

汽车行业用户观点主题与情感识别

比赛网址: <https://www.datafountain.cn/competitions/329/details/data-evaluation>

注意: 由于比赛已经结束, 所有的结果都是基于本地划分训练集、测试集的方式进行评估的。原始训练集合共有 12573 条数据, 将其转化为多标签形式以后共有 10654 条数据, 随机抽出 3197 条 (30%) 作测试集, 剩下的 7457 条作训练集。

一、问题描述与评估方法

1.1 问题描述

通过给定用户的观点, 识别用户观点的主题内容, 并给出这些主题内容的情感倾向。一个观点当中可以包含多个主题。

主题被分为 10 类, 包括: 动力、价格、内饰、配置、安全性、外观、操控、油耗、空间、舒适性。

情感分为 3 类, 分别用数字 0、1、-1 表示中立、正向、负向。

1.2 评估方法

本赛题采用 F1-Score 评价方式。按照“主题+情感分析”识别数量和结果 (是否正确) 来进行判断, 参赛者需要识别文本中可能包含的多个“主题”。

(1) 匹配识别结果:

Tp: 判断正确的数量;

Fp: 判断错误或多判的数量;

Fn: 漏判的数量;

关于 Tp、Fp、Fn 的统计规则说明如下:

当提交的一条数据结果包含“主题+情感值”, 如果参赛者对“主题+情感”的判断结果完全正确则计入 Tp, 如果对“主题”或“情感值”的判断结果错误则计入 Fp;

如果参赛者识别出的“主题+情感值”数量少于测试样本中实际包含的数量, 或未对某个测试样本数据给出结果, 缺少的数量计入 Fn;

如果参赛者识别出的“主题+情感值”数量多于测试样本中实际包含的数量, 超出的数量计入 Fp;

(2) 计算准确率 (P) 与召回率 (R):

$$P = Tp / (Tp + Fp)$$

$$R = Tp / (Tp + Fn)$$

(3) 最终以 F1 参数作为评分标准:

$$F1 = 2 * P * R / (P + R)$$

二、解决方法 (Ensemble)

分两步进行解决, 现根据用户的观点预测观点当中包含的主题 (多标签分类), 再根据观点当中的主题预测情感倾向 (ABSA)。

1.1 主题识别

主题识别是一个多标签问题, 采用的方法是将其转化为针对每一个标签的二分类问题, 再利用交叉验证的方法找寻 threshold 来判断改标签是否存在。

由于官方给出的训练集合太小了 (1W+), 这个训练集合又是和汽车行业高度相关的, 导致深度学习方法在这个集合上不能学习出很好的词表征, 引入外部词向量效果和直接训练差不多, 所以就没有用外部的词向量, 主要利用机器学习+深度学习的方法进行分类。

主要用到了以下方法:

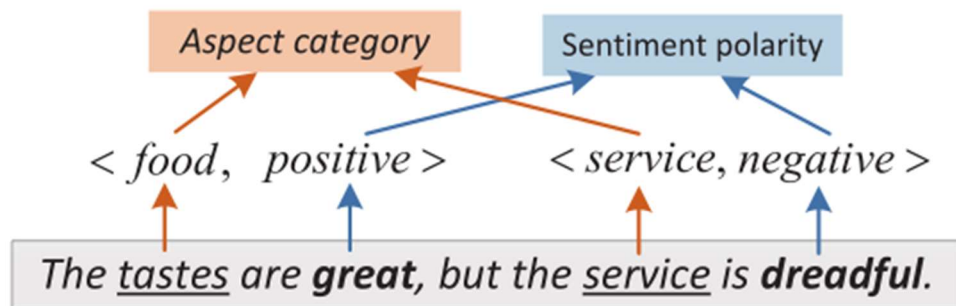
- 1) TextCnn
- 2) TextRnn
- 3) TextRCnn
- 4) Adaboost
- 5) gdbt
- 6) svm
- 7) randomforest
- 8) logisticregression

深度学习方法主要参考了[1], 用 TextCnn、TextRnn、TextRCnn 三种方法 (太复杂的模型在这种小样本集合上非常容易过拟合)。这些方法模型结构都非常简单, 这里就不多加介绍了 (一层 CNN 或者 RNN + FC)。

通过五折交叉验证, 利用交叉验证的结果寻找最优的 threshold 来判断是否存在此标签, 最终的结果是这些模型结果的平均值 (没有用 stacking, 发现 stacking 结果没有直接求平均好)。

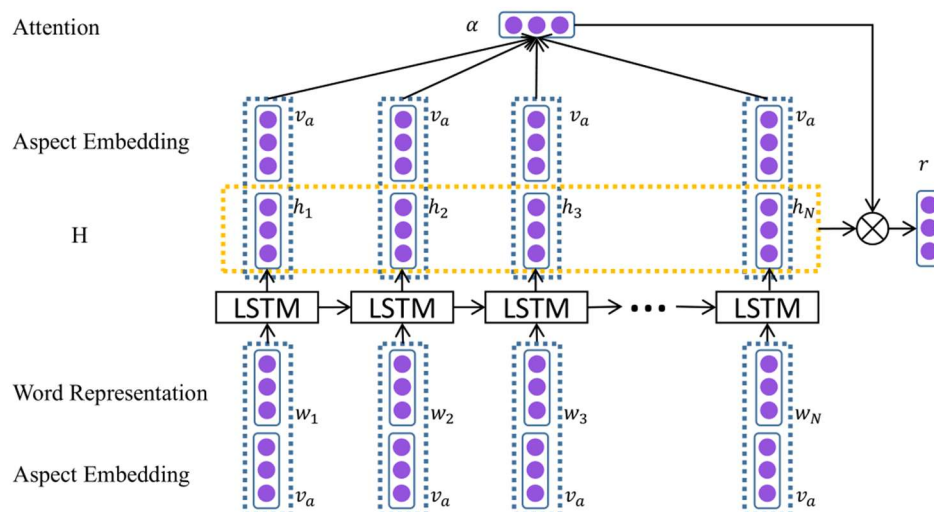
1.2 情感识别

ABSA 问题:

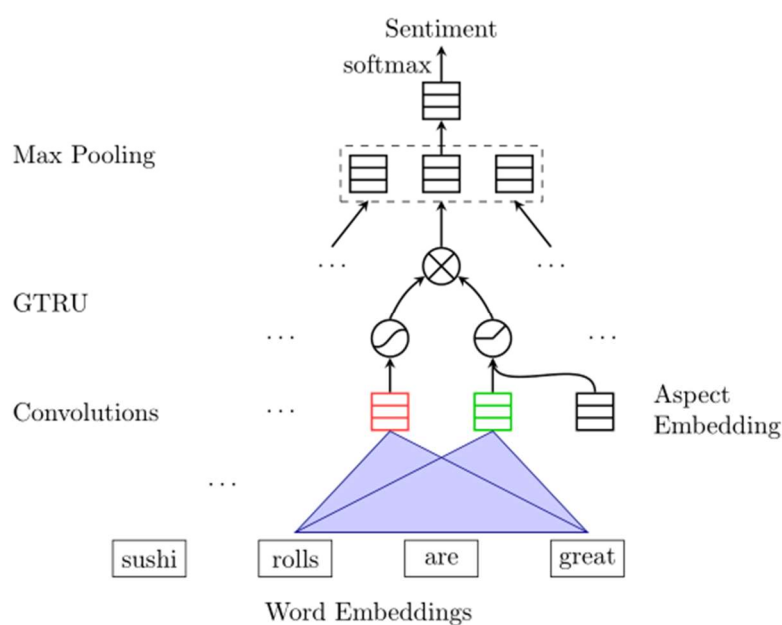


不同主题对应有不同的情感，传统的机器学习方法不太好捕捉不同主题的情感（当时参加比赛的时候是将标签加到句子的末尾当做一个特征进行预测），这里主要参考了第一名的做法，一共用到了三种模型：

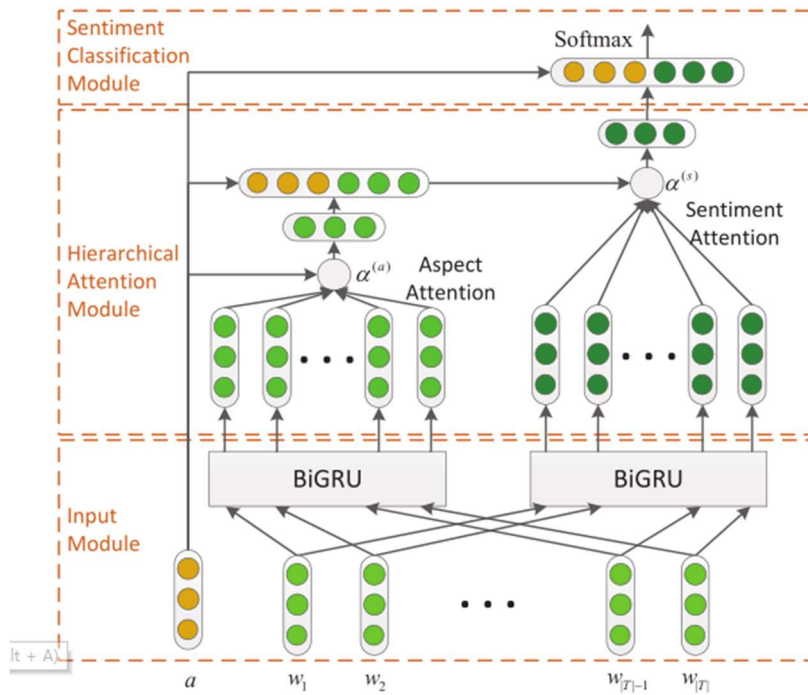
1) ATAE-LSTM[2]



2) GCAE[3]



3) HEAT[4]



其主要思想均是将主题编码后去关注与主题相关的特征，来获取情感极性。最终平均这三种模型的预测结果作为最终的预测结果。

三、BERT (fine-tune)

wait...

四、结果 (local)

4.1 主题识别

TextCnn	TextRnn	TextRCnn	svm	LogisticRegression	RandomForest	GDBT	Adaboost	Ensemble	BERT
0.8931	0.8901	0.8941	0.897	0.8863	0.9014	0.9018	0.8966	0.9092	

4.2 情感识别

情感识别是基于主题识别的最佳结果进行测试的：

ATAE-LSTM	GCAE	HEAT	BERT

五、参考

- [1] https://github.com/brightmart/text_classification
- [2] Wang, Y., Huang, M., Zhao, L., and Zhu, X. (2016).Attention-based LSTM for Aspect-level Sentiment Classification.Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing(EMNLP), pages 606–615.
- [3] Xue, W. and Li, T. (2018).Aspect Based Sentiment Analysis with Gated Convolutional Networks.
- [4] Cheng, J., Zhao, S., Zhang, J., King, I., Zhang, X., and Wang, H. (2017).Aspect-level Sentiment Classification with HEAT (HiErarchical ATtention) Network.Proceedings of the 26th ACM International Conference on Information and Knowledge Management, pages 97–106.
- [5] Kenton, M.-w. C., Kristina, L., and Devlin, J. (2017).BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.