

# SYS1874 for SMP2020-EWECT

作者：施云生

大连理工大学

## 摘要

SMP2020-EWECT 是一项基于微博文本数据进行文本情绪分析的比赛, 该比赛的任务包括对于通用微博文本以及疫情时期微博文本的情绪分析。在本次比赛中, 我们在 RoBERTa 的基础模型上进行改进, 构建了 FE-RoBERTa (Feature Ensemble-RoBERTa) 网络模型。该模型在最终测试集上测试时取得了综合第 5 名, 其通用微博 Macro\_F1、通用微博 Accuracy、疫情微博 Macro\_F1、疫情微博 Accuracy 成绩分别为: 0.7815 (4)、0.8054 (4)、0.6859 (5)、0.8140 (2)。

关键词: 自然语言处理, 情感分析, 社交媒体情绪分类, 特征融合, 模型融合

## 1. 引言

SMP2020-EWECT 是一项基于微博文本数据进行文本情绪分析的比赛, 该比赛的任务包括对于通用微博文本以及疫情时期微博文本的情绪分析。第一部分为通用微博数据集, 该数据集内的微博内容是随机获取到微博内容, 不针对特定的话题, 覆盖的范围较广; 第二部分为疫情微博数据集, 该数据集内的微博内容是在疫情期间使用相关关键字筛选获得的疫情微博, 其内容与新冠疫情相关。每条微博被标注为以下六个类别之一: `neural` (无情绪)、`happy` (积极)、`angry` (愤怒)、

sad (悲伤)、fear (恐惧)、surprise (惊奇)。通用微博训练数据集包括 27,768 条微博，验证集包含 2,000 条微博，测试数据集包含 5,000 条微博。疫情微博训练数据集包括 8,606 条微博，验证集包含 2,000 条微博，测试数据集包含 3,000 条微博。

在实验中我们发现，本次比赛主要存在的难点包括：1) 数据量较少，模型容易过拟合；2) 不同类别数据之间数量差异较大，即类别不均衡问题严重；3) 情感分析问题本身较为困难，尤其是在微博文本内容中同类情感的表现方式多样。

本报告为本次比赛中我们所用算法及模型的介绍，具体将分为如下三个方面：基础模型，特征组合以及训练细节。

## 2. 模型及方法介绍

所用模型以 RoBERTa 模型 (RoBERTa-Large) 为基础。文本特征被送入 24 层的编码器中特征隐藏层，最后将编码器所提取特征送入分类器中预测最终结果。考虑到情感分析问题本身的难度，我们融合了 RoBERTa-Large 中的浅层特征与深层特征，用于最终预测。

在本章节中，我们将首先简要介绍所用的基础模型 RoBERTa，接着介绍我们的改进部分，最后是训练设置以及模型 ensemble 等细节的介绍。

### 2.1 基础模型

我们使用了 RoBERTa-Large 作为基础模型。RoBERTa-Large 包含 24 个隐藏层。每一个隐藏层都包含相应的 self-attention 结构，用于利用不同位置的特征来改良文本特征。最终，我们将所得的编码器特征送入包含两层全连接层的分类器

(linear-tanh-linear)中预测最终结果。

## 2.2 特征融合

考虑到情感的表现方式的多样性,即同时存在与情感紧密相关的特定词汇或文字在原始文本中直接出现的情况(比如在愤怒、悲伤等较为极端的情况),以及存在情感较为内敛,需要从文本具体内容中提取情感信息的情况,我们在基础模型中同时提取了浅层特征和深层特征用于最终的预测,即把解码器隐藏层中10、15、20以及最终的输出(24)层的特征取出。

## 2.4 其他技巧

### 1. 模型训练

我们把预训练的 RoBERTa-Large 在比赛数据集中进行调整。学习率设置为  $2e-5$ , 权重衰减设置为  $1e-5$ , 句子最大长度为 256, batch 大小为 8, 网络在通用微博以及疫情微博数据上均只训练 1 个 epoch。

### 2. 模型融合

对于通用微博以及疫情微博,我们分别训练了 10 个和 6 个相同结构的模型,最终结果为这些模型预测结果的平均。

## 3. 实验结果及分析

在线验证集性能比较

模型	F1(Usual)	Acc(Usual)	F1(Virus)	Acc(Virus)
RoBERTa	0.7792	0.8005	0.6822	0.8010
FM-RoBERTa	<b>0.7840</b>	0.8050	<b>0.6970</b>	0.8030
FM-RoBERT	0.7834	<b>0.8055</b>	0.6787	0.8060

(Cross-Train)				
FM-RoBERTa +Focal Loss	0.7791	0.8015	0.6778	0.7995
FM-RoBERTa (x20)	0.7837	0.8045	0.6912	<b>0.8095</b>
FM-RoBERTa (Clean)	0.7804	0.8020	0.6822	0.8060

对比 RoBERTa 和 FM-RoBERTa 的结果发现，加入了更多浅层特征后网络性能的提升较为明显，尤其是在 F1 的提升。

除此之外，我们仍测试了不同训练方式的影响。FM-RoBERT(Cross-Train) 表示在使用各自数据训练网络模型后，再交叉训练两个模型，即把在通用（疫情）微博数据训练过的模型在疫情（通用）微博数据上进行训练作为疫情（通用）微博的最终预测模型。该训练方案得到了更高的准确度，但其 F1 却有所下降。考虑到类别不均衡问题，我们尝试使用 Focal Loss 进行训练，但在我们的实验中结果下降明显。另外，我们比较了大量模型 ensemble 的结果，即 FM-RoBERT(20)，我们使用了 20 组训练模型进行 ensemble，发现除了在疫情微博的准确度上提升明显以外，其他指标的效果反而不如现有设置（通用微博 x10，疫情微博 x6）。原始微博文本数据中包含了大量人名、微博 id 等信息，我们尝试清除文本中的这类信息，其效果如 FM-RoBERT(Clean)，即同样会有较为明显的下降，我们认为这类数据加入之后反而性能更好的原因在于其相当于为有效文本加入了一定的扰动，减轻了网络过拟合的问题。

在最终测试集上，我们取得了综合第 5 名，其通用微博 Macro\_F1、通用微博 Accuracy、疫情微博 Macro\_F1、疫情微博 Accuracy 成绩分别为：0.7815 (4)、0.8054 (4)、0.6859 (5)、0.8140 (2)。

## 4. 总结

本次评测中我们所用特征融合方法提升明显，能让网络更有效、更准确地提取出情感相关信息。但在本次比赛数据中的其余两个问题，即数据量以及过拟合问题，我们的实验仍未能找出更好的方法。因此，这一任务以及其相关的数据集未来仍存在很大的发展空间。