



论文摘要

随着目前全球疫情的蔓延,网络在线面试的需求程度上升,对智能面试评价工具的需求也随之上升。面试官在面试中通常会对面试者进行人格特性、沟通技能等方面的画像评估,并在此基础上形成综合评价。因此,在本文中,我们提出基于关键词注意力机制的多层次人格特性细粒度预测方法,充分结合面试者表述文本的多尺度上下文特征,对人格特性进行准确预测,并据此得到综合面试评分。基于真实面试场景数据的实验结果表明,该方法能有效地评价面试者的不同人格特性得分,并准确的预测面试者总体得分。

系统框架及算法原理

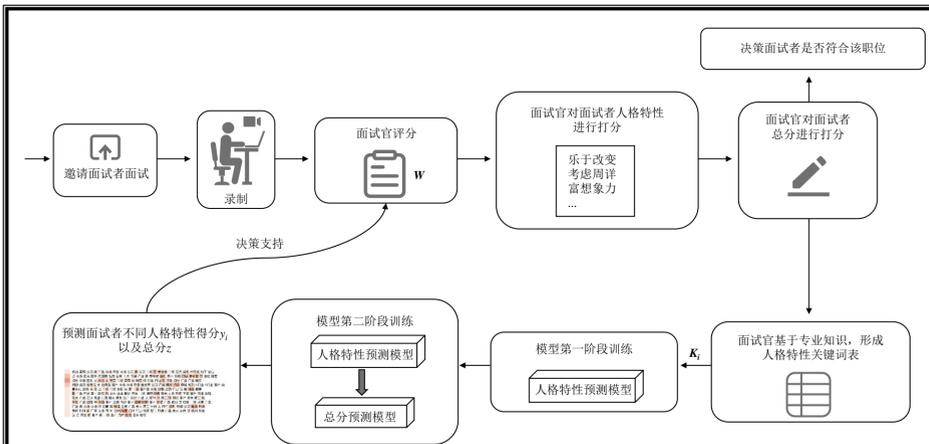


图1 总体任务结构

总体任务结构如图1所示:本文将自动面试评价任务形式化描述为一个两阶段学习的回归任务:给定一个面试者表述内容的文本序列 W 和关键词序列 K_i ,模型第一阶段预测与给定关键词序列对应的面试者不同人格特性得分 y_i ,模型第二阶段对面试者的总分 z 进行预测。

图2为本文提出方法中的第一阶段模型,使用关键词-问题注意力的层次网络来模拟面试者表述文本中的层次结构,捕获句子中的重要单词以及面试者表述中更贴切问题的回答,并预测人格特性得分 y_i 。

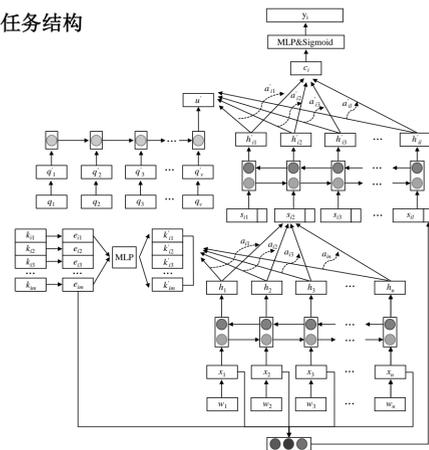


图2 基于关键词-问题注意力的人格特性预测模型

系统框架及算法原理(续)

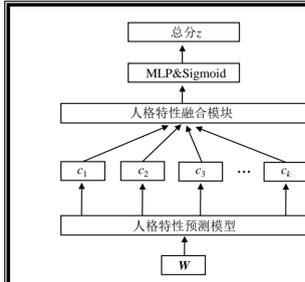


图3 基于人格特性向量融合的总分预测模型

如图3所示,模型在第二阶段训练中,将根据第一阶段抽取出来的人格特性向量,对面试者的总分 z 进行预测,其中人格特性向量的融合分为先融合与后融合两种模型。

实验仿真

我们通过与人力资源公司进行合作,搭建了异步在线面试平台,模拟了真实场景下的结构化面试过程。面试共分为3道题,每道题设定5分钟答题时间,表1、表2展示了我们的模型与基线模型的性能对比,图4展示了我们的模型与基线模型的细粒度人格特性预测的对比,表3展示了在第二阶段训练中,不同特征融合方法的对比。

表1 不同模型性能对比

Method	Personality Traits			Total Score		
	PCC	SCC	QWK	PCC	SCC	QWK
BiLSTM	0.465	0.440	0.376	0.485	0.399	0.344
HA-LSTM	0.458	0.430	0.369	0.519	0.434	0.367
KA-LSTM	0.595	0.588	0.546	0.544	0.447	0.392
HKQA-LSTM	0.624	0.619	0.578	0.783	0.706	0.679

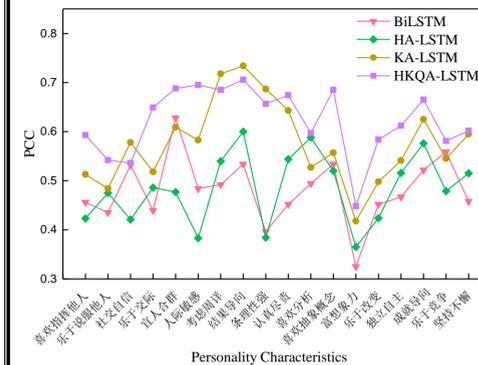


图4 不同模型细粒度评分对比

表2 不同模型预测结果MSE

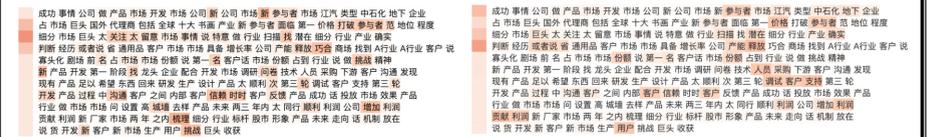
Method	Personality Traits	Total Score
BiLSTM	1.732	0.778
HA-LSTM	1.764	0.695
KA-LSTM	1.543	0.643
HKQA-LSTM	1.376	0.314

表3 不同特征融合方法预测精度

Method	Total Score		
	Precision	Recall	Macro-F1
HKQA-LSTM-Early	0.794	0.848	0.816
HKQA-LSTM-Late	0.824	0.872	0.845

实验仿真(续)

从图5的可视化例子中可以看出,我们的模型可以有效地关注文本信息中重要的单词,忽略掉不重要的单词。如图5(a)中,为了判断面试者的“乐于改变”人格特性,模型更加关注“打破”、“梳理”、“挑战”等词。而图5(b)中,由于考虑的是面试者的“考虑周详”人格特性,模型则将注意力更多的放在“用户”、“客户”、“调试”等词。

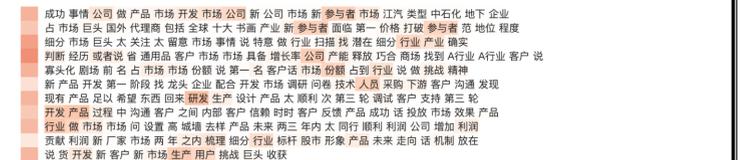


(a) “乐于改变”人格特性可视化

(b) “考虑周详”人格特性可视化

图5 注意力权重可视化

图6展示了模型预测误差较大的一项人格特性项“富想象力”。在该例子中,模型对不同句子的注意力比较平均,并将单词注意力更多地放在了重复率较高的一些名词上。一个可能的原因是“富想象力”本身是一项较为抽象的人格特性,因此面试官在打分的过程中会受到面试者的声音、表情动作等其他方面的影响。



“富想象力”人格特性可视化

图6 负例分析

论文结论

本文提出了一种基于关键词注意力机制的细粒度面试评价方法,模型利用外部的关键词以及问题作为模型不同层次上的注意力导向,对面试者面试过程中所展现出来的不同人格特性进行预测。

模型对第一阶段训练得到的人格特性表征尝试了不同的融合方式,得到第二阶段的最终模型。在数据集上的实验结果表明,本文提出的模型能有效地对面试者的不同人格特性进行细粒度评价,筛选出同一文本中不同的重要词句,并在此基础上对面试者进行总体评分,有效帮助面试官做出决策。下一步工作中,我们将引入声音、视觉等其他模态,并结合半监督学习方法进一步提升细粒度评分性能。