

基于层次注意力机制的多任务疾病进展模型

潘祖江¹ 刘宁¹ 张伟² 王建勇¹

1 清华大学计算机科学与技术系 北京 100084

2 华东师范大学计算机科学与技术学院 上海 200333

(pzj17@mails.tsinghua.edu.cn)

摘要 阿尔茨海默症是一种不可逆的神经退化疾病,由于脑组织的退化而产生的严重的认知问题。目前已有许多临床实验和研究计划来研究阿尔茨海默症的病理学,这些实验和计划会产生一些可以用来分析的数据。文中着重结合多种临床特征,对阿尔茨海默症进行自动诊断,并预测潜在的预后风险,进而提出了一个基于层次注意力机制的多任务疾病进展模型。该模型将疾病自动诊断任务作为主任务,疾病预后预测任务作为辅任务,以提升模型的泛化能力,进而提升疾病自动诊断任务的效果。其应用了两层的注意力机制,注意力分别应用在特征层和就诊记录层,使得模型可以对不同的特征以及不同的就诊记录有不同的注意力。在 ADNI(Alzheimer's Disease Neuroimaging Initiative)数据集上进行实验,并将所提模型与多个基准模型进行比较,实验结果表明,提出的模型具有更好的效果,为临床实际应用提供了更好的鲁棒性。

关键词: 注意力机制;多任务学习;自动诊断;预后预测;阿尔茨海默症

中图法分类号 TP391.4

MTHAM: Multitask Disease Progression Modeling Based on Hierarchical Attention Mechanism

PAN Zu-jiang¹, LIU Ning¹, ZHANG Wei² and WANG Jian-yong¹

1 Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

2 School of Computer Science and Technology, East China Normal University, Shanghai 200333, China

Abstract Alzheimer's disease (AD) is an irreversible neurodegenerative disease. The degeneration of brain tissue causes serious cognitive problems and eventually leads to death. There are many clinical trials and research projects to study AD pathology and produce some data for analysis. This paper focuses on the diagnosis of AD and the prediction of potential prognosis in combination with a variety of clinical features. In this paper, a multi-task disease progression model based on hierarchical attention mechanism is proposed. The task of disease automatic diagnosis is regarded as the main task, and the task of disease prognosis is regarded as the auxiliary task to improve the generalization ability of the model, and then improve the performance of disease automatic diagnosis task. In this paper, two layers of attention mechanism are applied in the feature layer and the medical record layer respectively, so that the model can pay different attention to different features and different medical records. The validation experiment is carried out on ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset. Compared with several benchmark models, the experimental results show that the proposed method has better performance and provides better robustness for clinical application.

Keywords Attention mechanism, Multi-task learning, Automatic diagnosis, Prognosis prediction, Alzheimer's disease

1 引言

《2018年世界阿尔茨海默症报告》^[1]指出,世界上每隔3s就会产生一例新的痴呆症病例,2018年全世界有5000万人受到痴呆症的折磨,而到2050年将达到1.52亿人。预计2018年全球痴呆症治疗支出为1万亿美元,到2030年将增至2万亿美元。而阿尔茨海默症是最著名的痴呆症类型,占有痴呆症病例的50%~60%。它会破坏脑细胞和神经元,

扰乱信息在大脑中的传递,是一种不可逆的神经退化疾病,脑组织的退化将产生严重的认知问题,最终导致患者死亡。目前已有许多临床实验和研究计划来研究阿尔茨海默症的病理学,并产生了一些可以用来分析的数据。目前还没有有效的治疗方法可以逆转阿尔茨海默症造成的损害,只能通过辅助技术来延缓病情的发展。因此,尽早对阿尔茨海默症进行诊断是至关重要的,可使治疗尽早开始,帮助患者维持认知功能。目前,阿尔茨海默症自动诊断的方法主要是使用基于多

收稿日期:2019-08-30 返修日期:2019-10-22 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金重点项目(61532010)

This work was supported by the Key Program of National Natural Science Foundation of China (61532010).

通信作者:张伟(wzhang@sei.ecnu.edu.cn)

模态异构数据的方法。ADNI 数据集中包含临床指标、医学影像、遗传基因数据和生物标志物等多模态异构数据,基于多模态异构数据的方法尝试将多模态数据结合起来对疾病进行建模。例如,Wang 等使用深层矩阵分解来融合多模态数据进行疾病建模^[2],探索不同模态之间的非线性相互作用,并利用该相互作用传递知识和提高预测性能。Dai 等结合医学影像、脑脊液等不同模态数据的临床特点,提出将流形学习和随机向量功能连接网络(Random Vector Functional-link Network, RVFL)结合起来^[3-4],从而提高分类识别能力。

由于 ADNI 数据集天生具有时序性,因此对其时序数据进行建模可以获得更多的信息,从而提升病情进展模型的性能。本文着重结合多种临床特征,提出了一个基于层次注意力机制的多任务疾病进展模型,将疾病自动诊断任务作为主任务,疾病预后预测任务作为辅任务,提升模型的泛化能力,进而提升疾病自动诊断任务的性能。该模型应用了两层注意力机制,即分别应用在特征层和就诊记录层,使得模型可以对不同的特征以及不同的就诊记录有不同的注意力。最后,在 ADNI 数据集上进行实验,并将本文模型与多个基准模型进行比较,实验结果表明,本文模型比基准模型具有更好的效果,为临床实际应用提供了更好的鲁棒性。

2 相关工作

2.1 阿尔茨海默症的研究进展

阿尔茨海默症的自动诊断和预后预测都可以看作一个三类问题。患者的就诊记录由各种医学测试变量组成,这些变量可以表示为特征点(存储为向量)。传统的统计机器学习算法和当前的深度学习算法^[5-6]都可用于将特定的医学记录(特征点)分类为相应的类别,例如正常认知老化(Cognitive Normal, CN)、轻度认知障碍(Mild Cognitive Impairment, MCI)和阿尔茨海默症(Alzheimer's Disease, AD)。用于阿尔茨海默症自动诊断的算法包括随机森林、支持向量机、K 近邻算法^[7]以及人工神经网络。Lebedev 等利用随机森林算法对阿尔茨海默症进行诊断^[8]。López 等利用支持向量机(Support Vector Machine, SVM)检测 AD 的早期症状^[9]。Shi 等在早期开发了用于诊断 AD 和 MCI 的非线性度量学习模型^[10]。Dai 等提出了一种改进的特征嵌入等距映射算法,并对类似任务采用了集成学习算法^[11]。文献[3]提出了一种将流形学习和 RVFL^[4]相结合的混合系统,以更好地从临床数据中获取高维非线性信息^[12-13]。

2.2 基于注意机制的模型

注意力机制主要是用于进行信息筛选,从输入信息中选取相关的信息。注意力机制可以分为两步:1)计算注意力分布 α ;2)根据 α 来计算输入信息的加权平均^[14]。

基于注意力机制的模型已成功地应用于自然语言处理、图像识别以及语音识别等不同类型的深度学习任务^[15-18]。从直观上看,机器翻译的注意力机制的工作方式如下:对于原序列中的每一个词 x_i ($i \in 1, \dots, S$),长短期记忆网络(Long short-term Memory Network, LSTM)生成隐状态 h_i ($i \in 1, \dots, S$)来表示句子中的单词。对于目标语言中的第 j 个词,计算各个编码器隐藏层状态 $h_1 \sim h_S$ 与解码器状态 s_{j-1} 之间

的相关程度,并进行 SoftMax 归一化得到每个隐藏层向量的权重 α_i ,然后计算上下文向量 $c_j = \sum_i \alpha_i h_i$,并用它来预测目标语言中的第 j 个单词。

反向时间注意力模型(REverse Time Attention model, RETAIN)通过以相反的时间顺序查看电子健康病历(Electronic Health Record, EHR)数据来模仿医生的做法,以使最近的临床就诊受到更高的关注,并使用注意力生成机制利用序列信息^[19]。文献[16]采用层次注意机制对文章进行建模,即分别对词级和句级应用注意力机制,得到一个上下文向量,然后执行文章的情感分类任务。本文受该思想启发,第一次将层次注意力机制应用于临床时序建模任务,分别在特征层和不同的时间步骤上应用注意力机制,学习到整体表示的上下文向量,最后执行疾病自动诊断和疾病预后预测任务。

2.3 基于多任务学习的模型

多任务学习(Multi-task Learning, MTL)已经在机器学习的许多应用中取得了成功,如自然语言处理^[20-21]、语音识别^[22]、计算机视觉^[23]和药物发现^[24]。Ruder 提出可以把多任务学习看作一种归纳转移^[25]。归纳转移可以通过引入归纳偏差来辅助改进模型,使得模型更倾向于某些假设。例如,归纳偏差的一种常见形式是正则化,正则化可以导致稀疏解。在 MTL 的情况下,辅助任务提供了归纳偏差,这使得模型更倾向于解释多个任务的假设,提高模型的泛化能力。即使在普通情况下每个任务只优化一个损失,增加一个辅助任务也有助于改进主要任务的性能。在深度学习的背景下,多任务学习通常通过隐藏层的硬参数或软参数共享来完成。本文使用的多任务学习框架为硬参数共享。

硬参数共享是神经网络中最常用的 MTL 方法,Caruana 于 1997 年已对其进行总结^[26]。其通过共享所有任务之间的隐藏层来实现,同时保留各个任务特定的输出层,如图 1 所示。硬参数共享大大降低了过拟合的风险,事实上,Baxter 提出,共享参数的过拟合风险 N 阶小于对单个任务的参数过拟合风险,其中 N 是任务的数目^[27]。

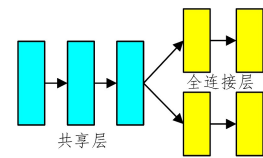


图 1 深度学习中 MTL 的硬参数共享

Fig. 1 Hard parameter sharing for MTL in deep learning

3 多任务层次注意力进展模型

基于层次注意力机制的多任务疾病进展模型针对阿尔茨海默症的疾病自动诊断和疾病预后预测任务进行多任务建模,将疾病自动诊断任务作为主任务,疾病预后预测作为辅任务,通过多任务学习来提高模型的泛化能力,进而提高主任务的性能表现。图 2 展示了该模型的总体结构。从 1 到 T 的给定时间步的模型输入,特征集集合为 $\{f_{it}\}_{i \in R}$, f_{it} ($i \in 1, \dots, |R|$) 表示某个患者 p 第 i 次就诊记录的第 t 个医学测试特征, $|R|$ 表示医学测试特征的数量。本文首先学习当前时间步输入层不同特征的特定权重 α_{it} ,继而得到一个整体表示,其在本文模型中被称为低层注意力。然后将学习到的输入表示作为

LSTM的输入,在每个时间步 LSTM 都会产生相应的隐藏状态 h_t 。对于根据患者一系列就诊记录得到的隐藏状态集 $\{h_t\}_{t=1}^T$,本文尝试学习对应于当前记录的权重 α_t ,此处 T 为病人就诊记录的长度。随后,本文通过权重向量 α 和 $\{h_t\}_{t=1}^T$ 得到上下文向量 c ,其在本文模型中被称为高层注意力。接下来,本文使用上下文状态 c 和当前状态 h_T 作为注意力隐藏状态 \tilde{h}_T ,将其作为两个不同的完全连接层的输入特征进行多任务学习,其中一个完全连接层用于疾病自动诊断,另一个完全连接层用于疾病预后预测。本文的整个模型是端到端训练的。

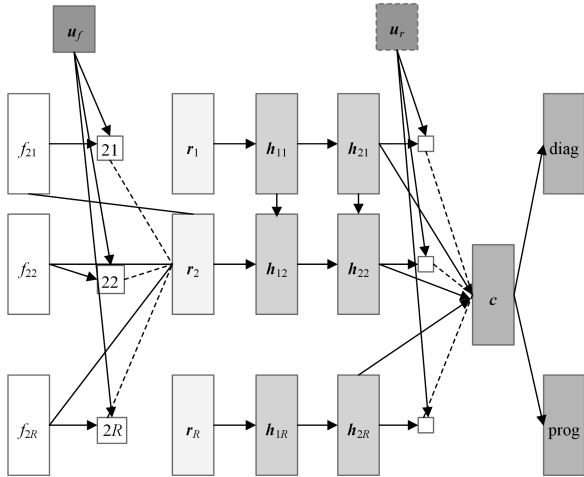


图2 模型框架

Fig. 2 Model architecture

对于由一组不同特征 f_{it} ($t \in 1, \dots, R$) 组成的就诊记录 x_i ,用注意力机制学习当前就诊记录的表示得到 r_i 。对于每个时间步骤和输入 r_i ,LSTM 都有一个对应的隐藏状态 h_i ($i \in 1, \dots, T$),本文有选择地对隐藏状态 $h_1 - h_T$ 施加注意力,最终得到上下文向量 c 。 c 是患者从 x_1 到 x_T 的就诊记录的最终表示,本文中 c 同时作为 T 时刻的疾病自动诊断任务和 $T+1$ 时刻的疾病预后预测任务的输入特征。

3.1 底层注意力机制

对于由一组不同特征 f_{it} ($t \in 1, \dots, R$) 组成的就诊记录 x_i ,并非所有特征都对就诊记录的表示有同等的作用,为了学到特定就诊记录的良好表示,本文使用注意力机制对当前就诊记录进行学习,从而对更重要的特征施加更多的注意力。特别地:

$$u_{it} = \tanh(W_f f_{it} + b_f) \quad (1)$$

$$\alpha_{it} = \frac{\exp(u_{it}^T u_f)}{\sum_t \exp(u_{it}^T u_f)} \quad (2)$$

$$r_i = \sum_t \alpha_{it} f_{it} \quad (3)$$

本文首先将特征 f_{it} 传入单层 MLP 从而学到 u_{it} 作为特征 f_{it} 的表示,然后通过测量 u_{it} 和特征级上下文向量 u_f 的相似度来计算特征的重要性,并通过 SoftMax 函数得到归一化的特征重要性权重 α_{it} 。之后,基于特征重要性权重计算特征的加权和,进而得到该次医疗记录的表示向量 r_i 。

3.2 长短期记忆网络

循环神经网络可以更好地利用传统神经网络无法建模的时序信息,但同时也带来了更大的技术挑战——长期依赖性。长短期记忆网络的出现很好地解决了这个问题。在许多任务

中,使用 LSTM 结构的循环神经网络的性能优于标准的循环神经网络。近年来,门控循环单元 (Gated Recurrent Unit, GRU) 被提出,但是实验表明它与 LSTM 的性能相当,故本文依旧采用长短期记忆网络。LSTM 有 3 个“门”结构和 1 个特殊的网络结构^[28]。

$$f_t = \sigma(W_f [h_{t-1}, r_t]) \quad (4)$$

$$i_t = \sigma(W_i [h_{t-1}, r_t]) \quad (5)$$

$$o_t = \sigma(W_o [h_{t-1}, r_t]) \quad (6)$$

遗忘门根据当前输入 r_t 和上一个时间步输出 h_{t-1} 计算一个维度为 n 的向量 f_t , r_t 为底层注意力机制中获得的医疗记录的表示, f_t 中每个维度的值在 $(0, 1)$ 之内。将前一个时间步的状态 c_{t-1} 按位乘以 f_t 向量,那么在 f_t 值接近 0 的维度上的信息将被“遗忘”,而 f_t 值接近 1 的维度上的信息将被保留。输入门决定将 r_t 和 h_{t-1} 中的哪些信息更新到状态 c_{t-1} 中从而生成一个新的状态 c_t 。输出门根据最新的状态 c_t 、前一时间步的输出 h_{t-1} 和当前输入 x_t 确定当前时间步的输出。

$$z_t = \tanh(W_z [h_{t-1}, r_t]) \quad (7)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot z_t \quad (8)$$

$$h_t = o_t \cdot \tanh c_t \quad (9)$$

其中, \tanh 和 σ 是激活函数, i, f, o 分别为输入门、遗忘门以及输出门, W 为需要学习的权重, h_t 为学习到的 t 时刻之前信息的表示。

3.3 高层注意力机制

对于每个时间步骤,LSTM 都有一个对应的输出隐藏状态 h_i ($i \in 1, \dots, T$)。本文的任务是根据 x_1 到 x_T 的就诊记录,对 x_T 时间步骤执行自动诊断任务和对 x_{T+1} 时间步执行疾病预后预测任务。本文直观地认为,如果只使用与 x_T 时间步相对应的输出隐藏状态 h_T 来进行后续的诊断和预后任务,将会丢失一些有效的信息。因此,本文有选择地注意对应于访问 $x_1 - x_T$ 的输出隐藏状态 $h_1 - h_T$,并且最终得到上下文向量作为该时间序列信息的表示。

$$u_i = \tanh(W_h h_i + b_h) \quad (10)$$

$$\alpha_i = \frac{\exp(u_i^T u_h)}{\sum_i \exp(u_i^T u_h)} \quad (11)$$

$$c = \sum_i \alpha_i h_i \quad (12)$$

其中, c 是总结就诊 $x_1 - x_T$ 的所有信息的上下文向量。

3.4 诊断和预后

上下文向量是患者 $x_1 - x_T$ 的就诊记录的层次表示,本文将上下文向量和隐状态作为注意力隐藏状态,并同时用作本文的疾病自动诊断任务和疾病预后预测任务的输入特征:

$$\tilde{h}_T = [c, h_T] \quad (13)$$

$$y_{\text{diag}} = \text{softmax}(W_{\text{diag}} \tilde{h}_T + b_{\text{diag}}) \quad (14)$$

$$y_{\text{prog}} = \text{softmax}(W_{\text{prog}} \tilde{h}_T + b_{\text{prog}}) \quad (15)$$

3.5 模型训练

由于模型训练是一个分类问题,本文使用正确标签的对数损失作为训练损失函数,该模型使用多任务学习,因此本文将诊断步骤损失和预测步骤损失相结合:

$$L_{\text{loss}} = L_{\text{diag}} + L_{\text{prog}} = -\ln(y_{\text{diag}}) - \ln(y_{\text{prog}}) \quad (16)$$

本文选择 Adam 作为模型优化器,Adam (Adaptive Moment Estimation) 算法能够自动调整学习率,计算高效,可以快速得到优良的结果^[29]。同时,为了在随机梯度下降的鲁棒

性和批量梯度下降的效率之间取得平衡,本文采用了 Mini-batch 梯度下降策略,将 batch-size 设置为 64,以提高学习速度,保证模型收敛的效果。

4 实验

4.1 数据描述和数据预处理

本文使用的数据来自阿尔茨海默症神经影像学数据库 ADNI。ADNI 是一项纵向多中心研究,旨在收集临床、影像、遗传和生生物标志物,用于阿尔茨海默症的早期检测并跟踪疾病进展^[30]。ADNI 项目实施以来,它极大地促进了老年痴呆症的科学研究,并使世界各地的研究人员能够共享数据。

参与者被分为 3 组:正常老化、轻度认知损害(MCI)和阿尔茨海默症(AD)。ADNI 在北美不同地点招募年龄为 55~90 岁的参与者。在获得受试者的同意后,受试者将接受一系列测试,并在接下来的几年中每隔 6 个月重复一次,这些测试包括临床评估、神经心理测试、遗传测试、腰椎穿刺、MRI 和 PET 扫描。ADNI 的总体目标是验证阿尔茨海默症临床实验的生物标志物,并试图找出可以用来识别 AD 早期进展的敏感和特殊的标志物。这将有助于研究人员和临床医生开发新的治疗方法并监控其效率,缩短和减少临床实验的时间和成本。

ADNI 有几个不同的阶段:ADNI-1, ADNI-GO, ADNI-2, ADNI-3。参加者超过 2000 人,年龄分布如图 3 所示。

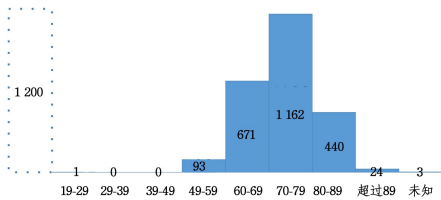


图 3 受试者年龄分布

Fig. 3 Age distribution of subjects

对于 ADNI 数据集,本文首先去除所有标签缺失的项,然后选取缺失率小于 30% 的医学测试特征,对于有缺失值的项,使用均值进行填充。由于不同患者的就诊次数不同,本文采用滑动窗口的技术来归一化输入长度。

4.2 基准模型

本文的模型 MTHAM 将与多个基准模型进行比较,既包括传统的方法如随机森林、支持向量机(Support Vector Machine, SVM),也包括深度学习的方法如多层感知机(Multi-layer Perceptron, MLP)和循环神经网络(Recurrent Neural Network, RNN)的变种。

- (1)随机森林: $n_estimators$ 设置为 100,评价标准为交叉熵。
- (2)支持向量机:SVM 核函数设置为径向基内核。
- (3)多层感知机:MLP 使用了两层隐藏层,隐藏层单元分别为 200, 30。
- (4)RNN:由两个隐藏层单元大小为 128 的 LSTM 构成。
- (5)RNN+attention(没有多任务):在 RNN 的基础上加上注意力机制。

模型的效果通过 Accuracy 值来评估,Accuracy 值表示分类正确的样例数占总样例数的比例。

4.3 实验结果

本文将 MTHAM 与 RF(随机森林)、SVM(支持向量

机)、MLP(多层感知机)、RNN、RNN+attention 等模型进行比较,实验结果如表 1 所列。通过实验结果可以看出,在疾病自动诊断任务上 MTHAM 表现良好,在 Accuracy 方面 MTHAM 比简单的 RNN 提升了 2.18%,比 MLP 高了 3.95%;基于深度学习的方法的效果显著优于基于传统机器学习的方法。分析可知,RNN,RNN+attention 和 MTHAM 模型相比 RF,SVM 以及 MLP 模型更多地利用了时序信息,使得模型达到了更好的效果。RNN+attention 相比 RNN 增加了对输出层隐状态权重的学习,获得了病人就诊记录序列更精确的表示,因此效果得到了提升。MTHAM 加入了疾病预后预测的辅任务,使得整个模型的泛化能力更好,进而提升了主任务疾病自动诊断的效果。

表 1 不同模型的实验结果

Table 1 Experimental results of different models

| Model | Accuracy |
|---------------|----------|
| RF | 0.8554 |
| SVM | 0.8569 |
| MLP | 0.8703 |
| RNN | 0.8880 |
| RNN+attention | 0.9028 |
| MTHAM | 0.9098 |

为了进一步说明所提模型分类效果,本文给出了 MTHAM 对阿尔茨海默症不同类别的 ROC 曲线。可以看到,macro-AUC(ROC 曲线下区域)值达到了 0.9749, micro-AUC 值达到了 0.9738。其中,macro-AUC 聚合所有类别的贡献获得一个平均值;micro-AUC 值通过计算每个类别标签的 AUC 值,然后计算它们的加权平均值来获得。由图 4 可知,MTHAM 的分类效果良好,NC 和 AD 类别的准确率更高,尽管正确分类 MCI 类别较难,但 AUC 仍达到了 0.9568。

Some extension of Receiver operating characteristic to multi-class

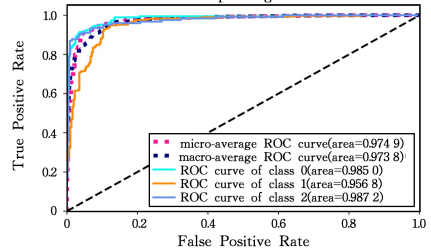


图 4 NC, MCI, AD 的 micro-ROC, macro-ROC 曲线

Fig. 4 micro-ROC, macro-ROC curves of NC, MCI, AD

结束语 本文提出了一种基于层次注意力机制的多任务疾病进展模型。该模型应用了多任务学习,将疾病自动诊断作为主任务,疾病预后预测作为辅任务,进而提升模型的泛化能力,提高主任务的性能。实验结果表明了该方法的有效性,相比其他模型在疾病进展建模问题上能更好地捕获病人病情的进展特征,具有更好的效果。在下一步的工作中,可以考虑完全应用注意力模型来替代 LSTM 模型,以更有效地学习病人病症的表示。

参考文献

- [1] PATTERSON C. World Alzheimer Report 2018—The state of the art of dementia research: New frontiers[R]. Alzheimer's Disease International (ADD): London, UK, 2018.

- [2] WANG Q,SUN M,ZHAN L,et al. Multi-Modality Disease Modeling via Collective Deep Matrix Factorization[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2017:1155-1164.
- [3] DAI P,GWADRY-SRIDHAR F,BAUER M,et al. Healthy cognitive aging: A hybrid random vector functional-link model for the analysis of alzheimer's disease[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [4] HUANG G B,ZHU Q Y,SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1/2/3):489-501.
- [5] HINTON G,DENG L,YU D,et al. Deep neural networks for acoustic modeling in speech recognition[J]. IEEE Signal processing magazine,2012,29(6):82-97.
- [6] LECUN Y,BENGIO Y,HINTON G. Deep learning[J]. Nature, 2015,521(7553):436.
- [7] TUFAIL A B,ABIDI A,SIDDIQUI A M,et al. Automatic classification of initial categories of Alzheimer's disease from structural MRI phase images: a comparison of PSVM, KNN and ANN methods[J]. Age,2012,2012:1731.
- [8] LEBEDEV A V,WESTMAN E,VAN WESTEN G J P,et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness[J]. NeuroImage: Clinical,2014,6:115-125.
- [9] LÓPEZ M,RAMÍREZ J,GÓRRIZ J M,et al. Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease[J]. Neurocomputing,2011,74(8):1260-1271.
- [10] SHI B,CHEN Y,HOBBS K,et al. Nonlinear Metric Learning for Alzheimer's Disease Diagnosis with Integration of Longitudinal Neuroimaging Features[C]// BMVC. 2015.
- [11] DAI P,GWADRY-SRIDHAR F,BAUER M,et al. Bagging ensembles for the diagnosis and prognostication of alzheimer's disease[C]// Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [12] XING E P,JORDAN M I,RUSSELL S J,et al. Distance metric learning with application to clustering with side-information [C]// Advances in neural information processing systems. 2003: 521-528.
- [13] LIN T,ZHA H. Riemannian manifold learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(5):796-809.
- [14] 邱锡鹏. 神经网络与深度学习[OL]. [2017-04-21]. <https://nndl.github.io/>.
- [15] BAHDANAU D,CHO K,BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473,2014.
- [16] YANG Z,YANG D,DYER C,et al. Hierarchical attention networks for document classification[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:1480-1489.
- [17] MNIH V,HEESS N,GRAVES A. Recurrent models of visual attention[C]// Advances in Neural Information Processing Systems. 2014:2204-2212.
- [18] YANG Y,YANG L,ZOU Y B,et al. Humor Recognition Based on Linguistic Features and Hierarchical Attention Mechanism [J]. Computer Engineering,2020,46(8):64-71.
- [19] CHOI E,BAHADORI M T,SUN J,et al. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism[C]// Advances in Neural Information Processing Systems. 2016:3504-3512.
- [20] COLLOBERT R,WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning[C]// Proceedings of the 25th International Conference on Machine Learning. ACM,2008:160-167.
- [21] ZHANG W J. An Online Multi-Task Learning Algorithm Based on Weight Matrix Decomposition [J]. Computer Engineering, 2019,45(8):190-197.
- [22] DENG L,HINTON G,KINGSBURY B. New types of deep neural network learning for speech recognition and related applications: An overview[C]// 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 8599-8603.
- [23] GIRSHICK R. Fast r-cnn[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015:1440-1448.
- [24] RAMSUNDAR B,KEARNES S,RILEY P,et al. Massively multitask networks for drug discovery[J]. arXiv:1502.02072, 2015.
- [25] RUDER S. An overview of multi - task learning in deep neural networks[J]. arXiv:1706.05098,2017.
- [26] CARUANA R. Multitask learning[J]. Machine Learning, 1997, 28(1):41-75.
- [27] BAXTER J. A Bayesian/information theoretic model of learning to learn via multiple task sampling[J]. Machine Learning, 1997, 28(1):7-39.
- [28] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.
- [29] KINGMA D P,BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980,2014.
- [30] WEINER M W,VEITCH D P,AISEN P S,et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception[J]. Alzheimer's & Dementia, 2013, 9(5): e111-e194.



PAN Zu-jiang, born in 1994, postgraduate. His research interests include data mining and machine learnig.



ZHANG Wei, born in 1988, Ph.D, associate researcher. His research interests include user data modeling and so on.