

• ABSTRACT

- 基于流式会话的推荐（SSR）是一项具有挑战性的任务，需要推荐系统在流式场景中执行基于会话的推荐（SR）。
- 在电子商务和社交媒体的实际应用中，在一定时间内生成的一系列用户项目交互被分组为一个会话，并且这些会话以流的形式连续到达。
- 最近的SR研究大多集中在静态设置上，首先获取训练数据，然后用于训练基于会话的推荐模型。他们需要对整个数据集进行几次训练，这在流式环境中是不可行的。此外，由于对用户信息的忽视或简单使用，它们很难很好地捕捉到用户的长期兴趣。
- 在本文中，作者提出了一个全局属性图（GAG）神经网络模型与Wasserstein库的SSR问题。
 - 当新会话到达时，基于当前会话及其关联用户构造具有全局属性的会话图。因此，GAG可以同时考虑全局属性和当前会话，以了解会话和用户的更全面表示，从而在推荐中产生更好的性能。
 - 为了适应流会话场景，建议使用Wasserstein库来帮助保存历史数据的代表性草图。

• INTRODUCTION

- 基于会话的推荐（SR），其目的是基于会话向用户推荐项目。然而，它们大多侧重于静态设置，这不适合于实际情况。在实践中，会话是作为流动态生成的，这就迫切需要基于流会话的推荐（SSR）。
 - SSR的一般过程是使用历史会话训练推荐模型以保持用户的长期兴趣，然后使用流会话进行在线更新以适应其最近的偏好。
 - 目前对SR的大多数研究都集中在静态场景上，其中推荐模型以批处理方式进行训练。由于用户的偏好随着时间的推移而变化，因此不可能为即将到来的新会话应用静态模型。
 - 由于内存空间的限制，通过吸收每一个新会话来叠加训练数据是不现实的。
 - 训练时间将是流媒体场景中的另一个关注点，在流媒体场景中，推荐模型预计会迅速更新。一个典型的随机共振模型需要长时间的训练才能收敛，这在短时间内是无法保证的。
- 有几种方法利用储层技术来进行流水作业。例如，文中图一的交互数据以相同的概率存储在库中，然后采样用于模型的在线训练。
 - 这种处理流式会话的方式会导致SSR模型遭受会话信息的丢失，因为数据样本作为离散交互进行存储和绘制。
 - 传统流式任务的库被设计为捕获矩阵分解信息，而不是会话的序列模式。
 - 在线学习很难适应新数据的基于会话的推荐任务。对于在线学习，当新课程开始时，模型将进行相应更新，以捕捉最新课程中最近的过渡模式。
 - 这些模型很容易过度拟合新数据，无法维持用户从历史数据中获得的长期偏好。
 - SSR模型有效利用用户信息，从而获得长期和短期偏好的综合表示。
- 最近，Guo通过评估每个会话的信息量，采用加权抽样方案将存储技术应用于SSR任务。

- 由于该方法需要使用预先计算的项目特征向量为会话中的每个项目生成信息性得分，所以该方法不能推广到其他模型。
- 该模型直接将基于会话的推荐方法与矩阵分解模块相结合，很难了解SSR问题中用户与项目之间的复杂相关性。
- 为了解决上述问题，作者提出了一个全局属性图（GAG）神经网络模型，并将Wasserstein库作为SSR的解决方案。
 - 为了最大限度地利用用户嵌入并维护SSR的长期偏好信息，作者首先将用户的会话序列转换为会话图，其中用户嵌入作为与交互项嵌入相关联的全局属性。
 - 基于全局属性会话图，GAG模型执行图卷积以学习更新的全局属性，并将其传递给排名模块以输出推荐列表。
 - 在GAG模型中，全局属性被应用于有效协助整个会话和会话内项目的联合代表学习。
 - 为了开发SSR问题的通用库，作者提出了Wasserstein库，它根据生成的推荐列表和用户真实交互之间的Wasserstein距离存储和采样会话数据。
 - 在采样过程中，Wasserstein库对推荐结果具有较高Wasserstein距离的会话进行采样。

• 总结

- 提出了GAG模型，通过将用户嵌入视为会话图的全局属性，有效地记忆用户的长期偏好，并将其纳入SSR的嵌入向量中，从而在学习表示时实现更高的表达能力。
- Wasserstein库旨在主动选择信息量最大的训练集，以便在流环境中更新模型。此外，Wasserstein库是一种有效但通用的在线学习方法，可轻松应用于其他流式会话数据。
- 在两个真实数据集上的大量实验结果表明，所提出的GAG模型和Wasserstein库达到了迄今为止最好的性能。

• RELATED WORK

• Session-based Recommendation

- 顺序推荐主要基于马尔可夫链模型，该模型学习序列数据中项目的相关性。
 - Zim-dars等人利用概率决策树模型，提出对项目转换模式的状态进行编码。
 - Shani等人利用马尔可夫决策过程（MDP）计算项目转移概率。
- 深度学习模型随着递归神经网络的繁荣而流行。
 - Hidasi等人提出了GRU4REC，它应用GRU将数据作为时间序列处理。最近的一些方法使用注意机制来避免时间顺序。
 - NARM将GRU堆叠为编码器以提取信息，然后将自我关注层汇总为会话嵌入。
 - 为了进一步缓解时间序列的偏差，STAMP用注意层替换了循环编码器。
 - 一些方法利用GNN对会话信息进行编码，以防止对会话顺序的误导。
 - SSRM考虑特定用户的历史会话，并应用注意机制将它们结合起来。

• Streaming Recommendation

- 在线学习侧重于使用新数据更新旧模型，以获取用户最近的兴趣。
 - He等人提出了一种替代最小二乘法的元素方法来处理缺失的数据。

- Jugovac等人应用了基于重播的评估协议，用新闻推荐中的新到达事件和文章更新模型。
 - 尽管上面的模型通过使用新的交互更新模型来捕获用户最近的兴趣，但它们无法记住历史交互。
- 随机抽样是一种通过引入存储库来存储用户的长期交互来解决忽略历史问题的技术。
 - Diaz Aviles等人在矩阵分解方法上应用了基于主动学习原则的抽样策略来更新模型。
 - 郭等人使用相同的reservoir技术来处理流媒体会话。
- Graph Neural Networks
 - GNN基本上是在一个简单的情况下应用于有向图。
 - 近年来，许多GNN方法的工作原理与消息传递网络非常相似，它们在节点的邻域上执行聚合以计算节点嵌入。后来，全局属性被引入GNN层，以维护物理系统中的图形级功能。
- METHOD
 - 问题解释
 - 推荐系统需要基于 $\{S_u, 0, S_u, 1, \dots, S_u, t\}$ 推荐项目 v_{t+1} ，这是用户从历史到当前的所有会话。项目 v_{t+1} 应最符合用户的偏好。同时，会话以高速到达，这意味着计算资源仅限于计算。因此，算法应该能够有效地处理历史会话和当前会话。通常，只向用户推荐排名前K的项目。
 - 概述
 - 作者提出了一种新的全局属性图（GAG）神经网络模型，主要通过将用户信息转换为全局属性并将其合并到会话图中来解决SSR问题。
 - GAG模型的架构包括两个关键组件：用于生成推荐的GAG模型和用于流数据学习的Wasserstein库。
 - 全局属性会话图
 - 在第一阶段，会话序列被转换成具有全局属性的会话图，以便通过GNN处理每个会话。由于会话序列的自然顺序，作者将其转换为加权有向图。
 - 此外，作者将用户的一般信息作为全局属性 u 合并到会话图中， $G_s = (u, V_s, E_s)$ ， $G_s \in G$ 、这是所有的会议记录。
 - 边的权重定义为该边在会话中出现的频率。
 - 全局属性图神经网络
 - 通过构造全局属性会话图 G_s ，提出了GAG模型，利用节点特征、边缘特征和全局属性对 G_s 进行图卷积。当GAG模型以会话图作为输入时，计算将从边（节点）开始，然后进入全局属性。
 - 推荐
 - GAG执行推荐的最后一个阶段是基于输入会话和用户配置文件的表示生成候选项。计算每个项目的分数并形成分数向量 $\hat{z} \in R_n$ ，其中 n 是项目集的大小。
 - 用于流模型的Wasserstein库

- 作者的脱机模型扩展到流媒体设置。其目的是用新到达的会话数据更新模型，同时保留从历史会话中学习到的知识。
- 传统上，在线学习方法仅使用新数据更新模型，这总是会导致忽略过去的数据。
 - 为了防止模型失去历史数据，作者利用reservoir来保持对历史数据的长期记忆。
 - 储层技术广泛应用于流数据库管理系统中。
- 应用储层的目的是维护所有历史数据的代表性草图。
 - 作者对历史数据进行随机抽样，以选择存储在水库中的数据。
 - 这种生成存储库的方法实际上是从当前数据集中随机采样，它可以成功地维持模型的长期记忆。
- 尽管reservoir库可以按照上面介绍的方式进行更新，但随着时间的推移，包含新到达会话的概率往往较小，reservoir库有可能忽略最近的数据。
 - 最近的数据对于预测用户的不同偏好至关重要。此外，新用户和新项目不断暴露于系统中。因此，当新会话数据 C_{new} 到达时，我们使用 C_{new} 和库 C 更新预先训练的模型。
- 上述储层取样程序使模型能够根据新旧数据不断更新。然而，它在现实中只能勉强实现良好的性能。
 - 原因是 M 已经很好地学习了 C 中的大多数训练序列，这导致 C_{rand} 主要包含无用的训练样本。
 - 实际上，如果当前模型对会话的预测较差，则更值得使用此会话更新模型，因为它要么包含用户的最新偏好，要么存在一些当前模型无法很好地了解的项目转换模式。这种会话称为模型的信息会话，该会话对模型更新更为重要。
 - 因此，我们提出了一种Wasserstein库建设策略，该策略对会话进行采样，该会话在推荐项上的输出概率分布与用户的真实交互之间具有更高的Wasserstein距离，并且具有更高的概率。
 - 直观地说，当会话的输出概率具有较高的Wasserstein距离时，与具有较低Wasserstein距离的会话相比，输出将包含更多的信息。

• 训练

- 由于推荐任务被视为整个项目集的分类问题，我们可以在预测的推荐分布 \hat{y} 和实际交互 y 之间应用多类交叉熵损失。

• EXPERIMENT SETUP

• 数据集

- LastFM是一个真实世界的音乐推荐数据集，由Celma Herrada发布。在这项工作中，我们主要关注音乐艺术家推荐。
 - 如郭等人所示，我们也考虑对艺术家提出建议，并选择10000种最流行的。根据时间顺序，作者在8小时内将来自同一用户的事务分组为一个会话。之后，将筛选出包含20个以上或少于2个事务的会话。预处理后总共有298919个会话。
- Gowalla 2是一个从社交网络收集的感兴趣点真实世界数据集，供用户登录。

- 与郭等人相同，30000个最受欢迎的地方用于实验，1天内的签到被定义为一个会话。同样，包含超过20个事务或少于2个事务的会话将被过滤掉。最后，作者在实验期间有198680个会话。

• 度量

- 继Chang等人和Guo等人之后，为了模拟数据到达情况的流化情况，按照所有数据的时间顺序将数据集分为两个比例（60%和40%）。
 - 第一部分定义为训练集（Dtrain），第二部分定义为候选集（Dcandidate）。
 - 具体而言，（Dtrain）用于将GAG模型训练为离线数据。对于（Dcandidate），它是为了模拟在线流式会话数据而设计的。
 - 特别是对于（Dcandidate），它进一步按时间顺序分为五个相同大小的测试集 Dtest1、...、Dtest5。并在测试期间按时间顺序提供给模型。
 - 在当前测试集上进行测试后，模型将根据它进行更新，更新后的模型将用于下一个测试集的测试。这样的在线更新是为流媒体场合而设计的。
 - 为了评估模型的性能，根据用户选择前几个推荐项目的性质，这里应用了前20个推荐项目，主要基于Recall@K和MRR@K。

• 参照物

- 在实验中，作者主要将GAG模型与以下有代表性的参照方法进行比较：
 - POP总是选择所有用户中最受欢迎的项目来推荐给其他用户。这是一个简单而有力的参照物。
 - S-POP推荐当前会话中出现的最流行的项目，而不是整个项目集。
 - BPR-MF是一种主要利用成对排名损失的方法。此外，矩阵分解也进行了修改，以适应中基于会话的建议。
 - GRU4REC利用GRU层学习匿名设置中的会话嵌入。
 - NARM在整个会话项目级别上添加了一个注意层，以便在匿名设置中对会话信息进行编码。
 - FGNN利用GNN学习匿名会话的嵌入以提出建议。该方法不考虑用户信息。
 - SSRM是解决SSR问题的一种最先进的方法，它应用一个库对历史会话进行采样，以帮助当前会话嵌入学习。

• 训练细节

- 在模型的实现中，为了比较的公平性，作者将所有MLP设置为1层，嵌入大小为200。
- 使用学习率为0.003的Adam，并将批量大小设置为100，以训练GAG模型。
- reservoir的尺寸设置为 $|D|/100$ ，每个C上的窗口尺寸设置为 $|C_{new}|/2$ 。

• EXPERIMENT RESULTS

• 与参照方法的比较

- General Comparison
 - 显然，所提出的GAG-50模型在所有情况下都优于所有参照方法。随着嵌入大小增加到200，GAG模型达到了最先进的结果。两者都显示了GAG模型的优越性。

- 传统方法（如POP和S-POP）的性能要差得多，这两种方法都向用户推荐最流行的项目。POP从整个项目集中推荐最受欢迎的项目，而S-POP则选择当前会话中最受欢迎的项目。
- POP修复了推荐列表，该列表无法检测不同会话中用户行为的不同模式。但是，S-POP仍然是一种基于会话的强大参照方法，因为它可以捕获会话中项目的重复出现模式。
- 此外，对于对整个用户项交互矩阵进行矩阵分解的浅层方法BPR-MF，由于能够执行个性化推荐，因此与POP相比具有更高的性能。然而，BPR-MF在SSR问题上仍然无法超越S-POP，因为S-POP可以进一步提取特定于会话的信息。
- 对于深度学习模型，GRU4REC是一种利用GRU将会话作为序列进行处理并输出会话嵌入以提出建议的方法。它在大多数情况下都优于传统方法，这证明了基于深度学习的方法的优越性。
- 此外，与GRU4REC相比，利用注意机制的方法，如NARM和SSRM，获得了很大的改进，这表明注意机制学习项目相互依赖性的能力。特别是，SSRM专门针对SSR问题，是实验中最强的基线。
- 最近，图形神经网络被证明具有很强的结构化数据建模能力。例如，FGNN是SR的一种最先进的方法。在本实验中，我们可以看到FGNN实现了与SSRM相当的性能。与GAG-50相比，FGNN的性能更差，因为它只针对SR设计，不能对用户信息进行建模。
- In-depth Comparison
 - 从结果来看，GAG和GAG-50在更高标准的推荐中仍然具有优势。与基于GNN的方法相比，SSRM在top-5和top-10性能上都有较大的下降，这意味着图结构和GNN更适合于会话表示和泛化能力。相反，注意机制无法区分会话中的项目转换模式。
- 全局属性的影响
 - 消融研究
 - 参照方法
 - FGNN在节点更新层和读出功能（正常GNN模型中的读出功能表示图形级输出功能）中使用GNNLayer（不连接用户信息）。并作为基本的参照方法。
 - GAG-FGNN：作者用FGNN的节点更新层替换节点更新功能，并在GAG中维护全局属性更新功能，以评估用户信息在全局属性更新中的集成。
 - GAG NoGA：在这个变体中，作者在节点更新过程中保留全局属性，而在全局属性更新函数中删除它。
 - 使用全局属性的每个模块都对推荐性能有贡献。一般来说，FGNN是最差的，因为它忽略了全局属性。GAG-FGNN和GAG NoGA都通过引入全局属性进行了改进。GAG-FGNN在全局属性更新函数中使用全局属性，而GAG-NoGA在节点更新函数中合并全局属性。
 - 比较这两个变量，结果证明应用于节点更新过程的全局属性对推荐性能的影响大于其自我更新。
 - Wasserstein库的影响
 - reservoir库由两个主要设计组成：

- 采样过程基于会话的预测分布和真实交互之间的Wasserstein距离；
- 包含新项目或新用户的会话将直接添加到培训样本中。

- 消融研究

- 变体

- GAG静态：这种方法简化了模型的在线训练。
 - GAG Ranui：该变体仅对当前reservoir库和新到达的reservoir库的联合集执行随机采样。这是reservoir库最常见的设计。
 - GAG FixNew：此变体直接将包含新项目或新用户的新会话添加到训练数据中。对于其余部分，它仍然执行随机采样。
 - GAG WassUni：该方法根据Wasserstein距离从当前reservoir库和新会话的联合集中采集训练数据。
 - 对于静态推荐版本gagstatic，它的性能会随着时间的推移而降低，因为用户的偏好会发生变化，并且流会话包含新项目和新用户。
 - 在大多数情况下，随机取样版本变量GAG-Ranui的性能比使用专门储层取样策略的其他方法差。
 - 与GAG-FixNew和GAG-Ranui相比，GAG-FixNew在Gowalla上的性能更好，而在LastFM上的Dtest4和Dtest5的长期预测能力有所下降。
 - 由于数据集这两部分中的新项和新用户不具有代表性，因此模型会被误导。将完整的GAG模型与GAG WassUni进行比较，可以看出新用户和项目的加入有助于在线训练。
 - 此外，还证明了Wasserstein距离的有效性。通过基于Wasserstein距离的抽样，GAG-WassUni在大多数情况下都优于GAG-RanUni。同样，GAG也比GAG-FixNew得分更高，后者随机抽取训练数据。

- 参数敏感度

- 嵌入大小

- 当嵌入大小设置为50或100时，前面的方法可以获得最佳结果。
 - 当嵌入大小设置为200时，GAG模型在所有情况下都具有最高的性能。与50和400相比，100是一个相对强大的变体。当嵌入大小设置为50和400时，它们在各自的方法中不具有代表性且参数化过度，这导致很难训练出强大的模型。

- 层数

- GAG层的数量控制模型的深度。作者用不同层数的：GAG，GAG-2和GAG-3来测试模型。
 - 一般来说，由于梯度爆炸，GNN模型总是受到模型深度增加的影响。
 - 在实验中，GAG模型的性能随着模型的深入而降低，这与通常的观察结果一致。此外，会话的连接性小于传统的图形数据，这也限制了更深层次的GNN模型的能力。

- CONCLUSION

- 在本文中，作者提出了一个带有Wasserstein储备的GAG模型来执行SSR。通过引入全局属性和GAG层来解决如何保护用户的长期利益的问题。

- 作者设计了一个有效且通用的Wasserstein库，它根据会话推荐结果与真实交互之间的Wasserstein距离对会话进行采样。
- 在未来，研究如何将跨会话信息整合到SSR问题中具有重要意义。