面向基于会话的推荐的时间感知双曲图注意力网络

Xiaohan Li*†、Yuqing Liu*‡、Zheng Liu‡、Philip S. Yu‡
†沃尔玛全球科技，美国加利福尼亚州桑尼维尔
xiaohan.li@walmart.com
‡伊利诺伊大学芝加哥分校，美国伊利诺伊州芝加哥{
yliu363, zliu212, psyu}@uic.edu

arXiv:2301.03780v1 [cs.IR] 10 Jan 2023

摘要——**基于会话的推荐（SBR）旨在根据用户之前的浏览会话，预测用户接下来感兴趣的物品。现有方法将会话建模为图或序列，根据用户交互过的物品估计其兴趣，进而生成推荐结果。近年来，基于图的方法在基于会话的推荐任务中取得了出色的性能。然而，这些方法均未考虑时间信息，而时间信息是基于会话的推荐中的关键特征，因为它反映了时效性或新近程度。此外，会话图呈现层次结构，且已被证明适用于双曲几何。但很少有论文在双曲空间中设计模型，这一方向仍在探索之中。**

本文提出时间感知双曲图注意力网络（TA-HGAT）——一种新颖的双曲图神经网络框架，用于构建考虑时间信息的基于会话的推荐模型。具体而言，TA-HGAT包含三个组成部分。首先，双曲投影模块将物品特征转换到双曲空间。其次，时间感知图注意力模块对物品之间的时间间隔以及用户当前兴趣进行建模。第三，模型末端的演化损失根据给定的时间戳，准确预测推荐物品。TA-HGAT构建于双曲空间中，以学习会话图的层次结构。实验结果表明，与两个真实世界数据集上的十种基线模型相比，所提出的TA-HGAT性能最佳。

索引词—**推荐系统，图神经网络，双曲嵌入**

I. 引言

推荐系统一直是帮助用户克服互联网信息过载问题的有效方案。基于这一理念开发了许多应用，包括在线零售 [1]，音乐流媒体 [2],和内容分享 [3]。为了更好地了解用户，对其浏览会话进行建模是一种有效的方法，因为会话反映了用户当前的兴趣。基于会话的推荐（SBR）通过对用户会话进行建模，预测用户接下来感兴趣的物品。深度学习模型，包括循环神经网络（RNN） [4], [5], 记忆网络 [6], 和图神经网络（GNN） [7], [8] ，已应用于这一问题，并取得了最先进的性能。

近年来，处理基于会话的推荐最具影响力的研究是基于图神经网络的方法。这些基于图神经网络的方法 [7]–[12] 将每个会话视为一张图，以学习物品之间的内部关系及其复杂转移。最具代表性的模型是 SR-GNN [7], ，这是首个将图神经网络应用于基于会话的推荐并取得最先进性能的研究。在 SR-GNN 的基础上， [10], [12] 通过注意力层对其进行改进。 [8], [11]考虑会话图中的物品顺序来构建模型。 [9], [13], [14] 将全局物品关系、物品类别和用户表示等额外信息纳入考量，以设计更全面的模型。HCGR [15]将会话图映射到双曲空间中，以提取层次信息。

尽管现有的基于图神经网络的方法已取得令人满意的性能，但仍存在两项局限。首先，与基于序列的模型不同，图结构无法明确展现物品之间的时间信息。时间间隔是一项关键特征，能够显著提升推荐性能 [16], [17], ，但现有的基于图的会话推荐模型忽略了这一点。此外，尽管将会话建模为图有助于学习物品的复杂转移 [7], ，但由于图结构下会话的起点和终点并不明确，会话图中的物品序列关系也不清晰。其次，根据 [18], [19],，图数据呈现出潜在的非欧几里得结构，因此，在欧几里得空间中学习这种几何结构并非恰当的选择。因此，近期一些研究 [15], [20], [21] 表明，推荐系统的真实世界数据集通常呈现树状层次结构，而双曲空间能够有效捕捉此类层次信息。因此，值得尝试在双曲空间中学习会话图。

双曲空间能够对层次结构数据进行建模，因为其扩展速度比欧几里得空间更快。双曲空间可以呈指数级扩展，而欧几里得空间只能呈多项式级扩展。已有研究 [15]展示了会话图的层次结构。然而，在双曲空间中对会话图进行建模仍处于探索阶段。首先，时间间隔表示两个物品之间的相关性。由于双曲嵌入更适合会话图，因此有必要定义一个新框架，用于识别会话图边上的时间间隔。其次，在图中学习用户当前兴趣至关重要，但在双曲空间中实现这一点并不容易。以往的研究

*两位作者对本研究贡献相同。

[6], [7] 基于会话中的最后一个物品构建欧几里得空间模型。最后一个物品代表用户当前兴趣，因此在预测下一个物品时发挥着重要作用。然而，由于双曲空间中的运算比欧几里得空间中的运算更复杂，因此在双曲空间中对这一特征进行建模更具挑战性。第三，考虑时间信息后，我们不仅可以提供下一物品推荐，还可以基于特定时间戳提供推荐结果。

为应对上述挑战，我们提出了时间感知双曲图注意力网络（TA-HGAT），这是一种考虑全面时间相关特征的双曲图神经网络。具体而言，我们通过双曲投影层将物品的原始特征投影到庞加莱球空间中。随后，我们设计了一种时间感知双曲注意力机制，在双曲空间中共同学习时间间隔和用户当前兴趣。该机制包含两个模块：结合时间间隔的双曲自注意力，以及结合用户当前兴趣的双曲软注意力。最后，模型通过演化损失进行训练，以预测用户在特定时间戳可能感兴趣的物品。以上三个组成部分均基于全双曲图神经网络框架。

在此，我们将贡献总结如下：

- 据我们所知，这是首篇在双曲空间中建模时间信息以提升推荐系统性能的论文。我们超越了传统的欧几里得机器学习模型，以更精细的方式对用户的时间相关特征进行建模。

- 我们提出了 TA-HGAT，这是一种基于双曲图神经网络的框架，包含三个主要组件：双曲投影、时间感知双曲注意力和演化损失。这三个组件在端到端图神经网络中协同工作，用于建模物品的时间间隔和用户当前兴趣。最终，我们的模型能够提供特定时间推荐。

- 我们在两个真实世界数据集上进行了实验，并将模型与十种基线模型进行了比较。实验结果表明，TA-HGAT 在平均倒数排名和精确率方面表现有效。

II. 预备知识

A. 图神经网络

图神经网络 [22], [23] 旨在处理结构化图数据。在图神经网络中，聚合是提取结构知识的核心操作。通过聚合邻居节点的信息，中心节点可以获取经由边从邻居节点传递而来的知识，并学习节点嵌入。图神经网络已被证明能够有效学习节点嵌入，因此被广泛应用于许多节点相关任务，例如节点分类 [23], 、图分类 [24],和链接预测 [22]。

基于聚合操作，图 $G = (\mathcal{V}, \mathcal{E})$ 上 GNN 的前向传播旨在学习嵌入节点 $v_i \in \mathcal{V}$ 的嵌入，方法是聚合其邻居节点的信息。我们假设每个节点 i 的初始节点嵌入为 $\mathbf{h}_i^{(0)}$，通常就是该节点的特征。在 GNN 的每个隐藏层中，中心节点 $\mathbf{h}_i^{(l)}$ 的嵌入由上一隐藏层中邻居节点的聚合嵌入 $\mathbf{h}_i^{(l-1)}$ 学习得到。该过程用数学公式描述如下：

$$\mathbf{h}_i^{(l)} = \sigma\Big(\mathbf{W}^{(l)}(\underset{j\in\mathcal{N}_i}{\mathrm{AGG}}(\mathbf{h}_j^{(l-1)})\Big), \tag{1}$$

其中，$\mathcal{N}_i$ 表示图中节点 i 的所有邻居节点集合，包括节点 i 本身。聚合函数 AGG(·) 将邻居信息整合在一起。随后应用非线性激活函数 σ（例如 Sigmoid 函数或 LeakyReLU），以生成第 l层中节点 i 的嵌入。

基于上述介绍的基础 GNN，GAT [25]被提出，以利用自注意力机制[26]改进 GNN。具体而言，对于节点 i的所有邻居节点，我们需要学习其注意力系数，以计算聚合过程中每个邻居节点的重要性。假设节点对 (i, j) 的注意力系数为 α_{ij}，学习 α_{ij} 的过程如下：

$$\alpha_{ij} = softmax(\mathbf{d}_{ij}) = \frac{\exp(\mathbf{d}_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(\mathbf{d}_{ik})}, \tag{2}$$

其中，$\mathbf{d}_{ij}$ 表示节点 i 与 j之间的相关性。此处的$\mathbf{d}_{ij}$ 可以是节点 i 和 j 的联合嵌入，例如节点嵌入的拼接，或节点对的相似度。

B. 双曲空间

定义上，双曲空间是具有负曲率的齐性空间。它是一个光滑的黎曼流形，可通过多种双曲几何模型建模，包括庞加莱球模型 [27], 克莱因模型 [28], 洛伦兹模型 [29], 等。本文选择庞加莱球模型，因为两点之间的距离呈指数增长，这与会话图的层次结构十分契合。形式上，d维庞加莱球空间 $\mathbb{P}_c^d$ 定义为

$$\mathcal{P}_c^d = \{\mathbf{x} \in \mathbb{R}^d, c\|\mathbf{x}\|<1\}, \tag{3}$$

其中，c 是球的半径，$\mathbf{x}$ 是流形$\mathcal{P}$上的任意一点。若 $c = 0$，则 $\mathbb{P}_c^d = \mathbb{R}^d$，且该球等于欧几里得曲面。本文中，我们设定 $c = 1$。切空间 $\mathcal{T}_{\mathbf{x}}\mathcal{P}$ 是一个 d维向量空间，用于近似$\mathbf{x}$附近的 $\mathcal{P}$，且与欧几里得空间同构。通过指数映射，欧几里得空间中的向量可以映射到双曲空间。对数映射是指数映射的逆映射，可将向量投影回欧几里得空间。

在双曲空间中，神经网络的基本数学运算（例如加法和乘法）不同于欧几里得空间中的运算。在本文中，我们选择 M¨obius 变换作为研究双曲几何的代数运算。对于一对随机向量（$\mathbf{a}$，$\mathbf{b}$），我们将模型中使用的运算列举如下：

- M¨obius 加法 $\oplus$ [30] 用于对 $\mathbf{a}$ 和 $\mathbf{b}$执行加法运算。

$$\mathbf{a} \oplus \mathbf{b} = \frac{(1 + 2\langle \mathbf{a}, \mathbf{b} \rangle + \|\mathbf{b}\|^2)\mathbf{a} + (1 - \|\mathbf{a}\|^2)\mathbf{b}}{1 + 2\langle \mathbf{a}, \mathbf{b} \rangle + \|\mathbf{a}\|^2\|\mathbf{b}\|^2}. \quad (4)$$

- 莫比乌斯矩阵-向量乘法 $\otimes$ [31] 用于通过矩阵$\mathbf{W}$变换$\mathbf{a}$ 。

$$\mathbf{W} \otimes \mathbf{a} = \tanh(\frac{\|\mathbf{W}\mathbf{a}\|}{\|\mathbf{a}\|} \tanh^{-1}(\|\mathbf{a}\|)), \quad (5)$$

- 莫比乌斯标量乘法 $\otimes$ 是标量 α 与向量$\mathbf{b}$的乘法。

$$\alpha \otimes \mathbf{b} = \tanh(\alpha \tanh^{-1}(\|\mathbf{b}\|)) \frac{\mathbf{b}}{\|\mathbf{b}\|} \quad (6)$$

- 指数映射将$\mathbf{a}$ 从欧几里得空间变换到双曲空间中的选定点$\mathbf{x}$ 。

$$\exp_{\mathbf{x}}(\mathbf{a}) = \mathbf{x} \oplus (\tanh(\frac{\lambda_{\mathbf{x}}\|\mathbf{a}\|}{2}) \frac{\mathbf{a}}{\|\mathbf{a}\|}), \quad (7)$$

- 对数映射将向量$\mathbf{a}$ 投影回欧几里得空间。

$$\log_{\mathbf{x}}(\mathbf{a}) = \frac{2}{\lambda_{\mathbf{x}}} \operatorname{arctanh}(\| -\mathbf{x} \oplus \mathbf{a}\|) \frac{-\mathbf{x} \oplus \mathbf{a}}{\| -\mathbf{x} \oplus \mathbf{a}\|} \quad (8)$$

- $\lambda_{\mathbf{x}}$ 是共形因子。

$$\lambda_{\mathbf{x}} = \frac{2}{1 - \|\mathbf{x}\|^2}. \quad (9)$$

III. 模型

本节介绍我们提出的时间感知双曲图注意力网络（TA-HGAT）框架，该网络旨在对双曲会话图中的时间信息进行建模。首先，我们定义基于会话的推荐任务；随后，阐述模型的三个主要组成部分：双曲投影、时间感知双曲注意力和双曲演化损失。这三个组成部分利用时间相关特征训练模型，并根据特定时间戳提供推荐结果。TA-HGAT 的整体结构如图 1 所示。

A. 问题定义

基于会话的推荐（SBR）旨在根据用户-物品交互会话，预测用户接下来会点击的物品。通常，这类方法会对用户的短期浏览会话数据进行建模，以学习用户当前的兴趣。下面我们将 SBR 问题表述为数学形式。

在 SBR 问题中，会话表示为 $S = \{v_1, v_2, \cdots, v_n\}$ 按时间戳排序。 S 中的每个 v 都是一个物品，物品集合为 V_s，由该会话中的所有不重复物品组成。为了将会话建模为有向图，我们将所有物品作为节点，并将物品间顺序依赖关系作为边，以构建会话图。该图表示为 $G_s = (V_s, E_s)$，其中 V_s，E_s 分别是节点集和边集。每条边连接两个连续物品，表示为 $e = (v_{t-1}, v_t)$。我们的目标是学习物品和会话的嵌入，并生成用户在下一个时间戳可能感兴趣物品的排序。

B. 双曲投影

在图神经网络中，每个节点都需要以初始嵌入作为输入。针对基于会话的推荐，图神经网络的输入是物品特征，例如类别或描述。物品 i 的初始嵌入为 $\mathbf{h}_i^0$。然而，大多数特征嵌入方法都是基于欧几里得空间的。为了使物品特征能够应用于双曲空间，我们使用式 7 中定义的指数映射，将物品的初始嵌入投影到双曲空间。具体而言，投影过程可表示为

$$\mathbf{m}_i = \exp_{\mathbf{x}}(\mathbf{h}_i^0), \quad (10)$$

其中$\mathbf{m}_i$ 是映射到双曲空间中的嵌入，$\mathbf{x}$是切空间中选取的点。

为了获得节点特征的高层潜在表示，我们还添加了一个由权重矩阵 $\mathbf{W}_1 \in \mathbb{R}^{d' \times d}$参数化的线性变换，其中 d' 是$\mathbf{m}_i$的维度， d 是节点最终嵌入的维度。请注意，$\mathbf{W}_1$ 是所有节点共享的权重矩阵。我们采用式 4 中定义的莫比乌斯矩阵-向量乘法来变换$\mathbf{m}_i$ ，其过程为

$$\mathbf{h}_i^1 = \mathbf{W}_1 \otimes \mathbf{m}_i, \quad (11)$$

其中，$\mathbf{h}_i^1$ 是变换后的嵌入，并在后续步骤中用作初始节点嵌入。

C. 时间感知双曲注意力

根据 [15], [20], [32], [33], ，在双曲空间中嵌入用户和物品是对基于图的推荐系统的一项重大改进。然而，这些研究都没有在双曲空间中对时间间隔和用户当前兴趣进行建模。我们提出的模型 TA-HGAT 首次尝试解决这一问题，其核心组件是时间感知双曲注意力。它由两个注意力层组成：1）聚合过程中的双曲自注意力，用于考虑物品之间的时间间隔；2）会话嵌入学习中的双曲软注意力，用于建模用户当前兴趣。

1）带时间间隔的双曲自注意力：根据第 II-A 节，图注意力的一个关键步骤是为每个节点对（i， j）学习注意力系数 α_{ij} 。 α_{ij} 表示邻居节点对中心节点的重要性。为了学习 α_{ij}，不同于采用线性变换 [25] 或内积 [26], 的传统注意力网络，我们在此使用双曲空间中节点嵌入之间的距离。具体而言，我们将节点对（i， j）的距离表示为（$\mathbf{h}_i$， $\mathbf{h}j$），其计算方式如下

$$d(\mathbf{h}_i^l, \mathbf{h}_j^l) = \operatorname{arcosh}(1 + 2\frac{\|\mathbf{h}_i^l - \mathbf{h}_j^l\|^2}{(1 - \|\mathbf{h}_i^l\|^2)(1 - \|\mathbf{h}_j^l\|^2)}). \quad (12)$$

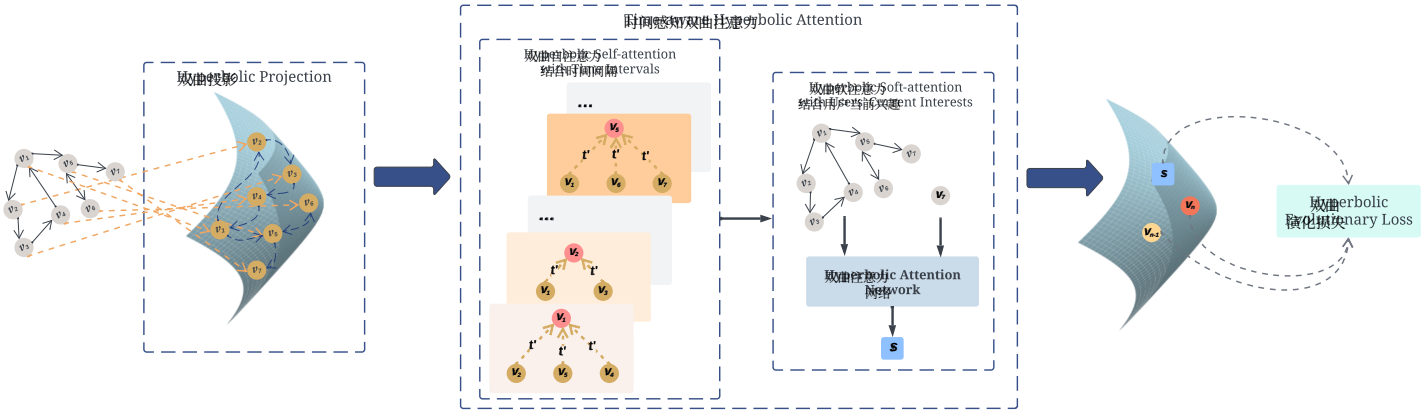


图 1. TA-HGAT 示意图。首先，根据会话序列构建有向会话图，然后将嵌入从欧几里得空间投影到双曲空间。接着，采用双曲自注意力聚合邻居信息和时间间隔 t'。之后，使用双曲软注意力机制将每个会话图表示为会话嵌入。最后，TA-HGAT 为每个会话预测最有可能在下一个时间戳被点击的 top-k 个物品。

然后，利用节点距离，我们进一步学习节点 i 与其所有邻居节点（包括其自身）之间的注意力系数 $\alpha_{ij}\,\mathcal{N}_i$，计算方式如下

$$\alpha_{ij} = softmax(d_{ij}) = \frac{\exp(d_{ij})}{\sum_{k\in\mathcal{N}_i}\exp(d_{ik})}, \tag{13}$$

我们使用双曲空间中的距离计算注意力系数，原因在于其有两项优势。首先，欧几里得空间中的注意力系数通常通过线性变换 [25]或内积 [26],计算，而这两种方式都不满足三角不等式。在双曲空间中，学习得到的注意力系数能够满足这一条件，并保留节点之间的传递性。其次，节点 i 与自身的注意力系数为 $\alpha_{ii} = d\,(\mathbf{h}_i,\ \mathbf{h}_i)\ = 0$，因此中心节点本身的影响不会影响注意力系数的计算。

得到注意力系数后，下一步是聚合节点嵌入，以学习下一层的中心节点嵌入。这里，学习到的注意力系数作为权重，作用于邻居节点嵌入。该过程表示为

$$\mathbf{h}_i^{l+1} = \sigma(\sum_{j\in\mathcal{N}_i}^{\oplus}\alpha_{ij}\otimes\mathbf{h}_j^l), \tag{14}$$

其中，$\sum^{\oplus}$ 表示加权邻居节点嵌入的M¨obius加法，σ 是非线性函数，例如Sigmoid函数和LeakyReLU。与公式11不同，公式14中的 $\otimes$ 是公式6中定义的M¨obius标量乘法。

为了将时间信息整合到注意力层中，核心思路是将时间间隔纳入聚合过程。具体而言，我们将时间间隔转换为双曲空间中的向量，并将时间向量与邻居节点嵌入结合起来进行聚合。由于时间间隔是连续值，我们将时间间隔值通过映射函数投影为向量。映射过程如下：

$$\mathbf{h}_{t'} = \mathbf{w}_t \otimes (t^+ - t), \tag{15}$$

其中，$t' = t^+ - t$ 表示时间间隔，$\otimes$ 在这里表示M¨obius矩阵向量乘法，而$\mathbf{w}_t$ 是将时间间隔投影到向量的变换向量。在本文中，如果两个物品之间有多个时间间隔，我们选择最近的一个。这一过程在建模前的数据预处理阶段完成。

受TransE启发， [34],时间感知双曲注意力通过时间信息将邻居节点嵌入变换为中心节点嵌入，因此节点嵌入与时间嵌入的联合嵌入由M¨obius加法生成，表示为$\mathbf{h}_j^l \oplus \mathbf{h}_{t'}$。

在式 14 中，中心节点 i 的所有邻居节点都通过 M¨obius 加法进行聚合。由于 M¨obius 加法较为复杂，且比欧几里得空间中的加法消耗更多计算资源，因此我们在此使用对数映射将嵌入投影到切空间（欧几里得空间）中进行聚合运算，从而简化式 14 的计算。随后，使用指数映射将嵌入重新投影回双曲流形。因此，式 14 中的聚合过程可以改写为

$$\mathbf{h}_i^{l+1} = \exp\left(\sigma\Big(\sum_{j\in\mathcal{N}_i}\log(\alpha_{ij}\otimes(\mathbf{h}_j^l\oplus\mathbf{h}_{t'}))\Big)\right). \tag{16}$$

2）结合用户当前兴趣的双曲软注意力：在上述过程中，我们利用邻居节点和时间间隔更新节点 i 的嵌入。为了基于学到的节点嵌入进行推荐，我们还需要通过聚合所有节点嵌入来获取会话图的全局嵌入。我们不只是简单地将所有节点嵌入相加，还提供了另一种方案：基于最近交互过的物品，在考虑用户当前兴趣的同时学习图嵌入。

理解用户当前兴趣是基于会话的推荐中的主要任务之一。在先前的研究中 [6], [7], [12],，会话中的最后一个物品被认为是此任务中最相关的特征。为了学习最后一个物品 p 与会话中其他每个物品之间的相关性，我们采用软注意力机制，为物品 p 与所有其他物品生成注意力系数，以表示各物品相对于当前时间戳的重要性。全局会话嵌入 $\mathbf{h}_s$ 的学习过程为

$$\beta_{pq} = \mathbf{x}^\intercal \otimes \sigma\Big(\mathbf{W}_2 \otimes \mathbf{h}_p) \oplus (\mathbf{W}_3 \otimes \mathbf{h}_q) \oplus \mathbf{c}\Big), \tag{17}$$

$$\mathbf{h}_s = \exp\left(\sigma\Big(\sum_{q \in V_s} \log(\beta_{pq} \otimes \mathbf{h}_q)\Big)\right), \tag{18}$$

其中，β_{pq} 是会话 S中物品 p 对另一个物品 q 的注意力系数。$\mathbf{x} \in \mathbb{R}^d$ 和$\mathbf{W}_2$,$\mathbf{W}_3 \in \mathbb{R}^{d\times d}$ 是权重矩阵。$\mathbf{h}_s$ 是会话嵌入，其中包含会话图结构、时间信息和用户当前意图，因此我们可以在下一步使用$\mathbf{h}_s$ 来推断用户的下一次交互。

D. 双曲演化损失

这里介绍如何利用演化损失，根据特定时间戳提供推荐结果。与其他研究 [3], [35], 不同，我们的演化损失也完全基于双曲空间。

1) 演化公式：演化损失的核心思想是根据未来时间戳预测未来会话和下一物品嵌入，然后生成推荐结果。演化损失的预测结果不像基于循环神经网络的模型 [4], [5] 那样依赖序列，而是基于 TA-HGAT 学得的最终嵌入。

由于$\mathbf{h}_s$ 是预测的未来会话嵌入，我们还需要一个估计的未来会话嵌入，以衡量预测嵌入是否准确。假设会话嵌入平滑增长。会话的嵌入向量在连续空间中演化。因此，我们设计了一个投影函数，根据先前嵌入与时间间隔的逐元素乘积来推断未来会话嵌入。当前时间 t 之后、未来时间 t^+ 时会话 S 的嵌入投影定义如下:

$$\widehat{\mathbf{h}}_s^{t^+} = \sigma\Big(\mathbf{h}_s^t \odot (\mathbf{1} \oplus \mathbf{h}_{t'})\Big), \tag{19}$$

其中，$\mathbf{1} \in \mathbb{R}^d$ 是一个所有元素均为 1 的向量，$\odot$ 是 Möbius 逐元素乘积。$\mathbf{h}_{t'}$ 是时间间隔向量，其学习方式与公式 15 相同。$\mathbf{1}$ 向量用于提供上一个会话嵌入与下一个会话嵌入之间的最小差异。通过这个投影函数，未来会话嵌入会随着时间间隔平滑地变化。

学习到会话的投影嵌入 $\widehat{\ }\mathbf{h}_s^{t^+}$ S后，下一步是应用另一个投影函数，生成下一物品 v的未来嵌入，记作 $\widehat{\ }\mathbf{h}_v^{t^+}$。投影后的未来物品嵌入由三个部分组成：前面步骤中学习到的投影会话嵌入、最后一个物品嵌入和时间间隔。在此，我们将下一物品 v 的投影公式定义为

$$\widehat{\mathbf{h}}_v^{t^+} = \sigma_v\Big((\mathbf{W}_4 \otimes \widehat{\mathbf{h}}_s^{t^+}) \oplus (\mathbf{W}_5 \otimes \mathbf{h}_{v_n}) \oplus \mathbf{h}_{t'}\Big), \tag{20}$$

其中，$\mathbf{W}_4$ 和$\mathbf{W}_5$ 表示权重矩阵。

2) 损失函数：通过上述投影函数，我们可以得到会话和下一个物品的未来嵌入估计值，并将其用作损失函数中的真实嵌入。为训练模型，损失函数旨在最小化模型生成的嵌入$\mathbf{h}_s^t$、$\mathbf{h}^{v_n}$ 与估计的真实嵌入 $\widehat{\ }\mathbf{h}_s^{t^+}$ 、 $\widehat{\ }\mathbf{h}_v^{t^+}$ 在每个交互时刻 t的距离。此外，为避免过拟合，还需要对物品嵌入施加另一项约束。我们约束最近两个物品 v_{n-1} 和 v_n 的嵌入之间的距离，以确保最后一个物品的嵌入与前一个物品保持一致。该约束假设最后一个物品和下一个物品反映相似的用户意图，并且会话嵌入在短时间内趋于稳定。最后，损失函数如下:

$$\begin{aligned}\mathcal{L} = \sum_{(s,v,t)\in\{S_i\}_{i=0}^I} & d(\widehat{\mathbf{h}}_v^{t^+}, \mathbf{h}_{v_n}) \oplus \Big(\lambda_s \otimes d(\widehat{\mathbf{h}}_s^{t^+}, \mathbf{h}_s^t)\Big) \oplus \\ & \Big(\lambda_v \otimes d(\mathbf{h}_{v_n}, \mathbf{h}_{v_{n-1}})\Big),\end{aligned} \tag{21}$$

其中，$\{S_t\}_{i=0}^I$ 表示数据集中的所有会话，λ_s和 λ_v 为平滑系数，用于防止在更新过程中会话和物品的嵌入偏移过大。$d(\cdot)$是双曲距离函数，其定义见公式 12。

为了向用户生成推荐结果，我们计算损失函数得到的预测物品嵌入与所有其他物品嵌入之间的双曲距离。然后，将距离最近的前k 个物品作为向该用户推荐的物品。

与传统的 BPR损失 [36], 相比，演化损失更适用于时间感知推荐，因为它考虑了时间间隔。因此，该损失 [3], 能够对变化轨迹进行建模，并根据特定时间戳更准确地推荐下一个物品。

IV. 实验

本节介绍我们在两个公开数据集上的实验结果，并将所提出的 TA-HGAT 与十种最先进的基线模型进行比较。我们的实验旨在回答以下研究问题:

- **RQ1**：TA-HGAT 与其他最先进的基于会话的推荐模型相比表现如何?
- **RQ2**：时间感知双曲注意力的两个模块，即结合时间间隔的双曲自注意力和结合用户当前兴趣的双曲软注意力，会如何影响 TA-HGAT 的性能?
- **RQ3**：双曲演化损失与其他损失函数相比表现如何?
- **RQ4**：不同超参数（即嵌入维度）有何影响?

表I 各数据集的物品数量、训练会话数、测试会话数、平均长度和点击次数。

数据集	物品	训练会话	测试会话	平均长度	点击次数
Diginetica	43,097	719,470	60,858	5.12	982,961
Yoochoose1/64	16,766	369,859	55,898	6.16	557,248
Yoochoose1/4	29,618	5,917,746	55,898	5.71	8,326,407

A. 实验设置

1）数据集：我们在两个广泛使用的公开数据集 Yoochoose 和 Diginetica 上开展实验。这些数据集的统计信息列于表I中。

- **Yoochoose**[1] 是 RecSys 挑战赛 2015 发布的一个公开数据集，包含来自yoo- choose.com、为期 6 个月的点击流。
- **Diginetica**[2] 来源于 CIKM 杯 2016。我们使用物品类别来初始化物品嵌入。

2) 评估指标： 在对比实验中，我们使用平均倒数排名（MRR@K）和精确率（P@K）评估模型的性能。

MRR@K考虑目标物品在推荐物品列表中的位置。如果目标物品不在排序列表的前-k 位，则其值为 0；否则按以下方式计算：

$$MRR@K = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Rank(v_t)}, \tag{22}$$

其中， v_t 是目标物品， N 是数据集中的测试序列数量。

P@K 用于衡量目标物品是否包含在前-k 个推荐物品中，其计算公式为

$$P@K = \frac{n_h it}{N} \tag{23}$$

3）实现： 我们的模型 [3] 使用 PyTorch 1.12.1 [37] 和 CUDA 10.2 实现。在测试阶段，我们将会话最后一个时间戳与测试物品时间戳之间的间隔作为输入的一部分，以获得推荐列表。这一设置不同于其他基线模型，因为它们无法处理时间信息。事实上，这一设置符合行业实际情况，因为我们的模型能够在用户登录网站后立即提供推荐，而且我们可以轻松获取实时区间。

B. 性能比较（*RQ1*）

为证明 TA-HGAT 的有效性，我们在两个公开数据集上进行了实验，并将该模型与十种最先进的基线模型进行比较。

[1]https://www.kaggle.com/datasets/chadgostopp/recsys-challenge-2015
2https://competitions.codalab.org/competitions/11161[3]数据集和代码将在论文录用后开放

表 II Diginetica 和 Yoochoose 数据集上的实验：基于排名列表前 20 项的平均倒数排名（MRR@20）和精确率（P@20），比较 TA-HGAT 与十种基线模型。每个数据集和指标中，粗体和带下划线的数字分别表示最佳和次佳结果。“改进”指相较于所有基线模型的最小改进幅度。

模型	Diginetica		Yoochoose 1/64		Yoochoose 1/4	
	MRR@20	P@20	MRR@20	P@20	MRR@20	P@20
S-POP	13.68	21.06	18.35	30.44	17.75	27.08
FPMC	8.92	31.55	15.01	45.62	-	-
GRU4REC	8.33	29.45	22.89	60.64	22.60	59.53
NARM	16.17	49.70	28.63	68.32	29.23	69.73
STAMP	14.32	45.64	29.67	68.74	30.00	70.44
SR-GNN	17.59	50.73	30.94	70.57	31.89	71.36
TAGNN	18.03	51.31	31.12	71.02	32.03	71.51
HCGR	18.51	52.47	31.46	71.13	32.39	71.66
NISER+	18.72	53.39	31.61	71.27	31.80	71.80
SGNN-HN	<u>19.45</u>	<u>55.67</u>	<u>32.61</u>	<u>72.06</u>	<u>32.55</u>	<u>72.85</u>
TA-HGAT	**19.73**	**56.28**	**32.90**	**72.75**	**32.94**	**73.56**
改进	1.44%	1.10 %	0.89%	0.96%	1.20%	0.97%

1）基线模型：

- **S-POP** 将每个会话中最受欢迎的物品作为推荐列表。
- **FPMC** [38] 是一种用于序列推荐的马尔可夫链方法，由于无法获取用户特征，因此在基于会话的推荐中仅使用物品序列。
- **GRU4REC** [39] 是首个将循环神经网络应用于基于会话的推荐、以学习物品序列依赖关系的工作。
- **NARM** [5] 利用注意力机制，通过全局和局部编码器对序列行为及用户的主要意图进行建模。
- **STAMP** [6] 采用注意力与记忆机制来学习用户偏好，并将会话中的最后一个物品视为近期意图，以生成推荐结果。
- **SR-GNN** [7] 是首个将会话建模为图的工作。它采用门控图神经网络来学习会话中的复杂物品转移。
- **TAGNN** [12] 通过针对不同目标物品学习兴趣表示向量来改进 SR-GNN，从而提升模型性能。
- **NISER+** [40] 通过 L2归一化和丢弃法缓解过拟合问题，从而应对基于会话的推荐中的长尾问题。
- **SGNN-HN** [41] 采用星形图神经网络来考虑彼此没有直接连接的物品。
- **HCGR** [15] 在双曲空间中对会话图进行建模，并利用多行为信息提升性能。在我们的实验中，由于数据集未提供行为信息，且我们针对的是更一般的场景，因此没有使用行为信息。

2）结果分析：对比研究的完整实验结果如表 II 所示。从结果中，我们得出以下观察：

- 我们提出的 TA-HGAT 在所有数据集和指标上均优于所有基线模型

这证明了该模型的有效性。此外，另一种基于双曲图的会话推荐模型 HCGR 的表现优于基于图的会话推荐模型 SR-GNN，但不及我们的模型。与 HCGR 相比，我们的模型在三个数据集上的 MRR@20 和 P@20 指标平均分别提升了 4.3% 和 4.1%。HCGR 优于 SR-GNN，这表明双曲嵌入与会话图相契合。而 TA-HGAT 相较于 HCGR 的性能提升，则体现了时间信息在会话行为推荐任务中的重要性。

- 从表 II 中还可以观察到，与 Yoochoose 相比，我们的模型在 Diginetica 数据集上的表现更好。平均而言，TA-HGAT 在 Diginetica 上的表现分别在 MRR@20 和 P@20 指标上优于 Yoochoose 37.8% 和 14.0%。这一现象可能源于物品的初始特征。在 Diginetica 中，每个物品都有其类别标签，我们将这一特征转换为独热向量，作为该物品的初始嵌入。在 HCGR 中，我们将初始特征建模为双曲空间中的特征向量，如公式 10 和 11 所示。Diginetica 与 Yoochoose 之间的性能差异表明，双曲嵌入对物品特征具有更强的表达能力。

C. 消融实验（*RQ2*）

在 TA-HGAT 中，时间感知双曲注意力包含两个主要模块：带有时间间隔的双曲自注意力，以及结合用户当前兴趣的双曲软注意力。本节分别评估这两个模块的有效性，以展示与移除这两个模块的消融模型相比所带来的提升。

我们设置了四种独立的消融模型，以比较各注意力层的有效性。第一种消融模型是*no-att*，我们移除了所有注意力层，仅直接执行聚合操作。第二种和第三种分别是*self-att* 和*soft-att*，这两种消融模型分别只包含自注意力层和软注意力层。第四种是*TA-HGAT*，即完整模型。图2展示了这些消融模型在 Diginetica 和 Yoochoose 数据集上的对比结果。从图2中，我们观察到以下结果：

- 在 Diginetica 和 Yoochoose 数据集上，non-att 的表现最差，而 TA-HGAT 的表现最好。结果表明注意力层是有效的。这是因为 TA-HGAT 充分利用了时间信息。与不含时间信息的图神经网络相比，我们的模型建立了物品之间的时间间隔关系，并且还考虑了用户当前兴趣。因此，丰富的信息有助于模型取得更好的结果。

- 自注意力的效果优于软注意力，这说明时间间隔相较于用户当前兴趣更有意义。这一现象可能是因为用户当前兴趣更为复杂，因此

表 III 不同损失函数的性能比较。

Loss	Diginetica		Yoochoose 1/64		Yoochoose 1/4	
	MRR@20	P@20	MRR@20	P@20	MRR@20	P@20
Softmax	19.38	55.81	32.67	72.10	32.72	72.95
BPR	19.43	55.97	32.53	72.28	32.66	72.84
TA-HGAT	**19.73**	**56.28**	**32.90**	**72.75**	**32.94**	**73.56**

最后一个物品无法充分代表这些兴趣。相比之下，时间间隔是一种更直观的特征，因此我们提出的双曲自注意力层能够有效处理这一信息。

D. 损失函数比较（*RQ3*）

本节将我们提出的双曲演化损失与传统损失函数进行比较，即 BPR [36] 和 softmax 损失 [7]。由于模型输出中学习到的会话嵌入位于双曲空间，因此在应用 BPR 和 softmax 损失之前，我们需要使用对数映射将嵌入投影回欧几里得空间。

比较结果如表 III 所示。表中将双曲演化损失记为 TA-HGAT。从表中可以看出，BPR 和 softmax 损失的性能相近，但与其他损失相比，我们提出的双曲演化损失有明显提升。这一观察结果表明，在为双曲空间设计的会话行为推荐任务模型中，考虑特定时间戳是有效的。

E. 超参数分析（*RQ4*）

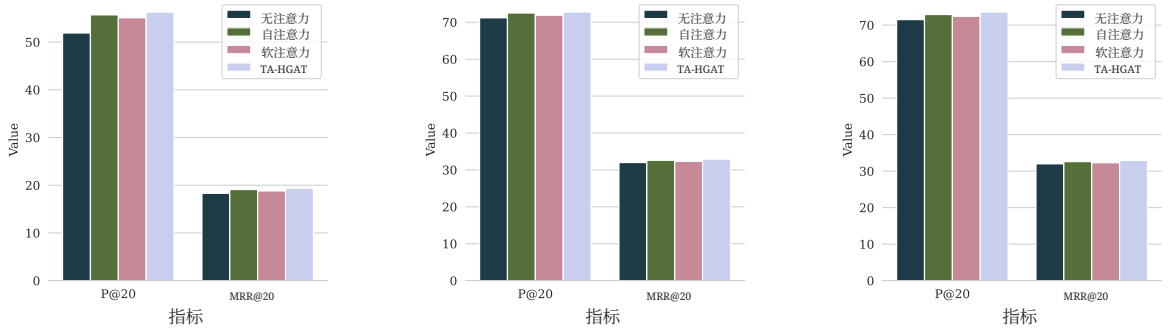
嵌入维度是我们提出模型中的一个超参数，因此本节测试不同嵌入维度的影响。嵌入维度的取值范围为 20 到 100。超参数分析结果如图 3 所示。

可以看出，合适的嵌入维度对于学习物品和会话表示至关重要。从图 3 可以看出，Diginetica 和 Yoochoose 1/64 在嵌入维度为 60 时均取得最佳性能，而 Yoochoose 1/4 的最佳结果出现在嵌入维度为 80 时。由于 Yoochoose 1/4 的规模远大于另外两个数据集，这表明规模更大的数据集需要更大的嵌入空间。

V. 相关工作

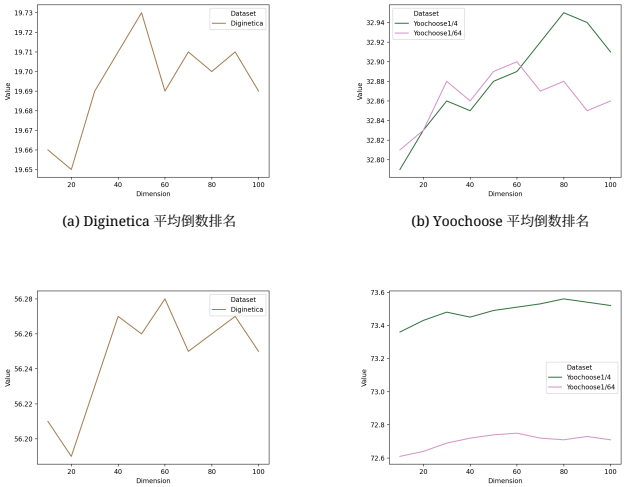
A. 双曲空间

近期研究表明，许多类型的复杂数据都呈现出高度非欧几里得结构 [19]。在许多领域，例如自然语言 [42],计算机视觉 [43],和医疗保健 [44]，，数据通常具有树状结构，或可用层次结构表示。由于这类数据包含内在的层次结构，因此难以在欧几里得空间中捕捉此类表示。为解决这一问题，当前研究越来越关注在黎曼空间中构建神经网络，例如



(a) Diginetica (b) Yoochoose 1/64 (c) Yoochoose 1/4

图 2。TA-HGAT 的消融实验。“non-att”表示不含注意力层的模型。“Self-att”和“Soft-att”分别仅由自注意力和软注意力层组成。TA-HGAT 是完整模型。



(a) Diginetica 平均倒数排名 (b) Yoochoose 平均倒数排名

(c) Diginetica 精确率 (d) Yoochoose 精确率

图 3. 嵌入维度的超参数分析。

双曲空间，它是一种具有恒定负曲率的齐性空间 [27]。与欧几里得空间相比，双曲空间中的球体积随半径呈指数增长，而非多项式增长。凭借强大的表示能力，双曲空间已应用于许多领域。例如， [45]以无监督方式从文本语料库中学习双曲空间中的词嵌入和句子嵌入。 [46]表明，双曲嵌入有益于视觉数据。 [47]提出了双曲图卷积神经网络，将图卷积网络的表达能力与双曲几何相结合，以学习图表示。这些研究展现了双曲空间在学习复杂数据的层次结构方面的潜力和优势。

鉴于双曲空间在这些领域中的表现，研究人员自然会想到将双曲学习应用于推荐系统。 [48] 论证了在神经推荐系统中使用双曲表示的合理性。 [49] 提出了 HyperML，通过度量学习方法弥合推荐系统中欧几里得几何与双曲几何之间的差距。 [21] 提出了一种用于协同过滤的双曲图卷积网络模型。 [50] 提出了 HyperSoRec，这是一种结合多方面学习的新型社交推荐图神经网络（GNN）框架。

B. 基于会话的推荐

基于会话的推荐（SBR）因其在建模用户当前兴趣方面的有效性，日益受到工业界和学术界的关注。在近期的 SBR 研究中，主要有三类将深度学习应用于 SBR 并取得最先进性能的方法，分别是基于序列的方法 [4], [51]，基于注意力的方法 [5], [6]和基于图的模型 [7]。GRU4REC [4]是基于序列的 SBR 模型中最具代表性的工作。它采用 RNN 的一种变体 GRU 对物品序列进行建模，并进行下一物品预测。在 GRU4REC 之后，一些其他论文 [39], [51], [52] 通过数据增强、层次结构和 top-k增益对其进行改进。基于注意力的模型旨在学习会话中不同物品的重要性，并使模型聚焦于重要物品。NARM [5] 利用注意力机制对会话的局部和全局特征进行建模，以学习用户兴趣。STAMP [6] 结合注意力模型和记忆网络，以学习会话的短期偏好。基于图的模型将物品连接成图，以学习物品之间的复杂转移。SR-GNN [7] 是首个将会话建模为图的工作。它利用门控图神经网络（GGNN）对会话图进行建模，并取得了最先进的性能。在 SR-GNN 的基础上，[10], [12]通过注意力层改进 SR-GNN。 [8], [11] 在模型中考虑会话图中的物品顺序。 [9], [13], [14]将全局物品关系、物品类别和用户表示等额外信息纳入考量，以设计更全面的模型。然而，所有这些方法都未能考虑会话图的层次几何和时间信息。

C. 基于GNN的推荐模型

图神经网络已被证明在不同研究领域中发挥作用[53]–[58]。也有许多研究在推荐系统的数据建模中考虑图结构。一般而言，关于图结构和嵌入空间，主要有两种方法。一种是在欧几里得空间中对用户-物品交互图进行建模。其中，[59]–[61]在用户-物品图上执行图卷积，以探索用户与物品之间的交互。[62]–[64]利用图神经网络中的逐层邻域聚合来捕获高阶连接。[65] 分别预训练用户-用户图和物品-物品图，以学习用户-物品交互图的初始嵌入。[35] 利用动态图和演化损失对用户-物品交互的变化进行建模。这些工作应用图神经网络从高维图数据中学习，并生成低维节点嵌入，无需特征工程，但学得的嵌入都位于欧几里得空间中，而有些图数据在表征学习中可能更适合采用其他几何空间。

另一种方法是在双曲空间中对推荐图进行建模，以学习层次几何。HGCF [21]利用用户—物品图，通过双曲图卷积网络学习节点嵌入。Wang等人 [20] 提出了一种全双曲图卷积网络，其中所有操作均在双曲空间中进行。Xu等人 [32] 在知识图谱中对商品图进行建模，并在双曲空间中学习节点嵌入。HCGR[15]是一种新颖的双曲对比图表征学习方法，用于实现基于会话的推荐。这些模型都没有利用会话图中的时间相关信息来提高推荐准确率。本文提出了一种新颖的框架，将时间感知图注意力机制融入双曲空间，专门用于基于会话的推荐。

VI. 结论

基于会话的推荐（SBR）旨在根据用户以往的会话预测其接下来感兴趣的物品。现有研究对会话中的图结构进行建模，并取得了最先进的性能。然而，这些研究未能考虑会话中的层次几何和时间信息。本文提出TA-HGAT，这是一种基于双曲图神经网络的模型，考虑了物品之间的时间间隔以及用户当前兴趣。实验结果表明，在两个真实世界数据集上，TA-HGAT的表现优于其他基于会话的推荐模型。

未来的工作中，我们将把模型扩展到更通用的推荐系统。时间间隔不仅存在于基于会话的推荐问题中，也存在于几乎所有推荐系统中。因此，我们希望测试模型在其他推荐问题上的表现，例如下一篮子推荐和兴趣点推荐，在这些问题中，时间信息对于提供推荐结果起着至关重要的作用。

VII. 致谢

本研究部分得到了美国国家科学基金会资助，项目编号为III-1763325、III-1909323、III-2106758 和 SaTC-1930941。

参考文献

[1] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, 《用于点击率预测的深度兴趣网络》，载于 SIGKDD。ACM，2018 年，第 1059–1068 页。[2] A. Van den Oord, S. Dieleman, and B. Schrauwen, 《深度内容音乐推荐》，《神经信息处理系统进展》，第 26 卷，2013 年。[3] S. Kumar, X. Zhang, and J. Leskovec, 《预测时序交互网络中的动态嵌入轨迹》，载于 SIGKDD。ACM，2019 年，第 1269–1278 页。[4] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, 《基于循环神经网络的会话推荐》，arXiv 预印本 arXiv:1511.06939，2015 年。[5] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, 《神经注意力基于会话的推荐》，载于 CIKM。ACM，2017 年，第 1419–1428 页。[6] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, 《STAMP：用于基于会话的推荐的短期注意力/记忆优先模型》，载于 SIGKDD。ACM，2018 年，第 1831–1839 页。[7] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, 《基于图神经网络的会话推荐》，载于 AAAI 人工智能会议论文集，第 33 卷，第 01 期，2019 年，第 346–353 页。[8] T. Chen and R. C.-W. Wong, 《处理图神经网络在基于会话的推荐中的信息损失》，载于 SIGKDD。ACM，2020 年，第 1172–1180 页。[9] Z. Wang, W. Wei, G. Cong, X.-L. Li, X.-L. Mao, and M. Qiu, 《全局上下文增强的基于会话的推荐图神经网络》，载于 SIGIR。ACM，2020 年，第 169–178 页。[10] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou, 《用于基于会话的推荐的图上下文化自注意力网络》，载于 IJCAI，第 19 卷，2019 年，第 3940–3946 页。[11] R. Qiu, J. Li, Z. Huang, and H. Yin, 《重新思考基于图神经网络的会话推荐中的物品顺序》，载于 CIKM。ACM，2019 年，第 579–588 页。[12] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, and T. Tan, 《TAGNN：用于基于会话的推荐的目标注意力图神经网络》，载于 SIGIR。ACM，2020 年，第 1921–1924 页。[13] L. Liu, L. Wang, and T. Lian, 《Case4SR：利用类别序列图增强基于会话的推荐》，《Knowledge-Based Systems》，第 212 卷，文章编号 106558，2021 年。[14] Y. Pang, L. Wu, Q. Shen, Y. Zhang, Z. Wei, F. Xu, E. Chang, B. Long, and J. Pei, 《用于个性化基于会话的推荐的异构全局图神经网络》，载于 WSDM。ACM，2022 年，第 775–783 页。[15] N. Guo, X. Liu, S. Li, Q. Ma, Y. Zhao, B. Han, L. Zheng, K. Gao, and X. Guo, 《HCGR：用于基于会话的推荐的双曲对比图表征学习》，arXiv 预印本 arXiv:2107 05366，2021 年。。[16] J. Li, Y. Wang, and J. McAuley, 《用于序列推荐的时间间隔感知自注意力》，载于 WSDM，2020 年，第 322–330 页。[17] W. Ye, S. Wang, X. Chen, X. Wang, Z. Qin, and D. Yin, 《时间至关重要：结合复杂时序信息的序列推荐》，载于 SIGIR，2020 年，第 1459–1468 页。[18] R. C. Wilson, E. R. Hancock, E. Pekalska, and R. P. Duin, 《数据的球面与双曲嵌入》，《IEEE模式分析与机器智能汇刊》，第 36 卷，第 11 期，第 2255–2269 页，2014 年。[19] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, 《几何深度学习：超越欧几里得数据》，《IEEE信号处理杂志》，第 34 卷，第 4 期，第 18–42 页，2017 年。[20] L. Wang, F. Hu, S. Wu, and L. Wang, 《用于推荐的全双曲图卷积网络》，载于 CIKM，2021 年，第 3483–3487 页。[21] J. Sun, Z. Cheng, S. Z uberi, F. P´erez, and M. Volkovs, 《HGCF：用于协同过滤的双曲图卷积网络》，载于 2021 年 Web会议，2021 年，第 593–601 页。[22] T. N. Kipf and M. Welling, 《基于图卷积网络的半监督分类》，ICLR，2016 年。[23] W. Hamilton, Z. Ying, and J. Leskovec, 《大型图上的归纳式表示学习》，载于《神经信息处理系统进展》，2017 年，第 1024–1034 页。[24] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, 《利用可微池化进行层次化图表示学习》，《神经信息处理系统进展》，第 31 卷，2018 年。

[25] P. Veli ˇckovi ´c, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Ben- gio, 《图注意力网络》，arXiv 预印本 arXiv:1710.10903，2017 年。[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, 《注意力就是你所需要的一切》，《神经信息处理系统进展》，第 30 卷，第 5998–6008 页，2017 年。[27] M. Nickel and D. Kiela, 《用于学习层次表示的庞加莱嵌入》，《神经信息处理系统进展》，第 30 卷，2017 年。´C. Gulcehre, M. Denil, M. Malinowski, A. Razavi, R. Pascanu, K. M. Hermann, P. Battaglia, V. Bapst, D. Raposo, A. Santoro 等, 《双曲注意力网络》，arXiv 预印本 arXiv:1805.09786，2018 年。[28] M. Nickel and D. Kiela, 《在双曲几何的洛伦兹模型中学习连续层次结构》，载于《国际机器学习会议》。PMLR，2018 年，第 3779–3788 页。[29] Q. Liu, M. Nickel, and D. Kiela, 《双曲图神经网络》，《神经信息处理系统进展》，第 32 卷，2019 年。[30] O. Ganea, G. B[31] ecigneul, and T. Hofmann, 《双曲神经网络》，《神经信息处理系统进展》，第 31 卷，2018 年。´D. Xu, C. Ruan, E. Korpeoglu, S. Kumar, and K. Achan, 《面向电子商务的商品知识图谱嵌入》，载于 WSDM。ACM，2020 年，第 672–680 页。[32] Y. Li, H. Chen, X. Sun, Z. Sun, L. Li, L. Cui, P. S. Yu, and G. Xu, 《用于序列推荐的双曲超图》，载于 CIKM，2021 年，第 988–997 页。[33] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, 《用于建模多关系数据的平移嵌入》，载于《神经信息处理系统进展》，2013 年，第 2787–2795 页。[34] X. Li, M. Zhang, S. Wu, Z. Liu, L. Wang, and S. Y. Philip, 《动态图协同过滤》，载于 ICDM。IEEE，2020 年，第 322–331 页。[35] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, 《BPR：基于隐式反馈的贝叶斯个性化排序》，arXiv 预印本 arXiv:1205.2618，2012 年。[36] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga 等, 《PyTorch：一种命令式风格的高性能深度学习库》，《神经信息处理系统进展》，第 32 卷，2019 年。[37] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, 《用于下一篮子推荐的个性化马尔可夫链分解》，载于第 19 届万维网国际会议论文集，2010 年，第 811–820 页。[38] B. Hidasi and A. Karatzoglou, 《基于会话推荐的循环神经网络与 Top-k 增益》，载于 CIKM，2018 年，第 843–852 页。[39] P. Gupta, D. Garg, P. Malhotra, L. Vig, and G. M. Shroff, 《NISER：结合图神经网络的归一化物品与会话表示》，arXiv 预印本 arXiv:1909.04276，2019 年。[40] Z. Pan, F. Cai, W. Chen, H. Chen, and M. de Rijke, 《用于基于会话的推荐的星形图神经网络》，载于 CIKM，2020 年，第 1195–1204 页。[41] A. Tifrea, G. B[42] ecigneul, and O.-E. Ganea, 《庞加莱 GloVe：双曲词嵌入》，arXiv 预印本 arXiv:1810.06546，2018 年。´M.-M. Yau and S. N. Srihari, 《用于多维数字图像的层次数据结构》，《ACM通讯》，第 26 卷，第 7 期，第 504–515 页，1983 年。\S. J. Wilkens, J. Janes, and A. I. Su, 《HIERS：利用拓扑化学图进行层次骨架聚类》，《药物化学杂志》，第 48 卷，第 9 期，第 3182–3193 页，2005 年。[43] B. Dhingra, C. Shallue, M. Norouzi, A. Dai, and G. Dahl, 《在双曲空间中嵌入文本》，载于第十二届自然语言处理图方法研讨会（TextGraphs-12）论文集，2018 年，第 59–69 页。[44] V. Khrulkov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempit-sky, 《双曲图像嵌入》，载于 CVPR，2020 年，第 6418–6428 页。[47] I. Chami、Z. Ying、C. R´e 和 J. Leskovec，“双曲图卷积神经网络”，神经信息处理系统进展，第 32 卷，2019 年。[48] B. P. Chamberlain、S. R. Hardwick、D. R. Wardrope、F. Dzogang、F. Daolio 和 S. Vargas，“可扩展双曲推荐系统”，*arXiv* 预印本 *arXiv:1902.08648*，2019 年。[49] L. Vinh Tran、Y. Tay、S. Zhang、G. Cong 和 X. Li，“HyperML：一种面向推荐系统、在双曲空间中提升度量学习的方法”，发表于*WSDM*，2020 年，第 609–617 页。[50] H. Wang、D. Lian、H. Tong、Q. Liu、Z. Huang 和 E. Chen，“Hy-persorec：利用多方面的双曲用户和物品表示实现社交感知推荐”，*ACM*信息系统汇刊 *(TOIS)*，第 40 卷，第 2 期，第 1–28 页，2021 年。[51] Y. K. Tan、X. Xu 和 Y. Liu，“改进用于基于会话推荐的循环神经网络”，发表于第 *1* 届推荐系统深度学习研讨会论文集，2016 年，第 17–22 页。[52] M. Quadrana、A. Karatzoglou、B. Hidasi 和 P. Cremonesi，“利用分层循环神经网络个性化基于会话的推荐”，发表于第十一届 *ACM*推荐系统会议论文集，2017 年，第 130–137 页。[53] H. Peng、J. Li、Q. Gong、Y. Song、Y. Ning、K. Lai 和 P. S. Yu，“利用异构图卷积网络进行细粒度事件分类”，*arXiv* 预印本 *arXiv:1906.04580*，2019 年。[54] Z. Liu、X. Li、Z. You、T. Yang、W. Fan 和 P. Yu，“通过多关系双曲图神经网络改进医疗分诊聊天机器人诊断”，发表于*SIGIR*。ACM，2021 年，第 1965–1969 页。[55] Y. Dou、Z. Liu、L. Sun、Y. Deng、H. Peng 和 P. S. Yu，“增强图神经网络欺诈检测器抵御伪装欺诈者的能力”，发表于 *CIKM*。ACM，2020 年，第 315–324 页。[56] Z. Liu、X. Li、H. Peng、L. He 和 S. Y. Philip，“电子健康记录上的异构相似性图神经网络”，发表于*2020 IEEE*国际大数据会议（*Big Data*）。IEEE，2020 年，第 1196–1205 页。[57] J. Li、H. Peng、Y. Cao、Y. Dou、H. Zhang、P. Yu 和 L. He，“高阶属性增强型异构图神经网络”，*IEEE*知识与数据工程汇刊，2021 年。[58] Y. Hei、R. Yang、H. Peng、L. Wang、X. Xu、J. Liu、H. Liu、J. Xu 和 L. Sun，“Hawk：通过异构图注意力网络快速检测安卓恶意软件”，*IEEE*神经网络与学习系统汇刊，2021 年。[59] L. Zheng、C.-T. Lu、F. Jiang、J. Zhang 和 P. S. Yu，“谱协同过滤”，发表于第 *12* 届 *ACM*推荐系统会议论文集，2018 年，第 311–319 页。[60] R. v. d. Berg、T. N. Kipf 和 M. Welling，“图卷积矩阵补全”，*arXiv* 预印本 *arXiv:1706.02263*，2017 年。[61] W. Yu 和 Z. Qin，“利用低通协同滤波器进行推荐的图卷积网络”，发表于*ICML*。PMLR，2020 年，第 10 936–10 945 页。[62] X. Wang、X. He、M. Wang、F. Feng 和 T.-S. Chua，“神经图协同过滤”，发表于*SIGIR*。ACM，2019 年，第 165–174 页。[63] X. He、K. Deng、X. Wang、Y. Li、Y. Zhang 和 M. Wang，“LightGCN：简化并增强用于推荐的图卷积网络”，发表于 *SIGIR*。ACM，2020 年，第 639–648 页。[64] Z. Liu、X. Li、Z. Fan、S. Guo、K. Achan 和 S. Y. Philip，“利用多意图翻译图神经网络进行购物篮推荐”，发表于*2020 IEEE*国际大数据会议（*Big Data*）。IEEE，2020 年，第 728–737 页。[65] X. Li、Z. Liu、S. Guo、Z. Liu、H. Peng、S. Y. Philip 和 K. Achan，“通过强化注意力多关系图神经网络预训练推荐系统”，发表于*2021 IEEE*国际大数据会议（*Big Data*）。IEEE，2021 年，第 457–468 页。