

# 基于动态注意力 GRU 的特定目标情感分类

李丽双\*, 周安桥, 刘阳, 钱爽, 耿浩彭

大连理工大学计算机科学与技术学院, 大连 116024

\* 通信作者. E-mail: lilishuang314@163.com

收稿日期: 2018–10–18; 修回日期: 2019–02–22; 接受日期: 2019–05–05; 网络出版日期: 2019–08–07

国家自然科学基金 (批准号: 61672126) 资助项目

**摘要** 特定目标情感分类是一个极具挑战的任务, 其目的是确定上下文语境中某些实体所表达的情感. 由于目标实体的情感依赖于实体本身以及对上下文的理解, 在分类时应该同时考虑句子和目标实体, 并且需要将目标实体与上下文语义进行充分的整合. 因此, 本文提出: (1) 句子、目标实体联合编码方法; (2) 基于动态注意力 DAGRU (dynamic attention gated recurrent unit) 的特定目标情感分析方法. 联合编码是指同时对句子和目标实体进行语义编码, 能为目标实体增加上下文语义; 动态注意力机制能动态地改变模型对上下文单词的注意力以及目标实体的表示, 从而更有效地获取上下文情感特征, 使模型更准确地识别不同目标实体的情感类别. 本文模型在 SemEval2014 的两个数据集 Laptop, Restaurant 上进行实验, 实验结果表明, 基于动态注意力 DAGRU 的模型相比基于标准注意力的模型结果有显著提高.

**关键词** 注意力机制, GRU, 情感分析, 深度学习, 自然语言处理

## 1 引言

特定目标情感分类 (aspect-based sentiment analysis, ABSA) 是情感分析领域的基础任务<sup>[1~3]</sup>, 旨在利用自然语言处理技术分析社交网络短文本的情感倾向, 帮助人们了解社交评论背后的观点和情感. 在该任务中, 给定一个句子以及该句子中的特定目标实体 (aspect), 其中的目标实体每个都包含情感极性, 任务的目的是推断出各个目标实体的情感极性 (包括积极、消极、中性). 表 1 列出了情感分析的两个句子, 每个句子包含若干目标实体, 表中第 2 列为目标实体, 第 3 列是目标实体的情感极性. 以表 1 中的句子 (a) 为例, 目标实体 “food” 的情感极性为积极的 (positive), 而 “service” 为消极的 (negative).

情感分类一般采用基于浅层机器学习方法<sup>[4~6]</sup>和深度学习方法. 例如, Kiritchenko 等<sup>[5]</sup>利用浅层机器学习方法设计了一套人工特征 (词袋模型 BOW、情感词典, 以及其他语义特征等), 利用这些特

**引用格式:** 李丽双, 周安桥, 刘阳, 等. 基于动态注意力 GRU 的特定目标情感分类. 中国科学: 信息科学, 2019, 49: 1019–1030, doi: 10.1360/N112018-00280

Li L S, Zhou A Q, Liu Y, et al. Aspect-based sentiment analysis based on dynamic attention GRU (in Chinese). Sci Sin Inform, 2019, 49: 1019–1030, doi: 10.1360/N112018-00280

表 1 情感分类实例

Table 1 Examples of sentimental classification

Sentence	Target entity	Sentiment polarity
(a) Great food but the service was dreadful	Food	Positive
	Service	Negative
(b) Except Patrick, all other actors don't play well	Patrick	Positive

征通过训练支持向量机 (SVM) 来进行情感分析. 浅层机器学习方法依靠大量复杂的特征集提高模型性能, 但特征设计和抽取会耗费大量人力物力.

相比之下, 深度学习能自动组合底层特征形成更复杂的高层特征, 避免大量人工特征使用. 例如, Tang 等<sup>[7]</sup> 利用两个 LSTM (long short-term memory) 对上文和下文两部分分别进行语义编码得到两个语义向量, 再用这两个语义向量拼接的向量进行情感分类. 该方法通过 LSTM 利用句子的上下文信息, 但 LSTM 对长句表现不佳, 这种做法可能丢失远距离词的情感信息.

为了解决远距离词的情感信息获取问题, 注意力机制 (attention mechanism) 被引入情感分析任务中. 例如, Wang 等<sup>[8]</sup> 利用单向 LSTM 获得各个单词的上文信息, 再利用注意力机制加权求和获得句子表示, 通过该句子表示进行情感预测. 其存在的问题主要有两个: (1) 单向 LSTM 只能获取上文信息, 无法获取下文信息; (2) 简单的注意力机制使注意力过于分散, 难以获得所有情感信息, 且容易引入过量噪声. Tang 等<sup>[9]</sup> 认为特征词的注意力值应该和特征词与目标实体的相对距离有关, 在注意力机制中加入了相对距离的信息, 而且使用多注意力机制获得最终实体表示, 同样通过该实体表示预测情感. 这种多注意力机制能提高情感信息在实体表示的比重, 弱化噪声的影响, 但未对句子上下文语义进行建模. Chen 等<sup>[10]</sup> 设计了 RAM (recurrent attention on memory) 解决上下文语义获取和注意力分散问题, 通过多注意力机制融合长难句中重要的特征, 并通过 GRU (gated recurrent unit) 组合多注意力的结果, 将目标实体情感编码成情感向量, 最后用这个情感向量预测目标实体情感.

上述方法中采用的多注意力机制虽然能捕获更多情感信息, 但每次注意力未考虑上一次注意力的有用信息. 直观地看, 多注意力机制中上一次注意力能够包含排除噪声的一些信息, 若能有效利用这些信息减少本次注意力注意到的噪声, 则本次注意力的结果也将更准确. 例如, 对于表 1 中的句子 (b), 单词 “except” 和短语 “don't play well” 一起对目标实体 “Patrick” 产生积极的情感. 假如多注意力机制上一次错误地注意到 “all other”, 那么本次注意力利用上次的错误信息就能减少 “all other” 在本次注意力中的影响.

本文基于该思想提出基于句子、实体联合编码以及动态注意力机制 DAGRU (dynamic attention gated recurrent unit) 的情感分析模型. 首先, 使用一个双向 GRU 同时对句子、实体进行语义编码 (即联合编码), 得到实体表示和整句表示. 这里, 双向 GRU 用来建模整句的上下文信息, 而联合编码策略能让实体表示也拥有上下文信息, 加强了实体表示. 然后, 将整句表示作为记忆体 (memory), 利用动态 DAGRU 多注意力机制提取记忆体的情感信息并融入实体表示中. 最后, 在最终实体表示上预测目标实体情感.

本文提出的动态注意力机制可以循环地利用 DAGRU 单元更新注意力结果和实体表示, 最终的实体表示就能够包含更为准确的情感信息. Chen 等<sup>[10]</sup> 采用 GRU 单元组合多次注意力的结果, 但未考虑多次注意力结果之间的关系. 如上文提到的例子, 上一次注意结果有助于提高本次注意力结果的准确性, 因此本文模型更新注意力结果时, 利用多层感知机提取了上次注意力信息.

该模型在 SemEval2014<sup>[3]</sup> 的两个数据集: Laptop 和 Restaurant 上进行实验, 实验结果表明, 基于

动态注意力 DAGRU 的模型相比基于标准注意力的模型结果有显著提高, 准确率分别达到 76.33% 和 81.96%.

## 2 相关工作

特定目标情感分类属于细粒度的情感分析任务, 旨在识别句子中的目标实体情感极性. 本文工作针对 SemEval2014 的目标实体情感极性识别任务进行<sup>[3]</sup>, 该任务确定句子中每个目标实体的情感极性, 研究方法一般分为浅层机器学习方法和深度学习方法.

机器学习方法利用语义分析、情感词典等外部资源获得特征, 再结合机器学习算法进行情感预测. 比如, Kiritchenko 等<sup>[5]</sup> 利用词袋模型 BOW、情感词典, 以及语义解析等外部资源构建特征, 通过训练支持向量机 (SVM) 来进行情感分类. 这类方法表现较好, 但结果依赖于特征设计与抽取, 需要消耗大量人力物力.

为克服上述缺点, 很多深度学习方法的模型被提出, 多种神经网络被用于这些模型并取得一定的成果, 比如, 递归神经网络 (recursive neural network)<sup>[11~13]</sup>、循环神经网络 (recurrent neural networks)<sup>[14]</sup> 和树形 LSTM (tree-structured LSTM)<sup>[15]</sup> 等. 这些基于神经网络的方法仅利用了上下文信息, 未考虑极性判断的关键 - 目标实体.

Jiang 等<sup>[16]</sup> 首次阐述了情感分类中目标实体的重要性, 并证明了情感分类错误的 40% 是未考虑目标实体所导致的. 为了加入目标实体的信息, Tang 等<sup>[7]</sup> 提出 TD-LSTM, 使用两个 LSTM 分别对目标实体的上下文文本进行建模, 以此考虑实体. Wang 等<sup>[8]</sup> 提出基于 attention 的 LSTM, 通过 LSTM 编码句子, 利用 attention 捕获编码后的句子表示中的重要信息.

由于简单注意力机制会使注意力过于分散, 一些复杂 attention 机制被提出. Tang 等<sup>[9]</sup> 设计了一个深度记忆网络, 它由多个收集目标信息的 attention 计算层组成, 通过线性方式组合这些信息, 改善了 attention 的准确性. Ma 等<sup>[17]</sup> 使用两个 attention 网络交互地捕获目标实体与上下文重要的情感信息并形成各自的交互表示, 通过两个交互表示进行情感预测. Chen 等<sup>[10]</sup> 提出利用循环 attention 结构捕获情感信息, 并通过 GRU 对 attention 结果进行非线性组合, 使模型更好地获取复杂文本的情感信息. Huang 等<sup>[18]</sup> 借助机器阅读理解中的 attention over attention (AOA) 机制使模型更好地抽取复杂文本的语义信息. Zheng 等<sup>[19]</sup> 提出使用 rotatory attention mechanism (LCR-Rot) 轮流地从目标实体与上下文中抽取交互信息, 以此进行情感分类, 有效地减少了池化函数损失的信息.

本文为了考虑目标实体以及解决注意力分散问题, 提出了一个基于动态注意力 DAGRU 的情感分类模型. 模型首先通过联合编码同时考虑目标实体与上下文得到各自的交互表示, 再通过 DAGRU 将句子中重要信息融入目标实体表示并过滤目标实体中的噪声, 经多次循环后得到最终实体表示, 最后利用 Softmax 分类器进行分类.

## 3 基于动态注意力 DAGRU 的模型

本文所提出的基于动态注意力 DAGRU 模型总体框架图如图 1 所示, 主要由以下 5 部分组成:

- (1) 输入层. 处理该模型的输入 (包括单词序列和位置序列), 以及对这些输入进行向量嵌入操作.
- (2) 语义编码层. 双向 GRU, 对输入进行上下文语义编码.
- (3) 记忆体 (memory). 编码层所有隐层状态输出.

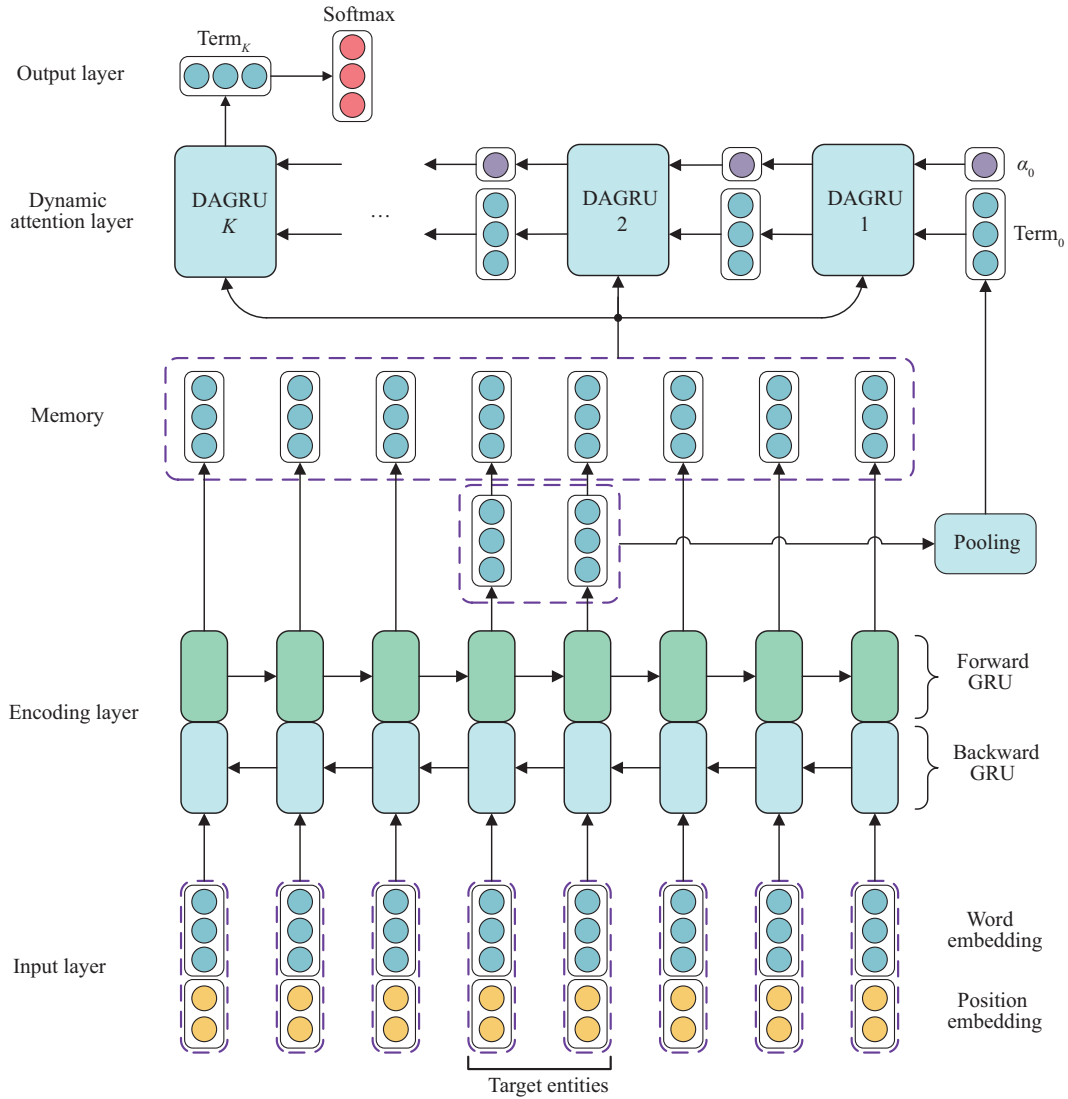


图 1 (网络版彩图) 基于动态注意力 DAGRU 模型总体框架图  
 Figure 1 (Color online) The architecture of DAGRU model

(4) 动态注意力层. 动态地改变内部 attention, 同时向实体表示内融入不同 attention 值的句子信息, 得到最终实体表示.

(5) 输出层. 利用动态注意力层的输出, 即利用最终实体表示进行情感预测.

### 3.1 输入层

该模型有两个输入, 一个是单词序列, 另一个是单词序列中单词相对目标实体的位置序列. 为了获得两者的语义信息, 需要将输入序列转化为向量形式, 利用向量嵌入可完成该操作. 句子嵌入成词向量后的表示为  $S = \{w_1, w_2, \dots, w_i, \dots, w_n\}$ , 其中  $w_i \in \mathbb{R}^{d_s}$  为句子中第  $i$  个单词的向量表示 ( $i \in [1, n]$ ),  $n$  为句子长度,  $d_s$  为词向量维度. 对句子中目标实体 (目标实体可能不止一个词), 假设为  $\text{Term} = \{w_{\tau+1}, w_{\tau+2}, \dots, w_{\tau+m}\}$ , 其中,  $m$  为目标实体长度,  $\tau + 1$  为目标实体在句中起始位置. 位置序

列嵌入后的表示为  $P = \{p_1, p_2, \dots, p_j, \dots, p_n\}$ , 其中  $p_j \in \mathbb{R}^{d_p}$  为第  $j$  个单词与目标实体相对位置的向量表示 ( $j \in [1, n]$ ),  $d_p$  为位置向量维度. 将这两个向量序列进行拼接, 作为编码层的输入.

### 3.2 语义编码层与记忆体 memory

循环神经网络 (recurrent neural networks, RNN) 是自然语言处理常用的神经网络模型, 在处理文本序列方面有独特优势. 但传统 RNN 难以处理长期依赖. 针对这一问题, Cho 等<sup>[20]</sup> 提出了 GRU, 通过特殊的门结构解决长期依赖问题. GRU 功能是将输入序列映射成一个新的序列, 这个新序列包含高层语义信息. 假定输入序列为  $X = [x_1, x_2, \dots, x_T]$ , 输出序列为  $H = [h_1, h_2, \dots, h_T]$ , 其中  $T$  为序列长度, 则 GRU 可如下形式化描述:

$$\begin{aligned} r_t &= \sigma(W_r[x_t, h_{t-1}] + b_r), \\ z_t &= \sigma(W_z[x_t, h_{t-1}] + b_z), \\ \tilde{h}_t &= \text{act}(W_h[x_t, r_t \odot h_{t-1}] + b_z), \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \end{aligned} \quad (1)$$

其中,  $r_t$  为重置门,  $z_t$  为更新门,  $\tilde{h}_t$  为候选输出,  $\sigma$  为 sigmoid 函数,  $\text{act}$  为 GRU 的激活函数, 下标  $t$  代表序列当前时刻,  $t-1$  代表上一时刻.

但单向 GRU 只能获得句子中单词的上文语义, 不能得到下文语义, 所以本文语义编码层采用双向 GRU (BiGRU) 编码单词序列的上下文语义, 且输出由前向 GRU 和后向 GRU 拼接得到. 为了方便表示, 这里分别定义序列  $X$  的单向 GRU 和 BiGRU 语义编码公式如下:

$$h_t^X = \text{GRU}_X(h_{t-1}^X, x_t), \quad h_t^X = \text{BiGRU}_X(h_{t-1}^X, x_t). \quad (2)$$

本文模型是对句子、实体联合编码, 因此并未显式输入目标实体, 而是直接从整个句子语义编码中提取. 具体操作是先将单词序列和位置序列进行拼接, 然后再通过 BiGRU, 如下所示:

$$h_t^S = \text{BiGRU}_S(h_{t-1}^S, [w_t, p_t]), \quad (3)$$

其中,  $w_t, p_t$  分别为输入层句子的第  $t$  个词向量和位置序列的第  $t$  个位置向量 ( $t \in [1, n]$ );  $h_t^S \in \mathbb{R}^{2d}$  为句子编码第  $t$  个隐层状态,  $d$  为双向 GRU 隐层单元数量.

值得注意的是, 式 (3) 已经完成了句子、实体联合编码, 编码仅使用一个 BiGRU, 已编码的目标实体表示包含在句子中, 后续经过特定操作取出. 这里, 联合编码的句子表示作为记忆体  $M$ , 其作用是存储着整个句子的信息, 以供后续操作提取信息. 记忆体  $M$  与目标实体表示  $T$  形式化描述如下所示:

$$M = \{h_1^S, h_2^S, \dots, h_t^S, \dots, h_n^S\}, \quad T = \{h_{i+1}^S, h_{i+2}^S, \dots, h_{i+m}^S\}. \quad (4)$$

### 3.3 基于 DAGRU 动态注意力层

考虑到错误的 attention 可能引入不相关的信息, 本文提出动态 attention 机制, 通过考虑以前的 attention 信息, 排除以前误注意到的干扰项, 减少实体表示中的噪声, 提升目标实体最终表示的准确性, 使情感预测更准确. 图 1 顶部描述了动态注意力层的总体流程. 首先, 将目标实体序列进行池化, 其结果作为初始状态的实体表示  $\text{Term}_0 \in \mathbb{R}^{2d}$ , 与初始化的注意力值  $\alpha_0$  一起送入 DAGRU. 经过  $K$  次 DAGRU 后, 得到最终实体表示  $\text{Term}_K$ . 公式描述如下所示:

$$\text{Term}_0 = \text{Pooling}(T), \quad [\text{Term}_k, \alpha_k] = \text{DARGU}_{\text{Term}}(M, [\text{Term}_{k-1}, \alpha_{k-1}]), \quad (5)$$

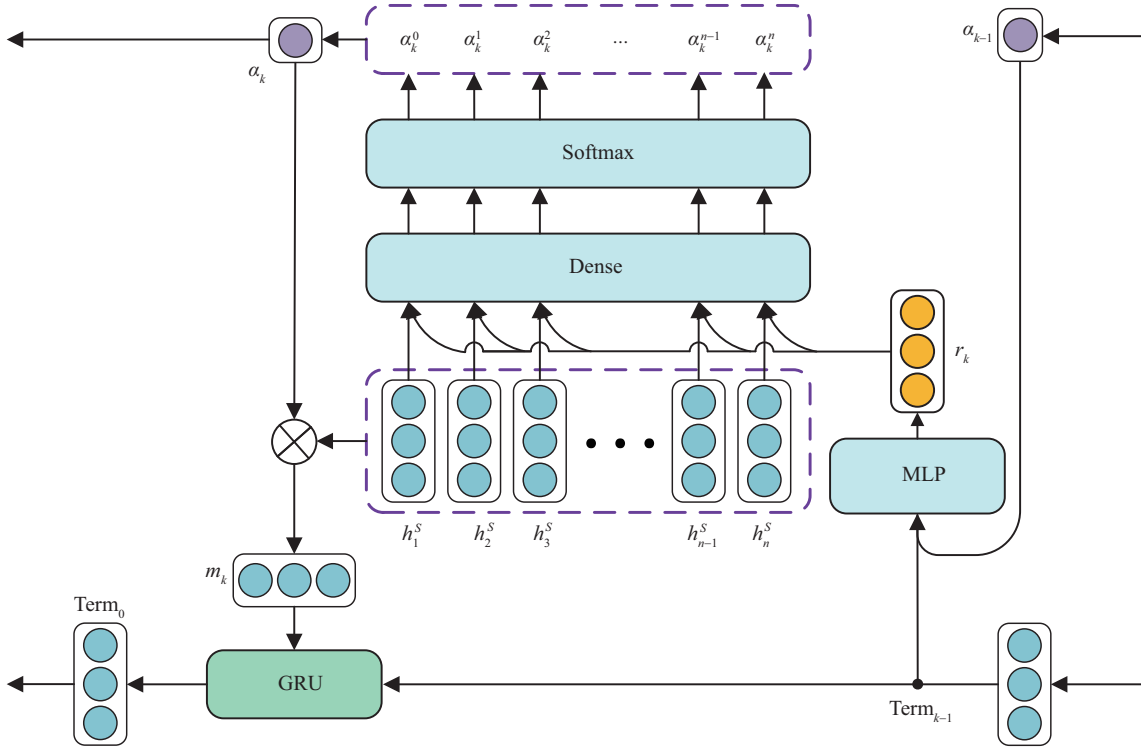


图 2 (网络版彩图) DAGRU 单元图  
Figure 2 (Color online) DAGRU unit

其中, Pooling 为池化函数,  $\text{Term}_k, \alpha_k$  为第  $k$  次 DAGRU 的两个隐层状态输出 ( $k \in [1, K]$ ). 本文实验初始化  $\alpha_0$  时,  $\alpha_0^i = 1/n, i = 1, 2, \dots, n$ .

DAGRU 的实现参考阅读理解任务中 DCN<sup>[21]</sup> 模型指针动态更新过程, 图 2 描述了 DAGRU 内部实现细节. 首先  $\text{Term}_{k-1}$  与  $\alpha_{k-1}$  一起通过一个多层感知机 (MLP), 得到包含上一时刻实体和注意力信息的向量  $r_k$ , 形式化描述如下:

$$r_k = \text{MLP}([\text{Term}_{k-1}, \alpha_{k-1}]). \quad (6)$$

本文实验时使用的 MLP 具体表达式为

$$\text{MLP}([\text{Term}_{k-1}, \alpha_{k-1}]) = \text{act}_M(W_M[\text{Term}_{k-1}, \alpha_{k-1}] + b_M), \quad (7)$$

其中,  $W_M$  为可训练权重,  $b_M$  为偏置量,  $\text{act}_M$  为多层感知机的激活函数.

接着让  $r_k$  与序列  $M = \{h_1^S, h_2^S, \dots, h_i^S, \dots, h_n^S\}$  一起通过 Dense 层得到记忆体中每个词的得分  $s_k^t$ , 经过 softmax 归一化获得  $\alpha_k$ , 即新的注意力值; 再利用新注意力值对记忆体每个元素进行加权求和得到新记忆体向量表示  $m_k$ ; 最后输入 GRU, 将新的上下文信息融入  $\text{Term}_{k-1}$  得到  $\text{Term}_k$ :

$$s_k^t = \text{act}_D(W_D[h_t^S, r_k] + b_D), \quad \alpha_k^i = \frac{\exp(s_k^i)}{\sum_{j=1}^n \exp(s_k^j)}, \quad (8)$$

$$\alpha_k = \{\alpha_k^1, \alpha_k^2, \dots, \alpha_k^n\}, \quad m_k = \sum_{i=1}^n \alpha_k^i h_i^S, \quad \text{Term}_k = \text{GRU}(\text{Term}_{k-1}, m_k),$$

表 2 数据集样本分布  
Table 2 The statistics of the datasets

Sentiment polarity	Laptop		Restaurant	
	Train	Test	Train	Test
Positive	994	341	2164	728
Neutral	464	169	637	196
Negative	870	128	807	196
Total	2328	638	3608	1120

其中,  $W_D$  为可训练权重,  $b_D$  为偏置量,  $\text{act}_D$  为 Dense 层的激活函数.

经过  $K$  次 DAGRU 后, 得到实体最终表示  $\text{Term}_K$  和最终注意力值  $\alpha_k$ , 其中  $\text{Term}_K$  作为最终情感预测的特征.

### 3.4 输出层与模型训练

输出层的工作就是利用实体表示预测情感, 其输入为动态注意力层输出之一  $\text{Term}_k$ . 假设在句子  $S$  中, 目标实体  $\text{Term}$  的情感极性为  $y$  的概率为  $p(y|S, \text{Term})$ , 则预测情感类别的计算公式如下:

$$\begin{aligned} p(y|S, \text{Term}) &= \text{Softmax}(W_S \cdot \text{Term}_K + b_S), \\ \hat{y} &= \text{argmax}_{y \in C} \{p(y|S, \text{Term})\}, \end{aligned} \quad (9)$$

其中,  $W_S$  为可训练权重,  $b_S$  为偏置量,  $C = \{\text{positive}, \text{negative}, \text{neutral}\}$  为情感类别集合.

本文模型使用交叉熵损失函数进行训练, 损失函数公式如下:

$$L = - \sum_i \sum_j \bar{y}_i^j \log(p(y_i = j|S, \text{Term})), \quad (10)$$

其中,  $i$  为第  $i$  个样本下标 ( $i \in [1, N]$ ,  $N$  为样本总数),  $j$  为情感类别下标 ( $j \in C$ ),  $\bar{y}_i$  为第  $i$  个样本的真实标签.

## 4 实验与分析

### 4.1 数据集与预处理

本文在 SemEval2014 的两个数据集: Laptop, Restaurant 上进行实验. 目标实体的情感极性分为积极、中性和消极, 各数据集上具体数量统计如表 2 所示.

在进行实验前, 对原始数据集进行了以下预处理:

(1) Restaurant 数据集的情感极性还包括矛盾 (conflict), 由于相关文献都不考虑这类情感, 因此本文也移除了这类样本.

(2) 数据集的分词工具采用 Keras 自带的 Tokenizer 分词工具.

(3) 由词与目标实体相隔的单词数生成相对位置序列, 在实体之前为负值, 实体之后为正值.

表 3 超参数设置表  
Table 3 Hyperparameter configuration

Parameter	Value
BiGRU hidden units	64
DAGRUs hidden units	128
Dropout	0.5
Recurrent dropout	0.5
DAGRUs layers $K$	5
Batch size	64
Learning rate	0.001

表 4 DAGRU 层数与语义编码影响表<sup>a)</sup>  
Table 4 The impacts of DAGRU layers and encoding<sup>a)</sup>

Layers	Laptop (%)		Restaurant (%)	
	Joint encoding	Individual encoding	Joint encoding	Individual encoding
0	74.92	56.27	80.09	67.50
1	75.70	72.57	80.80	80.18
2	76.02	72.88	81.07	80.18
3	75.86	72.88	81.16	80.80
4	<b>76.49</b>	73.04	81.25	<b>81.61</b>
5	76.33	<b>73.51</b>	<b>81.96</b>	81.07

a) The bold number represents the highest result.

## 4.2 模型初始化

本文词向量采用预训练好的 GloVe<sup>[22]</sup> 词向量, 向量维度选取  $d = 300$ ; 所有不在词向量词典中的词, 随机初始化为服从  $[-1, 1]$  均匀分布的 300 维随机向量. 位置向量嵌入矩阵, 由服从  $[-1, 1]$  的均匀分布随机初始化, 维度为 50. 模型中所有权重矩阵元素随机初始化为服从 glorot 均匀分布的值, 所有偏置量全部置为 0 向量.

## 4.3 模型超参数设置

本文模型使用 Keras 实现, 实验中使用的超参数值如表 3 所示.

另外, 模型中除 MLP 激活函数使用线性函数外, 其他均采用 ReLU 激活函数; 为了避免过拟合, 所有 DAGRU 使用同一套参数; 模型训练使用 RMSprop<sup>[23]</sup> 优化方法.

## 4.4 DAGRU 层数影响

实验首先测试了动态注意力机制 DAGRU 对模型的作用, 通过去除 DAGRU 或者改变 DAGRU 的层数探索其影响, 评价指标为准确率 (accuracy), 结果如表 4 所示.

实验中, 对层数从 0 到 5 进行了实验比较. 层数为 0 表示去除 DAGRU 模块的情况. 首先仅观察基于联合编码的数据, 可以看出:



(1) 在两个数据集上, 添加一层 DAGRU, 准确率分别从 74.92% 和 80.09% 提高到 75.70% 和 80.80%, 分别提升了 0.78% 和 0.71%, 这说明使用单注意力机制便能获得大量情感信息.

(2) 在两个数据集上, 从 1 到 4 层, 随着层数的增加, 结果皆逐步上升. 对 Laptop 数据集, 最高结果是层数为 4 时的 76.49%, 当层数达到 5 层时, 结果为 76.33%, 比层数为 4 时下降 0.16%. 对 Restaurant 数据集, 当层数为 5 时达到最高 81.96%. 对于这两个数据集, 多层 DAGRU 的最高值比单层的结果分别提升了 0.53% 和 1.16%. 这说明动态注意力机制能提取更多的情感信息, 使实体表达更准确. 总的性能比去除 DAGRU 模块时分别提高 1.57% 和 1.87%.

为了进一步验证动态 DAGRU 的作用, 本文也设置了对独立编码的实验, **独立编码是使用两个 BiGRU 分别对目标实体和句子编码**. 总体上与联合编码的结论一致, 即动态注意力机制能提取大量情感信息, 其最高结果比单层 DAGRU 分别提升 0.94% 和 1.43%. 但值得注意的是, 在两个数据集上, 去除 DAGRU 模块后结果与单层 DAGRU 的相差均较大, 分别低 16.30% 和 12.68%. 这是因为在无 DAGRU 模块的情况下, 相当于直接用实体表示去预测情感, 而情感分类任务需要同时考虑句子和目标实体, 该情况下相当于仅考虑目标实体, 没有任何上下文信息, 因此独立编码结果必然很低.

通过上述分析, 目标实体表示经过动态注意力层的更新, 能提高最终分类结果, 并且每经过一层动态注意力机制, 结果都得到提高. 因此, 可以得出: (1) 动态注意力每一层都能从记忆体内抓取有用的情感信息; (2) 动态注意力**不同层必定能一定程度上抓取到不同的情感信息**, 不同层的注意力结果会不同. 然而, 经过进一步的实验发现, 动态注意力层虽然能通过动态计算注意力抓取有用的情感信息融入实体表示中, 但**层数过多也会降低结果**, 这可能有以下两点原因: (1) 每层注意力计算都会为结果引入一定的噪声; (2) 若前面的动态注意力层已提取了大部分有用信息, 则后面的动态注意力层则只会为实体表示引入噪声.

#### 4.5 联合编码影响

联合编码与独立编码的对比实验结果如表 4 所示, 可以看出:

(1) 在没有 DAGRU 模块的情况下, 在两个数据集上, 联合编码比独立编码提高了 18.65% 和 12.59%. 这说明联合编码的实体表示同时考虑了句子信息, 已经包含了句子中的一些情感信息.

(2) 引入 DAGRU 模块后, 联合编码与独立编码的结果差距迅速减小. 在 Laptop 数据集上, 在 DAGRU 模块层数相同 (层数不小于 1) 的情况下, 两者结果差距保持在 3% 左右, 两者最高结果相差 2.98%; 而在 Restaurant 数据集上, 相同层数时的差距并未维持在同一水平, 甚至在层数为 4 时独立编码结果高于联合编码. 不同的数据集上出现这两种不同的现象可能是由于数据集分布不一致, 但总体上, 联合编码的结果高于独立编码, 且两个数据集的联合编码最高值比独立编码高 2.98% 和 0.35%. **说明联合编码带来的上下文信息包含着注意力机制较难注意到的信息, 从而能更好地判别实体的情感**.

#### 4.6 与其他先进工作的对比

本文与目前在 Laptop 和 Restaurant 数据集上表现较好的模型进行比较, 且在两个数据集上 DAGRU 层数统一取 5 层, 评价指标仍为准确率 (accuracy), 如表 5 所示. 从表中可看出:

(1) 深度学习模型在无人工特征的情况下就能获得较好结果, 本文模型的结果高于浅层机器学习方法的最高值. 深度学习模型 TD-LSTM<sup>[7]</sup> 未使用人工特征, 其在 Laptop 和 Restaurant 数据集上准确率分别达到 75.63% 和 68.13%. 而浅层机器学习方法依赖大量人工特征来提升模型性能, 比如, Kiritchenko 等<sup>[5]</sup> 提出的 Feature-enhanced SVM, 使用了 n-gram、句法解析树, 以及情感词典等特征,

表 5 模型总体性能比较表  
Table 5 Performance comparison of the models

Model	Laptop (%)	Restaurant (%)
Majority	53.45	65.00
Simple-SVM	66.97	73.22
Feature-enhanced SVM (Kiritchenko et al.) [5]	72.10	80.89
TD-LSTM (Tang et al.) [7]	68.13	75.63
AE-LSTM (Wang et al.) [8]	68.90	76.60
ATAE-LSTM (Wang et al.) [8]	68.70	77.20
MemNet (Tang et al.) [9]	70.33	79.98
IAN (Ma et al.) [17]	72.10	78.60
RAM (Chen et al.) [10]	74.49	80.23
AOA-LSTM (Huang et al.) [18]	74.50	81.20
LCR-Rot (Zheng and Xia) [19]	75.24	81.34
<b>DAGRU (proposed model)</b>	<b>76.33</b>	<b>81.96</b>

在两个数据集上分别达到 72.1% 和 80.9% 的准确率, 是目前浅层机器学习方法的最好结果. 本文的深度学习模型未采用任何人工特征, 准确率比浅层机器学习方法最高值 72.10% 和 80.89% 分别提高 4.23% 和 1.07%.

(2) 本文模型在基于多注意力机制的模型中结果最高. 表 5 中基于多注意力机制的模型有 Tang 等 [9] 提出的模型 MemNet 以及 Chen 等 [10] 提出的模型 RAM. MemNet 模型通过简单的线性操作组合不同注意力的实体表示, 其结果分别为 70.33% 和 79.98%. RAM 模型利用 GRU 作为载体, 组合不同注意力的情感向量, 其结果分别为 74.49% 和 80.23%. 本文模型同样属于多注意力机制, 综合性能比具有最高准确率的模型 RAM 分别提升 1.84% 和 1.73%.

(3) 本文模型在所有深度学习模型中也达到最高. 目前最好的深度学习模型是 Zheng 等 [19] 提出的 LCR-Rot 模型, 在 Laptop 和 Restaurant 数据集上均为最好结果, 分别为 75.24% 和 81.34%. 本文模型在两个数据集上相比该模型分别高出 1.09% 和 0.62%.

## 5 结论与展望

特定目标情感分类是情感分析的分支任务, 其目的是确定上下文语境中某些实体所表达的情感. 本文针对该任务提出: (1) 句子、目标实体联合编码方法; (2) 基于动态注意力 DAGRU 的特定目标情感分析方法. 在 Laptop, Restaurant 上的实验结果表明, 基于动态注意力 DAGRU 的模型相比基于普通注意力的模型结果有显著提高, 准确率分别达到 76.33% 和 81.96%, 达到当前先进水平. 主要结论如下:

- (1) 联合编码同时考虑了目标实体以及上下文信息, 使目标实体包含一些情感信息, 同时能从上下文提取许多注意力机制提取不到的信息;
- (2) 动态注意力 DAGRU 能够动态改变 attention 值以及实体表示, 提高注意力和实体表示的准确性.

尽管当前结果在作者认知范围内达到最高, 但还有许多可以考虑的研究内容, 本文接下来将考虑

如下两方面的研究: (1) 由于一个句子中其他实体的情感会影响当前实体情感的判断, 因此可以考虑排除句子中其他实体的情感信息; (2) 既然实体存在于句子中, 那么可以考虑不显式分离实体与句子, 直接对隐含实体的整个句子进行情感分析.

## 参考文献

- 1 Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2008, 2: 1–135
- 2 Liu B. Sentiment analysis and opinion mining. *Synth Lect Human Lang Technol*, 2012, 5: 1–167
- 3 Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: aspect based sentiment analysis. In: *Proceedings of the 8th International Workshop on Semantic Evaluation*, Dublin, 2014. 27–35
- 4 Wagner J, Arora P, Cortes S, et al. DCU: aspect-based polarity classification for semeval task 4. In: *Proceedings of the 8th International Workshop on Semantic Evaluation*, Dublin, 2014. 223–229
- 5 Kiritchenko S, Zhu X D, Cherry C, et al. NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: *Proceedings of the 8th International Workshop on Semantic Evaluation*, Dublin, 2014. 437–442
- 6 Vo D, Zhang Y. Target-dependent twitter sentiment classification with rich automatic features. In: *Proceedings of IJCAI*, Buenos Aires, 2015. 1347–1353
- 7 Tang D Y, Qin B, Feng X C, et al. Effective lstms for target-dependent sentiment classification. In: *Proceedings of COLING*, Osaka, 2016. 3298–3307
- 8 Wang Y Q, Huang M L, Zhao L, et al. Attention-based lstm for aspect-level sentiment classification. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 2016. 606–615
- 9 Tang D Y, Qin B, Liu T. Aspect level sentiment classification with deep memory network. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Austin, 2016. 214–224
- 10 Chen P, Sun Z Q, Bing L D, et al. Recurrent attention network on memory for aspect sentiment analysis. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 2017. 452–461
- 11 Socher R, Pennington J, Huang H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Edinburgh, 2011. 151–161
- 12 Dong L, Wei F R, Tan C Q, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification. In: *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics*, Baltimore, 2014. 49–54
- 13 Qian Q, Tian B, Huang M L, et al. Learning tag embeddings and tag-specific composition functions in recursive neural network. In: *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics*, Beijing, 2015. 1365–1374
- 14 Mikolov T, Karafi'at M, Burget L, et al. Recurrent neural network based language model. In: *Proceedings of the Interspeech*, Makuhari, 2010. 1045–1048
- 15 Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks. In: *Proceedings of Annual Meeting of the Association for Computational Linguistics*, Beijing, 2015. 1556–1566
- 16 Jiang L, Yu M, Zhou M, et al. Target-dependent twitter sentiment classification. In: *Proceedings of Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, 2011. 151–160
- 17 Ma D H, Li S J, Zhang X D, et al. Interactive attention networks for aspect-level sentiment classification. In: *Proceedings of IJCAI*, Melbourne, 2017. 4068–4074
- 18 Huang B X, Ou Y L, Carley K M. Aspect level sentiment classification with attention-over-attention neural networks. 2018. ArXiv:1804.06536
- 19 Zheng S L, Xia R. Left-center-right separated neural network for aspect-based sentiment analysis with rotatory attention. 2018. ArXiv:1802.00892
- 20 Cho K, Merriënboer B V, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Doha, 2014. 1724–1734
- 21 Xiong C M, Zhong V, Socher R. Dynamic coattention networks for question answering. 2017. ArXiv:1611.01604v3

- 22 Pennington J, Socher R, Manning C D. Glove: global vectors for word representation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, Doha, 2014. 1532–1543
- 23 Tieleman T, Hinton G. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neur Netw Mach Learn, 2012, 4: 26–30

## Aspect-based sentiment analysis based on dynamic attention GRU

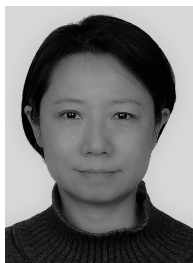
Lishuang LI\*, Anqiao ZHOU, Yang LIU, Shuang QIAN & Haopeng GENG

*College of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China*

\* Corresponding author. E-mail: lilishuang314@163.com

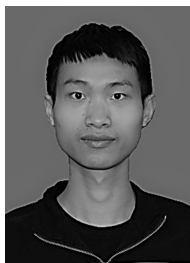
**Abstract** Aspect-level sentiment analysis is a fine-grained task that aims to identify the sentiment polarity (i.e., negative, neutral, or positive) of a specific target opinion in its context. Since the sentiment polarity of a target depends on the target itself and the semantics of the context, the target and the sentence should be treated equally and modeled interactively. For aspect-level sentiment analysis, we propose (1) a method to encode the aspect and sentence simultaneously, and (2) a neural network based on a dynamic attention gated recurrent unit. The simultaneous encoding manner can generate the target representation, which contains more contextual clues. The dynamic attention mechanism can achieve the attention values of contextual words and further generate the target representation dynamically. Experimental results achieved on a SemEval 2014 dataset (Laptop and Restaurant) show that our approach achieves a significant improvement in the accuracy rates over the standard attention-based models.

**Keywords** attention mechanism, GRU, sentiment analysis, deep learning, natural language processing



**Lishuang LI** was born in 1967. She received her B.Sc., M.Sc., and Ph.D. degrees from the Dalian University of Technology in 1989, 1992, and 2013, respectively. She is currently a professor at the School of Computer Science and Technology at the Dalian University of Technology. Her research interests include text mining, natural language processing, and machine translation. In recent years, she has focused

on text mining for biomedical literatures and information extraction from huge biomedical resources.



**Anqiao ZHOU** was born in 1994. He received his bachelor's degree in machine science from Southwest Jiaotong University, Emei, in 2017. Currently, he is a master's candidate at the Dalian University of Technology. His research interest is natural language processing.