

一种自适应记忆神经网络多跳读取与覆盖度机制结合的药物推荐模型

王延达¹, 陈炜通², 皮德常¹, 岳琳²

(1. 南京航空航天大学计算机科学与技术学院, 江苏南京 211106;

2. 昆士兰大学信息技术与电子工程学院, 澳大利亚昆士兰布里斯班 4072)

摘要: 药物推荐的目标是依据病人的电子医疗记录生成药物处方, 为医生提供临床决策支持. 提取电子医疗记录中蕴含的时序模式以及上下文信息, 是成功推荐药物的关键. 以往研究忽略了病人之间医疗记录数据量存在差异, 无法根据不同病人自身情况, 调整数据读取过程中的关注重点以及数据读取迭代次数. 针对上述问题, 本文提出一种选择性覆盖度机制与自适应记忆神经网络读取结合的药物推荐模型. 模型使用记忆神经网络存储病人健康状况对应的时序模式编码结果, 利用覆盖度机制进行迭代读取过程中的数据过滤与注意力权重调整. 同时模型依据病人自身情况, 自适应决定记忆神经网络读取次数. 基于真实临床数据的实验结果显示, 本模型能够自适应地提取电子医疗记录中的重要数据, 构建有效的病人健康状况表示向量, 进而完成药物推荐.

关键词: 药物推荐; 记忆神经网络; 注意力机制; 覆盖度机制; 自适应多跳读取

中图分类号: TP183

文献标识码: A

文章编号: 0372-2112(2022)04-0943-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210968

Adaptive Multi-Hop Reading on Memory Neural Network with Selective Coverage Mechanism for Medication Recommendation

WANG Yan-da¹, CHEN Wei-tong², PI De-chang¹, YUE Lin²

(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China;

2. School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Queensland 4072, Australia)

Abstract: Medication recommendation aims to make effective prescriptions based on electronic healthcare records (EHRs) of patients, and assists caregivers in clinical decision making. Obtaining temporal patterns of patient conditions as well as contextual information contained in EHRs are the key issues for the success of recommendation. Existing methods do not take the difference in the amount of medical records of different patients into account, and fails to change the focus or number of iterations during information extraction according to personalized patient conditions. To address these problems, the medication recommendation model adaptive multi-hop reading with selective coverage mechanism (AMHSC) is proposed. The model stores encoded temporal patterns with memory neural networks (MemNN), and applies the selective coverage mechanism to balance attention weights over selected information during the attentive multi-hop reading on MemNN. Meanwhile, AMHSC adaptively determines the number of reading hops on MemNN according to personalized patient conditions. Experiments on real-world clinical dataset demonstrate that AMHSC successfully derives important information from EHRs to build informative patient representations for medication recommendation.

Key words: medication recommendation; memory neural network; attention mechanism; coverage mechanism; adaptive multi-hop reading

1 引言

随着机器学习技术的快速发展, 相关研究成果在

多个学科交叉融合领域取得了令人瞩目的研究成果, 例如智能交通、智慧城市以及智慧医疗等^[1-5]. 其中, 智

慧医疗旨在借助人工智能技术,以病人临床数据为基础,做出疾病诊断与治疗建议,为医生提供临床决策支持,提升诊疗质量与治疗效率.与此同时,电子医疗记录(Electronic Healthcare Records, EHRs)系统的广泛应用,为临床决策支持相关研究的开展提供了充足的数据. EHRs 记录了病人入院后各项生理指标的变化情况,以及使用的药物和接受的治疗手段等数据,能够满足不同临床应用研究对数据的需求.借助深度学习对 EHRs 开展分析,协助医护人员应对复杂多变的病人健康状况,是提升诊疗质量的重要手段.作为构建智慧医疗体系的重要一环,相关研究成果是提升医疗效率、节省医疗资源及应对人口老龄化的关键所在.

针对临床决策支持的研究克服了 EHRs 带来的种种挑战,已经在疾病诊断、死亡预测以及药物推荐等问题上取得了诸多进展^[6-9],涉及的研究议题包括 EHRs 信息提取、EHRs 表示学习以及输出预测等. EHRs 信息提取的目标,是从数据中提取结构化的医学概念,例如疾病和药品名称,并通过对缩写进行扩展和标准化来统一概念名称,作为概念之间关系提取的基础,用以明确不同概念之间的影响方式,例如药品 A 能够缓解疾病 B 的症状. EHRs 包含两方面的内容表示学习,即针对医学概念的表示学习和针对病人的表示学习.针对医学概念的表示学习,目标是以不同概念之间的相似性为基础,提取具有临床意义的概念表示向量.针对病人的表示学习,则是以病人对应的各项概念表示向量为基础,构建其对应的综合表示向量,作为输出预测的依据.输出预测依据病人表示向量对病人对应的临床输出进行预测,例如诊断病人所患疾病或疾病发作时间.根据模型使用数据的时间跨度,相关研究可以分为两类,即不考虑时序模式的静态预测和考虑时序模式的趋势预测.

药物推荐作为临床决策支持相关研究中的重要课题^[7,10],目标是以 EHRs 为依据,预测病人的药物使用情况,作为医生开具处方的参考,其所涉及的研究内容包括获取 EHRs 中蕴含的病人健康状况变化规律,消解因药物反应(Drug-Drug Interaction, DDI)造成的副作用,以及利用反事实推理评估药物治疗效果.为获取病人健康状况变化规律,以往研究大多使用循环神经网络(Recurrent Neural Network, RNN)对 EHRs 中的时序模式进行编码,并基于注意力机制对存储历史编码结果的记忆神经网络(Memory Neural Network, MemNN)进行多跳读取,提取其中的上下文信息^[7].为消解 DDI 引起的副作用,相关研究使用 DDI 信息构建图结构数据^[10],并利用图卷积网络(Graph Convolutional Networks, GCN)^[11]获取节点表示向量,为模型提供药物不

良反应信息.另一些研究基于强化学习(Reinforcement Learning, RL)或生成对抗网络(Generative Adversarial Network, GAN)对药物推荐过程进行修正,避免生成会引起不良反应的药物组合^[7,12].反事实推理通常将病人接受某项治疗方案的情形称为事实,将没有接受治疗的情形称之为反事实,并通过对比两种场景下病人群体的输出结果,明确药物与疾病之间的因果关系^[13].干扰因素(Confounder)是该过程面临的主要问题^[14],已有研究的应对方法包括基于逆治疗概率加权(Inverse Probability of Treatment Weighting, IPTW)的边际结构模型(Marginal Structural Models, MSMs)、循环边际结构网络(Recurrent Marginal Structural Networks, RMSNs)以及反事实循环网络(Counterfactual Recurrent Network, CRN)^[15,16].

在药物推荐涉及的各项研究内容中,准确获取 EHRs 中蕴含的病人健康状况变化规律以构建病人表示向量是药物推荐成功与否的关键所在.然而,以往方法使用的常规 MemNN 多跳读取使用同一固定数值作为所有病人的 MemNN 读取次数,忽略了不同病人所对应的入院次数以及所患疾病等数据存在差异这一情况.当 MemNN 读取次数不足时,模型无法获得全部的上下文信息,而对 MemNN 的不必要读取,则会导致读取结果中包含冗余数据.此外,常规注意力机制对注意力权重的分布没有限制,应用于 MemNN 多跳读取时很有可能重复关注 MemNN 的相同位置,从而读取冗余数据.为神经机器翻译(Neural Machine Translation, NMT)设计的覆盖度机制(Coverage Mechanism)是解决该问题的一种方案^[17].但是,NMT 与药物推荐对注意力机制的使用存在显著差异.NMT 任务要求将注意力权重分散到所有待翻译词汇,确保语句中的所有词汇均被准确翻译.药物推荐模型则需要通过变换注意力机制的关注重点,有选择地读取 MemNN 中的上下文信息.因此,对于药物推荐问题,合理的 MemNN 读取方法是将注意力权重的调整与分布限定在选定的重要数据范围内.

为解决上述问题,本文提出自适应 MemNN 多跳读取机制与选择性覆盖度机制,结合二者构建药物推荐模型 AMHSC(Adaptive Multi-Hop reading with Selective Coverage mechanism).该模型根据病人自身情况决定对应的 MemNN 多跳读取次数,取代了常规方法使用单一数值作为全局读取次数的做法.同时,AMHSC 模型使用门控循环单元(Gated Recurrent Unit, GRU)对多跳读取过程中 MemNN 数据的覆盖度进行记录和更新,基于覆盖度信息对 MemNN 数据进行选择,并对注意力权重进行调整,保证注意力权重合理地分散于所选定的重要数据范围内.两种机制的结合保证了 AMHSC 对

EHRs的合理读取,避免了重要数据缺失或冗余数据造成的消极影响.同时,为每个病人构建独有的MemNN并自适应地决定MemNN读取次数的做法,能够考虑病人独有的生理特征,准确获取病人健康状况表示向量,

作为药物推荐的依据.

2 AMHSC 模型

AMHSC模型整体框架如图1所示.

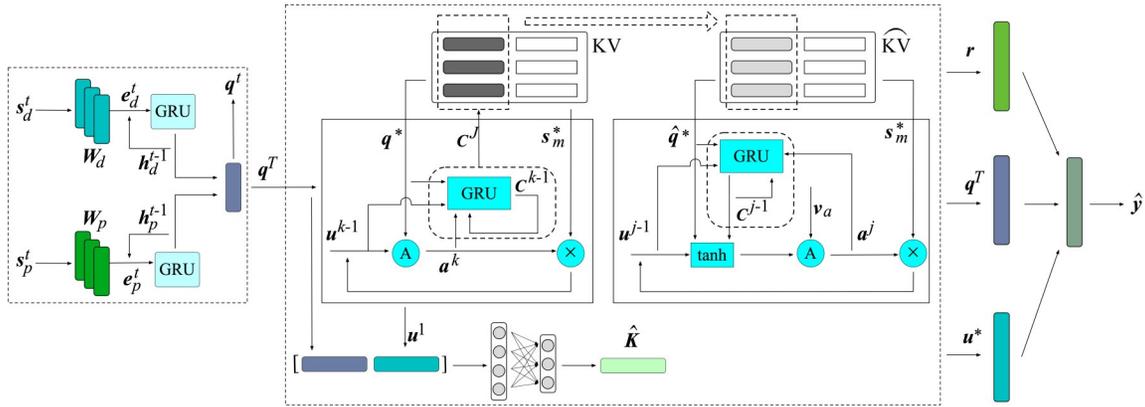


图1 AMHSC模型整体框架

AMHSC模型首先对病人的历史EHRs进行编码,获取病人健康状况表示向量 q^t ,并在此过程中构建记忆神经网络KV.然后,模型针对KV开展基于注意力机制的多跳读取,并对覆盖度进行更新.在进行 J 轮读取后,依据覆盖度 C^J 对KV进行过滤以获取 \widehat{KV} .接着,模型针对 \widehat{KV} 进行基于注意力机制的多跳读取,并在计算注意力权重时将覆盖度作为依据.在依据推荐结果获取多跳读取的最佳次数后,将其作为监督信息,用于训练MemNN读取次数的自适应决定模块,得到最佳读取次数预测结果 \hat{K} ,完成最后的药物推荐.

2.1 基本符号与问题定义

病人对应的EHRs可看作是一个序列 $x=[x^1, x^2, \dots, x^T]$,其中 T 为总入院次数, x^t 为第 t 次入院对应的数据.原始EHRs使用ICD(International Classification of Diseases)代码记录病人的疾病诊断结果和手术治疗方案,使用(National Drug Codes)NDC代码记录病人服用的药物.这些代码在分别经过字典化处理后,均被转化为唯一的整数数值,得到对应的整数集合 S_d, S_p 和 S_m ,其中包含的整数数量分别为 $|S_d|, |S_p|$ 和 $|S_m|$. x^t 中包含了第 t 次入院时的疾病诊断结果、手术治疗方案以及服用的药物,分别对应三个整数集合 $d^t \in S_d, p^t \in S_p$ 和 $m^t \in S_m$.在使用 x 作为输入数据时,模型会将 d^t, p^t 和 m^t 分别转化为multi-hot向量 s_d^t, s_p^t 和 s_m^t .

基于EHRs编码结果构建的MemNN中包含 $T-1$ 对键-值,第 t 对键-值对应第 t 次入院时的数据,分别是由连续型数值构成的病人健康状况表示向量 q^t 和以整数集合形式存在的药物 m^t .使用注意力机制对MemNN进行多跳读取时,模型会计算当前的查询条件 u^{k-1} 与MemNN中各个键对应的注意力权重 a^k ,以此权重为依

据,从MemNN中读取内容 o^k ,用以完成查询条件从 u^{k-1} 到 u^k 的更新.同时,模型会依据读取过程中的各项数据,对MemNN中 q^t 对应的覆盖度信息 C_i^{k-1} 进行更新,得到 C_i^k .

AMHSC模型针对药品的共同使用情况构建关系图,基于药品的临床医学关系获取对应的表示向量.模型使用矩阵 CG 表示关系图, CG 维度为 $|S_m| \times |S_m|$.若某次入院记录对应的EHRs中同时包含药品 i 与 j ,则 $CG_{i,j} = CG_{j,i} = 1$,否则 $CG_{i,j} = CG_{j,i} = 0$.

给定病人对应的EHRs,AMHSC模型的目标是预测病人需要服用的药物 m^T .

2.2 EHRs编码

对EHRs进行编码的目的在于获取病人每次入院对应的健康状况表示向量,作为构建MemNN和药物推荐的基础.首先,模型分别借助嵌入矩阵 $W_d \in \mathbb{R}^{|S_d| \times \dim}$ 和 $W_p \in \mathbb{R}^{|S_p| \times \dim}$,将 s_d^t 与 s_p^t 转换为对应的特征向量,如式(1)和式(2)所示,其中 \dim 表示转化后特征向量的长度.得到的特征向量 e_d^t 与 e_p^t 将作为GRU当前时刻的输入数据,依据式(3)和式(4)进行更新,获取 s_d^t 与 s_p^t 中蕴含的时序模式.二者对应的GRU状态 h_d^t 与 h_p^t 将在被连接为单一特征向量后,经线性转换得到第 t 次入院对应的病人健康状况表示向量 q^t .随后, q^t 和对应的药品集合 m^t 一起组成一对键-值,存储于MemNN中.在完成对所有EHRs的编码后,MemNN中共包含 $T-1$ 对键-值,分别对应第1次至第 $T-1$ 次入院的数据.

$$e_d^t = W_d s_d^t \quad (1)$$

$$e_p^t = W_p s_p^t \quad (2)$$

$$h_d^t = \text{GRU}_d(h_d^{t-1}, e_d^t) \quad (3)$$

$$\mathbf{h}_p^t = \text{GRU}_p(\mathbf{h}_p^{t-1}, \mathbf{e}_p^t) \quad (4)$$

对 MemNN 进行的多跳读取分两阶段进行. 第一阶段, 模型基于常规注意力机制读取数据, 不考虑覆盖度信息对注意力权重计算的影响. 对于第 k 轮读取, 模型首先依据上一轮读取结果 \mathbf{u}^{k-1} 与 MemNN 的键 \mathbf{q}^t ($t < T$) 之间的交互信息, 计算注意力权重 \mathbf{a}^k , 计算过程如式 (5) 和式 (6) 所示. 随后, 模型使用权重 \mathbf{a}^k 从 MemNN 读取药品数据, 并依据各药品对应的表示向量 \mathbf{Z}^k 对读取结果进行转化, 得到本次读取的内容 \mathbf{o}^k , 如式 (7) 所示. 读取结果 \mathbf{o}^k 会被用于对 \mathbf{u}^{k-1} 依照式 (8) 进行更新, 并将读取结果 \mathbf{u}^k 作为下一轮读取的查询依据.

$$\mathbf{a}_t^k = \frac{\exp(\mathbf{e}_t^k)}{\sum_{t'}^{T-1} \exp(\mathbf{e}_{t'}^k)} \quad (5)$$

$$\mathbf{e}_t^k = (\mathbf{u}^{k-1})^T \mathbf{q}^t \quad (6)$$

$$\mathbf{o}^k = \mathbf{a}^k \mathbf{s}_m^* \mathbf{Z}^k \quad (7)$$

$$\mathbf{u}^k = \mathbf{u}^{k-1} + \mathbf{o}^k \quad (8)$$

AMHSC 模型使用 GCN 处理药品关系矩阵 \mathbf{CG} , 获取式 (7) 使用的药品表示向量 \mathbf{Z}^k , 计算过程如式 (9) 所示. 其中, $\hat{\mathbf{D}}$ 为对角矩阵, 计算方式为 $\hat{\mathbf{D}}_{i,i} = \sum_j \mathbf{CG}_{i,j}$, \mathbf{I} 则为单位矩阵. 模型会对 \mathbf{CG} 使用两层 GCN 操作, 所获取的节点特征即为药品的表示向量, 计算过程如式 (10) 所示.

$$\hat{\mathbf{A}} = \hat{\mathbf{D}}^{-\frac{1}{2}} (\mathbf{CG} + \mathbf{I}) \hat{\mathbf{D}}^{-\frac{1}{2}} \quad (9)$$

$$\mathbf{Z} = \hat{\mathbf{A}} \tanh(\hat{\mathbf{A}} \mathbf{W}_e) \mathbf{W}_1 \quad (10)$$

第二阶段 MemNN 多跳读取过程中, 模型使用覆盖度机制对注意力权重的计算进行调整, 将其作为注意力权重的计算依据之一. 具体计算细节将在 2.3 节介绍.

2.3 结合选择性覆盖度机制的 MemNN 多跳读取

AMHSC 模型使用覆盖度信息记录 MemNN 中键-值对被读取的程度, 并基于覆盖度信息完成两项工作: 筛选 MemNN 中的重要数据, 以及对 MemNN 读取过程中的注意力权重进行调整, 确保模型关注所有重要数据.

2.3.1 基于 GRU 的覆盖度信息更新

基于 GRU 对覆盖度信息进行更新的过程如图 2 所示. 对于第 k 轮读取过程, 模型对截至 $k-1$ 轮的覆盖度信息 \mathbf{C}^{k-1} 更新时使用的数据包括 \mathbf{q}^t , \mathbf{u}^{k-1} 以及注意力权重 \mathbf{a}^k . \mathbf{a}^k 是第 k 轮读取时 \mathbf{u}^{k-1} 与 \mathbf{q}^t ($t < T$) 之间的注意力权重, 其计算过程如式 (11) 及式 (12) 所示, 其中 \mathbf{C}_t^{k-1} 表示第 t 对键-值对应的覆盖度信息. 相比于常规注意力权重只考虑 \mathbf{q}^t 与 \mathbf{u}^{k-1} 的做法, 该计算过程通过对 concat 注意力机制的修改^[18], 将覆盖度信息纳入考虑之中.

$$\mathbf{a}_t^k = \frac{\exp(\mathbf{e}_t^k)}{\sum_{t'}^{T-1} \exp(\mathbf{e}_{t'}^k)} \quad (11)$$

$$\mathbf{e}_t^k = \mathbf{v}_c^T \tanh(\mathbf{U}_c \mathbf{u}^{k-1} + \mathbf{V}_c \mathbf{q}^t + \mathbf{W}_c \mathbf{C}_t^{k-1}) \quad (12)$$

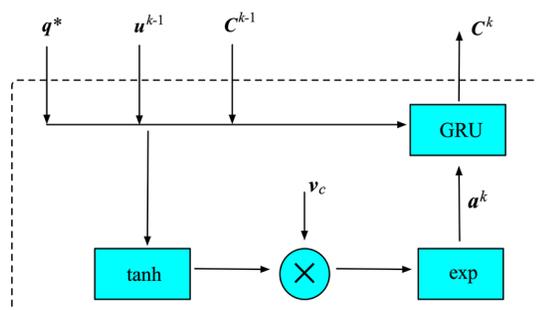


图2 基于 GRU 的覆盖度信息计算流程

将 \mathbf{C}^{k-1} 更新至 \mathbf{C}^k 的过程如式 (13) 至式 (16) 所示, 其中 $\sigma(\cdot)$ 为 sigmoid 函数. 根据 GRU 的计算过程可知, 式 (13) 和式 (14) 首先分别基于 sigmoid 函数构建更新门 (update gate) 和重置门 (reset gate), 用于决定原覆盖度信息 \mathbf{C}^{k-1} 中被保留以及被去除的数据量. 接着, 式 (15) 基于输入数据以及 \mathbf{C}^{k-1} 生成当前时刻的新数据 $\tilde{\mathbf{C}}_t^k$. 最后, 式 (16) 基于更新门结合 \mathbf{C}^{k-1} 与 $\tilde{\mathbf{C}}_t^k$, 生成更新后的覆盖度信息 \mathbf{C}^k .

$$\mathbf{z}_t^k = \sigma(\mathbf{U}^z \mathbf{C}_t^{k-1} + \mathbf{W}_a^z \mathbf{a}_t^k + \mathbf{W}_q^z \mathbf{q}^t + \mathbf{W}_u^z \mathbf{u}^{k-1}) \quad (13)$$

$$\mathbf{r}_t^k = \sigma(\mathbf{U}^r \mathbf{C}_t^{k-1} + \mathbf{W}_a^r \mathbf{a}_t^k + \mathbf{W}_q^r \mathbf{q}^t + \mathbf{W}_u^r \mathbf{u}^{k-1}) \quad (14)$$

$$\tilde{\mathbf{C}}_t^k = \tanh(\mathbf{r}_t^k \circ \mathbf{U} \mathbf{C}_t^{k-1} + \mathbf{W}_a \mathbf{a}_t^k + \mathbf{W}_q \mathbf{q}^t + \mathbf{W}_u \mathbf{u}^{k-1}) \quad (15)$$

$$\mathbf{C}_t^k = \mathbf{z}_t^k \circ \mathbf{C}_t^{k-1} + (1 - \mathbf{z}_t^k) \circ \tilde{\mathbf{C}}_t^k \quad (16)$$

2.3.2 选择性覆盖度机制

AMHSC 模型根据病人历史入院次数决定是否使用选择性覆盖度机制. 当病人入院次数较少时, 对应 EHRs 包含的数据量有限, 模型需要充分利用所有数据, 因此不对 MemNN 进行数据过滤和注意力权重调整, 仅依据常规注意力权重进行 MemNN 多跳读取 (具体见第 2.2 节). 当病人入院次数超过 L 时, 模型启动选择性覆盖度机制, 依据覆盖度信息对存储在 MemNN 中的数据进行过滤, 并对注意力权重进行调整, 避免注意力机制重复关注 MemNN 相同位置的数据.

选择性覆盖度机制工作流程如图 3 所示, 以 MemNN 数据选择为界, 分为两步进行.

首先, 模型依照第 2.2 节的描述, 基于常规注意力机制对 MemNN 进行多跳读取. 在此过程中, 模型会使用 GRU 对键-值对的覆盖度信息 \mathbf{C}_t^k 进行记录和更新, 但不会使用 \mathbf{C}_t^k 调整注意力权重. \mathbf{C}_t^k 记录了 MemNN 中不同键-值对被读取和使用情况, 反应了相应数据对于药物推荐的重要性. 因此, 在对 MemNN 进行 J 轮读

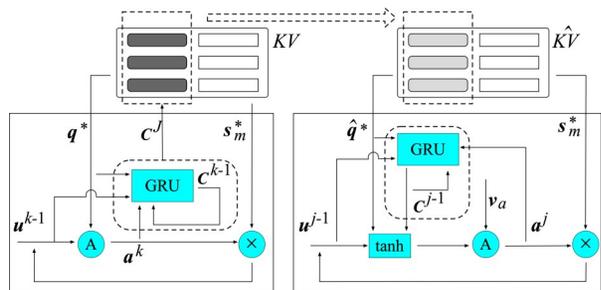


图3 结合选择性覆盖度机制的MemNN多跳读取

取后,模型以 C_i^j 为依据计算 q^j 中各元素对应的门 (element-wise gate), 利用门控机制对 q^j 进行数据筛选, 以达到保留重要数据并去除冗余数据的目的. 计算过程如式(17)所示:

$$\hat{q}^j = \sigma(C_i^j) \circ q^j \quad (17)$$

完成 MemNN 数据筛选后,模型继续基于注意力权重对 MemNN 进行读取,并使用覆盖度信息对注意力权重进行调整,保证其适当地分散于 MemNN 不同位置. 计算过程如式(18)和式(19)所示. 得到的注意力权重 \hat{a}^j 将被用于从过滤后的 MemNN 中读取数据 $\hat{\delta}^j$, $\hat{\delta}^j$ 会被用于将 \hat{u}^{j-1} 更新为 \hat{u}^j , 计算过程如式(20)和式(21)所示:

$$\hat{a}_i^j = \frac{\exp(\hat{e}_i^j)}{\sum_t \exp(\hat{e}_t^j)} \quad (18)$$

$$\hat{e}_i^j = \hat{v}_a^T \tanh(\hat{U}_c u^{j-1} + \hat{V}_c \hat{q}^j + \hat{W}_c C_i^{j-1}) \quad (19)$$

$$\hat{\delta}^j = \hat{a}^j \hat{s}_m^* Z^j \quad (20)$$

$$\hat{u}^j = \hat{u}^{j-1} + \hat{\delta}^j \quad (21)$$

2.4 自适应 MemNN 多跳读取

自适应 MemNN 多跳读取根据病人自身情况,确定对应的 MemNN 读取次数,用以取代常规方法使用单一数值作为所有病人 MemNN 读取次数的做法. 不同病人对应的入院次数、住院时间以及所患疾病种类等数据存在差异,对应的 EHRs 中包含的数据量不尽相同. 根据病人自身 EHRs 的数据量确定 MemNN 读取次数,能够在获取充足数据完善病人健康状况表示向量的同时,避免读取冗余数据.

监督信息的缺失是实现自适应读取的难点,即缺少病人对应的最佳 MemNN 读取次数. 为解决该问题, AMHSC 模型首先使用单一固定数值作为所有病人的 MemNN 读取次数,用以完成数据读取以及药物推荐. 然后,依据每轮 MemNN 读取结果生成药物,评估推荐性能,以最佳性能对应的读取次数作为自适应读取的监督信息,用于模型训练.

2.4.1 监督信息获取

在使用固定数值 K 作为所有病人的 MemNN 读取次数时,模型首先按照 2.2 节介绍的常规读取方式,基

于注意力机制对 MemNN 进行 J 轮读取. 若病人住院次数超过 L 次,模型会使用 2.3 节介绍的选择性覆盖度机制继续对 MemNN 进行读取,直至完成 K 轮读取. 否则,模型将继续使用常规读取模式,完成剩余读取.

在完成 K 轮读取后,模型将得到最终读取结果 u^K . 该读取结果将与病人最后一次入院对应的健康状况表示向量 q^T , 以及 q^T 与药品表示向量 Z^K 之间的关联信息 r 一起,作为生成最终药品推荐结果的依据. 关联信息 r 反映了病人健康状况与药品之间的关联,其计算过程同样依据注意力机制进行. 模型首先依据式(22)和式(23)计算注意力权重 a^r , 然后依照式(24)使用 a^r 整合药品表示向量 Z^K , 得到结果 r .

$$a_i^r = \frac{\exp(e_i^r)}{\sum_{i=1}^{|S_m|} \exp(e_i^r)} \quad (22)$$

$$e_i^r = (q^T)^T W_r Z_i^K \quad (23)$$

$$r = a^r Z^K \quad (24)$$

获取上述 q^T , u^K 以及 r 三项数据后,模型对三者进行连接,并利用全连接层处理,作为 sigmoid 函数的输入数据,用以计算每种药品被推荐的概率,得到概率向量 \hat{y} . 计算过程如式(25)所示,其中 $f(\cdot)$ 表示全连接层, $\sigma(\cdot)$ 为 sigmoid 函数. 若药品 i 对应的概率 \hat{y}_i 大于预先设定的概率阈值 pr , 则该药品会被选取作为推荐药品之一,包含在最终生成的药品集合中,否则该药品不会被推荐.

$$\hat{y} = \sigma\left(f\left([q^T, u^K, r]\right)\right) \quad (25)$$

在以固定数值 K 作为所有病人 MemNN 读取次数完成模型训练后, AMHSC 模型会依据每一轮 MemNN 读取结果生成药物集合,用于与真实药物集合 m^T 进行对比. 对比时使用 F1 作为评价指标,选取对应最高数值的 MemNN 读取次数 K' 作为病人对应的 MemNN 最佳读取次数. 生成药物集合的过程如式(26)所示,其中 u^k 是 MemNN 第 k 轮读取得到的结果, r^k 是 q^T 与第 k 轮读取后的药品表示向量 Z^k 之间的关联信息,计算过程参照式(22)至式(24)进行.

$$\hat{y} = \sigma\left(f\left([q^T, u^k, r^k]\right)\right) \quad (26)$$

2.4.2 自适应 MemNN 读取

为完成自适应 MemNN 读取, AMHSC 模型需要依据病人自身数据,预测对应的 MemNN 最佳读取次数. 模型首先以 q^T 为初始查询条件,进行基于常规注意力机制的 MemNN 读取,得到读取结果 u^1 . u^1 在经过全连接层和 sigmoid 函数处理后,得到概率向量 \hat{K} , 处理过程如式(27)所示. 概率向量 \hat{K} 维度为 K , 包含的元素 \hat{K}_i 表

示 $i+1$ 作为 MemNN 最佳读取次数的概率. 随后, 模型选取 $\hat{\mathbf{K}}$ 中概率最高的数值, 以其对应位置作为 MemNN 自适应读取次数, 完成 MemNN 读取以及药物推荐.

$$\hat{\mathbf{K}} = \sigma(f(\mathbf{u}^1)) \quad (27)$$

2.5 模型训练

AMHSC 训练过程分为两步. 首先, 模型使用单一固定数值作为所有病人的 MemNN 读取次数, 进行药物推荐, 并根据每一轮 MemNN 读取后对应的推荐结果, 确定每一位病人对应的最佳 MemNN 读取次数. 然后, 使用得到的最佳 MemNN 读取次数作为监督信息, 训练模型自适应地决定每一位病人的 MemNN 读取次数, 完成最终的模型训练.

第一步训练使用的损失函数如式(28)所示, 由二元交叉熵损失函数(Binary Cross Entropy Loss) \mathcal{L}_b 以及多标签边际损失函数(Multi-label Margin Loss) \mathcal{L}_m 的加权和组成. \mathcal{L}_b 计算过程如式(29)所示, 其中 \mathbf{y} 即为 \mathbf{s}_m^T , $\hat{\mathbf{y}}$ 则是药品对应的预测概率向量. \mathcal{L}_m 计算过程如式(30)所示, 公式中包含的 $\hat{\mathbf{Y}}$ 是模型预测得到的药品集合. 在使用常规损失函数 \mathcal{L}_b 的基础上使用 \mathcal{L}_m 的原因在于, \mathcal{L}_m 能够扩大真实标签对应的预测概率与错误标签对应的预测概率之间的差距, 提升正确标签对应的预测概率并降低错误标签对应的概率, 从而能使用预先设定的概率阈值为界限, 更容易的区分二者.

$$\mathcal{L} = \alpha_1 \mathcal{L}_b + \alpha_2 \mathcal{L}_m \quad (28)$$

$$\mathcal{L}_b = - \sum_{i=1}^{|\mathcal{S}_m|} \left[y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log (1 - \sigma(\hat{y}_i)) \right] \quad (29)$$

$$\mathcal{L}_m = \sum_{i=1}^{|\mathcal{S}_m|} \sum_{j=1}^{|\hat{\mathbf{Y}}|} \frac{\max(0, 1 - (\hat{\mathbf{y}}[\hat{\mathbf{Y}}_j] - \hat{\mathbf{y}}[i]))}{|\mathcal{S}_m|} \quad (30)$$

第二步训练使用的损失函数如式(31)所示, 其中 $\bar{\mathbf{K}}$ 是由自适应读取监督信息 \mathbf{K} 转化得到的 one-hot 向量, 即 $\bar{\mathbf{K}}_{K-1} = 1$, 其余位置元素数值为零. 该损失函数只会对式(27)描述的全连接层进行训练, 模型其余部分的参数不会收到影响, 仍然为使用固定 MemNN 读取次数 K 时得到的训练结果. 这么做的目的是保证基于固定读取次数获取的监督信息始终有效, 可用于自适应模型的训练.

$$\mathcal{L}_s = - \sum_{i=1}^{|\hat{\mathbf{K}}|} \left[\bar{\mathbf{K}}_i \log \sigma(\hat{\mathbf{K}}_i) + (1 - \bar{\mathbf{K}}_i) \log (1 - \sigma(\hat{\mathbf{K}}_i)) \right] \quad (31)$$

3 实验结果与分析

本文所有实验均基于真实的临床医疗数据集 Medical Information Mart for Intensive Care-III v1.4

(MIMIC-III)^[19]完成. AMHSC 模型使用 PyTorch 编写*, 其训练借助英伟达 TITAN XP GPU 完成.

3.1 数据集

MIMIC-III v1.4 收集了超过四万名病人住院期间的各项临床数据, 从不同角度描述和记录了病人健康状况变化过程, 涉及的数据包括生命指征、诊断结果、服用药物以及手术治疗等. AMHSC 模型选取具有至少两次住院记录的病人, 选取的数据包括病人的疾病诊断结果、接受的手术治疗方案以及服用的药物等. 该过程共计选取了 6 350 名病人的数据, 这些病人被随机划分为训练集、测试集以及验证集, 各集合占比分别为 80%、10% 和 10%, 其中不同性别及年龄段病人统计情况如表 1 所示.

表 1 不同性别及年龄段病人数量统计信息

性别	年龄	训练集	测试集	验证集	合计
女	0~19	21	0	0	21
	20~39	131	17	17	165
	40~59	538	81	73	692
	60~79	972	116	119	1 207
	≥80	566	73	71	710
男	0~19	36	0	0	36
	20~39	145	28	16	189
	40~59	790	107	97	994
	60~79	1 299	138	171	1 608
	≥80	582	75	71	728

3.2 对比方法

选取以下方法与 AMHSC 模型进行对比:

(1) ARMR^[7]: ARMR 使用 MemNN 对历史 EHRs 的编码结果进行存储, 并对 MemNN 进行基于注意力机制的多跳读取, 通过每次读取时注意力权重的变化关注不同数据, 获取 EHRs 中蕴含的上下文信息, 用于药物推荐.

(2) DMNC^[20]: EHRs 中包含多种从不同角度描述病人健康状况的数据, DMNC 模型使用 MemNN 存储不同角度数据之间的交互信息, 并将其用于自身数据的更新, 以此提升药物推荐性能.

(3) GAMENet^[10]: GAMENet 使用 MemNN 存储历史 EHRs 的编码结果, 基于注意力机制从中读取数据, 并基于图数据表示药物之间的相互联系, 使用 GCN 获取其表示向量用于药物推荐.

(4) Leap^[12]: Leap 将药物推荐看作序列决策问题, 基于药物不良反应信息制定奖励函数, 使用强化学习对序列式的药物推荐过程进行修正.

*<https://github.com/yanda-wang/AMHSC>

(5)RETAIN^[21]:RETAIN 基于双层级注意力机制,从历史医疗记录中获取对当前药物推荐具有重大影响的历史数据,用于提升药物推荐性能.

3.3 实验结果

AMHSC 模型及各对比方法的性能,通过以下三项评价指标进行比较:Jaccard 相似度(Jaccard)、Precision Recall AUC(PRAUC)及 F1 值.

3.3.1 性能比较与分析

AMHSC 模型与各对比方法的比较结果如表 2 所示.由表 2 可知,AMHSC 模型性能在各指标上均优于对比方法,其 Jaccard, PRAUC 和 F1 值分别为 0.5080, 0.7704 和 0.6646.在几种不同对比方法中,DMNC 使用了与 AMHSC 相似的数据编码方法,对数据中从不同角度描述病人健康状况的临床记录分别进行建模及结合,但其性能相较于 AMHSC 有一定差距,这说明以键-值对形式的 MemNN 对历史数据建模,有利于提升药物推荐性能.此外,GAMENet 采用了类似的键-值对形式 MemNN 对历史 EHRs 进行存储,并基于注意力机制从中读取数据.但是,对于所有病人,GAMENet 均只进行一次 MemNN 读取,而不是由模型根据不同病人的健康状况,自适应地决定读取次数.这种设定使得 GAMENet 的性能与 AMHSC 相比有一定差距,同时也与进行 MemNN 多跳读取的 ARMR 存在差距. ARMR 模型与 AMHSC 模型均使用键-值对形式的 MemNN 对历史 EHRs 进行建模,并通过 MemNN 进行多跳读取获取数据.但是,ARMR 没有使用覆盖度对注意力权重进行调整,注意力权重有可能重复关注 MemNN 中若干个键-值对,使得所读取数据包含冗余信息.同时,ARMR 针对所有病人均进行相同次数的 MemNN 读取,忽略了不同病人 EHRs 数据量不同,从而需要进行不同次数 MemNN 读取.上述原因导致 ARMR 的性能虽然优于其他对比方法,但与 AMHSC 相比仍然存在差距.

表 2 AMHSC 模型与对比方法性能比较结果

模型名称	Jaccard	PRAUC	F1
ARMR	0.502 6	0.761 3	0.655 9
DMNC	0.493 3	0.726 9	0.651 1
GAMENet	0.497 0	0.758 9	0.651 9
Leap	0.448 4	0.645 7	0.610 9
RETAIN	0.489 7	0.749 9	0.649 4
AMHSC	0.508 0	0.770 4	0.664 6

使用不同性别及年龄段病人所对应数据对 AMHSC 模型的有效性进行测试,结果如表 3 所示.针对模型有效性的测试结果表明,AMHSC 模型对于 MIMIC 数据集中不同群体的病人均可以取得较好的药物推荐效果.其中,60~79 年龄段女性病人对应的 Jaccard, PRAUC 以及 F1 三项指标的数值最高,分别达

到 0.5320, 0.7873 和 0.6878;其次为针对 40~59 年龄段女性病人的测试结果,对应的 Jaccard, PRAUC 以及 F1 达到了 0.5167, 0.7733 和 0.6719.相比之下,应用于 20~39 岁以及 80 岁以上女性病人群体时所取得的效果稍差.模型针对男性病人的药物推荐效果呈现出类似的规律,即针对 40~59 岁以及 60~79 岁病人取得的推荐效果,要优于对 20~39 以及 80 岁以上病人取得的推荐效果.同时,对比不同性别病人的数据可知,模型对于女性病人表现出更加良好的整体推荐性能.

表 3 AMHSC 针对不同性别及年龄段病人的药物推荐性能

性别	年龄	Jaccard	PRAUC	F1
女	20~39	0.501 8	0.752 4	0.656 1
女	40~59	0.516 7	0.773 3	0.671 9
女	60~79	0.532 0	0.787 3	0.687 8
女	≥80	0.488 3	0.755 8	0.649 3
男	20~39	0.487 2	0.765 0	0.650 1
男	40~59	0.509 9	0.765 6	0.665 3
男	60~79	0.501 8	0.770 7	0.658 3
男	≥80	0.496 6	0.764 3	0.653 0

为进一步测试当 AMHSC 被应用于不同类型疾病时的有效性,选取了 9 种人体不同器官或系统特有的疾病,并从 MIMIC 数据集中选取患有相应疾病的病人,使用他们对应的数据对 AMHSC 进行测试,结果如表 4 所示,模型对于 MIMIC 数据集中患有不同类型疾病的病人均可以取得较好的药物推荐效果.模型对绝大多数疾病可以取得 0.67 以上的 F1,其中最高数值 0.6860 出现在败血症病人对应的推荐结果中,糖尿病和脑血管疾病对应的 F1 数值也分别达到了 0.6852 以及 0.6811.此外,针对心脏病、高血压、肺部疾病以及胰腺疾病的 F1 数值也在 0.67 以上,仅有肝脏疾病和胃部疾病的 F1 数值小于 0.66.

表 4 AMHSC 针对患有不同类别疾病病人的药物推荐性能

疾病种类	Jaccard	PRAUC	F1
脑血管疾病	0.525 3	0.786 1	0.681 1
糖尿病	0.529 0	0.783 3	0.685 2
心脏病	0.523 2	0.783 3	0.678 8
高血压	0.516 2	0.776 4	0.672 3
肝脏疾病	0.494 7	0.756 1	0.651 7
肺部疾病	0.517 7	0.779 2	0.674 4
胰腺疾病	0.514 5	0.771 1	0.672 5
败血症	0.529 9	0.794 5	0.686 0
胃部疾病	0.488 6	0.758 1	0.646 9

3.3.2 AMHSC 模型消融实验

为评估 AMHSC 中各组件的必要性,模型被拆解为

不同形式的变型. 每次拆解会去除模型中包含的某一组件, 并保留其余组件. 去除单一组件后, 组件负责的相应功能也随之消除, 无法在药物推荐过程中发挥作用. 因此, 各种变型与 AMHSC 的对比, 能够说明不同组件对于模型的重要性, 以及是否对提升药物推荐性能发挥了积极作用. 同时, 为说明自适应决定 MemNN 读取次数的重要性, 各变型会被进一步去除该功能, 改为对所有病人使用相同的 MemNN 读取次数. 拆解后得到的不同变型具体情况如下.

(1) Single-GRU: AMHSC 模型使用 GRU 对诊断结果以及病人接受的手术治疗手段分别进行编码, 然后对编码结果进行整合, 而 Single-GRU 则将两种数据作为单一序列数据处理.

(2) No-Coverage: 该变型放弃使用覆盖度信息对注意力权重进行调整, 在 MemNN 多跳读取过程中不会对注意力权重施加任何限制条件.

(3) Arbitrary Coverage (Arb-Coverage): AMHSC 模型根据病人历史入院次数决定是否使用覆盖度信息对读取 MemNN 时使用的注意力权重进行调整, 而 Arb-Coverage 忽略对历史入院次数的限制, 处理所有病人数据时均使用覆盖度信息对注意力权重进行调整.

(4) No-Selection: AMHSC 模型会依据覆盖度信息, 对 MemNN 中的数据进行选择, 随后依据覆盖度信息对过滤后数据对应的注意力权重进行调整, 而 No-Selection 变型则舍弃数据过滤这一步骤, 对原始 MemNN 的注意力权重直接进行调整.

(5) No-GCN: AMHSC 模型使用 GCN 获取药品表示向量, 对 MemNN 读取结果进行编码, 而 No-GCN 使用嵌入矩阵替代 GCN, 舍弃药品之间的关系对获取表示向量的影响.

(6) No-Adaptability (No-Adap): AMHSC 模型依据病人自身数据, 自适应地决定 MemNN 读取次数, 而 No-Adap 不具备自适应 MemNN 读取功能, 使用单一固定数值作为所有病人的读取次数.

上述变型与 AMHSC 的比较结果如表 5 所示.

实验结果显示, AMHSC 能够取得最佳的药物推荐效果, 对应的 Jaccard, PRAUC 以及 F1 指标数值分别为 0.508 0, 0.770 4 和 0.664 6, 证明了模型各组件对于药物推荐的必要性. 具体来说, No-Coverage 缺少覆盖度信息对注意力权重进行调整, 在 Jaccard 和 F1 两项指标上

表 5 AMHSC 模型与不同变型性能比较结果

模型名称	Jaccard	PRAUC	F1
Single-GRU	0.491 8	0.743 4	0.650 3
No-Coverage	0.490 8	0.763 4	0.648 9
Arb-Coverage	0.491 1	0.747 4	0.649 4
No-Selection	0.497 2	0.746 5	0.655 8
No-GCN	0.494 5	0.757 0	0.652 7
No-Adap	0.504 7	0.770 5	0.661 8
AMHSC	0.508 0	0.770 4	0.664 6

取得的数值最低, 分别为 0.490 8 和 0.648 9. 与之相对的 Arb-Coverage 则是对所有病人均使用覆盖度机制, 忽略了病人自身数据量对使用覆盖度机制的影响, 得到的性能同样欠佳. 上述结果说明覆盖度机制对 MemNN 读取效果影响显著, 但需要具备足够数据量时才能够发挥预期作用. 合理使用覆盖度机制的 Single-GRU 相较于上述两种变型在 Jaccard 和 F1 两项指标上有所提升, 但仍然与 AMHSC 有一定差距. 这说明 EHRs 中包含的数据从不同角度对病人健康状况变化过程进行了记录, 相关时序模式应当被分别处理后再予以合并. No-Adap 在各项指标上的表现与 AMHSC 最为接近, 是除 AMHSC 模型外推荐性能最佳的变型. 但是, 使用单一固定数值作为所有病人 MemNN 读取次数的做法, 令其表现没有能超越 AMHSC. 其余两种变型 No-Selection 与 No-GCN 的表现比较相似, 各项指标对应数据均不如 AMHSC. 上述结果验证了 AMHSC 模型中各组件对提升模型性能的必要性, 同时也说明根据病人数据自适应决定 MemNN 读取次数有助于获取更佳的药物推荐效果.

3.3.3 MemNN 读取次数对推荐性能的影响

AMHSC 模型自适应决定的 MemNN 读取次数统计情况如表 6 所示, 表中汇总了不同 MemNN 读取次数对应的病例数目. 由该表可知, 对 MemNN 进行单次读取的情况占据绝大多数, 对应病例数目为 3 251, 其次是进行两次读取的情况, 对应病例数目为 1 180. 剩余读取次数对应的病例数目呈现出一定起伏, 但对病例数目相对较少, 均没有超过 500. 该统计结果显示, 由于自身数据量的差异, 病人对应的最佳 MemNN 读取次数并不相同, 应当根据病人具体情况, 自适应地决定 MemNN 读取次数.

表 6 最佳 MemNN 读取次数对应病例数目

读取次数	1	2	3	4	5	6	7	8	9	10	11	12	13	14
病例数	3 251	1 180	483	326	251	217	204	192	208	184	191	193	177	186

若对所有病人均使用同一数值作为 MemNN 读取次数, 则模型 No-Adap 推荐性能随数值变化的情况如

表 7 所示. 表 7 包含了 MemNN 读取次数由 1 次至 14 次时, No-Adap 模型在各项性能评价指标上的表现. 与

表 7 不同 MemNN 读取次数下 No-Adap 对应推荐性能

读取次数	Jaccard	PRAUC	F1
1	0.506 9	0.770 3	0.663 7
2	0.507 7	0.770 2	0.664 4
3	0.507 8	0.770 2	0.664 4
4	0.507 6	0.770 4	0.664 3
5	0.507 8	0.770 4	0.664 4
6	0.507 6	0.770 4	0.664 2
7	0.503 7	0.770 3	0.664 0
8	0.506 7	0.770 2	0.663 4
9	0.506 2	0.770 2	0.663 1
10	0.506 0	0.770 1	0.662 9
11	0.505 6	0.770 0	0.662 5
12	0.505 4	0.770 0	0.662 3
13	0.505 1	0.769 8	0.662 1
14	0.504 7	0.770 5	0.661 8

AMHSC 的性能对比可知,采用自适应方法的 AMHSC 模型能够取得最佳的药物推荐结果.此外,随着统一的固定 MemNN 读取次数的增加, No-Adap 模型性能呈现先上升后下降的趋势.在 MemNN 读取次数为 5 时,其取得的效果相对最佳.随后 No-Adap 模型性能整体呈现出下降趋势,并在 MemNN 读取次数为 14 时,表现相对最差.这种对比和变化说明,根据病人自身特定情况,自适应决定 MemNN 读取次数,能够从 MemNN 中获取最佳推理信息,用于完善病人健康状况表示向量,提升药物推荐性能.

4 结论

本文针对药物推荐问题,提出了选择性覆盖度机制与自适应 MemNN 读取机制,在此基础上构建了一种新颖的药物推荐模型 AMHSC. AMHSC 模型使用 GRU 对 MemNN 所存储数据对应的覆盖度信息进行记录和更新,作为选取 MemNN 重要数据的依据,并利用覆盖度信息对 MemNN 读取过程中的注意力权重进行调整,保证其适当地分散于 MemNN 中的重要数据范围内,避免了常规注意力机制重复关注 MemNN 相同区域的问题.此外, AMHSC 模型根据每个病人自身数据的不同情况,自适应地决定病人对应的 MemNN 读取次数,保证模型在充分利用重要数据的同时,避免从 MemNN 读取冗余数据.基于真实临床数据集 MIMIC- III 对模型的有效性进行测试,结果显示, AMHSC 在各项指标上的表现均优于对比方法.同时,消融实验证明模型各组件发挥了预期作用,保证了数据读取和推理过程的有效性和高效性.

AMHSC 模型对于疾病与药物之间对应关系的利用,以及对病人健康状况和用药过程个性化差异的考虑仍然存在不足,未来仍需要进一步研究.利用疾病与药物之间的对应关系直接生成备选方案,能够提升药物推荐效率与准确性.该过程需要考虑如何获取药物与疾病之间的对应关系,以及如何有效利用该对应关系.同时,由于病理的复杂性以及病人自身独特的生理特性,疾病症状的表现往往复杂多变,对应的药物使用情况也呈现出高度的个性化差异.对这种个性化差异的考虑能够进一步提升推荐模型性能.

未来研究中,计划以医学知识数据库中疾病与药物的映射关系以及药物不良反应相关知识等为依据,构建疾病与药物之间的对应关系.同时,尝试构建图形式的数据,以节点表示疾病或者药物,以节点之间的边表示对应关系,使用图神经网络或知识图谱获取节点表示向量,为模型提供疾病与药物对应关系的相关知识.针对病人的个性化差异,计划在依据病人所患疾病对病人健康状况进行编码的同时,使用病人对应的生理监测数据以及各项化验指标作为模型输入,例如心率、血氧含量和血液白细胞数目等,从而使模型能够以病人健康状况的变化为依据,考虑个体之间的差异性和疾病所表现症状的复杂性.

参考文献

- [1] YAO Z, WANG Y, LONG M, et al. Unsupervised transfer learning for spatiotemporal predictive networks[C]//Proceedings of the 37th International Conference on Machine Learning. Virtual Event: PMLR, 2020: 10778-10788.
 - [2] WANG Y, HU M, ZHOU Y, et al. Unobtrusive and automatic classification of multiple people's abnormal respiratory patterns in real time using deep neural network and depth camera[J]. IEEE Internet of Things Journal, 2020, 7(9): 8559-8571.
 - [3] FANG Z, KUANG K, LIN Y, et al. Concept-based explanation for fine-grained images and its application in infectious keratitis classification[C]//Proceedings of the 28th ACM International Conference on Multimedia. Seattle: Association for Computing Machinery, 2020: 700-708.
 - [4] HAN B, HAN Y, GAO X, et al. Boundary constraint factor embedded localizing active contour model for medical image segmentation[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10(10): 3853-3862.
 - [5] 李居朋, 王颖慧, 李刚. 医学图像关键点检测深度学习方法研究与挑战[J]. 电子学报, 2022, 50(1): 226-237.
- LI J P, WANG Y H, LI G. Research and challenges of

- medical image landmark detection based on deep learning [J]. *Acta Electronica Sinica*, 2022, 50(1): 226-237. (in Chinese)
- [6] YAN L, ZHANG H T, GONCALVES J, et al. An interpretable mortality prediction model for COVID-19 patients[J]. *Nature Machine Intelligence*, 2020, 2(5): 283-288.
- [7] WANG Y, CHEN W, PI D, et al. Adversarially regularized medication recommendation model with multi-hop memory network[J]. *Knowledge and Information Systems*, 2021, 63(1): 125-142.
- [8] 张小恒, 张馨月, 李勇明, 等. 面向帕金森病语音诊断的非监督两步式卷积稀疏迁移学习算法[J]. *电子学报*, 2022, 50(1): 177-184.
ZHANG X H, ZHANG X Y, LI Y M, et al. An unsupervised two-step convolution sparse transfer learning algorithm for Parkinson's disease speech diagnosis[J]. *Acta Electronica Sinica*, 2022, 50(1): 177-184. (in Chinese)
- [9] 赖小波, 许茂盛, 徐小媚. 多分类 CNN 的胶质母细胞瘤多模态 MR 图像分割[J]. *电子学报*, 2019, 47(8): 1738-1747.
LAI X B, XU M S, XU X M. Glioblastoma multiforme multi-modal MR image segmentation using multi-class CNN. *Acta Electronica Sinica*, 2019, 47(8): 1738-1747. (in Chinese)
- [10] SHANG J, XIAO C, MA T, et al. Gamenet: graph augmented memory networks for recommending medication combination[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii: AAAI Press, 2019: 1126-1133.
- [11] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//*5th International Conference on Learning Representations, Conference Track Proceedings*. Toulon: OpenReview.net, 2017: 1-14.
- [12] ZHANG Y, CHEN R, TANG J, et al. Leap: learning to prescribe effective and safe treatment combinations for multimorbidity[C]//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 2017: 1315-1324.
- [13] BICA I, ALAA A M, LAMBERT C, et al. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges[J]. *Clinical Pharmacology & Therapeutics*, 2021, 109(1): 87-100.
- [14] RICHENS J G, LEE C M, JOHRI S. Improving the accuracy of medical diagnosis with causal machine learning [J]. *Nature Communications*, 2020, 11(1): 3923.
- [15] BICA I, ALAA A M, JORDON J, et al. Estimating counterfactual treatment outcomes over time through adversarially balanced representations[C]//*8th International Conference on Learning Representations*. Addis Ababa: OpenReview.net, 2020: 1-13.
- [16] MANSOURNIA M A, DANAEI G, FOROUZANFAR M H, et al. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: analysis with marginal structural models[J]. *Epidemiology*, 2012, 23(4): 631-640.
- [17] MI H, SANKARAN B, WANG Z, et al. Coverage embedding models for neural machine translation[C]//*Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin: Association for Computational Linguistics, 2016: 955-960.
- [18] LUONG T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [C]//*Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon: Association for Computational Linguistics, 2015: 1412-1421.
- [19] JOHNSON A E W, POLLARD T J, SHEN L, et al. Mimic-iii, a freely accessible critical care database[J]. *Scientific Data*, 2016, 3(1): 160035.
- [20] LE H, TRAN T, VENKATESH S. Dual memory neural computer for asynchronous two-view sequential learning [C]//*Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: Association for Computing Machinery, 2018: 1637-1645.
- [21] CHOI E, BAHADORI M T, KULAS J A, et al. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc, 2016: 3512-3520.

作者简介



王延达 男, 1991年出生, 山东枣庄人. 南京航空航天大学计算机科学与技术学院博士研究生. 主要研究方向为数据挖掘及医疗数据分析.

E-mail: yandawang@nuaa.edu.cn



陈炜通 男,1985年出生,广东深圳人. 昆士兰大学信息技术与电子工程学院讲师. 主要研究方向为数据挖掘、机器学习、医疗数据分析.

E-mail: w.chen9@uq.edu.au



皮德常(通讯作者) 男,1971年出生,河南周口人. 南京航空航天大学计算机科学与技术学院教授,博士生导师. 主要研究方向为数据挖掘、大数据处理与可靠性分析.

E-mail: dc.pi@nuaa.edu.cn



岳琳 女,1985年出生,吉林长春人. 昆士兰大学信息技术与电子工程学院助理讲师. 主要研究方向为数据挖掘、机器学习、医疗数据分析.

E-mail: l.yue@uq.edu.au