

• INTRODUCTION

- 用户通常表现出特定的短期意图。
- 会话数据的自然特征精确地反映了用户的行为模式和当前偏好。
- 根据用户在最近会话中的行为对其当前意图进行建模通常会得到令人满意的推荐结果，这是基于会话的推荐（SR）的基本原则。
- SR系统的目标是根据用户在会话中最近的行为，预测目标用户在下一次交互中将交互的项目。
- SR系统算法
 - 基于马尔可夫模型
 - 基于深度神经网络（DNN）的模型。
 - 例子
 - FPMC的基本思想是基于马尔可夫链计算会话中项目之间的转移概率。
 - 递归神经网络（RNN）以不同的方式应用于学习用户的动态和时移偏好，表现出比传统方法更好的性能。
- SR系统算法存在的问题
 - 缺乏对用户微观行为的利用。
 - 大多数基于会话的模型仅从宏观角度对会话进行建模，即将会话视为一系列项目，而不考虑用户的不同操作。
 - 解决方案
 - 将用户对某个项目的特定操作视为一种微观行为，与项目级的宏观行为（即用户-项目交互）相比，该微观行为具有更细的粒度，并提供了对用户更深入的理解。
 - 项目知识利用不足-倾向于用户项交互的稀疏性问题。
 - 由于大多数以前的SR系统仅基于会话中的交互项序列对会话进行建模，因此当历史用户项交互稀疏时，尤其是对于冷启动项，它们无法充分学习会话表示。
 - 解决方案
 - 导入项目属性作为边信息，这些边信息通常是从开放知识图（KG）中提取出来的，并被识别为项目知识，以缓解数据稀疏问题。
 - 微观行为方面的项目转换也可以通过项目知识来表示。
- 为了解决上述问题，提出了一种新的SR模型MKM-SR，将用户微观行为（M）和项目知识（K）结合到基于会话（S）的多任务（M）推荐（R）学习中。
 - 在MKM-SR中，会话中用户的顺序模式是在微观行为层而不是项目层上建模的。
 - 会话由一系列用户微观行为构成。
 - 每个微观行为实际上是一个项目及其相应操作的组合。

- 学习项目嵌入和操作嵌入是学习微观行为嵌入的前提，并在此基础上生成会话表示。
- 为了实现这一目标，将会话的操作序列和项目序列分别输入选通递归单元（GRU）和选通图神经网络（GGNN）。
 - 在这一步中，由于操作和项目的不同特点，所以采用了不同的学习机制。
 - 此外，通过TransH结合项目知识来学习更好的项目嵌入，该模型已被证明是处理多对多/一关系的有效KG嵌入模型。
- 与以前使用知识嵌入作为预训练项目嵌入的模型不同，本文将知识嵌入学习作为辅助任务，并将其添加到多任务学习（MTL）范式中，其中主要任务是预测下一个交互项目。
 - 大量实验证明，MTL范式在提高SR绩效方面比以前的训练前范式更有效。

• 总结

- 为了提高SR性能，作者将用户微观行为纳入会话建模中，在细粒度级别捕获会话中连续项之间的转换模式，并进一步研究不同算法对微观行为建模的影响。
- 作者通过MTL范式将项目知识纳入模型中，MTL范式将知识嵌入学习作为SR的辅助（子）任务。此外，作者通过广泛的比较验证了MTL的最佳训练策略。
- 作者深入了解了模型设计机制的基本原理，并对两个现实数据集（KKBOX和JDATA）进行了广泛的评估，以证明作者的模型优于最先进的推荐模型。

• RELATED WORK

• 基于会话的顺序推荐

- 作为顺序推荐的一个子任务，SR的目标是根据会话中的隐式反馈预测匿名用户可能与之交互的连续项目。
- 两大类方法可以利用用户历史记录中的顺序信息

• 基于马尔可夫模型

• 原理

- 使用马尔可夫链捕获连续用户项交互之间的顺序模式。
- 最后一次点击时描述用户的最新偏好。

• 缺点

- 忽略之前的点击，并在长序列中丢弃有用的信息。

• 例子

- Rendel等人提出了一种混合模型FPMC，它结合了矩阵分解（MF）和马尔可夫链（MC）来为下一个购物篮建议的顺序行为建模。
 - FPMC的一个主要问题是，它仍然对用户意图采用静态表示。
- 最近，受DNN在NLP中建模序列能力的启发，一些基于DNN的解决方案已经开发出来，并展示了SR的最先进性能。

• 基于DNN的模型

• 原理

- 基于RNN的模型（包括LSTM和GRU）被广泛用于通过将历史交互编码为隐藏状态（向量）来捕获用户的一般兴趣和当前兴趣。

- 例子

- 作为将RNN用于SR的先驱，Hidasi等人提出了一个深度SR模型，该模型将项目编码到一个热嵌入中，然后将其输入GRU以实现推荐。
- Jing等人通过添加额外的机制来解决RNN固有的内存不足问题，进一步改进了基于RNN的解决方案。
- Hi-dasi等人提出并通过调整损耗函数来改善其模型性能。

- 总结

- 尽管这些模型在SR任务中表现出了良好的性能，但仍然有改进的余地，因为它们都忽略了会话中的用户微观行为。
- 他们忽略了项目和操作之间的不同转换模式。
- 在本文中，作者采用RNN和GNN同时对微观行为进行建模，不仅考虑了项目和操作的差异，而且保持了操作顺序的逻辑。

- 基于知识的推荐

- 传统上，知识包括各种项目属性，这些属性用作筛选用户喜爱项目的约束。
- 随着越来越多的开放链接数据的出现，许多研究人员利用KGs中丰富的知识作为辅助信息来提高推荐系统的性能。
- 近年来，许多研究人员利用DNN学习知识嵌入，并将其输入到下游推荐模型中。

- 例子

- Wang等人提出了一个新闻推荐的深度模型DKN，其中他们还使用了基于翻译的KG嵌入模型TransD来学习知识嵌入，以丰富新闻表示。

- 与MKM-SR的多任务学习方法不同，这些模型使用知识嵌入作为预先训练的项目嵌入。
- 另一个具有代表性的包含知识的深度推荐模型是RippleNet，其中用户代表通过KG（包括项目实体及其属性）之间的迭代传播来学习。
- KG也被用于顺序推荐。

- 例子

- FDSA是一个具有自注意力的特征级顺序推荐模型，其中项目特征可以被视为项目知识，用于丰富用户表示。

- 与作者的模型不同，这两个基于KG的顺序推荐模型在宏观（项目）层面而非微观层面对用户行为序列进行建模。
- 此外，作者的模型通过MTL范式将项目知识融入其中，MTL范式将知识嵌入学习作为SR的辅助（子）任务，确保用户微观行为和项目属性之间的信息共享，从而更好地学习表示。

- METHODOLOGY

- 问题定义

- 由于作者在本文中关注的是用户微观行为，而不是会话中的交互项，因此首先使用 $\{m_1, m_2, \dots, m_L\}$ 来表示会话 s 中的微观行为序列，其中 L 是序列的长度。
- mt ($1 \leq t \leq L$) 是第 t 个微观行为，实际上是一个项目及其相应操作的组合。

- 此外，作者纳入MKM-SR的项目知识以三元组的形式表示。
 - 形式上，知识三元组 $\langle i, r, a \rangle$ 表示a是项目i的属性r的值，该属性r通常被认为是一种关系。
 - 例子
 - $\langle \text{Sugar}, \text{歌曲艺术家}, \text{Maroon 5} \rangle$ 描述了Maroon 5是song Sugar的歌手。
- MKM-SR通过在会话中观察到的用户微观行为进行训练，并获得项目知识。
- 目标
 - 基于给定会话预测下一个交互项目。
- 实现思路
 - 作者的模型被提供给给定的会话s和（下一个交互的）候选项j，以在会话s和j之间生成匹配核心（概率） y^s_j 。
 - 使用 y^s_j ，可以为给定会话（一个样本）生成一个top-k排名列表。
 - 通常，得分最高的项目将被预测为下一个交互项目。
- 模型概述
 - 输入
 - 操作序列
 - 项目序列
 - 知识图谱
 - 首先，从给定的会话中同时提取一个项目序列和一个操作序列（长度皆为L），然后将其反馈给GGNN和GRU，分别学习项目嵌入和操作嵌入。
 - 这两种类型的嵌入组合了一系列用于生成会话表示的微行为嵌入。
 - 鉴于场景（序列）中的不同对象在表示该会话时具有不同的优先级，作者采用软注意机制来生成反映用户长期偏好的全局表示。得到s。
 - 具体而言，s的表示形式s是通过聚合一组微观行为嵌入来获得的。
 - 对于给定的会话s和候选项j，最终的score y^s_j 由多层感知器（简称MLP）计算，然后进行Softmax操作。（s表示和j嵌入都是d维向量）
 - 如前所述，项目知识有助于揭示会话中项目之间的语义关联。因此，作者通过设计不同损失函数的加权和，将知识学习作为辅助任务添加到MTL范式中。
 - 作者引入了TransH的损失函数作为知识学习的损失，因为它是一个KG嵌入模型，可以有效地建模多对多/一关系。
 - 采用交替训练策略来训练MKM-SR。
- 对会话信息进行编码
 - 基于SR的基本原理，获得会话表示的前提是了解会话中每个对象的嵌入。
 - 微观行为
 - 微观行为是一个项目和该项目上提交的操作的组合。
 - 在本文的设定中，会话序列中的对象是一种微观行为。

- 在MKM-SR中，首先分别学习项目嵌入和操作嵌入，然后将项目嵌入和操作嵌入连接起来作为微观行为的嵌入，而不是直接学习整个微观行为嵌入。
- Intuitions of Embedding Learning
 - 对于会话的项目序列，其转换模式实际上比先前基于RNN的顺序模型捕获的连续项目之间的单向转换更复杂。
 - 不仅后续项与序列中的前一项相关，而且前一项与后续项相关。
 - 这也是用户经常与自己以前已经交互过的项目交互的原因。
 - 这种转换模式依赖于双向上下文（前面的项和后面的项），而不是单向上下文，单向上下文可以用基于图的模型来建模，而不是像GRU这样的单向顺序模型。
 - 综上所述，作者打算采用GGNN对项目序列进行建模，以获得MKM-SR中的项目嵌入。
 - 尽管用户在会话中提交的操作也会组装一个序列，但它们的转换模式与项序列的转换模式不同。
 - GGNN不适合建模操作序列的原因。
 - 在大多数平台上，独特的操作类型非常有限。
 - 如果将操作序列转换为有向图，则一个操作可能以较大的概率在序列中重复出现，从而导致大多数节点（操作）具有相似的邻居组。
 - 因此，通过在这样的图上应用GGNN学习的大多数操作嵌入非常相似，这不能很好地描述用户偏好的多样性。
 - 两个连续操作之间的转换通常表现出某种顺序模式。
 - 综上所述，作者采用GRU而不是GGNN来学习操作嵌入。
- Learning Item Embeddings
 - 为了通过GGNN学习项目嵌入，首先需要将项目序列转换为有向图。
 - 给定会话中的微观级项目序列 $S_i = \{it_1, it_2, \dots, it_L\}$ ，其中每个对象都是微观行为中的项目，相应的有向图是 $G = (V, E)$ 。
 - 在 G 中，每个节点代表 S_i 中的一个唯一项，以及每个有向边 $(it_{k-1}, it_k) \in E$ ($2 \leq k \leq L$) 链接 S_i 中的两个连续项。
 - 为了改进模型 G ，作者进一步将其构造为加权有向图。
 -
- Learning Operation Embeddings
 - 作者采用了一个带有操作序列的GRU来学习操作嵌入。
 - GRU是标准RNN的改进版本，用于建模动态时间行为，旨在解决消失梯度问题。
 - 作者使用 $S_o = \{ot_1, ot_2, \dots, ot_L\}$ 来表示输入到GRU的操作序列。
 - 对于操作 ot_k ($1 \leq k \leq L$) 在 S_o 中，它的初始嵌入 ot_{0k} 也是通过嵌入矩阵中的查找操作获得的。
 - 然后，其学习到的嵌入 ot_k 是GRU输出的第 k 步中的隐藏状态（向量），其基于 ot_{0k} 和第 $(k-1)$ 步中的隐藏状态计算

-
- Generating Session Representations
 - 为了获得会话表示，应该聚合该会话中所有微观行为的嵌入。
 - 作者考虑了会话的局部偏好和全局偏好。会话的局部偏好通过嵌入最新的微观行为（即 m_L ）直接表示。
 - 为了表示会话的全局偏好，作者使用软注意机制为会话中嵌入的每个微观行为分配适当的权重，因为不同的微观行为具有不同的优先级。
 -

- 学习知识嵌入

- 在包含项目的KG中，经常观察到多对一和多对多关系。
- 在最先进的KG嵌入模型中，transH引入了超平面来有效处理多对多/一关系。
- 因此，作者引入TransH的训练损失来学习模型中的知识嵌入。
- 具体地说，对于每个属性关系 r ，作者首先在关系特定的超平面 w_r 中定位关系特定的转换向量 d_r 。
- 给定一个三元组 $\langle i, r, a \rangle$ ，项目 i 的嵌入 i 和属性 a 的嵌入 a 首先以 w_r 作为法向量投影到超平面。
-

- 多任务学习的目标

- 以前的许多基于知识的推荐模型通常提前学习知识嵌入，这些知识嵌入被用作预先训练的项目嵌入。在这种情况下，知识嵌入学习和推荐是两个独立的学习任务。
- 一般来说，如果两项任务相互关联，那么将两项学习任务合并到MTL范式中比分别实现各自的目标更有效。
- 在MTL中，一项任务的学习结果可以作为提示，指导另一项任务更好地学习。
- 学习知识嵌入可被视为预测用于SR预测任务的特征（项目嵌入）的辅助任务。
- 因此，在MKM-SR中，我们将知识嵌入学习作为辅助任务导入MTL范式，以辅助SR任务。
-

- EXPERIMENTS

- Experiment Settings

- 数据集
 - KKBOX
 - JDATA
 - 对于这两个数据集，作者将四项属性（关系）视为知识，并将其纳入模型中。
 - 将JDATA中会话的持续时间阈值设置为1小时，并将KKBOX中会话的索引间隔设置为2000（根据统计分析），以划分不同的会话。
 - 筛选出长度为1的会话以及数据集中出现次数少于3次的项目。

- 对于这两个数据集，作者将之前90%的用户行为作为训练集，并将随后（最近）10%的用户行为作为测试集。
- 在模型预测中，给定一个测试会话，模型首先计算所有项目的匹配分数，然后根据分数生成top-k列表。
- 为了证明整合项目知识以缓解冷启动项目问题的有效性，作者在数据集上添加了两个额外的操作，这与之前的SR评估不同。
 - 保留仅出现在测试集中的项目，即冷启动项目。
 - 通过只保留早期1%的用户行为来模拟稀疏的JDATA数据集（表示为Demo）。
 - 这样的稀疏数据集具有更大比例的冷启动项。
 - 在以前的SR模型中，这些冷启动项的嵌入是随机初始化的，并且不能在模型训练期间进行调整，因为它们不涉及任何训练样本。因此，关于这些项目的建议往往不能令人满意。

- 比较模型

- 为了强调MKM-SR的优越性能，作者将其与以下最先进的SR模型进行了比较：
 - FPMC
 - GRU4REC+BPR/CE
 - NARM
 - STAMP
 - SR-GNN
 - RIB
- 为了证明在作者的模型中纳入微观行为和知识的必要性和有效性，其进一步提出了MKM-SR的一些变体，如下所示进行比较。
 - KM-SR
 - M-SR
 - M(GRU/GGNN)-SR
 - M(GGNNx2)-SR

- 评价指标

- Hit@k
- MRR@k

- 超参数设置

- Global Performance Comparisons

- 比较结果表明，MKM-SR优于所有数据集中的所有基线和变量。
- 特别是在有冷启动项目和演示的数据集上，MKM-SR和KM-SR具有更显著的优势。
- 这些结果证明了整合知识以缓解冷启动项目稀疏性问题的效果。
- KKBOX比JDATA具有更多独特的操作，这有助于更好地捕获细粒度级别的用户偏好。

- 除了MKM-SR和M-SR之外，与不包含操作的基于GRU的基线相比，另一个包含用户操作RIB的模型在KKBOX中也比在JDATA中具有更显著的优势。
- 在JDATA数据集中，用户更有可能会话中与同一项交互。在这种情况下，GGNN比GRU更好地捕捉到成功项目之间的转换模式。
- Ablation Study
 - 与KM-SR和M-SR相比，MKM-SR的优势表明，微观行为（操作）和项目知识都应纳入w.r.t.以提高SR性能。
 - 此外，M-SR优于M（GRU）-SR和M（GGNN）-SR，这表明分别通过学习项目嵌入和学习操作嵌入来建模会话比直接学习微观行为嵌入更有效。
 - 项目序列的转换模式不同于操作序列的转换模式。将项目与操作组合为微观行为，然后仅通过特定模型学习微观行为序列的效果较差。
 - 此外，M-SR相对于M（GGNNx2）-SR的优势表明，操作序列应该由GRU而不是GGNN学习。
- Strategies of Incorporating Knowledge Learning
- MTL's Control Parameter Tuning
 - 最后，作者研究了MLT损失L中 λ_1 对MKM-SR最终推荐性能的影响。
 - 当 $\lambda_1=0.0001$ 时，MKM-SR得分最高。这意味着，作为一项辅助任务，知识嵌入学习如果被赋予更多的权重，将会干扰SR的主要任务。