文章编号：1003-0077（2021）00-0000-00

**对属性级情感分析的研究**

**摘要：**属性级情感分析是近几年来迅速发展起来的一个课题，受到了学界研究人员的广泛关注。本次实验我主要以时间为脉络，学习了该问题各大经典模型的原理并进行实验对比效果。本文着重选择了LSTM、GRU等经典文本分析模型，以LSTM为基础的众多改进模型，基于CNN的属性级情感分析模型、基于attention机制的属性级情感分析模型这四种模型，将这些模型应用于SemEval2014数据集，通过对比分析，比较不同模型之间的优劣，做出总结。

**关键词：**属性级情感分析；LSTM；SemEval2014; attention

**中图分类号：**TP391  **文献标识码：**A

#### Research on Aspect Based Sentiment Analysis

Ma Yiming1

(1.Dalian University of Technology ,Dalian ,Liaoning 116024, China)

**Abstract :** Aspect Based Sentiment Analysis is a subject that has developed rapidly in recent years and has attracted extensive attention from academic researchers. In this experiment, I mainly took time as the context, learned the principles of the major classical models of the problem, and conducted experiments to compare the results. This paper focuses on the selection of classic text analysis models such as LSTM and GRU, many improved models based on LSTM, Aspect Based Sentiment Analysis model based on CNN and Aspect Based Sentiment Analysis model based on attention mechanism. These models are applied to semeval2014 data set. Through comparative analysis, the advantages and disadvantages of different models are compared and summarized.

**Key words:** Aspect Based Sentiment Analysis; LSTM；SemEval2014; attention

**0 引言**

属性级情感分析（Aspect Based Sentiment Analysis, ABSA）[1]，是自然语言处理和计算机语言学中的重要任务之一。句子中的情感倾向对于理解句子的本意至关重要，属性级情感分析已经在业界和学术界引起了广泛的关注。属性级情感分析作为一种细粒度的情感分析任务，其含义为识别一条句子中对指定对象（Aspect）的情感褒贬信息。以商品为例，如果商家可以有效地分析出顾客对于商品的褒贬态度，那么对于其商业的运作是具有关键性指导作用的。给定一个句子和一个对象，我们的任务就是推断出句子对对象的情感是褒义、贬义、或者中性。比如说对于“这把尺子质量很好，但是太难看了。”这句话，对于尺子质量来说是褒义看法，如果是针对尺子的外形来说是贬义看法。而对于一个特定的句子中特定目标来说，其评价只会是褒义、贬义、中性中的一个。也就是说，基于目标的情感分析实质上是一个文本的三分类问题。这也就为此问题提供了用机器学习去解决的思路。

在这篇文章中，我将会在第1部分介绍对属性级情感分析的研究相关工作，包括从LSTM应用于基于目标的情感分析到此后的一系列模型。在第2部分我会对本次实验使用到的模型的原理进行概括说明。在本文的第3部分，将会展示本次实验的实验结果。第4部分是对本次实验的总结和模型评价。第5部分是本次实验的心得。

**1 相关工作**

属性级情感分析是文献中一种典型的文本分类问题。大多数现有研究使用监督机器学习方法构建情感分类器，如基于特征的支持向量机[2]。这些方法取得了一定的成效，但是大体上来说效果还是欠佳。因为这些算法只是从简单的分类上去考虑，而没有考虑到文本本身的特性。第一个将文本上下文特征引入到情感分析的人是Dong[1]，这篇文章将一个句子转化为了一个二叉树，我们的目标词语放在根部节点的叶子节点上。然后，利用RNN自下而上的传播到目标，将最终的结果输入到softmax分类器中进行分类，得到最后的情感分类。Tang[3]等人为了更好地提取目标词语与句子中其他词语的关系，利用LSTM创建了TD-LSTM与TC-LSTM两种模型，TD-LSTM使用两个LSTM分别去建模目标和左边的上下文以及目标和右边的上下文，将两个LSTM最后的隐含向量拼接送入到一个softmax分类器中进行分类。TC-LSTM将目标词向量的平均值与句子中每个词的词向量进行拼接然后进行和TD-LSTM一样的操作。之后，Qin[4]等人利用attention机制提出了MemNet, 句子中的每个词首先被转换成词向量表示，整个句子中所有词的词向量就组成了一个矩阵，称为memory。之后利用多层的线性层和attention层进行迭代得到的结果进行softmax分类。Wang[5]等人此后又提出了两种基于LSTM和attention机制的模型，AT-LSTM和ATAE-LSTM，这两种模型的特点是利用LSTM隐层的输出和attetion机制结合进行预测，ATAE-LSTM对输入变量进行了改进。Yang[6]等人将上述模型进行了简化，提出了一种Attention-based LSTM，即LSTM加上Attention机制。以上的模型主要关注点都在目标之上，Ma[7]等人提出了一种交互式注意力网络模型IAN，该模型不仅注重目标，也注重于文章本身，他利用LSTM分别对目标和句子的隐层输出拼接作为一个变量，更加注重相关因果。而Chen[8]等人对MemNet进行了改进，提出了RAM模型，他大体上与MemNet相似，但是对Memory的构建进行了改进。Xue[9]等人从CNN出发构建出了另一种模型，GCAE。该模型利用了n-gram特征，速度更快且鲁棒性更好。Kathleen M. Carley[10]等人借鉴了问答系统中的attention-over-attention机制提出了AOA-LSTM模型，该模型的精度更加的准确。可以说，从整体的发展上来说，目前基于方面的情感分析大多利用到了诸如LSTM[11]、GRU[12]、attention[13]机制等一系列基于上下文的模型和分析方法。这些方法在文本情感的推测上已经可以做到很高的精度。

在本篇报告中，我将以LSTM为主体，研究以LSTM为核心的众多基于上下文模型以及几个非LSTM的模型对于该数据集对比分析。

**2 选择研究模型及其原理**

**2.1 经典模型**

2.1.1 LSTM

LSTM全程为Long Short Term网络，即长短记忆神经网络。是一种 RNN 特殊的类型，可以学习长期依赖信息。LSTM 通过刻意的设计来避免长期依赖问题。记住长期的信息在实践中是 LSTM 的默认行为，而非需要付出很大代价才能获得的能力。

图示

描述已自动生成

图1 LSTM原理图

LSTM中最核心的内容是利用了门结构，即遗忘门、记忆门、输出门。三个门分别对应了三个参数的更新，遗忘门对应的参数是对过去信息的丢弃程度，记忆门是安定什么样的新信息被存放在细胞状态中，输出门则是基于以上两个门确定输出的值。

具体来说，其更新公示如下：



其中，、、分别是三个门的控制参数，为记忆细胞的值，为最终的输出值。LSTM通过这几个参数，巧妙的控制了过去信息与当前信息的平衡。

2.1.2 GRU

GRU（Gate Recurrent Unit）和长短期记忆网络LSTM(Long Short Term Memory)都是循环神经网络(Recurrent Neural Network, RNN)的一种变体。

图示

描述已自动生成

图2 GRU原理图

其更新公式如下：



GRU相比于LSTM来说更加的简洁，更新公示也更少，对于计算力的需求大大降低。因此，相比于LSTM来说计算速度方面来说GRU更加的占优势。

**2.2 基于LSTM的方面级情感分析**

2.2.1 TD-LSTM

TD-LSTM全程Target-Dependent Long Short-Term Memory。此神经网络是针对属性级情感分析这一任务对LSTM的改进。其原理图如下：

图示

描述已自动生成

图3 TD-LSTM原理图

TD-LSTM使用两个LSTM分别去建模target和左边的上下文以及target和右边的上下文，将两个LSTM最后的隐含向量拼接送入到一个softmax分类器中进行分类。

这种神经网络模型在建模的初衷上考虑到了上下文都很重要这一关键点，不能只从前往后推理，目标词语后边的句子对目标情感的分析也是十分重要的。

2.2.2 TC-LSTM

TC-LSTM与TD-LSTM是同一篇论文中提出的模型，与TD-LSTM相比，TC-LSTM将Target词向量的平均值与句子中每个词的词向量进行拼接然后进行和TD-LSTM一样的操作。下边是TC-LSTM的原理示意图：

图示, 示意图

描述已自动生成

图4 TC-LSTM原理图

2.2.3 AT-LSTM

AT-LSTM是在MemNet这一attention机制的网络之后提出的，与MemNet不同的是，它通过将aspect进行词嵌入后经过线性层，然后与LSTM隐层的状态拼接，隐层同时计算attention,最后的输出输入softmax预测类别。原理图如下：

图示

描述已自动生成

图5 AT-LSTM原理图

2.2.4 ATAE-LSTM

ATAE-LSTM是在AT-LSTM的基础上提出的，该模型为了更好地利用aspect，将每个词的embedding和aspect的embedding进行拼接，以此作为输入。原理图如下：

图示, 示意图

描述已自动生成

图6 ATAE-LSTM原理图

2.2.5 BILSTM-ATT

BILSTM-ATT模型将BILSTM与attetion机制结合起来。模型的结构主要包括5部分，分别是输入层、词嵌入层、BiLSTM层、Attention层和softmax输出层。其原理图如下：

图片包含 项链, 挂, 桌子, 束

描述已自动生成

图7 BILSTM-ATT原理图

2.2.6 IAN

相比于之前对输入变量直接进行context和target的拼接进行联合建模，IAN将两者进行了对称式处理。通过使用LSTM对target和context分别处理，将各个时刻的隐层变量进行加权平均取得平均状态，然后计算注意力权重大小，将最终分别得到的值进行拼接操作输入softmax。

图示, 示意图

描述已自动生成

图8 IAN原理图

**2.3基于卷积神经网络的方面级情感分析**

GCAE在某种程度上实际是利用卷积对门控制的模拟。有两个卷积在同时扫描句子，这两个卷积就是两个门单元。两个卷积输入不同，激活函数不同，分别对方面和情感两类信息进行了编码，得到了两个向量、，两个向量元素对应乘的结果，即为GCAE的。可视为情感词的权重。经过训练，通过relu函数后，模型会给与方面词较密切的情感词一个较高的权重，反之，若关系较远，则权重可能很小或者为0。

图示

描述已自动生成

图9 GCAE原理图

**2.4其他基于attention机制的方面级情感分析**

2.4.1 MemNet

MemNet是第一个把attention机制引入到属性级情感分析任务的模型，它通过将aspect进行词嵌入后经过线性层，然后隐层的状态拼接，隐层同时计算attention,最后的输出输入softmax预测类别。原理图如下：

图示

描述已自动生成

图10 MemNet原理图

2.4.2 RAM

RAM是对MemNet的改进，RAM使用BiLSTM计算得到的隐层状态作为memory，此外RAM利用GRU对aspect进行更新。

图片包含 图示

描述已自动生成

图11 RAM原理图

**3 实验结果**

**3.1 实验数据集与处理**

本次实验的数据集来自SemEval2014，SemEval是国际语义评测大会，是全球范围内影响力最强、规模最大、参赛人数最多的语义评测竞赛。Aspect based Sentiment Analysis(ABSA)任务中，大部分论文都用到了SemEval2014中的数据。

实验数据集用到了Laptops和Restaurants。文件给定了每个属性词对应的句子、属性词、属性词的位置、属性词的褒贬感情倾向。感情倾向有positive、neutral、negative。此外还包含了conflict状态。

知道了数据集的架构后便是对数据集的处理。因为神经网络不可能直接对文字进行操作，所以我们必须将文本转化为数字，同时转化为数字之后还要能还原为文本。这就需要我们构建一个哈希映射，或者说构建一个词典，即常说的wordtoid方法。然后就是转化为词向量，即词嵌入。

图形用户界面

中度可信度描述已自动生成

图12 数据集预处理流程图

本次词嵌入使用的是Glove词嵌入，利用了glove.840B.300d.txt文件。LSA和word2vec作为两大类方法的代表，一个是利用了全局特征的矩阵分解方法，一个是利用局部上下文的方法。GloVe模型就是将这两中特征合并到一起的，即使用了语料库的全局统计特征，也使用了局部的上下文特征（即滑动窗口）。该方法通过共现矩阵随机初始化等方法，使得生成网络损失最低，达到目的。

本次数据集的大小如下：

表1 数据集大小及标签分布

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | | Samples | Neg./Neu./POS. |
| Resturant14 | train | 1978 | 805/633/2164 |
| test | 600 | 196/196/728 |
| Laptop14 | train | 1462 | 866/460/987 |
| test | 411 | 128/169/341 |

**3.2 实验结果**

基于github上提供的开源代码，该开源代码已经将数据处理和模型训练封装完毕，我对以上提到过的模型在laptop14和restaruant14上进行了测试。评价标准包括了经典的accuracy和F1。其中，对于多类别问题来说，F1分为Macro-F1和Micro-F1两种。在开源代码的基础上，迭代60轮，其中，在Restautants上的实验结果如下：

表2 Restautants数据集预测结果

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | Accuracy | Macro-F1 | Micro-F1 |
| LSTM | 65.52 | 57.21 | 65.52 |
| GRU | 68.88 | 57.55 | 68.88 |
| TD-LSTM | 71.25 | 60.38 | 71.25 |
| TC-LSTM | 71.10 | 59.61 | 71.10 |
| AT-LSTM | 68.92 | 57.95 | 68.92 |
| ATAE-LSTM | 68.78 | 57.51 | 68.78 |
| AT -BILSTM | 70.65 | 59.28 | 70.65 |
| IAN | 70.56 | 58.94 | 70.56 |
| CNN | 63.03 | 51.13 | 63.03 |
| GCAE | 70.22 | 60.27 | 70.22 |
| MemNet | 71.23 | 60.63 | 71.23 |
| RAM | 70.21 | 59.91 | 70.21 |

其中，模型的参数为默认参数。在实验的过程中对诸如learning-rate、droupout等超参数进行了修改，发现常规的修改对最终的实验结果影响并不是很大，所以在此就不进行结果的粘贴了。

**3.3 实验分析**

3.3.1 经典模型与改进模型的对比

图表, 折线图

描述已自动生成

图13 各模型准确率对比图

首先，两个经典模型之间其实也有很大的差别，虽然GRU是对LSTM的简化，但是通过本次实验发现，GRU的预测准确率不一定比LSTM低，甚至有很大的提升。

此外，通过对经典模型如LSTM、GRU两个模型与其他模型对比发现，改善之后的模型效果普遍是要高于经典模型的。

从原理上来说，经典模型将所有的文本顺次传入神经网络，分析文本数据。但是这种方法只考虑到了正向的文本分析，而忽略了目标词之后的信息。改进后的模型通过诸如关键词拼接、注意力机制、双向传播等一系列的方法，加强了关键词与文本的互相关性，从而去深度挖掘句子中隐藏的情感信息。

3.3.2 基于LSTM的众模型对比

在众多的属性级情感分析预测模型中，基于LSTM的模型数量众多，在本次实验中，我尝试了LSTM、TD-LSTM、TC-LSTM、AT-LSTM、ATAE-LSTM、AT-BILSTM、IAN等一系列模型。

其中，TD-LSTM、TC-LSTM两者是使用了双向传递，将目标词前后的句子都进行传入训练。AT-LSTM、ATAE-LSTM、IAN利用了attention机制、AT-BILSTM将LSTM进行替换。

可以发现，虽然AT-LSTM、ATAE-LSTM、IAN等模型似乎从原理上来说更加注重了因果相关性，但是从实验结果上来看，TD-LSTM、TC-LSTM这样较为简单的思想似乎与他们起到的最终效果是差不多的。我想，这可能与某些超参数的调整、数据的预处理等有关。

3.3.3 基于LSTM的模型与其他模型的对比

在本次实验中，除了众多以LSTM为基础的模型外，我还尝试了CNN、GCAE、MemNet、RAM等模型。普通的CNN的效果很不理想，但是改进之后的GCAE从结果上来看效果是非常不错的。在我看来，GCAE很像是利用CNN对门控制的模拟，但是这个模拟可以说是自动调整的，相比于刻意的门控制设计来说更加的灵活，最后取得的效果也是比较理想的。而MemNet、RAM这两个从attention机制出发的效果好于基于attention的AT-LSTM和ATAE-LSTM，所以并不是说方法结合越多，效果就越好。

**4 总结及模型评价**

总的来说，本文以SemEval2014数据集中的Laptops和Restaurants为训练测试对象，在进行词典构建与词嵌入之后，利用github上的开源代码，对LSTM、GRU、TD-LSTM等众多模型进行了测试，并对经典模型、LSTM系列模型、CNN系列模型、attention机制模型进行了对比分析。

在对比分析后，我发现改进后的模型通过利用诸如关键词拼接、注意力机制、双向传播等一系列的方法对经典文本分析模型做到了较大的提升。这些方法从原理上加强了目标词与文本的结合相关度，加强了对句子情感的提取，进而达到了更好的结果。

**5 心得与需要改进之处**

属性级情感分析作为NLP中的经典问题，其对应的模型众多，对于像我这样的初学者来说是用来学习的一个很好的问题。

在本次实验学习的过程中，我了解到了文本数据与普通数据的不同，如何将文本转化为计算机可以利用的语言是NLP需要考虑的第一个问题。此外，本次实验中，我对词嵌入这一方法有了更进一步的了解。由于之前所做的任务大多数是与视觉相关的，所以对诸如LSTM等模型的了解不够深入。通过本次的实验，我对LSTM的结构，以及如何以LSTM为基础，加强对目标词语情感的挖掘进行了了解。

通过阅读论文发现，属性级情感分析方面的新模型大多都是14年左右提出的，该课题还是一个比较新颖的方向，最近几年比较兴起的一些模型或许在该课题上会取得不错的成果，比如transformer模型等。但是由于github上的论文中并不支持该模型，且我自己写的代码目前还存在一些问题，没有跑通，所以只是个人的猜测。

此外，在本次实验的过程中，所有的数据都是英文，在网上找到了一些中文的情感分析数据集，但是目前还没有开始实验，不知道这些在英文上表现不错的模型是否在中文数据集上也可以起到不错的作用。个人感觉中英文的区别主要在于前期的数据预处理部分，对中文进行词嵌入时便不能直接利用该开源代码，目前的策略是构建词典之后转化为对应数字，直接利用torch.nn中的torch.nn.Embedding进行词嵌入,已将此作为下一步的实验计划。

总的来说，本次对属性级情感分析的学习让我受益匪浅，整个研究流程让我对问题的研究方法有了进一步的理解，在实验的过程中，有一些模型由于时间原因没有很好的理解他们的原理，只是简单的使用代码进行了结果的验证，感觉还有很多需要进一步学习，希望下一次能做的更好。

**参考文献**

1. Dong, Li, et al. "Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification." Meeting of the Association for Computational Linguistics 2014.
2. Long, J., et al. "Target-dependent Twitter Sentiment Classification." The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA Association for Computational Linguistics, 2012.
3. D Tang, et al. "Effective LSTMs for Target-Dependent Sentiment Classification." Computer Science (2015).
4. Tang, D., B. Qin , and T. Liu . "Aspect Level Sentiment Classification with Deep Memory Network." 2016.
5. Wang, Y., et al. "Attention-based LSTM for Aspect-level Sentiment Classification." Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing 2016.
6. Yang, Min, et al. "Attention-based LSTM for target-dependent sentiment classification." (2017).
7. Ma, D., et al. "Interactive Attention Networks for Aspect-Level Sentiment Classification." (2017).
8. Peng, C., et al. "Recurrent Attention Network on Memory for Aspect Sentiment Analysis." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing 2017.
9. Wei, X., and L. Tao . "Aspect Based Sentiment Analysis with Gated Convolutional Networks." (2018).
10. Huang, B., Y. Ou , and K. M. Carley . "Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks." International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (2018).
11. Hochreiter, S, and J. Schmidhuber. "Long Short-Term Memory."Neural Computation 9.8(1997):1735-1780.
12. Cho, K.,et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." Computer Science (2014).
13. Mnih, Volodymyr, et al. "Recurrent Models of Visual Attention." Advances in Neural Information Processing Systems 3(2014).