



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

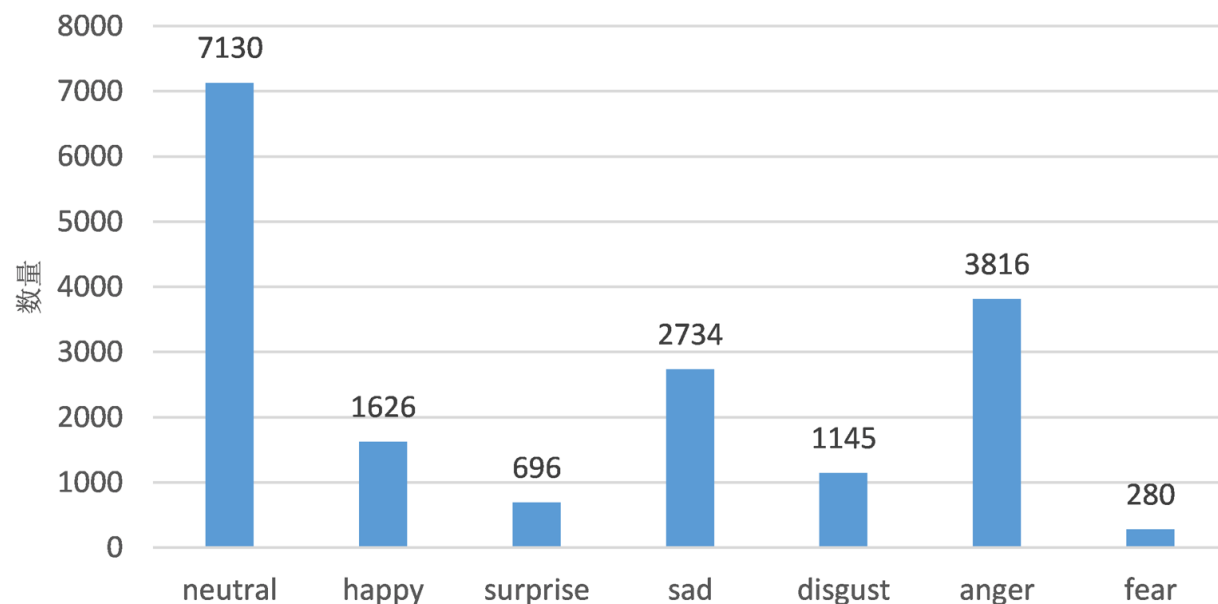
CCAC2023-MERC评测经验分享

赵志龙、胥卜凡、徐冰

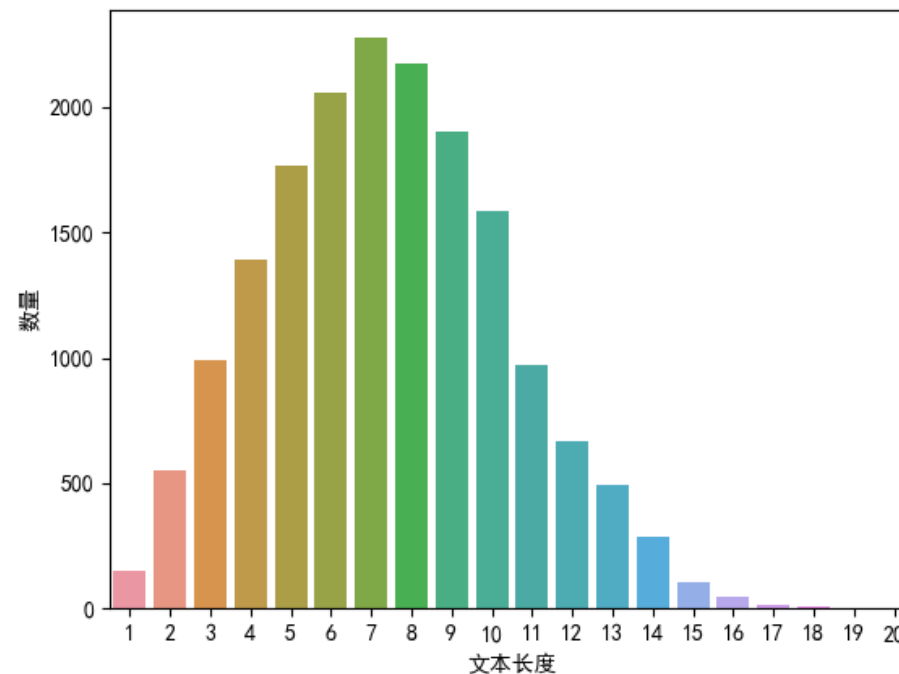
哈尔滨工业大学语言技术研究中心

对训练数据集进行分析，其语料来自 38 部中文影视剧，每段对话都有 A，B 两个说话者，对话文本长度主要集中在7个字左右，比较简短，同时存在类别不平衡的问题。

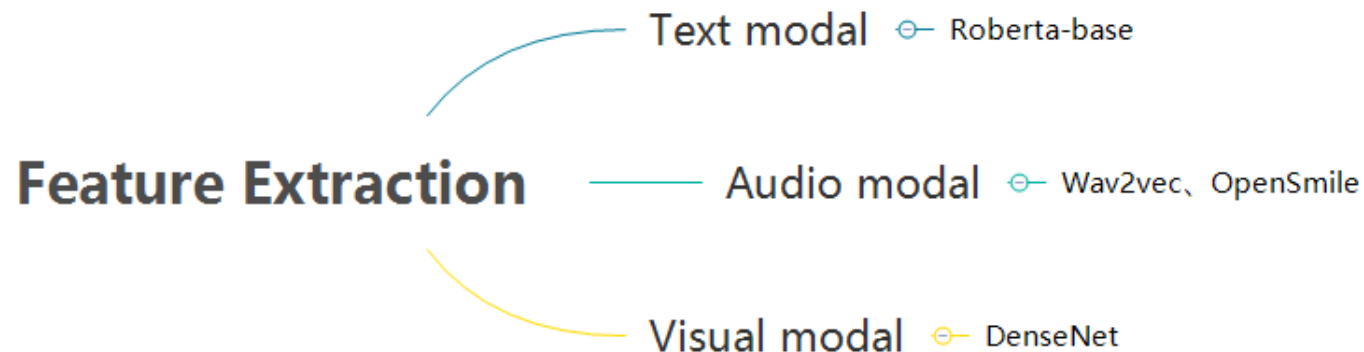
类别分布图



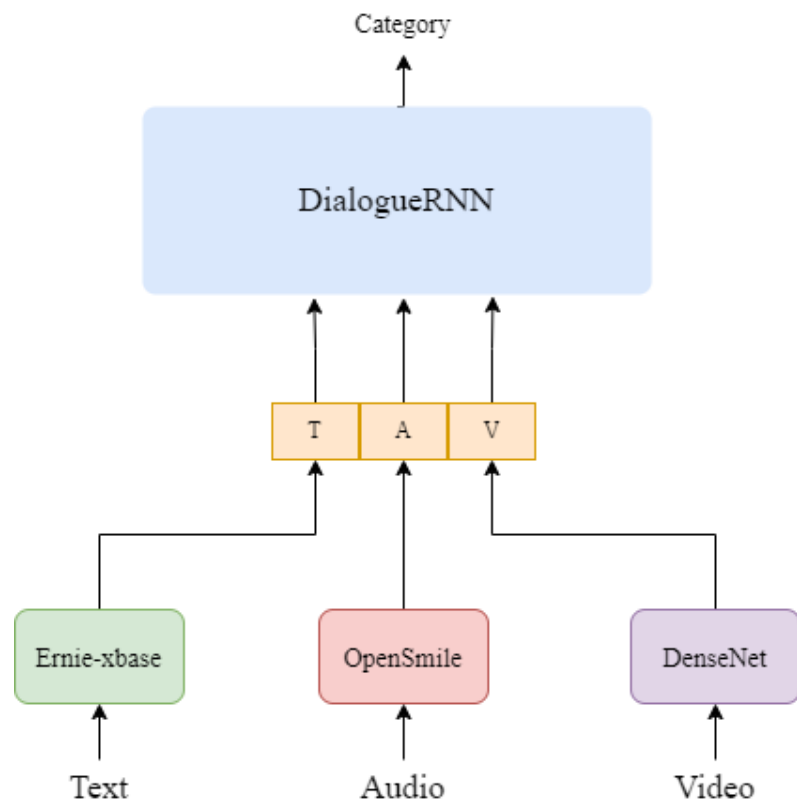
文本长度分布



- 参考官方给出的特征文件中多模态特征提取方法：
 - 文本模态：RoBERTa-base等预训练模型编码，在评测数据集上进行 finetune
 - 音频模态：Wav2vec或OpenSmile 提取音频特征
 - 视频模态：DenseNet提取面部级别特征信息



观察训练集数据特点可以发现，当前句的情感受到上文情感直接影响。而DialogueRNN模型优势在于对说话者，当前对话的上文以及上文中的情感进行了很好的建模。因此，我们采用DialogueRNN模型并在此基础上对模型输入特征和训练过程做优化。



由于数据集存在数据不平衡问题，因此可以对不同类别进行加权，具体实现上，可通过设置 `nn.CrossEntropyLoss` 的 `weights` 参数达到目的。

权重设置也有两种方案：

- 用样本数的倒数作为权重：

`class_nums=(100000,100,10)` \longrightarrow `weights=(1/100000,1/100,1/10)`

- 类别中最大样本数量除以当前类别样本的数量作为权重：

`class_nums=(100000,100,10)` \longrightarrow `weights=(1,1000,10000)`

我们采用了官方提供的四个特征文件，分别进行多模态实验，实验结果表明文本模态采用RoBERTa-base在数据集上微调后特征，音频模态采用OpenSmile 抽取句子级别的 IS10 情感特征，视频模态采用DenseNet提取的人脸特征整体效果最好。

表 2 特征文件选择实验

Method	F1(%)
finetune_RoBERTa+IS10+DenseNet	41.08
RoBERTa+ sen_avg_wav2vec+DenseNet	38.14
RoBERTa+ finetune_cls_wav2vec +DenseNet	33.24
finetune_RoBERTa+ finetune_cls_wav2vec +DenseNet	36.69



文本特征抽取我们抛弃了原始特征文件采用的 RoBERTa-base 预训练模型，改用百度提出的Ernie3.0-xbase 预训练模型，在评测数据集上进行 finetune 后对对话文本进行编码，下表展示了融合三种模态特征后在原始测试集中的表现：

表 3 文本特征优化实验

Method	P(%)	R(%)	F1(%)
RoBERTa-base	47.46	39.43	41.08
Ernie3.0-xbase	45.69	41.42	42.88(1.80 ↑)



对抗学习采用PGD和FGM两种方法对输入样本进行扰动:

➤ FGM对抗学习

$$r_{adv} = \epsilon \cdot \frac{g}{\|g\|_2}$$

$$g = \nabla_x L(\theta, x, y)$$

➤ PGD对抗学习

$$r_{adv|t+1} = \alpha \cdot \frac{g_t}{\|g_t\|_2}$$

$$g = \nabla_x L(\theta, x, y)$$

$$\|r\|_2 \leq \epsilon$$

表 4 对抗学习方法实验

Method	P(%)	R(%)	F1(%)
baseline	44.24	42.77	42.69
PGD	43.74	44.11	43.27(0.58↑)
FGM	46.37	43.04	43.98(1.29↑)

集成学习基准模型采用K折交叉验证模型，对这K个模型采用等权融合投票的方式：

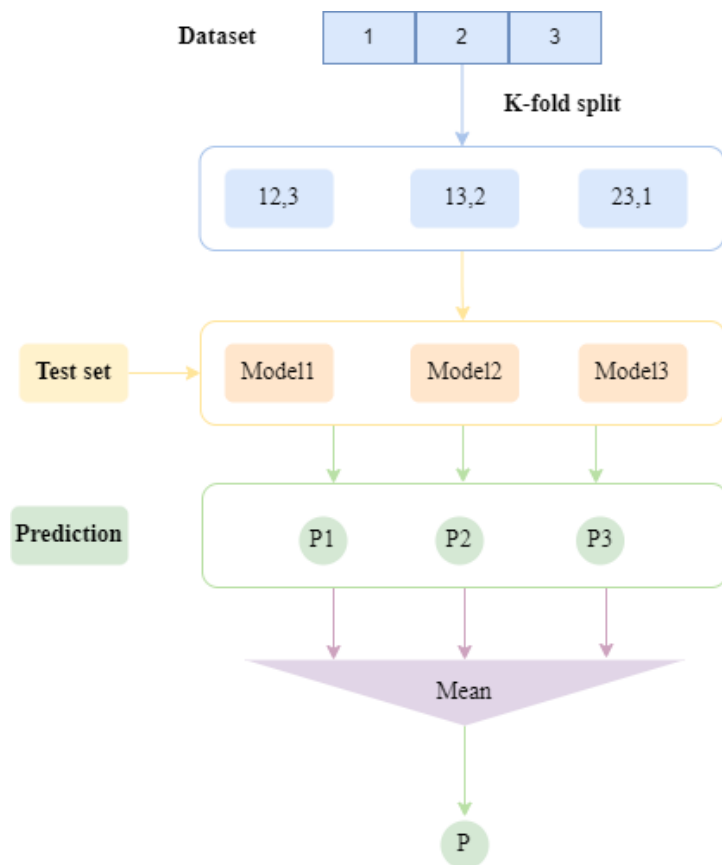


表 6 交叉验证模型表现

Method	P(%)	R(%)	F1(%)
Model1	57.22	53.94	54.91
Model2	61.87	56.40	56.49
Model3	60.59	56.31	57.42

此次评测，我们借鉴经典的对话情感分析模型DialogueRNN，通过文本特征编码优化，对抗训练，集成学习等优化策略，最终在给出测试集中F1值表现为50.38。

但遗憾的是，本次评测由于时间限制最终提交结果只采用了文本和音频两种模态信息，没有实现提取视频模态特征信息。除此之外，k折交叉验证并未做进一步的调参实验，这都是在本次评测中值得反思之处。

表 5 实验参数设置

超参数	值
K	3
Batch size	30
Learning rate	5e-4
Epoches	60
Drop_out	0.1



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

谢谢！