arXiv:2303.03049v1 [eess.AS] 6 Mar 2023

基于自发语音的阿尔茨海默病跨语言迁移学习检测

Bastiaan Tamm、Rik Vandenberghe、Hugo Van hamme*

鲁汶大学，比利时

摘要

阿尔茨海默病（AD）是一种进行性神经退行性疾病，通常与记忆缺陷和认知能力下降有关。随着人口老龄化，人们对认知障碍检测自动化方法的兴趣日益浓厚。近年来，利用自发语音进行阿尔茨海默病检测备受关注。虽然取得的结果颇有希望，但尚不确定所学到的语音特征能否跨语言泛化。为填补这一空白，ADReSS-M 挑战赛应运而生。本文介绍了我们参加 ICASSP-2023 信号处理大挑战（SPGC）的方案。该模型使用图片描述任务的 228 个英语样本进行训练，并仅用 8 个样本迁移至希腊语。我们在阿尔茨海默病检测中取得了 82.6% 的准确率，认知评分预测的均方根误差为 4.345，并在 24 名参赛者中排名第二。

索引词——阿尔茨海默病，跨语言

1. 引言

阿尔茨海默病（AD）是一种进行性神经退行性疾病，通常与记忆缺陷和认知衰退有关。它是最常见的痴呆症，也是 65 岁及以上人群的第五大死因 [1]。随着人口老龄化，人们对认知障碍检测的自动化方法产生了浓厚兴趣，尤其是那些成本低廉且易于扩展的方法。近年来备受关注的一种可能途径是分析自发语音，这是一种易于获取的媒介，可帮助我们了解大脑的运作。然而，大多数已提出的方法都没有探究哪些语音特征能够跨语言迁移以用于阿尔茨海默病检测 [2]。

因此，ADReSS-M——ICASSP-2023 信号处理大挑战赛（SPGC）——应运而生，旨在研究这一问题 [3]。该挑战赛的目标是利用图片描述任务中的英语语音训练模型，并将其应用于希腊语中的另一项图片描述任务。挑战赛包含两项任务：第一项是预测阿尔茨海默病诊断，第二项是预测参与者的简易精神状态检查（MMSE）得分。后者是一份广泛用于临床和研究环境中、用于测量认知障碍程度的问卷，总分为 30 分。

本文介绍了我们提交给该挑战赛的方案。[1]这些模型使用一系列声学特征和协变量（年龄、性别、教育程度）进行预测。模型首先在英语数据上训练，然后通过混合语言批次和参数平均迁移至希腊语。该方法在阿尔茨海默病检测任务中取得了 82.6% 的准确率，在认知评分预测任务中取得了 4.345 的成绩；相比之下，基线的相应成绩为 73.9% 和 4.955。凭借这一方案，我们在 24 名参赛者中排名第二。

2. 方法

2.1. 数据集

本次挑战使用了两个数据集，一个为英语数据集，一个为希腊语数据集，均包含健康对照组和阿尔茨海默病患者描述图片时的音频录音。挑战组织者将数据分为三个子集：英语训练集（n=237，122 名阿尔茨海默病患者）、希腊语样本集（n=8，4 名阿尔茨海默病患者）和希腊语测试集（n=46，22 名阿尔茨海默病患者）。测试集的统计数据根据提交结果的混淆矩阵得出，事先并不知晓。已知各子集在阿尔茨海默病、年龄和性别方面均保持平衡 [3]。

我们从英语训练集移除了一名健康对照组成员（无认知评分），并移除 8 名阿尔茨海默病患者以保持平衡（n=228，114 名阿尔茨海默病患者）。最后，对于 12 名未提供教育程度信息的对照组成员，我们将缺失值设为 12 年。

2.2. 预处理

每个音频文件被切分为十个等长片段。使用 openSMILE [5]计算每个片段的 25 维 eGeMAPS [4] 特征向量。

2.3. 模型

阿尔茨海默病检测模型和认知评分预测模型基于相同的架构。它们

*本研究由鲁汶大学特别研究基金资助，资助编号为 C24M/22/025。

[1]我们的代码：https://github.com/lcn-kul/madress-2023

都相当小：参数量分别仅为 767 和 468。每个模型的输入均为由十个 eGeMAPS 特征组成的序列，以及年龄、性别和教育程度这些协变量。估计得到的阿尔茨海默病概率也作为协变量纳入认知评分预测模型。为简化起见，协变量与 eGeMAPS 特征序列拼接在一起。

该架构由四个部分组成。首先，对输入特征进行批归一化。接着，将特征向下投影至较小的隐藏空间（分别为 12 维或 8 维），然后进行丢弃和 ReLU 激活。

随后，使用注意力池化压缩时间维度。注意力权重通过一个两层前馈网络计算，该网络的中间空间是隐藏空间的两倍大小，并在之后应用 softmax，使权重之和为 1。两层之间使用丢弃和 ReLU 激活。

最后一步是通过线性投影将向量映射到输出空间（分别为 2 维或 1 维）。对于认知评分预测模型，使用 sigmoid 函数将数值映射到 [0,1] 范围内。MMSE 标签也会归一化到相同范围。

2.4. 英语预训练

英语数据按 80% 训练集和 20% 验证集进行划分。模型使用英语训练数据进行训练，并在希腊语样本集上进行验证。为考虑随机初始化的影响，使用五个不同的随机种子进行五次训练。选择五次运行中验证损失最低的模型作为预训练模型。

2.5. 混合批次迁移学习

预训练模型在英语训练数据和 4 个希腊语样本（其中 2 个为阿尔茨海默病样本）上进行微调。每个小批次中的第五个样本都替换为一个希腊语样本。模型使用英语验证数据和 4 个以相同方式插入的留出的希腊语样本进行验证。

为提高鲁棒性，我们重复上述过程，但将 4 个希腊语训练样本替换为 4 个留出的样本。最后，由于这两个模型均由同一预训练模型初始化，因此可以对它们的参数取平均。

2.6. 训练细节

这些模型使用 Python 3.8 中的 PyTorch（v.1.11.0）和 PyTorch Lightning （ v.1.8.6 ）库实现。网络使用 AdamW 优化器进行训练，权重衰减设为 1e-2。学习率在前 100 步线性预热，之后固定为 3e-3。阿尔茨海默病检测模型使用交叉熵损失，认知评分预测模型使用均方误差损失。每个模型的批次大小为 32，共训练 30 个训练轮次，并选择验证损失最低的模型。

3. 结果

该挑战允许提交五次。因此，为了测试上述方法的鲁棒性，我们使用不同的随机种子完整运行了五次。阿尔茨海默病检测任务的测试准确率按升序排列为 71.7%、73.9%、76.1%、80.4% 和 82.6%。表现最佳的模型特异度为 91.7%，精确率为 88.9%，敏感度为 72.7%，F1 分数为 80.0%。

认知评分预测任务的均方根误差（RMSE）按降序排列分别为 4.837、4.816、4.716、4.713、4.345。请注意，这些模型将阿尔茨海默病检测模型输出的概率作为输入。由于事先并不知道哪一个阿尔茨海默病检测模型效果最好，因此对提交的 5 个阿尔茨海默病检测模型输出的概率取了平均值。

4. 结论

本文介绍了我们参加 ADReSS-M 挑战赛的方案。我们的方法优于最佳基线模型，且对模型初始化具有鲁棒性。我们工作的创新之处在于预训练和混合批次微调流程。实际架构极其简单，我们相信，通过改进特征提取网络，性能还能进一步提升。

5. 参考文献

[1] “2022 年阿尔茨海默病事实与数据”，《阿尔茨海默病与痴呆》，第 18 卷，第 4 期，第 700–789 页，2022 年。[2] Sofia de la Fuente Garcia、Craig W. Ritchie 和 Saturnino Luz，“用于监测阿尔茨海默病的人工智能、语音和语言处理方法：系统综述”，《阿尔茨海默病杂志》，第 78 卷，第 1547–1574 页，2020 年，第 4 期。[3] Saturnino Luz、Fasih Haider、Davida Fromm、Ioulietta Lazarou、Ioannis Kompatsiaris 和 Brian MacWhinney，“通过自发语音进行多语言阿尔茨海默病痴呆识别：信号处理重大挑战”，2023 年。[4] Florian Eyben 等，“用于语音研究和情感计算的日内瓦极简声学参数集（GeMAPS）”，《IEEE 情感计算汇刊》，第 7 卷，第 2 期，第 190–202 页，2015 年。[5]Florian Eyben、Martin W¨ollmer 和 Bj¨orn Schuller，“openSMILE：慕尼黑通用、快速的开源音频特征提取器”，载于《第 18 届 ACM 多媒体国际会议论文集》，美国纽约州纽约市，2010 年，MM ’10，第 1459–1462 页，美国计算机协会。