arXiv:2304.01242v1 [cs.CL] 3 Apr 2023

基于证据图多通道异质学习增强临床证据推荐

罗茂林和张翔

摘要—**临床证据涵盖患者、干预措施（如药物或物理治疗）、问题与结局之间的关联及其影响。临床证据推荐旨在为医疗从业者提供相关信息，支持其决策过程并生成新的证据。我们的具体任务聚焦于基于临床问题推荐证据。然而，某些临床问题与相关证据之间的直接连接往往较为稀疏，造成了链接稀疏性问题。此外，为了推荐适当的证据，必须联合利用证据之间的拓扑关系以及描述这些证据的文本信息。为应对这些挑战，我们定义了两个知识图谱：证据共指图和证据文本图，分别表示证据元素之间的拓扑关系和语言关系。我们还提出了一种多通道异质学习模型和一种融合注意力机制，以处理证据推荐中的共指—文本异质性。实验结果表明，我们的模型在公开数据集上的表现优于最先进的方法。**

索引词—**图注意力网络，推荐系统，循证医学。**

引言

循证医学（EBM）是指在患者护理决策中运用当前最佳证据 [1]。随机对照试验和荟萃分析是循证医学的主要工具，可提供群体结局方面的证据 [2]。将外部临床证据纳入临床决策，称为循证医学实践，有助于临床医生做出更明智的决策 [1], [3], [4]。尽管有诸多益处，但从海量信息中发现潜在证据对临床医生而言颇具挑战，而构建临床证据也是一个耗时的过程。

目前，研究主要聚焦于协助临床医生做出药物推荐决策。然而，这些方法仅根据疾病推荐药物名称，例如 SafeDrug、 [5] MedRec [6]和 DRMP [7]，，这可能不足以说服临床医生并指导其决策过程，因为他们需要更明确的临床证据。

在本研究中，我们利用 ClinicalTrials.gov 公共数据集中的临床证据[1]，根据其与患者临床问题的潜在相关性提供推荐。每项证据均使用三个要素进行描述：患者/问题（P）、干预措施/对照（I/C）和结局（O），符合 PICO 词汇体系[2]。例如，在图 1 中，我们考察了两项针对 COVID-19 问题的研究（NCT04365153 和 NCT04421404）。研究表明，卡那单抗注射液 600mg 或 COVID-19康复者血浆治疗可能预防进行性心脏和呼吸衰竭，或减缓 COVID-19 患者向重度低氧血症进展。另有一项研究（NCT04402970）并非专门针对 COVID-19，而是针对一种由 SARS-CoV-2 引起的高度相关综合征。该研究表明，吸入/雾化多纳酶α可能改善 SARS-CoV-2 相关综合征的临床结局指标。尽管最后这项证据与 COVID-19 并无直接相关性，但由于它与另外两项研究存在连接，因此对于治疗 COVID-19 感染仍可能具有价值。这些连接可能包括语言相似性，或共同指向相同 PICO要素，使其成为推荐的良好候选证据。

我们的证据推荐任务面临着巨大挑战。**信息稀疏性**是许多推荐任务中的普遍问题，包括电影、书籍、音乐、[8], 体育锻炼 [9], 和药物 [5], [6]推荐。为应对这一挑战，我们提出使用证据共指图（ECG）对共享PICO要素的两项临床研究之间的结构连接进行建模。然而，ECG的稀疏性很高，图密度仅为0.00156。为缓解这一问题，我们还引入证据文本图（ETG），对研究之间的语言连接进行建模。除了信息稀疏性这一挑战外，在结合结构信息和语言信息，为给定的临床状况推荐相关证据时，我们还需要处理**信息异质性**问题。为应对这一挑战，我们提出多通道异构注意力网络（MHAN），通过独立的图通道，对来自两个证据图的异构特征进行建模。我们在MHAN中引入多头融合注意力，以有效融合证据实体之间的异构相关性。最后，MHAN为给定问题中的每条证据分配相关性得分，并据此推荐排名前k 的研究。本文的主要贡献如下：

- 我们使用公开数据构建了两个临床证据图。通过

[1]https://clinicaltrials.gov/ct2/home

[2]https://linkeddata.cochrane.org/pico-ontology

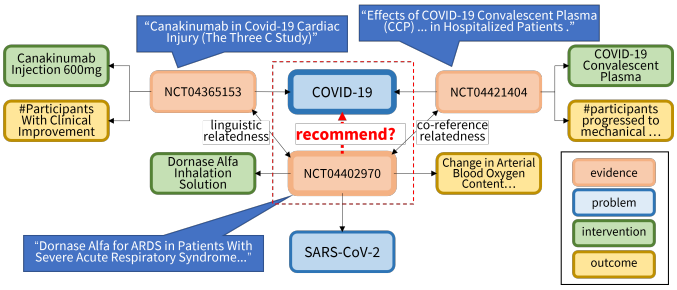


图 1. 临床证据推荐示例。利用可用的拓扑信息和文本信息，我们根据拓扑或文本信息方面与先前证据的潜在连接，推荐可能与 COVID-19 问题相关的临床证据。

补充临床研究之间稀疏的结构相关性与丰富的语言相关性，解决图稀疏性问题。

- 我们提出一种多头融合注意力机制，以有效融合不同图中证据实体的异质特征。
- 我们提出 MHAN 模型来学习证据与问题之间的相关性，并将相关性用作 top-k 证据推荐的指标。实验表明，与其他强模型相比，我们的模型取得了最先进的性能。

MHAN 的数据集和源代码已公开提供，网址为 https://github.com/wds-seu/MHAN。

II. 相关工作

据我们所知，此前尚无研究探讨临床证据推荐。尽管如此，针对电影、书籍、音乐和药物的推荐系统，已有大量研究。

在基于知识的推荐中，协同过滤广泛应用于推荐任务 [16]。LightGCN [17] 旨在聚合高跳邻居。ECFKG [18] 将 TransE [19] 应用于一个包含用户、物品、实体和关系的统一图，将推荐任务转化为对 $<u,\ interact,\ i>$ 三元组合理性的预测。KGAT [8]使用用户、物品、实体和关系的图嵌入。尽管该模型考虑了外部信息，并利用注意力机制取得更好的结果，但它没有考虑图的异质性。基于协同过滤的推荐模型（如 SimGCL [20], ）旨在学习更均匀的用户物品表示，从而能够隐式缓解流行度偏差。

知识图谱嵌入旨在将知识图谱中的实体和关系映射到低维空间。图神经网络（GNN），例如 GCN [10]和 GAT [12], ，利用对邻近节点的信息聚合来学习图表示。Wang 等人 [13]提出了 AM-GCN，以提升 GCN 的嵌入能力。这是因为 GCN 可能难以学习节点特征与拓扑结构之间复杂的关联信息。然而，在实际应用中，知识图谱可能具有异质性。RGCN [11] 被用作 MedRec [6]中的图嵌入模型，该模型在药物推荐方面实现了最先进的性能。HGT [14] 使用依赖节点类型和边类型的参数，根据元关系刻画每条边上的异质注意力，从而在知识嵌入任务中达到最先进水平。

当前基于知识的推荐研究深受图稀疏性问题困扰，这严重降低了为用户识别有价值物品的效果。与此同时，异构图神经网络因能够融合结构化信息和非结构化信息而日益流行。然而，异构网络在推荐系统中的应用尚未得到广泛研究。

III. 方法

A. 问题陈述

我们首先对临床证据进行形式化定义，并明确临床证据推荐任务。

1) 临床证据：“临床证据”是指一组摘要，其中介绍了医疗保健领域最有效的循证实践。为构建该集合，我们从 ClinicalTrials.gov 中提取所有相关证据，并将其汇编为一个名为 E 的集合。每条证据 $e \in E$ 都包含文字描述以及三个要素：所要解决的临床问题（p）、为治疗或预防该问题而采取的措施（t），以及临床试验中对患者结局的测量（o）。为涵盖临床试验中的比较（通常涉及一种或多种干预措施），我们将 PICO 词汇体系中的“比较”要素与干预措施合并。ClinicalTrials.gov 中的所有问题、干预措施和结局均分别归入不同集合： P 表示问题， T 表示干预措施， O 表示结局。

2) 临床证据推荐：推荐任务的目标是针对每个临床问题 p，计算每个 $e \in E$ 的得分，记为 score(p, e)。临床证据推荐任务需要针对给定的临床问题 p，找出得分 score(p, e_i) 最高的前 k 个实体（e_1, e_2, ..., e_k）。这些排名靠前的实体被视为与当前关于 p 的研究高度相关，因此对关注 p 的医疗从业者很有帮助。

B. 证据图的构建

为涵盖临床证据中的结构关联和语言关联，我们构建了两种证据图：证据共指图（ECG），以临床研究之间共有的PICO要素为核心；证据文本图（ETG），以这些研究之间的文本相似性为核心。

1) 证据共指图：证据共指图 $\mathcal{G}_{\mathcal{C}} = (\mathcal{U}_{\mathcal{C}}, \mathcal{R}_{\mathcal{C}}, \gamma, \phi)$ 是直接根据证据集构建的异构图。$\mathcal{U}_{\mathcal{C}}$ 是节点集，由证据集 E 以及 P、T 和 O 所指向的集合组成。$\mathcal{R}_{\mathcal{C}}$ 是边集，其中每条边 $r \in \mathcal{R}_{\mathcal{C}} = (u_i, u_j)$ 都将一个 $u_i \in E$ 连接到一个证据元素 $u_j \in P \cup T \cup O$。γ和 ϕ是两个映射函数，其中 $\gamma(u)$ 将节点 u 映射到其特定类型的 E、P、T和 O，并

$\phi(u_i, u_j)$ 将一条边映射到其可能的类型“*E-P*”、“*E-T*”或“*E-O*”。

2）证据文本图：证据文本图 $\mathcal{G}_\mathcal{T}$ =($\mathcal{U}_\mathcal{T}, \mathcal{R}_\mathcal{T}$) 是一个同质图。节点集 $\mathcal{U}_\mathcal{T}$包含 E中的所有证据。给定一个 $r \in \mathcal{R}_\mathcal{T} = (u_i, u_j)$，这表示证据 u_i 和 u_j 的描述之间具有较高的文本相似度。我们根据两个证据节点的 BERT [21] 嵌入之间的欧几里得距离计算其相似度，以反映它们的潜在语义相关性。

对于每个 $u \in \mathcal{U}_\mathcal{T}$，其文本描述都使用 BERT [21]编码为一个 768 维向量 $H[u]$。给定 $u_i, u_j \in \mathcal{U}_\mathcal{T}$，它们的嵌入 $H[u_i]$ 和$H[u_j]$ 之间的欧几里得距离记为 $D(H[u_i], H[u_j])$，其文本相似度η_{ij} 按公式 1 计算。我们设置相似度阈值 η_{thr} 以过滤掉无关紧要的语言连接：$(u_i, u_j) \in \mathcal{R}_\mathcal{T} \rightarrow \eta_{ij} \geq \eta_{thr}$

$$\eta_{ij} = 1 - \frac{D(H[u_i], H[u_j])}{\max(D(H[u])) - \min(D(H[u]))} \tag{1}$$

C. 证据图的嵌入

1) 证据共指图嵌入：我们使用 BERT预训练模型生成一个向量 $H^{(0)}[u]$，用于表示每个节点的文本描述 $u \in \mathcal{U}_\mathcal{C}$。该向量作为 u 的初始嵌入。为了为不同类型的节点和边保留专门的表示，我们将异质图变换器（HGT） [14] 引入模型中。我们采用迭代传播方法来更新节点表示，如下所示：

$$A_C = \underset{\forall u_i \in \mathcal{N}(u_j)}{\text{Softmax}} \left(\left(K(u_i, \gamma(u_i)) W^{ATT}_{\phi(e)} Q(u_j, \gamma(u_j)) \right) \cdot \frac{\mu}{\sqrt{d}} \right) \tag{2}$$

$$\widehat{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[u_j] = \underset{\forall u_i \in \mathcal{N}(u_j)}{\oplus} \left(A_C \cdot V(u_i, \gamma(u_i)) W^{MSG}_{\phi(e)} \right) \tag{3}$$

$$\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[u_j] = \sigma \left(\text{Linear}_{\gamma(u_j)} \widehat{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[u_j] \right) + H^{(l-1)}[u_j] \tag{4}$$

线性投影 $K(\cdot)$ $Q(\cdot)$ $V(\cdot)$由节点嵌入 $H^{(l-1)}$推导而来，并由节点类型 $\gamma(u)$索引。对于每种边类型 $\phi(r)$，矩阵 $W^{ATT}_{\phi\ r\ ()}$和$W^{MSG}_{\phi\ r\ ()}$都是独特的基于边的矩阵。张量 μ 表示关系三元组 $< \gamma(u_i)\ \phi(r)\ \gamma(u_j) >$，而 d 表示向量 $H[u]$ 的维度。通过应用 Softmax，计算出 $\mathcal{G}_\mathcal{C}$的节点注意力 A_C，随后进行消息传递，将目标节点的嵌入从 $H^{(l-1)}_{\mathcal{G}_\mathcal{C}}[u_j]$ 更新为 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[u_j]$。$\mathcal{N}(u_j)$表示所有指向目标节点 u_j的邻居节点。

2）证据文本图嵌入：证据文本图 $\mathcal{G}_\mathcal{T}$ 被表示为一个同质图，其中仅包含证据节点 $u \in E$。当节点 u_i 和 u_j 之间的相似度 η_{ij} 大于相似度阈值 η_{thr}时，两者之间便存在一条边。我们使用与各节点文本描述对应的 BERT 编码向量$H^{(0)}[u]$ 初始化其向量。为了学习每条证据基于其文本描述和语言关系的表示，我们将 GAT [12] 作为嵌入层的一部分，以迭代方式进行处理：

$$A_T = \underset{\forall e_i \in N(e_j)}{\text{Softmax}} \left(\vec{a} \left(W H^{(l-1)}[e_j] \parallel W H^{(l-1)}[e_i] \right) \right) \tag{5}$$

$$\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_j] = \sigma \left(\text{Mean} \left(A_T \cdot W H^{(l-1)}[e_i] \right) \right) \tag{6}$$

为了获得足够的文本表达能力，我们将 $H^{(l-1)}[E]$ 投影到适当层级的特征空间。随后，将源节点特征与目标节点特征拼接起来。A_T 通过 Softmax 计算得到节点注意力。我们将消息从源节点传递到目标节点，以更新其嵌入，其中 $\sigma(x) = \frac{1}{\exp(-x)+1}$ 为 ~~Sigmoid~~ 函数。

3）注意力机制：节点 u 尽管我们已经获得了证据图的 ECG 和 ETG 嵌入，但节点 E 在 $\mathcal{G}_\mathcal{C}$ 和 $\mathcal{G}_\mathcal{T}$中分别具有不同的嵌入 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[E]$ 和 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[E]$。我们需要融合两个知识图谱中节点 E 的信息，以形成统一实体表示 $H^{(l)}[E]$。

为了合并同一实体在多个图上的不同嵌入，AM-GCN [13] 引入了自适应注意力机制，该机制在半监督分类任务中取得了令人瞩目的成果。自适应注意力以自适应的方式学习各嵌入的重要性权重。融合节点 e_i 信息的公式如下：

$$A_{ad} = \text{Softmax} \left(\text{MLP} \left(\left[\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i] \right] \right) \right) \tag{7}$$

$$H^{(l)}[e_i] = A_{ad} \left[\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i] \right] \tag{8}$$

节点 e_i 在 $\mathcal{G}_\mathcal{C}$ 和 $\mathcal{G}_\mathcal{T}$中的嵌入分别为 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i] \in \mathbb{R}^d$ 和 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i] \in \mathbb{R}^d$，且维度相同。我们将它们堆叠在一起，形成$\left[\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i] \right] \in \mathbb{R}^{2\times d}$。通过一个激活函数为 $tanh$ () 的两层多层感知机，将堆叠后的矩阵投影为矩阵 $\in \mathbb{R}^{2\times 1}$。随后，通过计算 softmax函数，得到一组注意力系数 A_{ad}，用于 e_i 在 $\mathcal{G}_\mathcal{C}$ 和 $\mathcal{G}_\mathcal{T}$上的注意力。分别将注意力系数应用于 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i] \in \mathbb{R}^d$ 和 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i] \in \mathbb{R}^d$ 后，求和得到 e_i 表示的统一向量 $H^{(l)}[e_i] \in \mathbb{R}^d$。

除了以往研究提出的节点融合方法外，评估这些方法在推荐系统下游任务中的有效性也很重要。人们采用了多种融合方法（如拼接、求和和平均）来结合两个节点的信息。为保留节点的语义信息，我们采用基于注意力的方法，对嵌入 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{C}}[e_i]$ 和 $\widetilde{H}^{(l)}_{\mathcal{G}_\mathcal{T}}[e_i]$ 进行加权求和。为确保融合过程中每个图的重要性，我们提出了一个取值受限于 0 到 1 之间的共享矩阵 W。对于每个节点 e_i，其在两个图中的嵌入权重分别记为 W_i 和 $1 - W_i$。节点的融合表示

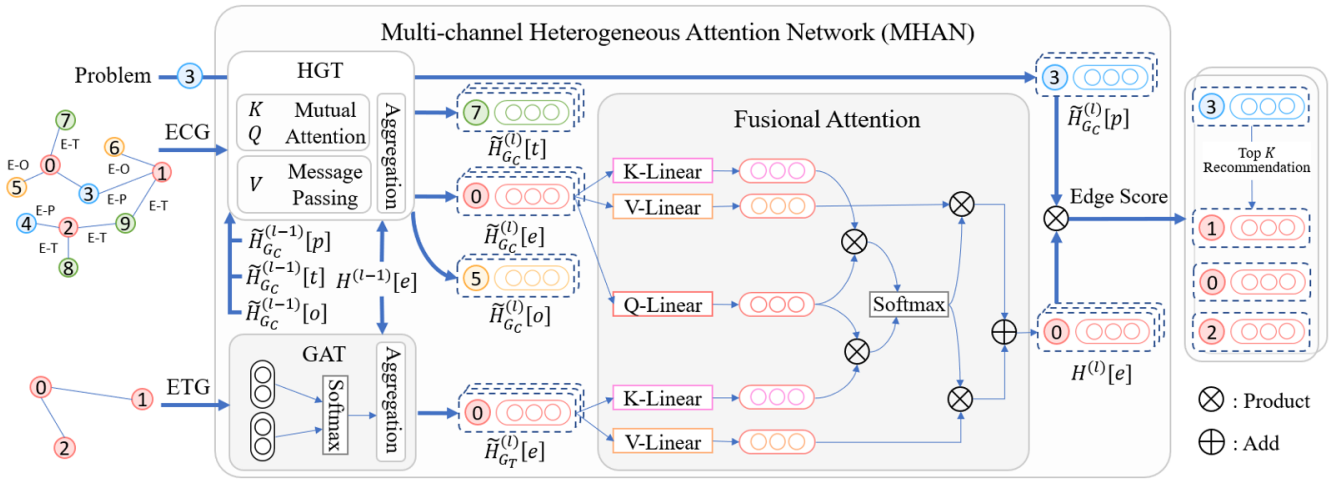


图 2. 多通道异构注意力网络（MHAN）的架构。给定一个感兴趣的问题 p，ECG $\mathcal{G}_C = (\mathcal{U}_C, \mathcal{R}_C, \gamma, \phi)$、ETG $\mathcal{G}_T = (\mathcal{U}_T, \mathcal{R}_T)$ 以及两个图中各实体的初始嵌入 $H^{(0)}$ 作为输入，MHAN 模型将学习 ECG 和 ETG 中每个实体的融合拓扑与语言表示 $H^{(l)}$。MHAN 包含三个组件：(1) ECG 嵌入组件 HGT；(2) ETG 嵌入组件 GAT；(3) 融合注意力。模型会为每个问题/证据对计算相关性排序分数，并将其用于下游推荐任务。在此示例中，推荐证据 0 和 1，是因为 ECG 中明确指出它们与问题 3 相关；同时也推荐证据 2，因为它与证据 1 共享一项干预措施，因此与感兴趣的问题存在隐含关联。

e_i 随后通过对共享矩阵 W 与拼接后的嵌入取哈达玛积得到。

$$H^{(l)}[e_i] = W[i] \odot \widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i] + (1 - W[i]) \odot \widetilde{H}^{(l)}_{\mathcal{G}_T}[e_i] \quad (9)$$

与依赖节点嵌入的注意力系数 a_{ad}相比，权重系数矩阵 W 使模型在学习每个节点嵌入的重要性时具有更大的灵活性。我们的实验表明，在推荐系统任务中，该方法优于自适应注意力。

然而，丢弃节点信息来计算注意力系数可能会导致过拟合问题。因此，我们仍倾向于基于节点嵌入计算注意力系数。

在 ECG 中，节点 e 从其邻居节点 p收集信息；而在 ETG 中，节点 e 仅从其邻居节点 e收集信息。对于根据一个 p 节点推荐一组 e 节点的推荐系统任务，我们提出了一种融合注意力方法。我们使用 p 作为查询来获取注意力系数。多头融合注意力机制的公式如下：

$$H^{(l)}[e_i] = \Big\|_{n=1}^{N} \text{Softmax}\left(\frac{Q^n(e_i)K^n(e_i)^T}{\sqrt{d/N}}\right) V^n(e_i) \quad (10)$$

$$Q^n(e_i) = \text{Q-Linear}^n\left(\widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i]\right)$$
$$K^n(e_i) = \text{K-Linear}^n\left(\left[\widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_T}[e_i]\right]\right)$$
$$V^n(e_i) = \text{V-Linear}^n\left(\left[\widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_T}[e_i]\right]\right)$$

我们通过线性投影 Q-Linear^n 将节点 e_i 的嵌入 $\widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i]$ 投影为第 N- 个 $Query$ 向量 $Q^n(e_i)$: $\mathbb{R}^d \to \mathbb{R}^{\frac{d}{N}}$，其中 N 是注意力头的数量，$\frac{d}{N}$ 是每个注意力头的向量维度。类似地，我们也通过线性投影分别将堆叠矩阵 $\left[\widetilde{H}^{(l)}_{\mathcal{G}_C}[e_i]\ ,\ \widetilde{H}^{(l)}_{\mathcal{G}_T}[e_i]\right] \in \mathbb{R}^{2\times d}$ 投影为第 N 个 Key $Value$ $\mathbb{R}^{2\times d} \to \mathbb{R}^{2\times \frac{d}{N}}$ 矩阵和矩阵：。接下来，我们计算$Q^n(e_i)$与每个 $K^n(e_i)$之间的相似度。我们将通过 softmax函数计算得到的注意力系数与$V^n(e_i)$相乘，得到每个注意力头的嵌入。最后，我们将N 个注意力头拼接在一起，得到 $H^{(l)}[e_i]$的嵌入。

所有节点 E 和 P，T，O在 ECG 中的嵌入共同构成第 l 层所有节点的表示。

$$H^{(l)} = \left\{H^{(l)}[E],\ \widetilde{H}^{(l)}_{\mathcal{G}_C}[P],\ \widetilde{H}^{(l)}_{\mathcal{G}_C}[T],\ \widetilde{H}^{(l)}_{\mathcal{G}_C}[O]\right\} \quad (11)$$

D. 证据推荐

我们定义 score(p，e)，用于表征每个 p 与 e之间的相关性排序分数。σ 是 Sigmoid 函数。如公式 12 所示，给定输入临床问题 p_i，我们推荐相关性得分最高的前 k 个实体（e_1，e_2，..... e_k）。

$$score(p_i, e_i) = \sigma\left(H[p_i] \cdot H[e_i]^T\right) \quad (12)$$

MHAN 以端到端方式进行训练。给定一个已在一条证据 e中被提及的临床问题 p，我们假设 e 与 p之间应具有较高的相关性得分，因为它们在数据集中存在显式关系。此外，score(p，e)应高于p 与一组 k 个随机采样且与 p不存在显式关系的证据 $\{e'\} \in E$ 之间的任一分数。我们将边际损失定义如下：

$\{e'\}$ 从任意噪声分布 $e' \sim P_n(e)$中.....。我们为每组边负采样 k e' 个节点

表 I实验数据集统计

Data	#p	#i	#o	#e	#边
ClinicalTrials.gov	180	361	629	247	1563

表 II各证据图中边数量的统计

ECG			ETG
具有问题	具有干预措施	具有结果	相关
449	485	629	522

连接 ting p 和 e。我们将边际损失定义如下：流程：

$$\mathcal{L} = \sum_{e_i \sim P_n(e), i=1,\ldots,k} \max\left(0, 1 - y_{p,e} + y_{p,e_i}\right) \tag{13}$$

IV. 实验

在本节中，我们评估 MHAN，以回答以下研究问题：

- **研究问题1：** 与各种最先进的推荐模型相比，MHAN 的表现如何？（见 B 节）
- **研究问题2：** 不同模块对整体模型性能有何贡献？（见 C 节）
- **研究问题3：** 我们能否简单地将 ECG 和 ETG 合并到单个通道中？（见 C 节）
- **研究问题 4：** 融合注意力与最先进的自适应注意力及其他非注意力机制相比，表现如何？（见 D 节）
- **研究问题 5：** ETG 的相似度阈值如何影响性能？（见 E 节）
- **研究问题 6：** 在不同的融合注意力头超参数设置下，MHAN 的表现如何？（见 G 节）
- **研究问题 7：** 如何解释推荐结果？（见 F 节）

A. 实验设置

1）数据集： 我们在 ClinicalTrials.gov 的临床证据数据集上评估了所提出的 MHAN 的性能。数据集的统计信息汇总于表 I。我们的数据集包含 247 项最新证据，涉及 180 个问题、361 项干预措施和 629 个结局。根据该数据集得出的 ECG 和 ETG 连接统计信息汇总于表 II。

2）评估指标： 我们采用推荐任务的通用指标，包括 $HR@K$ 和 $nDCG@K$ （K =3、5、7、9）。

$$HR@K = \frac{NumberofHits@K}{|GT|} \tag{14}$$

公式的分子表示在排名前 k 条推荐意见中正确推荐的样本数量之和，而分母 $|GT|$ 表示测试集中的样本总数。

$$nDCG@K = \frac{DCG@K}{IDCG@K} \tag{15}$$

$nDCG@K$ 旨在为列表中排名较高的物品赋予更高分数，同时对排名较低的物品进行惩罚。每个物品的得分按其截至该位置的折损累计增益（DCG）除以理想折损累计增益（$IDCG$）计算，后者是理想排序下的 DCG。换言之，如果相关证据实体在推荐列表中的排名较高，就会获得更高分数。

3）训练细节： 为确保研究结果可复现，我们将随机种子固定为2022年。训练集和测试集按照8:2的比例随机划分。我们训练了MHAN模型，其层数为2，隐藏维度为256，阈值为0.8，学习率为0.001。HGT和GAT嵌入组件使用4个注意力头，融合注意力使用16个注意力头。模型使用Python==3.9.13、pytorch-cu116和dgl-cu116实现。

所有基线方法的初始向量均采用与 MHAN 相同的 BERT 模型进行初始化。其初始参数根据各自论文中的建议设置，我们还进一步仔细调整参数，以获得最佳性能。

B. 与基线方法的比较

为评估我们的方法 MHAN 的性能，我们将 MHAN 与近期具有竞争力的最先进基线方法进行比较，这些方法涵盖推荐系统和基于知识的嵌入。

- **NFM** [15] 是一种基于神经网络的推荐模型，结合了因子分解机（FM）和深度学习模型各自的优势。
- **ECFKG** [18] 是一种协同过滤方法，利用知识图谱嵌入来提升推荐性能。
- **KGAT** [8]是一种推荐模型，利用用户、物品、实体和关系的图嵌入，通过注意力机制捕捉物品的协同信息和内容信息。
- **LightGCN** [17] 是一种高效的推荐方法，能够处理数据中的稀疏性，并利用数据的图结构提高推荐意见的质量。
- **SimGCL** [20] 是一种对比学习推荐方法，可学习更加均匀的用户—物品表示，从而隐式地缓解流行度偏差。
- **RGCN** [11] 是一种使用图卷积网络（GCN）学习图中实体表示的模型，能够处理实体之间的多种关系。
- **HGT** [14] 是一种图嵌入模型，采用基于 Transformer 的架构和多头注意力机制，对异构图中的节点交互进行建模。

证据推荐的结果见表 III。总体而言，基于协同过滤的方法 NFM 在稠密网络上的表现更好，但在稀疏的临床证据推荐中效果较差。两种推荐系统方法 KGAT 和 ECFKG 均考虑了外部节点信息，例如干预实体和结局实体。其中，KGAT 通过使用注意力机制取得了比 ECF KG 更好的结果。LightGCN 专为处理稀疏图而设计，而 SimGCL 使用对比学习方法。这些方法优于

表 III临床证据上的不同方法的实验结果

模型	指标							
	NDCG@3	NDCG@5	NDCG@7	NDCG@9	HR@3	HR@5	HR@7	HR@9
NFM （2017年）	0.102	0.105	0.135	0.144	0.111	0.133	0.200	0.222
ECFKG （2018年）	0.067	0.075	0.090	0.103	0.067	0.089	0.122	0.156
KGAT （2019年）	0.189	0.197	0.205	0.222	0.200	0.222	0.267	0.300
LightGCN （2020年）	0.150	0.151	0.168	0.185	0.189	0.244	0.256	0.267
SimGCL （2022年）	0.159	0.157	0.185	0.199	0.189	0.233	0.244	0.300
RGCN （2018年）	0.247	0.254	0.260	0.279	0.244	0.289	0.322	0.344
HGT （2020	0.277	0.274	0.280	0.289	0.233	0.300	0.333	0.356
MHAN年）	**0.284**	**0.319**	**0.326**	**0.325**	**0.256**	**0.344**	**0.378**	**0.400**

表 IV变体实验结果

	REeb	CRec	URec	MHAN
NDCG@3	0.157	0.277	0.275	**0.284**
NDCG@5	0.153	0.274	0.279	**0.319**
NDCG@7	0.153	0.280	0.283	**0.326**
NDCG@9	0.157	0.289	0.287	**0.325**
HR@3	0.144	0.233	0.233	**0.256**
HR@5	0.189	0.300	0.289	**0.344**
HR@7	0.189	0.333	0.333	**0.378**
HR@9	0.211	0.356	0.356	**0.400**

表 IV不同节点融合方法的性能

	自适应注意力	共享矩阵	融合注意力
NDCG@3	0.144	0.243	**0.284**
NDCG@5	0.155	0.250	**0.319**
NDCG@7	0.174	0.255	**0.326**
NDCG@9	0.192	0.263	**0.325**
HR@3	0.122	0.233	**0.256**
HR@5	0.144	0.278	**0.344**
HR@7	0.178	0.322	**0.378**
HR@9	0.211	0.333	**0.400**

传统推荐系统方法，但仍未超过 KGAT，因为它们未纳入外部信息。RGCN 和 HGT 将异构图 ECT 学习为嵌入表示，并基于公式 13 进行负采样。它们利用了图的异构性以及所有实体的信息；其中，HGT 通过采用基于 Transformer 的架构和多头注意力取得了更好的结果。

我们提出的 MHAN 模型优于所有基线模型，这可归因于我们的建模方法纳入了 ETG。ETG 中的语言学连接揭示了证据实体之间超越单纯共指关系的相互关系。我们采用两个独立通道，分别对异质 ECG 和同质 ETG 进行表示学习。通过多头融合注意力融合证据实体，充分考虑了异质信息和文本信息，从而缓解了图稀疏性，并解决了共指与文本之间的异质性问题。

C. 消融研究

本小节讨论 MHAN 各个组成部分对其功能的贡献。

1) 初始嵌入：我们利用语言学连接构建 ECG 和 ETG，但将基于 BERT 的初始嵌入替换为随机初始嵌入，以展示异构文本信息作为临床证据推荐初始嵌入的影响。如表 IV 所示，基于 BERT 的嵌入优于随机嵌入（REeb），这表明基于文本信息的嵌入对于学习这两个图的优质表示具有重要作用。

2) 多通道：我们将仅使用通道通过 ECG 进行推荐定义为 CRec。我们将来自 ETG 的三元组（“证据”、“相似”、“证据”）添加到 ECG 中，con-构建统一证据图（UEG）。我们将基于 UEG 的推荐定义为 URec。

根据表 IV，我们可以观察到，CRec 和 URec 的结果相当，但都低于使用 MHAN 对 ECG 和 ETG 进行双通道表示学习所获得的结果。这表明，语言学连接不同于传统的实体链接，因为它们在揭示潜在语言学连接并缓解图稀疏性的同时，也会引入干扰信息。另一方面，MHAN 通过两个通道进行嵌入，一个用于处理节点和边的类型异质性，另一个用于处理语言学连接的文本异质性，因此取得了更好的结果。

D. 节点融合研究

在本小节中，我们基于双通道的嵌入，对不同的节点信息融合方法进行比较。

为了计算注意力系数，自适应注意力方法依赖于两个图中证据实体 v 的嵌入，而共享矩阵方法不依赖节点嵌入。另一方面，融合注意力方法依赖于两个图的 v 嵌入，但依赖程度不同。

如表 V 所示，我们观察到，融合注意力方法优于共享矩阵方法，而共享矩阵方法优于自适应注意力，这表明自适应注意力方法未能合理利用节点嵌入来计算注意力系数。

在临床证据推荐任务中，我们希望实体 p_i 和 v_i之间（存在链接时）获得更高的分数 $score(p_i, v_i)$。在 ECT嵌入中，v 聚合了有关 p的信息；而在 ETG 嵌入中，v 仅聚合有关 v的信息。融合证据实体时，充分考虑下游任务十分重要。

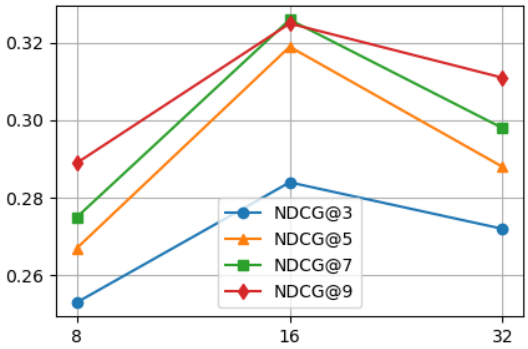


图 3. 不同融合注意力参数在归一化折损累计增益指标下的性能。

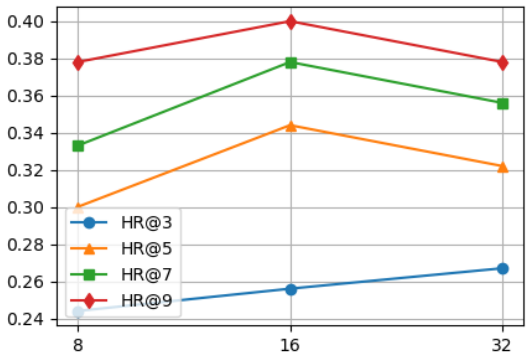


图 4. 不同融合注意力参数的性能及其命中率指标。

因此，采用融合注意力方法，并使用来自 ECT嵌入的 p 作为$query$ 向量来计算注意力系数，可取得最佳结果。

E. 相似度鲁棒性研究

表VI显示，我们将 ETG 的嵌入阈值设为从 0 到 1、步长为 0.1 的一系列值。当$\eta_{thr} = 1$时，ETG 的所有节点都只有自环；而当 $\eta_{thr} = 0$时，所有证据实体完全连接。结果表明，当 $\eta_{thr} = 0.8$时，整体性能最佳，但我们的模型在其他阈值下也表现良好。这说明，我们的融合注意力能够在融合节点时学习最相关的信息，并最大限度地降低无关信息的影响，使 MHAN 具有很强的鲁棒性。

F. 融合注意力头研究

我们详细讨论融合注意力头的数量 N 对模型性能的影响。通过将参数 N 从 8 调整到 32，我们在图 3 和图 4 中展示了其影响。结果显示，当 $N = 16$时，MHAN 的性能最佳。这是因为注意力头过多可能导致过拟合，而注意力头过少则可能造成注意力不足，无法有效捕捉相关信息。

G. 案例研究

我们展示了多个推荐结果案例，以分析模型的有效性。表VII展示了 MHAN 针对 COVID-19 问题推荐的前五个证据标题。其中，证据第1、2和4项涉及 COVID-19，属于测试集；证据第3和5项则不属于测试集，其问题也不涉及 COVID-19。由于这三条测试集证据的标题和描述信息与训练集中的内容相似，且它们在 ETG 中相互连接，有利于消息传递，因此我们的模型能够准确推荐它们。此外，模型还推荐了两条不属于测试集、但从标题来看与 COVID-19 相关的证据，这些证据与 COVID-19 相关。证据第3和5项的标题表明，它们与 COVID-19 研究具有医学相关性，因为 SARS-CoV-2 是导致 COVID-19 的病毒。尽管它们涉及的问题并非 COVID-19，我们的模型仍能识别它们。我们进一步考察了这两条证据涉及的问题，发现第3项与肺炎和 SARS-CoV-2 相关，而第5项与急性呼吸窘迫综合征和 SARS-CoV-2 相关。

表 VIII 展示了另一个案例，我们在此分析模型推荐意见的有效性。问题为难治性恶性实体瘤的证据 1、3 和 4 位于测试集中，而证据 2 不在其中。尽管如此，我们的模型仍推荐了证据 2，因为其研究结果、描述性信息或干预措施可能与该问题相似。这一推荐凸显了我们的模型有望启发人们发现有关难治性恶性实体瘤的新证据。

V. 结论

本文探讨了临床证据推荐任务，即根据给定问题预测一组证据。为缓解稀疏性，我们基于语言相似性构建了 ETG。我们采用双通道解决共指文本异质性，并通过实验表明，基于语言的链接并不等同于传统的边异质性。为了利用证据实体的文本信息，我们还将证据实体的文本信息用作图的初始嵌入向量。为融合不同通道聚合的节点信息，我们提出了融合注意力。我们证明了 MHAN 优于当前最先进方法。

基于我们的研究，后续工作可以从多个方向展开。一项有意义的工作是利用公开数据集中的文本信息提取更丰富的实体，以丰富图表示。我们的研究也具有实际意义，尽管距离在临床医疗中的实际应用还有一定距离，但它可以为循证医学领域潜在的证据或药物研究指明方向。

表VI性能的相似度

η_{thr}	NDCG@3	NDCG@5	NDCG@7	NDCG@9	HR@3	HR@5	HR@7	HR@9
0.0	0.281	0.288	0.296	0.300	**0.267**	0.333	0.356	0.367
0.1	0.274	0.284	0.290	0.300	0.256	0.322	0.356	0.367
0.2	0.263	0.285	0.288	0.300	0.256	0.322	0.356	0.367
0.3	0.268	0.287	0.287	0.304	0.256	0.322	0.356	0.378
0.4	0.263	0.281	0.289	0.296	0.256	0.322	0.356	0.367
0.5	0.272	0.278	0.287	0.292	0.256	0.322	0.344	0.378
0.6	0.264	0.277	0.287	0.305	**0.267**	0.322	0.344	0.389
0.7	0.259	0.280	0.290	0.297	0.256	0.333	0.367	0.389
0.8	**0.284**	**0.319**	**0.326**	**0.325**	0.256	**0.344**	**0.378**	**0.400**
0.9	0.260	0.274	0.295	0.301	0.256	0.311	0.344	0.378
1.0	0.259	0.295	0.309	0.316	0.256	0.311	0.344	0.378

表VII
COVID-19 排名前五的结果

问题：COVID-19**第1项：**一项随机对照适应性研究：比较
COVID-19康复者血浆（CCP）与非免疫血浆在限制住院患者冠状病毒相关并发症方面的效果**第2项：**卡那单抗用于减轻SARS-CoV-2相关急性心肌损伤伴炎症加重患者的心脏和呼吸功能恶化第***3***项：马弗利单抗用于减缓重症COVID-19肺炎及全身性过度炎症患者急性呼吸衰竭的进展**第4项：**一项随机、双盲、安慰剂对照试验：评估羟氯喹和阿奇霉素在预防COVID-19患者住院或死亡方面的疗效第***5***项：吸入用阿法脱氧核糖核酸酶治疗SARS-CoV-2患者的急性呼吸窘迫综合征

表 VIII
难治性恶性实体瘤的前四项结果

问题：难治性恶性实体瘤**编号1：**MATCH治疗子方案Y：针对携带AKT突变肿瘤患者的AZD5363编号***2***：BN-Brachyury、恩替司他、恩美曲妥珠单抗和M7824用于晚期乳腺癌**编号3：**MATCH治疗子方案Z1A：比美替尼用于携带NRAS突变的肿瘤患者（黑色素瘤除外）**编号4：**MATCH治疗子方案W：针对FGFR通路异常肿瘤患者的AZD4547 II期研究

参考文献

[1] Sackett, David L.《循证医学》。围产学研讨会。第21卷，第1期。WB Saunders，1997年。[2] Sackett D L, Rosenberg W M C, Gray J A M, 等。循证医学：它是什么，又不是什么[J]。《英国医学杂志》，1996年，312(7023)：71-72。[3] Andrews J, Guyatt G, Oxman A D, 等。GRADE指南：14. 从证据到推荐意见：推荐意见的重要性及其呈现方式[J]。《临床流行病学杂志》，2013年，66(7)：719-725。[4] Sutton, Reed T., 等。《临床决策支持系统概述：益处、风险与成功策略》。《NPJ数字医学》，3.1（2020年）：1-10。[5] Yang C, Xiao C, Ma F, 等。SafeDrug：用于安全药物推荐的双分子图编码器[ ] 2105 02711, J 。2021 arXiv预印本 arXiv: . 。[6] Zhang Y, Wu X, Fang Q, 等。用于药物推荐的知识增强型属性多任务学习[J]。《ACM信息系统汇刊》（TOIS），2022年。[7] Ren Y, Shi Y, Zhang K, 等。基于消息传播和DDI门控机制的药物推荐模型[J]。《IEEE生物医学与健康信息学杂志》，2022年。[8] Wang X, He X, Cao Y, 等。KGAT：用于推荐的知识图谱注意力网络[C]//第25届ACM SIGKDD知识发现与数据挖掘国际会议论文集。2019年：950-958。[9] Mahyari A, Pirolli P, LeBlanc J A。利用深度推荐系统从专家处实时学习：应用于移动健康领域的体育锻炼[J]。《IEEE生物医学与健康信息学杂志》，2022年，26(8)：4281-4290。[10] Kipf T N, Welling M。利用图卷积网络进行半监督分类[J]。arXiv预印本 arXiv:1609.02907，2016年。[11] Schlichtkrull M, Kipf T N, Bloem P, 等。利用图卷积网络对关系数据建模[C]//欧洲语义网会议。Springer，Cham，2018年：593-607。[12] Veli˘ckovi´c P, Cucurull G, Casanova A, 等。图注意力网络[J]。arXiv预印本 arXiv:1710.10903，2017年。[13] Wang X, Zhu M, Bo D, 等。AM-GCN：自适应多通道图卷积网络[C]//第26届ACM SIGKDD知识发现与数据挖掘国际会议论文集。2020年：1243-1253。[14] Hu Z, Dong Y, Wang K, 等。异构图变换器[C]//2020年万维网会议论文集。2020年：2704-2710。[15] He X, Liao L, Zhang H, 等。神经协同过滤[C]//第26届万维网国际会议论文集。2017年：173-182。[16] Rendle S, Gantner Z, Freudenthaler C, 等。利用因子分解机实现快速的上下文感知推荐[C]//第34届国际ACM SIGIR信息检索研究与发展会议论文集。2011年：635-644。[17] He X, Deng K, Wang X, 等。LightGCN：简化并增强用于推荐的图卷积网络[C]//第43届国际ACM SIGIR信息检索研究与发展会议论文集。2020年：639-648。[18] Ai Q, Azizi V, Chen X, 等。学习异构知识库嵌入以实现可解释推荐[ ] 2018, J 。11 9 137 Algorithms，( )：。[19] Bordes A, Usunier N, Garcia-Duran A, 等。用于多关系数据建模的嵌入平移[J]。《神经信息处理系统进展》，2013年，第26卷。[20] Yu J, Yin H, Xia X, 等。图增强是否必要？用于推荐的简单图对比学习[C]//第45届国际ACM SIGIR信息检索研究与发展会议论文集。2022年：1294-1303。[21] Devlin J, Chang M W, Lee K, 等。BERT：用于语言理解的深度双向变换器预训练[J]。arXiv预印本 arXiv:1810.04805，2018年。