的情感分类方法, 充分利用情感词典对特定领域情感词的敏感程度以及深度学习在数据稀疏问题和高层次文本特征提取方面的优势, 提出了基于多跳动态记忆网络和情感词典的情感分析模型 (SD-DMN, Sentiment Dictionary with Dynamic Memory Network)。在情感分析领域中引入多跳动态记忆网络, 首先将基于注意力的 GRU 结构与记忆单元结合构建

情节记忆模块，可以对单词权重进行调整消除权重偏置问题对分类结果的影响。其次将情节记忆模块进行叠加形成多跳结构可以在更短的路径上获得长距离依赖，记忆单元融合本层提取的特征以及历史特征信息，利用记忆单元与通过单层 GRU 编码得到的情感问题向量进行多次更新捕捉更深层次的非线性特征。同时构建扩充的网络情感词典作为动态记忆网络模型的辅助方法，充分发挥对网络情感词识别的敏感性优势。最后将两个方法得到的情感分值进行线性加权得到情感分类结果，减少深度学习模型中对于训练语料的依赖和参数设置的偏差以及情感词典词库的覆盖度对分类结果带来的负面影响，提高模型的性能。

## 1 相关工作

### 1.1 基于情感词典的情感分析

基于情感词典的情感倾向分析就是利用标注好的情感词典和语法规则，找到文本中的种子情感词以及与之相关的同义词和反义词最后得到情感词集合，以达到挖掘并判断主体对于某客体的内在评价倾向的目的<sup>[15,16]</sup>。情感词典最早由 Hatzivassiloglou 等<sup>[17]</sup>提出，之后大量的学者通过扩充情感词典以及改进语法规则提高模型性能，并应用于社交文本情感识别中。Georgios 等<sup>[18]</sup>提出了一种基于情感词典的无监督方法对推特上的社交短文本进行情感评估，在三个真实数据集上的准确率都高于传统机器学习方法。Ma 等<sup>[19]</sup>、林江豪等<sup>[20]</sup>以及王志涛等<sup>[21]</sup>将表情符号加入情感词典在社交平台上进行应用得到了更准确的情感分析结果。Maks

等<sup>[22]</sup>通过不同词性的组合建立了新的词典模型用于情感分析和观点挖掘。

### 1.2 融合注意力机制的情感分析

注意力机制可以使得神经网络能够具备专注于特征子集的能力，用来自动学习输入数据对输出数据的贡献大小从而将注意力集中在更为重要的数据上，是机器学习中的一种数据处理方法。Attention 机制其实最早于上世纪九十年代在视觉图像领域被提出来，但是直到 2014 年 Google Mind 团队<sup>[23]</sup>在 RNN 模型上使用了 Attention 机制来进行图像分类才让注意力机制开始成为深度学习领域的研究热点，与此同时也被广泛应用于情感分析领域。冯兴杰等<sup>[24]</sup>、胡荣磊等<sup>[25]</sup>将注意力机制与其他例如 CNN、LSTM 等神经网络模型进行拼接得到改进的情感分析模型，其情感分类准确率明显高于单独的神经网络模型。宋云峰等<sup>[26]</sup>利用跨模态注意力机制将 CNN 和 Bi-GRU 提取的单模态内部特征两两融合，通过实验证明情感分类的准确率和召回率均有所提升。但是这些研究没有将注意力机制融入模型改变其结构而是进行简单的拼接，因此学界开始关注注意力机制的深入融合。王盛玉等<sup>[27]</sup>在卷积神经网络的词嵌入层中增加注意力机制层以提取重要的局部特征，结果验证该模型拥有更高的情感分类准确率。

### 1.3 记忆网络

常见的拥有记忆功能的神经网络模型包括 LSTM、GRU、RNN 等模型，虽然在序列建模问题上有一定优势，在时间跨度较长时能具备不错的预测效果，但是把这些信息压缩成一个

稠密的向量会丢失很多信息，为了解决长距离记忆存储的问题，2015年 Facebook AI 实验室第一次提出了记忆神经网络 (MemNN)<sup>[28]</sup> 的概念。输入文本和问题经过 Input (输入模块) 被编码成特征向量的形式，输入文本被传送到 Generalization (记忆更新模块) 中进行记忆的

读写，文本信息基本大部分被保留在记忆插槽中，Output (输出模块) 根据问题对记忆插槽中的信息进行加权处理，然后抽取与问题相关性最高的记忆组合向量，最后经过 Response (响应模块) 将向量转化为需要的自然语言答案，如图 1 所示。

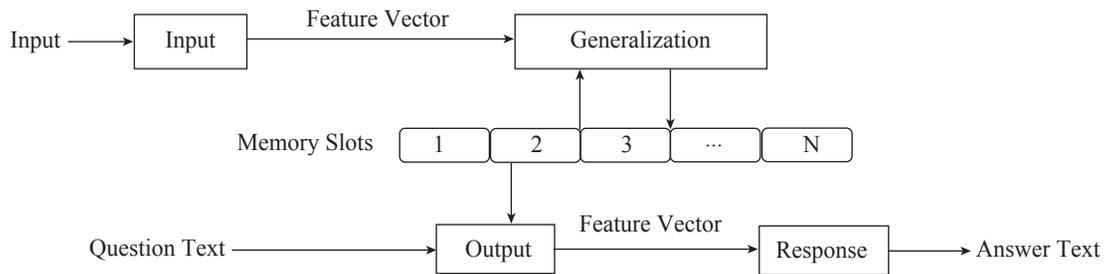


图 1 记忆网络

然而对于复杂的输入文本或者问题而言，通过简单的前馈可能无法有效的将记忆组件中所有需要的信息挖掘出来，因此 Kumar 等<sup>[29]</sup> 提出了一种动态的记忆网络 (DMN) 用于问答任务，模型可以根据需要任意访问数据充分捕捉与问题相关的句子特征信息。模型使用循环神经网络作为数据处理的工具，因此可以广泛应用于其他自然语言处理任务中，解决了 Memory Networks 使用词袋处理输入的局限性。模型包括输入、问题、情节记忆和回答四部分，使用单向 GRU 对文本与问题进行编码，在情节记忆模块中引入基于注意力机制的 GRU 模型，注意力机制从前一轮迭代的输入记忆结果以及问题向量中进行取舍提取特征向量以形成本轮的记忆，问题通过多次的迭代注意过程得到最终的答案。

## 2 整体框架

针对情感词典情感分类模型迁移性差以及深度学习模型过于依赖训练文本和参数的设置

问题，论文提出了一种结合情感词典和深度学习的情感分类模型，使得模型学习到更深层次的语义情感特征，改善模型的分类性能。模型主要分为三个部分：1) 利用多跳动态记忆网络对文本进行训练然后得到分类结果；2) 对情感词典进行扩充，基于情感词典和语义规则计算文本的情感分值；3) 将两种方法所得的分值进行线性加权得到最终的情感分值，完成情感分类任务。论文模型整体框架图如图 2 所示。

### 2.1 构建多跳动态记忆网络

论文的深度学习模型部分引入了优化的动态记忆网络作为模型的主干框架，将情感分类任务建模为一个问答系统使输入词向量进行多次前馈，每个词嵌入表示都会被赋予权重，权重与其和问题的相关性对应，经过多计算层的注意力机制提取情感特征，再依据记忆更新机制得到和问题最相关的最终句子特征表示。最后一个记忆模块所提取的句子情感分类特征表示送入 softmax 层获得最终的情感极性预测结果。

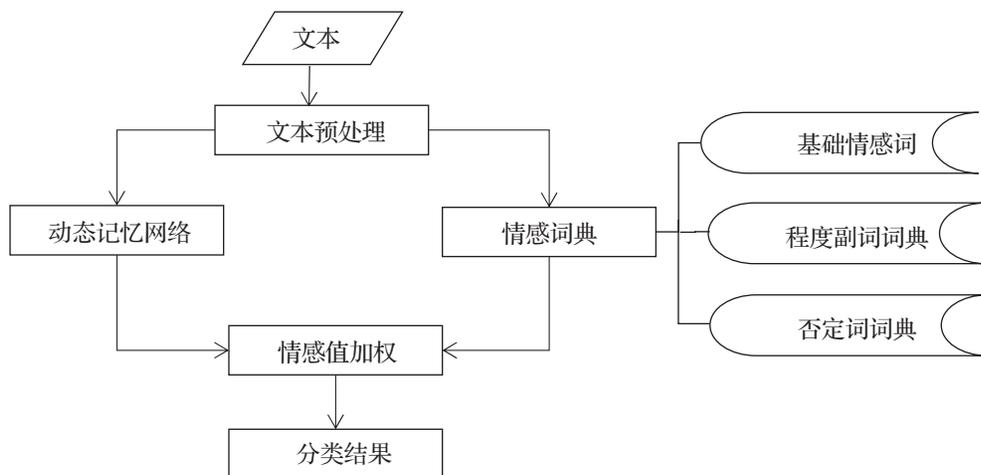


图2 论文模型整体框架图

考虑到社交文本大多都是短文本，也可能包括拥有多个情感词的复杂句式，论文引入了动态记忆网络以更好的捕捉时序信息。动态记忆网络是一个统一的神经网络框架，一般分为输入模块、问题模块、情节记忆模块和答案模块四个部分，常用于文本问答，论文模型把情感极性分类建模为一个问答系统，将情节记忆模块拆分为多跳注意力和记忆更新两个部分，并将答案模块单独列为情感分类层。具体来说首先设计与情感分类相关的问题输入到问题模块中，对其进行编码得到问题向量表示，输入模块接收原始数据，经过语义挖掘层中双向GRU处理后生成句子特征向量，将其输入到情节记忆模块中，多跳结构帮助注意力机制提取出与情感分类问题最相关的子信息，并进一步利用记忆更新机制高效的捕捉融合更精确的情感特征。多跳动态记忆网络是多个注意力机制和记忆更新模块的叠加，每经过一个注意力机制提取模块，都会根据最终的隐层状态、问题以及上一层提取的记忆计算本层的记忆，如此

一来便完成了一次记忆更新。多跳动态记忆网络结构如图3所示。

假设用于训练句子级情感分析的语料库为 $C_s$ ，语料库中包含多条评论语句，其中将拥有 $n$ 个词的文本语句 $S$ 表示为 $S=\{S_1, S_2 \cdots S_i \cdots S_n\}$ ， $S_i$ 表示句子中的第 $i$ 个单词，利用预训练好的word2vec词向量得到句子的词嵌入向量 $W=\{W_1, W_2 \cdots W_i \cdots W_n\}$ ， $W_i$ 表示句子中第 $i$ 个词的词向量表示。本模块的最终任务就是根据所建立的多跳动态记忆网络模型学习得到语句 $S$ 的情感分类结果。动态记忆网络各层的具体任务如下所示：

(1) 输入模块。该模块将原始数据转化为一组有序向量传递到情节记忆模块中，在经典的动态记忆网络模型中，常用单向GRU进行简单的输入处理，论文事先将原始输入传递到双向GRU模型中进行输入优化，最终得到一个有序的句子特征表示 $R=\{r_1, r_2, \cdots, r_i, \cdots, r_n\} \in \mathbb{R}^{N \times D_1}$ ，其中 $N$ 表示句子中包含的单词数量， $D_1$ 表示双向GRU隐藏层输出向量的维度。

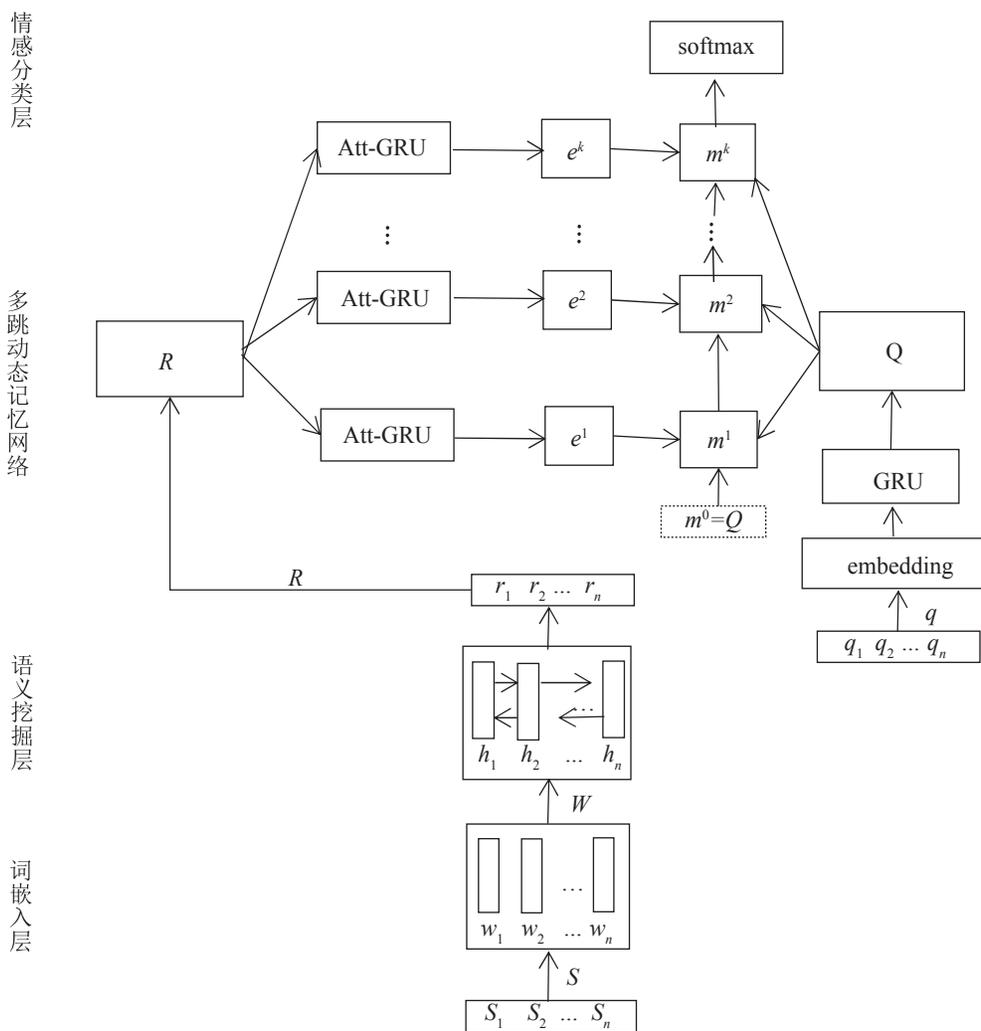


图3 多跳动态记忆网络结构

(2) 问题模块。论文的情感分类结果包括正向和负向两种，因此针对情感极性分析结果设计对应的问题”这句话的情感分析结果是？”将其映射为词向量后利用单向 GRU 对其编码得到向量  $Q = q_0 \in \mathbb{R}^{D_2}$ ，其中  $q_0$  为单向 GRU 最终的隐藏层状态， $D_2$  表示 GRU 网络隐藏层输出向量的维度。

(3) 情节记忆模块。情节记忆模块主要由多个基于 GRU 的记忆提取注意力机制以及捕捉特征的内存更新模块两部分叠加组成。这一模块假设经过  $k$  个注意力机制和记忆更新模块以

后得到的记忆单元  $m^k$  可以包含与情感分类相关的所有信息。

(a) 注意力机制

论文使用门控作为注意力机制，通过对输入信息的加权求和提取记忆，构建一个权重门  $g_i^k$  替代 GRU 中的更新门，通过权重门控制输入的事实因子  $c_i$ ，更好的结合问题、上一层的记忆以及本次输入的事实信息控制信息流动。其中，权重门的计算如下：

$$z_i^k = [c_i \odot Q; c_i \odot m^{k-1}; |c_i - Q|; |c_i - m^{k-1}|] \quad (1)$$

$$Z_i^k = W_3 \tanh(W_4 z_i^k + b_4) + b_3 \quad (2)$$

$$g_t^k = \frac{\exp(Z_t^k)}{\sum_{j=1}^{M_i} \exp(Z_j^k)} \quad (3)$$

其中 $\odot$ 表示元素乘积， $|\cdot|$ 是取绝对值，分号表示向量拼接， $m^{k-1}$ 是上一层的记忆信息， $c_t$ 是第 $t$ 个事实， $Q$ 为问题向量， $W_3$ 、 $W_4$ 和 $b_3$ 、 $b_4$ 为第 $k$ 层注意力机制的参数，公式1其实是在计算问题向量、输入向量与记忆单元之间的相似度，然后将其传入公式2所示的两层神经网络中并用softmax函数激活得到权重门 $g_t^k$ ，从上述三个公式可以看出权重门与输入门相比，虽然同样具备决定信息流动的能力，但是不仅考虑了输入和前一时间步的记忆信息，而且包括前一层注意力机制提取的记忆信息以及事实因子，可以更好地决定重要信息的通过。

构建权重门函数替代输入门之后的流程与标准GRU类似，进行隐藏层的更新，通过权重门和重置门得到GRU最终的隐藏层状态用于记忆更新。主要的计算过程如下：

$$r_t^k = \sigma(W_r^k c_t + U_r^k h_{t-1}^k + b_r^k) \quad (4)$$

$$\tilde{h}_t^k = \tanh(W_h^k c_t + r_t^k \odot U_h^k h_{t-1}^k + b_h^k) \quad (5)$$

$$h_t^k = g_t^k \odot \tilde{h}_t^k + (1 - g_t^k) h_{t-1}^k \quad (6)$$

$$e^k = h_n^k \quad (7)$$

上述运算式中的上标 $k$ 均表示第 $k$ 层， $W$ 、 $U$ 、 $b$ 是需要学习的网络参数，本文模型的GRU使用权重门 $g_t^k$ 更新其内部状态， $r_t^k$ 是第 $k$ 层的重置门，用来控制候选状态 $\tilde{h}_t^k$ 与上一时刻的隐藏层状态 $h_{t-1}^k$ 的依赖大小，公式(6)清楚的表现了权重门 $g_t^k$ 可以决定从候选状态 $\tilde{h}_t^k$ 中接收多少信息，并且能控制当前隐藏状态 $h_t^k$ 保留多少上一时刻隐藏状态 $h_{t-1}^k$ 中的信息。如公式(7)所示传送给最后一个用于情感分类的记忆模块的是GRU的最终隐藏状态 $e^k$ ，即经过第 $k$ 层基于注意力的GRU训练得到的最终隐藏状态，在本文中我们将 $e^k$ 称之为情节向量，可把它看成第 $k$ 层的事实聚焦，经过多层的迭代运算实现嵌入信息的注意力选择。基于注意力的GRU内部结构模型如图4所示。

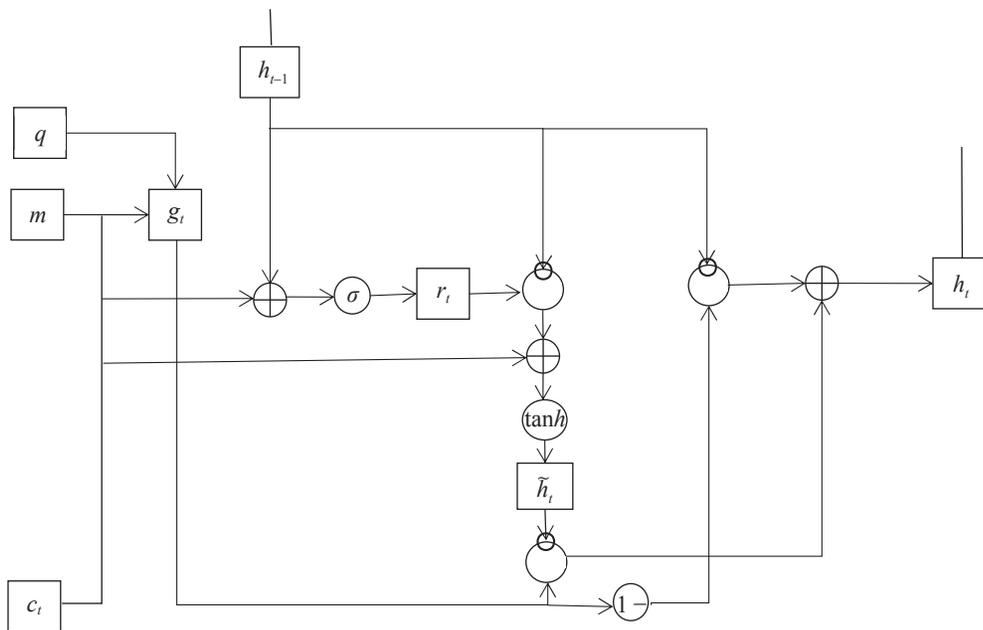


图4 基于注意力的GRU内部结构

(b) 记忆更新

每一层的注意力机制都会生成一个新的情节向量，用于本层记忆单元的更新，动态记忆网络的记忆更新通常使用单层 GRU，论文采用 ReLU 函数更新记忆，计算公式如下：

$$m^k = ReLU(W_m^k[m^{k-1}; e^k; Q] + b) \quad (8)$$

其中， $W_m^k \in \mathbb{R}^{D_2 \times D_2}$  是第  $k$  层记忆更新时激活函数的权重参数， $b \in \mathbb{R}^{D_2}$  是偏置项， $D_2$  表示隐层维度。由此可知，第  $k$  层的记忆单元  $m^{k-1}$  是由上一层的记忆单元  $m^k$ ，经过注意力机制得到的本层情节向量  $e^k$  以及问题向量  $Q$  计算所得，第一层动态记忆网络的输入记忆单元  $m^0=Q$ ，通过多计算层的注意力机制和记忆更新最后可以获得更精确的句子特征向量表示，使情感分类结果更为准确。

(4) 答案模块。在经过  $k$  层动态记忆网络训练后，将记忆更新模块最终的句子特征表示  $m^k$  传递给情感分类层，由于论文所设计的情感分类问题最终的答案是单个词，因此可以直接用 softmax 函数作为线性层进行情感极性预测，主要公式为：

$$\hat{y} = softmax(W_a m^k) \quad (9)$$

其中  $\hat{y}$  表示多层动态记忆网络情感分类结果，分别有正向和负向两种情感类别。

(5) 模型训练。本模块的训练以最小化情感极性预测序列的交叉熵误差为目标，为了防止过拟合引入了  $L_2$  正则化参数，交叉熵损失函数的公式如下：

$$L = -\frac{1}{N} \sum_{i=1}^D \sum_{j=0}^{C-1} y_{ij} \log(\hat{y}_{ij}) + \lambda \|\theta\|^2 \quad (10)$$

其中  $D$  是训练数据集， $C$  为情感极性类别数，论文共有正向和负向两类， $y_{ij}$  表示真实情感极性， $\hat{y}_{ij}$  是预测情感极性， $\lambda \|\theta\|^2$  为正则项，

$\lambda$  为  $L_2$  正则化权重参数， $\theta$  为模型中所有参数的集合。

2.2 构建情感词典

使用情感词典进行情感分析是一种比较传统的情感倾向分析方法，国内外已经有很多成熟的情感词典。论文在其他研究的基础上，整合了台湾大学 NTUSD 简体中文情感词典、知网 HowNet 情感词典、清华大学李军中文褒贬义词典、BosonNLP 网络情感词典以及一些人工定义的词汇，并且融合了百度搜索引擎和小鸡词典中获取的网络新词，最终整理得到正向词汇 14497 个，负向词汇 13906 个，程度副词 216 个，否定词 58 个。词典中主要包括基础情感词、程度副词以及否定词三类，主要设定如下：

(1) 基础情感词：在计算情感值时，将词典中的正向情感词设为 1，负向情感词设为 -1，然后依次寻找句子中所有的情感词进行求和。

(2) 程度副词：在文本分析中，情感词前出现的程度副词往往能在很大程度上影响情感的强烈程度。当文本匹配到词典中的基础情感词以后，自情感词向前寻找程度副词并与所建立的程度副词词典匹配得到相应的权重值，最后将基础情感词分值与程度副词对应的权重值相乘得到情感值。程度副词的划分如表 1 所示。

表 1 程度副词表

| 程度量级           | 程度副词        | 权值   |
|----------------|-------------|------|
| most           | 充分、极其、不折不扣… | 4    |
| very           | 特别、尤其、很…    | 3    |
| more           | 更加、较为、越来越…  | 2    |
| ish            | 略微、稍微、未免…   | 0.5  |
| insufficiently | 丝毫、不那么、相对…  | -0.3 |
| over           | 过分、太、极端…    | -0.5 |

(3) 否定词: 文本中是否存在否定词对语义情感倾向起着至关重要的作用。如果情感词之前出现否定词, 原先的情感倾向可能会出现截然相反的转变。判断基础情感词之前是否存在程度副词之后, 再判断情感词之前是否存在否定词。如果存在否定词且数量为奇数, 则进行反转取值<sup>[30]</sup>, 如果数量为偶数, 则情感值不变。论文所收录的否定词词表如表 2 所示。

表 2 否定词表

| 否定词                        | 权值 |
|----------------------------|----|
| 不、别、否、没有、不要、百搭、切莫、从不、杜绝... | -1 |

### 2.3 情感值计算

论文的情感度计算包括两个部分: (1) 基于情感词典的情感词加权法。(2) 基于动态记忆网络的 softmax 分类法。将两个方法得到的情感分值进行线性加权得到最终的情感分值, 形成文本情感分类结果。

用情感词加权法计算时, 将文本按照标点符号划分为若干分句。计算每个分句的情感分值, 然后将分句的情感值相加得到整个句子的情感值, 计算公式如下:

$$S = \sum_{i=1}^m \sum_{j=1}^n w_i \times u_i \times E_{ij} \quad (11)$$

其中,  $S$  代表整个句子的情感值,  $w_i$  表示第  $i$  个短句中程度副词的权重,  $u_i$  表示第  $i$  个短句中否定词的权重,  $E_{ij}$  表示第  $i$  个短句中第  $j$  个情感词的分值。

用深度学习模型进行情感分类时, 最终的 softmax 分类层得到的标签结果如下所示:

$$\hat{y} = \{p_1, p_2, \dots, p_n\} = \arg \max(p_i) \quad (12)$$

其中  $\hat{y}$  表示经过深度模型训练后的句子情感分类结果,  $p_i$  表示第  $i$  个标签情感分类估计的概率值。

基于深度学习的情感分类和基于情感词典的情感分类都有各自的优缺点, 将上述两种方法的分类结果进行加权可以提高情感识别的准确率。

$$E = \alpha S + (1 - \alpha) \hat{y} \quad (13)$$

其中,  $\alpha$  作为权重系数表示不同情感分类方法的重要程度。

## 3 实验

### 3.1 实验数据集

为了评估论文模型的效果在 WEIBO\_SENTI\_100K<sup>[31]</sup> 和 NLPCC2013<sup>[32]</sup> 两个数据集上进行实验。这两个数据集都是质量比较高的微博情感分析数据集, 其中 WEIBO\_SENTI\_100K 包含十万多条二分类微博文本, 而 NLPCC2013 包含一万多条文本数据, 有七种情绪分类, 本实验将喜悦、开心合并为正向情感标签数据, 将厌恶、难过、生气合并为负向情感标签数据。最终从经过处理的 WEIBO\_SENTI\_100K 和 NLPCC2013 两个二分类公开数据集中随机抽取一万条语料样本数据, 将其按照 4:1 划分为训练集和测试集, 两个数据集的具体分类情况如表 3 所示。

表 3 数据集各类情感类别分布情况

| 数据集              | 正向   |     | 负向   |      |
|------------------|------|-----|------|------|
|                  | 训练集  | 测试集 | 训练集  | 测试集  |
| WEIBO_SENTI_100K | 3971 | 977 | 4029 | 1023 |
| NLPCC2013        | 3978 | 974 | 4022 | 1026 |

### 3.2 评价指标和模型参数

本实验在参考了 Kumar 等<sup>[7]</sup>使用 DMN 完成问答任务实验中设置的参数基础上, 结合两个数据集自身的特征调整了模型的超参数, 以使得模型的训练效果更好。文本的嵌入维度为 80, 选取文本数据的最大长度作为标准长度, 对其余不足的文档进行补全并输入到 Bi-GRU 中, 隐藏层单元大小为 80, DMN 最终提取的特征向量维度为 80, 为了防止过拟合在全连接层中加入了 dropout 正则化技术, 主要的超参数设置如表 4 所示。

表 4 实验超参数设置

| 参数名称    | WEIBO_SENTI_100K | NLPCC2013 |
|---------|------------------|-----------|
|         | 参数值              |           |
| 批处理大小   | 100              | 100       |
| 隐藏层维度   | 80               | 80        |
| Dropout | 0.9              | 0.9       |
| hops    | 5                | 6         |
| 学习率     | 0.001            | 0.001     |
| L2正则项   | 0.001            | 0.001     |
| 最大迭代次数  | 256              | 256       |

论文使用分类准确率 ( $Acc$ )、精确率 ( $P$ )、召回率 ( $R$ ) 以及  $F1-score$  作为评测指标来评价模型性能并确定权重系数和各参数。各指标公式如下所示:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$F1-score = \frac{2PR}{P + R} \quad (17)$$

其中,  $TP$  表示模型预测正确的正样本,  $FN$  表示模型预测为 negative 的正样本,  $FP$  表示模型预测为 positive 的负样本,  $TN$  为模型预测正确的负样本。

### 3.3 权重因子选择

权重因子  $\alpha$  的作用是以加权平衡的方式通过最终的准确率判断两种模型在情感分类任务中的重要性程度从而使得模型具有一个更佳的性能。通过具体实验, 我们发现不同的权重因子选择对分类准确率的影响各不相同, 结果如表 5 所示。

表 5 权重因子  $\alpha$  对比实验

| $\alpha$ | WEIBO_SENTI_100K | NLPCC2013 |
|----------|------------------|-----------|
|          | 准确率              |           |
| 0.1      | 0.8705           | 0.8885    |
| 0.2      | 0.8775           | 0.89      |
| 0.3      | 0.8725           | 0.8775    |
| 0.4      | 0.867            | 0.8665    |
| 0.5      | 0.862            | 0.866     |

从表 5 中不难发现, 在其他条件相同的前提下, 对于两个不同的数据集来说, 情感词典与动态记忆网络相结合的综合模型基本上都比单独的模型分类效果好, 其中 NLPCC2013 的结果更佳, 一个可能的原因是它本身的情感倾向更细腻, 数据质量更高。随着权重因子越来越大, 情感识别的准确率呈现出先上升后下降的趋势, 而当权重因子  $\alpha$  为 0.2 时, 两个数据集都达到了他们最佳的情感分类准确率, 分别为 87.75% 和 89%, 因此我们选择 0.2 作为论文模型的权重因子。与此同时也可以证明, 相比

起基于情感词典的传统分类模型，系统更依赖于动态记忆网络模型。

### 3.4 模型性能对比

为了评价论文提出的 SD-DMN 模型在两个数据集上的有效性，实验中不仅将本文模型与单个情感词典及动态记忆网络模型进行对比，还在同等实验环境下与以下一些典型的机器学习以及深度学习基准模型进行对比：

(a) RAE：该模型可以对比较复杂的短语进行学习以了解其空间向量，是以递归自编码器为基础预测句子级情感分类的一种模型。

(b) Bayes：贝叶斯分类算法以贝叶斯定理为基础，首先确定特征属性，然后通过分类器分别计算样本属于每一个类别的概率，取其最大值作为预测结果。

(c) BiLSTM：该模型将正向长短期记忆网络和负向长短期记忆网络整合并融入了注意力机制，是比较常用的一种神经网络模型。

(d) CNN：卷积神经网络模型通过卷积操作捕捉并表示特征向量，以得到分类结果。

(e) 情感词典：利用建好的词库以及自行设定的情感语义规则计算情感分值并实现分类，论文中的情感词典结合了近年来普遍使用的几大情感词典以及最新的网络用语。

(f) DMN：是一个包括输入、输出、问题和情境记忆的问答模型，结合 Attention 机制选择出与问题最相关的输入向量，通过层层迭代生成记忆，最后结合记忆向量和问题向量生成答案。

论文模型与上述基准模型的对比实验结果如表 6 所示。

表 6 基于 WEIBO\_SENTI\_100K 的模型对比实验结果

| 数据集    | WEIBO_SENTI_100K |      |          |
|--------|------------------|------|----------|
|        | ACC              | R    | F1-score |
| RAE    | 71.9             | 72.1 | 72.0     |
| Bayes  | 70.3             | 68.8 | 63.2     |
| BiLSTM | 86.2             | 84.6 | 85.1     |
| CNN    | 84.8             | 84.9 | 84.9     |
| 情感词典   | 80.8             | 85.8 | 82.7     |
| DMN    | 86.4             | 86.3 | 86.6     |
| SD-DMN | 87.7             | 86.2 | 87.3     |

表 7 基于 NLPCC2013 的模型对比实验结果

| 数据集    | NLPCC2013 |      |          |
|--------|-----------|------|----------|
|        | ACC       | R    | F1-score |
| RAE    | 72.3      | 72.1 | 72.2     |
| Bayes  | 68.8      | 70.1 | 69.2     |
| BiLSTM | 85.1      | 85.0 | 84.8     |
| CNN    | 82.6      | 83.2 | 82.1     |
| 情感词典   | 84.8      | 87.8 | 85.2     |
| DMN    | 87.9      | 87.3 | 87.5     |
| SD-DMN | 89.1      | 87.9 | 88.6     |

从表 7 中可以直观的看出，机器学习方法（朴素贝叶斯、RAE）所得的情感分类效果明显低于深度学习方法（BiLSTM、CNN、DMN），这是因为深度学习模型拥有强大的特征捕捉能力，可以通过海量数据的训练学习更精确的情感特征，而机器学习模型只能刻画比较有限且简单的数据特征，因此分类效果不如深度学习方法好。

进一步对比表中数据可以发现论文所提出的 SD-DMN 模型在两个数据集中都取得了最高的情感分类准确率、召回率和 F1 值。其中，

论文模型在 WEIBO\_SENTI\_100K 的实验中取得了 87.7% 的准确率，相比于单独的情感词典和动态记忆网络模型分别提高了 6.9% 和 1.3%，F1 值也相应提高了 4.6% 和 0.7%；在 NLPCC2013 中取得了高达 89.1% 的准确率，同样的比单独的情感词典和动态记忆网络模型分别提高了 4.3% 和 1.2%。此外，论文模型与传统的机器学习方法和深度学习方法相比模型性能都有了一定的提高。在上述基准模型中分类效果最佳的是 BiLSTM，而本文模型的准确率、召回率和 F1 值在其基础上均有所提升。从上述实验结果可以发现，朴素贝叶斯作为机器学习中的经典模型分类效果不佳，主要原因可能是因为其算法原理比较简单相当于一个判断器的作用，由于短文本的语义稀疏性无法深度挖掘语义特征之间的

联系。RAE 模型无法衡量节点之间的权重从而无法保证微博信息提取的准确性。CNN 和 BiLSTM 模型的情感分类结果相对比较好，可以提取局部的关键特征有效捕捉情感词汇，但是单一的深度学习模型依赖于训练文本，对于新型词汇的敏感度不高，并且当句子中出现特殊情感词或者多个情感词时可能无法准确识别而出现误分类的情况。本文提出的混合模型可以更好的利用情感词典与深度学习模型的优势，在三个评价指标上都得到了提升，可以证明本文模型的有效性。

### 3.5 错误分析

虽然本文的模型与传统的情感分类模型相比性能得到了提高，但是在测试时仍然能发现一些误分类的样本，例如表 8 所示。

表 8 实验错误结果示例

|   | Content                              | DMN | 情感词典 | SD-DMN | 标签 |
|---|--------------------------------------|-----|------|--------|----|
| 1 | 求看动作血腥暴力悬疑推理科幻等类型影片[哈哈]              | 正面  | 负面   | 负面     | 正面 |
| 2 | 好嫩呢好嫩呢我批准你加入萌系了[哈哈]第二三张是被怪叔叔骚扰了吗[哈哈] | 正面  | 负面   | 负面     | 正面 |
| 3 | 三亚的晚餐酒会@黄静昆Daind @侯新庆 @现代可汗 @蓝色药_玩   | 负面  | 正面   | 负面     | 正面 |
| 4 | 仲有乜? 好? 还能说什么? 所谓的传媒, 什么样的素质呀?       | 正面  | 中性   | 正面     | 负面 |

上述示例是比较典型四类误分类样本：

#### (1) 未考虑情感方面项

可以发现第一个错误样本情感色彩浓烈，基于情感词典的情感分类模型只关注到“血腥”“暴力”这些负面情感词，但是忽略了情感词所对应的方面项是“影片”，虽然 DMN 模型识别到了句子的正倾向性，但最后混合模型的结果仍然偏向于负面。

#### (2) 玩笑语气

第二类错误样本是由于无法正确识别语

气导致的，例如样本 2 中的“怪叔叔”“骚扰”等词汇会误导情感词典将其归为负向情感样本，而基于深度学习的模型会联系上下文语义进行深层次的情感判断，混合模型最后的结果因为情感词典给出的负分值较大导致归类错误。

#### (3) 微博文本的特殊性

样本 3 与其他句子的不同点在于将圈出的用户昵称也涵盖在内，在数据预处理时只是单纯提取出中文字符的做法会把带有情感色彩的

昵称也涵盖在文本内，而昵称内容本不该影响句子的情感分类，这也是第三类样本情感倾向判断错误的主要原因。

#### (4) 无情感词

第四类样本没有明显的情感词，所以在情感词典中倾向于把样本认为是中性的，此时情感分类结果完全取决于深度学习模型，此时的混合模型无法起到平衡分值的作用。

从上述的分析中可以对现有模型做后续的改进。首先在预处理过程中可以提前识别并去除用户昵称；其次可以结合文本的表情符号判断文本的语气从而改变语气词的程度分值；最后现有情感词典所覆盖的情感词是有限的。由于很多像微博这样的社交平台的核

UGC(用户输出内容)，网络新词的迭代速度非常快且具有口语化、非正式化的特点使得模型的泛化能力降低，因此可以引入 PMI 算法计算新词与现有词之间的语义相似度，将新词加入情感词典中。

### 3.6 模型多跳结构分析

为了分析 SD-DMN 模型的多跳结构对于整个模型性能的有效性以及不同跳数设置对于不同数据集的影响，在两个数据集上取不同跳数做对比实验，以获得对于每个数据集最佳的跳数设置，提升模型性能。分别绘制出跳数在 1~8 范围中变化的两个数据集的情感分类准确率，准确率结果如图 5 和图 6 所示。

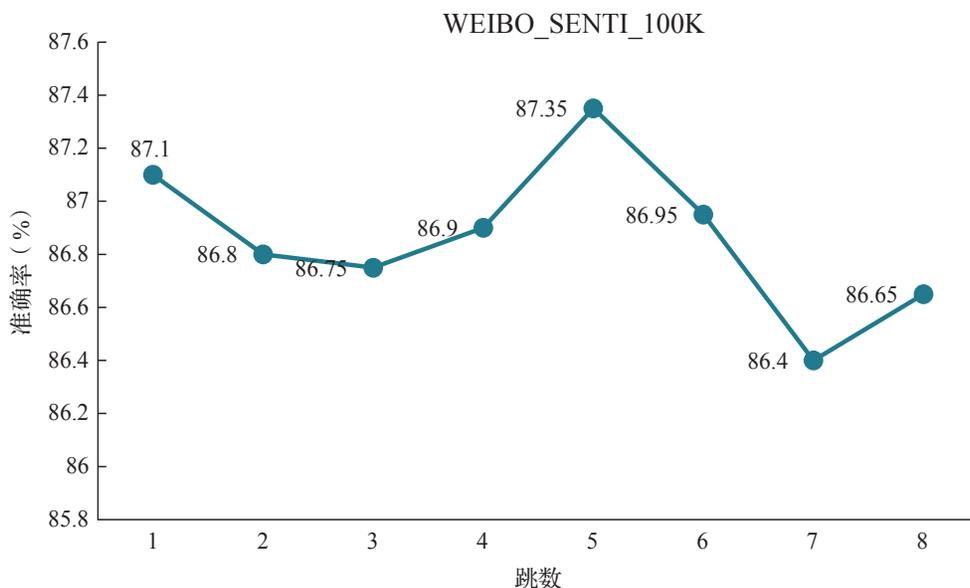


图 5 不同跳数下数据集 WEIBO\_SENTI\_100K 上的分类准确率

从两张实验结果图中我们可以发现，对于不同的数据集来说，模型的最优跳数是不同的，其中 WEIBO\_SENTI\_100K 数据集的最优跳数为 5，对应的准确率为 87.35%；NLPC2013 数据集的最优跳数为 6，对应的准确率为 88.59%。通

过两张图的对比我们可以发现，准确率的变化折线图是不规则的，并没有一个标准的形状，但是他们有共同的趋势，随着跳数的增加，模型的分类准确率会逐渐上升至一个最优值，随着跳数的继续增加，模型的分类准确率呈现下降的趋势，

这是因为如果记忆网络的层数过多可能会导致过拟合问题的出现。这表明，动态记忆网络的多跳结构可以优化模型的性能，特定记忆模块的叠加

结构可以让模型提取到更深层次、更准确的语义信息，但是一味的增加记忆模块数量会使得训练参数过多，反而降低模型的性能。

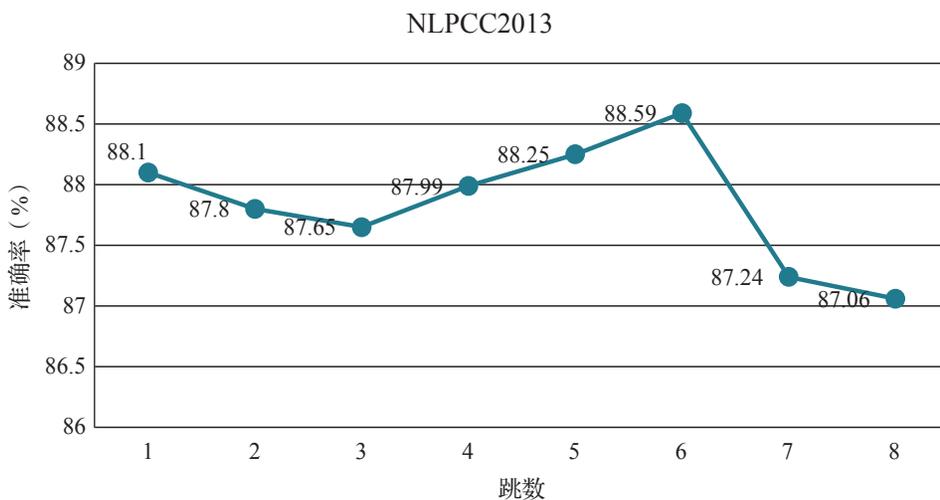


图 6 不同跳数下数据集 NLPC2013 上的分类准确率

#### 4 结束语

本文针对网络短文本的特点，通过阅读大量文献和已有研究成果，发现在文本情感分析方面主要的研究逐渐从传统情感词典分类转向了机器学习和深度学习，三种情感分类策略各有其独有的优势和缺陷，因此论文在其他研究的基础上，提出了融合情感词典和深度学习的情感分类模型。模型首先利用扩充的网络情感词典和设定的语义规则计算情感分值，其次将情感分析设计为问答任务，引入动态记忆网络，结合改进的多跳记忆结构进一步挖掘深层次的文本特征，最后将基于词典和基于优化动态记忆网络的情感分值进行线性加权得到最终的情感分类结果。经实验验证，论文模型在 WEIBO\_SENTI\_100K 和 NLPC2013 两个数据集上的分

类效果较流行的基线模型有一定的提高，并且多跳结构对于模型性能的提高也有一定的作用。虽然文本模型构建的初衷是基于社交文本的情感分类，但是可以利用混合模型的优势实现跨领域分类。动态记忆网络中的多个注意力机制可以充分捕捉文本中的重要情感词汇，根据领域文本的特点通过扩充和更改情感词典以及规则改进等方法可以将该混合模型迁移到金融、医疗等专业领域，大大降低不同领域特征分布的差异导致的误分类问题。

但是论文模型在未来仍然有可提升的空间，情感词典和语言 / 规则需要进一步的扩充和完善，例如需要考虑否定词、程度词与情感词的位置关系，以进一步提高词典分类策略的质量。此外，动态记忆网络的情境记忆模块可以加入词性等辅助信息让模型更高效的挖掘出复杂信

息, 优化模型性能。

## 参 考 文 献

- [1] 沈超, 王安宁, 方钊, 等. 基于在线评论数据的产品需求趋势挖掘 [J]. 中国管理科学, 2021, 29(5):211-220.
- [2] Rao Y, Xie H, Li J, et al. Social emotion classification of short text via topic-level maximum entropy model[J]. Information & Management, 2016, 53(8):978-986.
- [3] 李实, 叶强, 李一军, 等. 挖掘中文网络客户评论的产品特征及情感倾向 [J]. 计算机应用研究, 2010, 27(8):3016-3019.
- [4] 李君轶, 张妍妍. 大数据引领游客情感体验研究 [J]. 旅游学刊, 2017, 32(9):8-9.
- [5] 彭敏, 席俊杰, 代心媛, 等. 基于情感分析和 LDA 主题模型的协同过滤推荐算法 [J]. 中文信息学报, 2017, 31(2):194-203.
- [6] Ku L W, Chen H H. Mining opinions from the Web: Beyond relevance retrieval[J]. Journal of the American Society for Information Science & Technology, 2010, 58(12):1838-1850.
- [7] Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining[C]. Proceedings of the International Conference on Language Resources and Evaluation(LREC). Valletta, Malta: LREC. 2010: 2200-2204.
- [8] Ebert S, Vu N, Schütze H. A Linguistically Informed Convolutional Neural Network[C]. Proceedings of the 6<sup>th</sup> Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Lisbon: ACL Press, 2015: 109-114.
- [9] 陈国兰. 基于情感词典与语义规则的微博情感分析 [J]. 情报探索, 2016(2):1-6.
- [10] 杨开漠, 吴明芬, 陈涛. 广义文本情感分析综述 [J]. 计算机应用, 2019, 39(S2):11-19.
- [11] Ahmad M, Aftab S, Muhammad S S, et al. Machine Learning Techniques for Sentiment Analysis: A Review[J]. International Journal Of Multidisciplinary Sciences And Engineering, 2017, 8(3):27-32.
- [12] Singh J, Singh G, Singh R. Optimization of sentiment analysis using machine learning classifiers[J]. Human-centric Computing and Information Sciences, 2017, 7(1):32.
- [13] Kim Y. Convolutional neural networks for sentence classification [C]. Conference on Empirical Methods in Natural Language Processing, October 25-29, 2014, Doha, Qatar. 2014:1746-1751.
- [14] Zhu X, Yin S, Chen Z. Attention based bilstm-mcnn for sentiment analysis[C]. 2020 IEEE 5<sup>th</sup> International Conference on Cloud Computing and Big Data Analytics (ICCCBDA). IEEE, April 10-13, 2020, Chengd, China. 2020:170-174.
- [15] Kamps J, Marx M, Mokken R J, et al. Using WordNet to measure semantic orientations of adjectives[C]. International Conference on Language Resources & Evaluation. European Language Resources Association (ELRA), May24-30, 2004, Lisbon, Portugal. 2004:1115-1118.
- [16] Hu M, Liu B. Mining and summarizing customer reviews[J]. Proceeding of ACM Sigkdd Int.conf. knowledge Discovery & Data Mining, 2004:168-177.
- [17] Hatzivassiloglou V, Mckeown K R. Predicting the Semantic Orientation of Adjectives[J]. Proceedings of the ACL, 2002:174-181.
- [18] Georgios P, Mike T. Twitter, My Space, Digg: Unsupervised Sentiment Analysis in Social Media[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(4):66.
- [19] Ma B N, Huang Y F, Deng B X. Emotion dictionary construction based on emoticons[J]. Computer Engineering and Design, 2016, 37(5):1129-1133.
- [20] 林江豪, 顾也力, 周咏梅, 等. 基于表情符号的情感词典的构建研究 [J]. 计算机技术与发展, 2019, 29(6):181-185.
- [21] 王志涛, 於志文, 郭斌, 等. 基于词典和规则集的中文微博情感分析 [J]. 计算机工程与应用, 2015, 51(8):218-225.
- [22] Maks I, Vossen P. A lexicon model for deep sentiment analysis and opinion mining applications - ScienceDirect[J]. Decision Support Systems, 2012, 53(4):680-688.

- [23] Mnih V, Heess N M O, Graves A, et al. Recurrent models of visual attention[M]. MIT Press, 2014:2204-2212.
- [24] 冯兴杰, 张志伟, 史金钊. 基于卷积神经网络和注意力模型的文本情感分析 [J]. 计算机应用研究, 2018, 35(5):1434-1436.
- [25] 胡荣磊, 芮璐, 齐筱, 等. 基于循环神经网络和注意力模型的文本情感分析 [J]. 计算机应用研究, 2019, 36(11):3282-3285.
- [26] 宋云峰, 任鸽, 杨勇, 等. 基于注意力的多层次混合融合的多任务多模态情感分析 [J]. 计算机应用研究, 2022(3):1-6.
- [27] 王盛玉, 曾碧卿, 商齐, 等. 基于词注意力卷积神经网络模型的情感分析研究 [J]. 中文信息学报, 2018, 32(9):123-131.
- [28] Weston J, Chopra S, Bordes A, et al. Memory Networks [C]. The 3<sup>rd</sup> International Conference on Learning Representations(ICLR), 2015, May 7-9, San Diego, CA, USA, arXiv preprint arXiv:1410.3916.
- [29] Kumar A, Irsoy O, Ondruska P, et al. Ask me anything: dynamic memory networks for natural language processing[C]. Proceedings of the 33<sup>rd</sup> International Conference on Machine Learning. New York: ICML Press, June 19-24, 2016, New York City, NY, USA. 2016: 1378-1387.
- [30] Hao J, Wang X, Shi S, et al. Multi-granularity self-attention for neural machine translation[C]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9<sup>th</sup> International Joint Conference on Natural Language Processing (EMNLP -IJCNLP), November 3-7, 2019, AsiaWorld-Expo, Hong Kong, China. 2019: 887-897.
- [31] weibo\_senti\_100k[EB/OL]. [2018-04-02]. <https://github.com/SophonPlus/ChineseNlpCorpus>.
- [32] NLPCC2013[EB/OL]. [2017-04-25]. [http://tcci.ccf.org.cn/conference/2013/pages/page04\\_eva.html](http://tcci.ccf.org.cn/conference/2013/pages/page04_eva.html).