

目录

1. 实验环境说明	2
2. 实验任务分析	2
3. 实验方法设计	2
3.1 文字编码部分	3
3.2 主要模型	4
4. 实验结果	7
5. 讨论	8
6. 参考文献	9

1. 实验环境说明

Python 版本: 3.6.8

Cuda 版本: 9.0.176

Tensorflow 版本: 1.8.0

GPU 版本: Titan V

Cudnn 版本: 7.5.0

2. 实验任务分析

该任务属于文本情感分析任务中的方面级情感分析(Asspect Based Sentiment Analysis, ABSA)下, 与篇章级情感分析和句子级情感分析相比, 它的粒度最细, 其目的是分析出文本中涉及到的每一个方面上的情感倾向性 (共 20 个方面)。

ABSA 有两个子任务:

(1) 基于 aspect 类的情感分析(aspect-category sentiment analysis, ACSA), 主要任务是判断关于句子中的某个实体(entity)的情感极性。比如:

输入: 文本是“这块手表的颜色不错”, entity 是“颜色”

输出: 情感极性

(2) 基于 aspect 词的情感分析(aspect-term sentiment analysis, ATSA), 主要任务是判断判断句子在某方面的情感极性, 类似于一个多标签(multi-label)的任务。

输入: 文本是“这家餐厅的环境还不错, 可以菜做的不怎么样”

输出: 环境方面的情感极性, 交通便利方面的情感极性, 做菜方面的情感极性。

该任务属于基于 aspect 词的情感分析(aspect-term sentiment analysis, ATSA)。

3. 实验方法设计

由于最后预测的结果包含二十个维度, 因此先训练 20 个 LightGBM 模型, 根据特征重要性进行排序并选取 TOP7 作为各维度的关键词, 例如交通是否便利(traffic convenience)维

度的“地铁站 地铁 地理位置 位置 公交车 公交车站 公交站”，距离商圈远近(distance from business district)维度的“百货 商圈 商场 广场 购物中心 城 商业街”。

```
subjects = ['地铁站 地铁 地理位置 位置 公交车 公交车站 公交站',  
            '百货 商圈 商场 广场 购物中心 城 商业街',  
            '容易 位置 醒目 找到 找 地理位置 显眼',  
            '小时 排队 等 排 排号 队 号',  
            '态度 服务员 热情 服务态度 老板 服务 服务生',  
            '开车 停车费 停车位 停 停车场 车位 泊车',  
            '很快 催 慢 速度 分钟 上菜 等',  
            '小贵 不贵 价位 原价 块钱 价格 性价比',  
            '不划算 物有所值 不值 物美价廉 超值 性价比 实惠',  
            '活动 团 霸王餐 代金券 团购 优惠 券',  
            '装修 布置 灯光 古色古香 装饰 优雅 情调',  
            '安静 环境 装修 氛围 嘈杂 吵闹 音乐',  
            '大 宽敞 空间 面积 装修 拥挤 店面',  
            '整洁 干净 环境 卫生 苍蝇 不错 脏',  
            '吃不完 一份 量 量足 个头 好大 少',  
            '入味 吃 不错 味道 好吃 口味 好喝',  
            '造型 颜色 精致 卖相 好看 色香味 食欲',  
            '推荐 强烈推荐 值得 强推 一试 极力推荐 菜品',  
            '好 满意 纪念品 内地 之 肠 灌',  
            '还会 机会 再 不会 来 值得 推荐']
```

图 1 维度

3.1 文字编码部分

由《R-Net: Machine Reading Comprehension with Self-Matching Networks》[1]中，同时使用了 char-embedding 和 word-embedding，不同的是 char-embedding 是通过将 char 放入双向 GRU 之后，最终的是通过 GRU 的最终状态来得到。利用两个双向 RNN 网络分别对 question 和 passage 再编码。而之前多数都是用的 CNN 卷积和 highway。

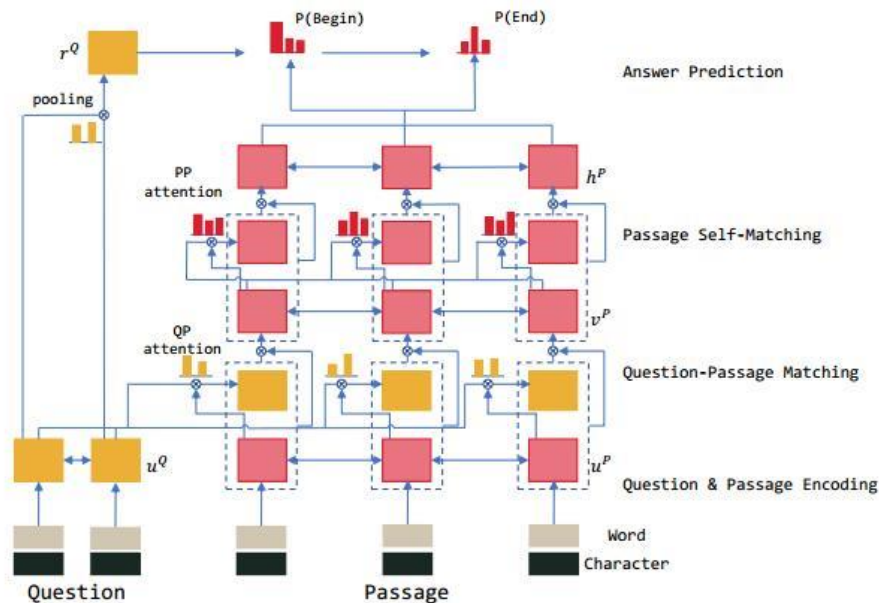


图 2 R-net

于是在本次实验之中我们也借鉴这种 word 和 char 结合的方式，变相地增加了输入的数据量。并且使用 w2v 对 word 进行编码，对 char 使用 BiRNN 进行编码。

举个例子：奶茶十分好喝。词向量表示：[W2V(奶茶), W2V(十分), W2V(好喝)]

经过 BiRNN 的字向量：[BiRNN(奶,茶), BiRNN(十,分), BiRNN(好,喝)]

最终向量表示：[Concat(W2V(奶茶), BiRNN(奶,茶)), Concat(W2V(十分), BiRNN(十,分)), Concat(W2V(好喝), BiRNN(好,喝))]

在进行连接操作之后，再把该向量输入一个单向 LSTM 网络，捕捉上下文的信息。

至此，就完成了整个句子的编码。

3.2 主要模型

由《Aspect Based Sentiment Analysis with Gated Convolutional Networks》[2]，此文关注 aspect 类别或目标实体的情感极性。许多基于 ABSA 的模型已经被开发，主要处理两个不同的子任务：aspect 分类情绪

aspect 分类情绪分析 (ACSA)

aspect 实体的情绪分析 (ATSA)

例如，在句子 "Average to good Thai food, but terrible delivery ."中，ATSA 会分析实体 Thai food 的情感极性；ACSA 会分析 aspect "service"情感极性，即使"service"没有出现在

句子中。

在 aspect 极性抽取任务之中，多数实验使用 LSTM 和 attention 机制来预测有关目标的情感极性，但模型往往很复杂并且需要很长的训练时间。本文提出了一个基于 CNN 和 Gating Mechanisms（门机制）的模型。一种新的 Tanh-ReLU 门单元能够根据给定的 aspect 或实体选择输出的情感特征。这个结构比应用于现有模型的 attention 层简单得多。同时，本文中模型的计算在训练中很容易并行化，因为卷积层不像 LSTM 层那样有时间依赖性，并且门单元也能够独立工作。此文在 SemEval 数据集上进行试验，表明了实验效果的提升。

本文模型在更短的训练时间下能得到更好的精确度。对于 ACSA 任务，模型在 embedding 层之上有两个单独的卷积层，这个 embedding 层的输出是由新型门单元的组合组成的。多个滤波器的卷积层可以有效地提取每个接受域上多个粒度的 n-gram 特征。门单元有两个非线性门，两个中的每一个都和一个卷积层链接。在给定的 aspect 信息下，对于情感的预测能够抽取 aspect-specific 情感信息。比如说，在句子"Average to good Thai food, but terrible delivery"中当 food 这个 aspect 被提供，门单元能够自动忽视在第二个子句中 delivery 这个 aspect 的消极情绪，并且只输出第一个子句的积极情绪。因为模型中的每个组成部分都能够并行，因此比 LSTM 和 attention 机制需要更少的训练时间。对于 ATSA 任务，其中 aspect 由多个单词组成，本文扩展了模型以包含目标表达式的另一个卷积层。原文模型如下图：

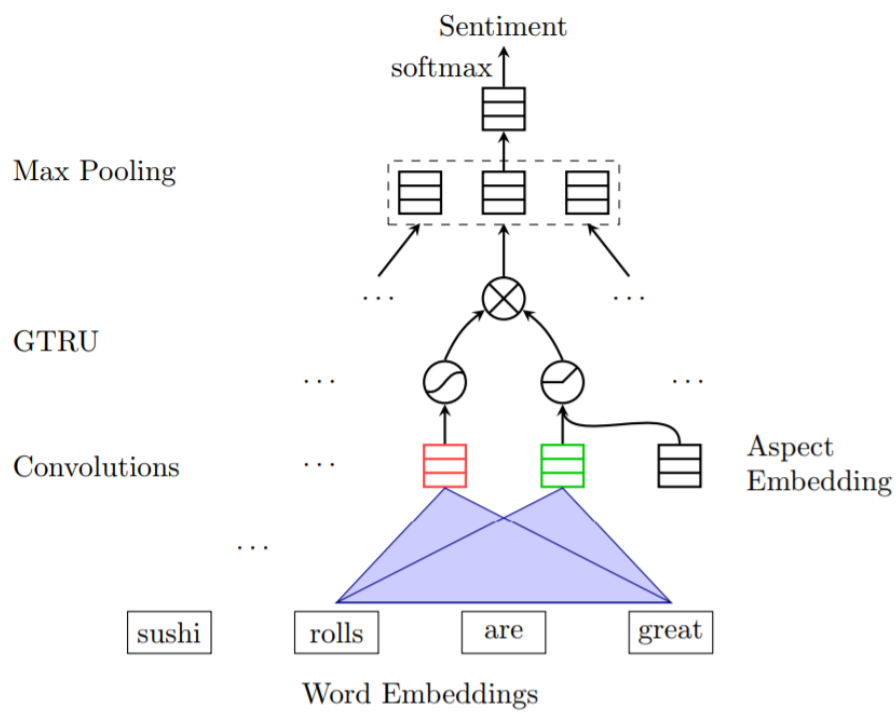


图 3 GCAE

在这个模型的基础上，在最后的线性层之前加上了 CNN 增加分类能力，最后在使用 softmax 分类。我们最终模型如下图：

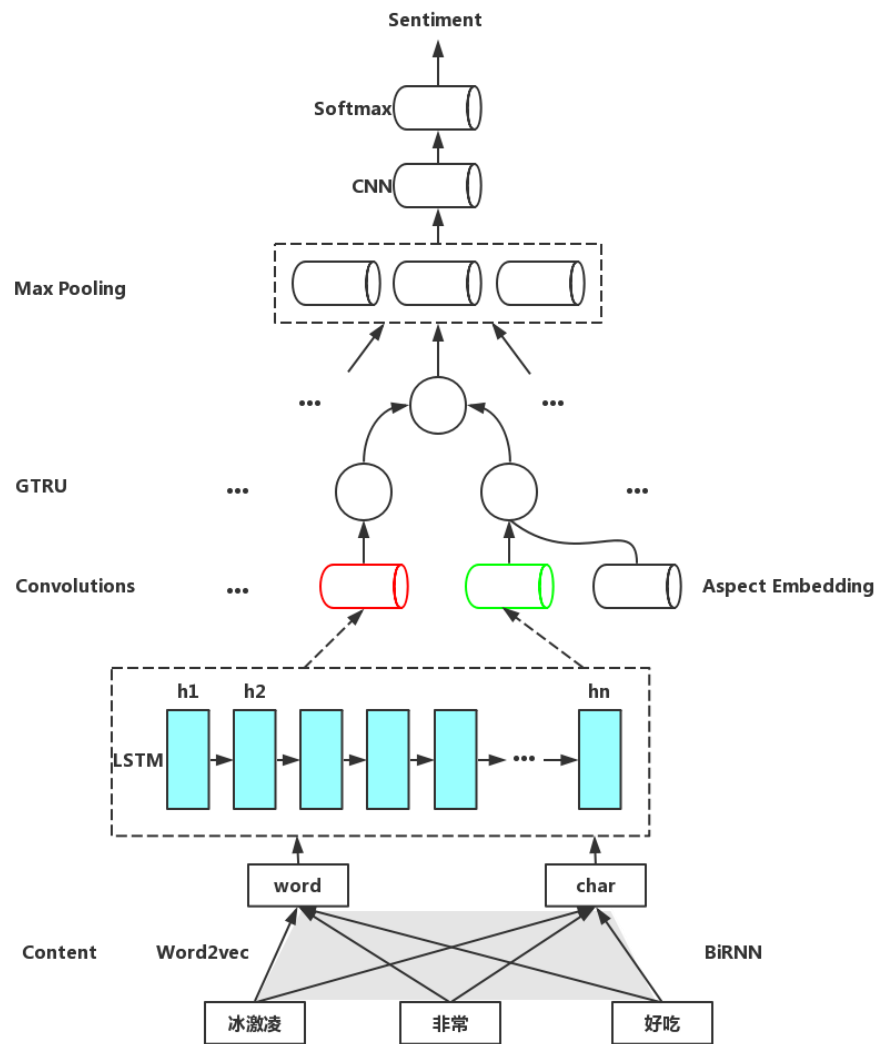


图 4 GCAE_expanded

4. 实验结果

表 1 模型及其预测结果

模型\参数	n_hidden	learning_rate	dropout_keep	F1 score
GCAE	400	0.0005	0.5	71%
	450	0.0005	0.5	70%
	400	0.0005	0.3	70%
	300	0.001	0.5	70%

fastText	-	-	-	49%
XGboost	-	-	-	45%
Adaboost	-	-	-	44%
SVM	-	-	-	41%
BP	-	-	-	24%
Naive Bayesian	-	-	-	23%

5. 讨论

在 GCAE 模型构建出来之前，还尝试过多种机器学习和深度学习的方法，结果如表 1 所示。其中 fastText 模型的训练速度较快，且结果也不逊于传统的机器学习模型，scikit-learn 中的 skift 可以作为第三方工具被使用。最终，在验证集上得到一个 f1 均值约为 0.49 的 fasttext 多分类模型(20 个)，存储在 fasttext_model.pkl 中，其中涉及到 learning_rate、word_ngrams、min_count 等参数，最后的结果将 learning_rate 设置为 0.5、word_ngrams 设置为 2（相较于设置为 1 的效果更好）、min_count 设置为 1（设置为 2 有一些负向影响）。

Boost 系列与 BP 和 Naive Bayesian 相比，最后得到的结果更好。猜测是测试集并不是随机抽样得来的，而是一些相对比较难的样本，这样测试集与验证集的先验分布不同，因此虽然 BP 和 Naive Bayesian 在验证集的分值很高（约为 0.7），然而在测试集却表现很差；而 boost 算法中会给判错的样本更高的权重，相当于把先验分布改变得更均匀，因此表现更好。

“训练集是真实样本总体的无偏采样”这个假设往往并不成立，我们未必能有效地基于训练集观测几率来推测出真实几率。因此在后续实验中，可以采用如下方法尝试解决这个问题：

①直接对训练集里的“未提及”样例进行欠采样，即去除一些该样例使得四类样本数目相近，然后再进行学习，代表性算法为 EasyEnsemble[3]，该算法是利用集成学习机制，将该样例划分为若干个集合供不同学习器使用，这样对每个学习器来看都进行了欠采样，但从全局来看却不会丢失重要信息。

②对训练集里“正面”、“中性”和“负面”样本进行“过采样”，即增加一些该样本使四类样本数目相近，但并不是简单地进行重复采样，这样会导致严重的过拟合。代表性算法为 SMOTE[4]，通过对该样例进行插值来产生额外的样例。

此外，Attention 对于文本的表达还是有很好的效果，之后可以 Attention 方式去更充分的提取文本特征，从而提升情感分析任务的精度。

6. 参考文献

- [1] Wang, W., Yang, N., Wei, F., Chang, B., & Zhou, M. (2017). R-NET: Machine reading comprehension with self-matching networks. *Natural Lang. Comput. Group, Microsoft Res. Asia, Beijing, China, Tech. Rep*, 5.
- [2] Xue, W., & Li, T. (2018). Aspect based sentiment analysis with gated convolutional networks. *arXiv preprint arXiv:1805.07043*.
- [3] Liu, T. Y. (2009, August). Easyensemble and feature selection for imbalance data sets. In *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing* (pp. 517-520). IEEE.
- [4] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.