

# 专业实践：情感分类技术实践

Authors: Sentiment Computing Group@HIT-SCIR

## 情感分类简介

文本情感分类是自然语言处理的一个分支，任务定义为：对于给定的一段文本，在预先定义的情感类别集合中，选择一个或多个最符合当前文本情感的类别，这里将任务简化为单类别分类任务。

## 情感分类技术实践 1：细粒度情感分析

请自行查找相关知识及代码框架，完成以下一个评测任务

项目名称：细粒度情感分析

项目介绍：本部分采用 AI Challenger 2018 细粒度情感分析比赛数据，该数据为餐饮领域的用户评论，预先定义 20 个评价对象，并针对每个评价对象进行四分

类，数据样例如下图所示。其中，-2 表示未提及，-1 表示负面情感，0 表示中性情感，1 表示正面情感。

“味道不错的面馆，性价比也相当之高，分量很足～女生吃小份，胃口小的，可能吃不完呢。环境在面馆来说算是好的，至少看上去堂子很亮，也比较干净，一般苍蝇馆子还是比不上这个卫生状况的。中午饭点的时候，人很多，人行道上也是要坐满的，隔壁的冒菜馆子，据说是一家，有时候也会开放出来坐吃面的人。”

交通是否便利(traffic convenience)	-2	装修情况(decoration)	1
距离商圈远近(distance from business district)	-2	嘈杂情况(noise)	-2
是否容易寻找(easy to find)	-2	就餐空间(space)	-2
排队等候时间(wait time)	-2	卫生情况(cleaness)	1
服务人员态度(waiter's attitude)	-2	分量(portion)	1
是否容易停车(parking convenience)	-2	口感(taste)	1
点菜/上菜速度(serving speed)	-2	外观(look)	-2
价格水平(price level)	-2	推荐程度(recommendation)	-2
性价比(cost-effective)	1	本次消费感受(overall experience)	1
折扣力度(discount)	-2	再次消费的意愿(willing to consume again)	-2

在该数据上完成分类任务，评价指标采用各评价对象宏平均 F1 值的平均值。

为简化该任务，您可以自行选择一个评价对象（例如服务人员态度）进行评价即可。

## 数据集获取

采用官方评测数据集

链接：<https://pan.baidu.com/s/1HAGV-UET6lof61MxGipK3g> 提取码：w1n0

## 方法实践

分别采用以下几种常见的深度学习模型，完成这个分类任务

1. TextCNN
2. LSTM
3. BERT

其中：采用 LSTM、BERT 进行分类，参考 F1 值分别为 60%，70%。

您可以调整各种超参数，例如：

- TextCNN 的卷积核个数及大小
- 给 LSTM 加层数及双向编码
- 调整各类隐藏层维度大小
- 学习率，batch\_size

等等来提升任务指标（F1 值）

## 推荐资源

请充分利用 Google 等搜索引擎及 GitHub 代码资料库来获取你所需要的信息，你所遇到的问题（例如环境配置）别人一般都遇到过，在尝试过所有手段后，你可以放心地向我们求助。

## GPU 资源

非 Bert 的小模型可以在您的 PC 上的游戏级别显卡上进行训练，耗时完全可以接受。若您没有相关的 GPU 资源，您可以前往 Google 的 Colab，申请免费的 GPU 资源。不用纠结于 GPU 资源有无，您完全可以在数据集的一个小子集上观察到 loss 的正常下降再在全集上训练

## 参考资料

- 机器学习-李宏毅
- Stanford CS224N
- 动手学深度学习-李沐

## 代码框架和第三方库

- 深度学习有一套较为固定的代码框架，您可以在 GitHub 上搜索一个类似的任务，

修改可读性强的代码作为您的基础代码框架。

建议：一定要自己动手实现，好的分类代码框架确实可能略微修改后就可以适配这个简单的分类任务。

- 第三方库：Bert 不用自己实现，推荐 huggingface 的 transformers 库，hugging face 也有自己的深度学习代码框架。

建议：TextCNN 和 LSTM 网络结构简单，可以用 pytorch 实现一遍。不推荐使用封装好的 LSTM,TextCNN 第三方库。

## 情感分类技术实践 2：大语言模型下的情感分类

经过指令微调的大语言模型能够通过 prompt 工程，以文本生成任务的形式完成许多通用 NLP 任务，也包括情感分类，请自行了解 prompt 工程的相关概念，通过调用 openAI [ChatGPT](#) 的官方 API，在上述餐饮评论情感分类数据集上尝试情感分类任务，并记录宏平均 F1 指标，观察与直接微调的模型所达到的指标有何差距。为简化任务，您只需要在情感分类数据集上采样 30-50 条数据，进行测试即可。

## 情感分类技术实践 3：利用大语言模型进行数据增强

在许多 NLP 任务中，高质量的数据标注是训练高质量模型的一个必要条件，然而这一个步骤往往开销巨大。我们同样可以用为大语言模型设计一些特定的 prompt，让其生成一个数据集，将其与原数据集进行混合，称为数据增强。我们仍以餐馆评论情感分类数据集为例，现在对其进行数据增强，即：再构造 30-50 条与原数据集格式一致的，包含文本及对应情感标签的数据。