

面向电子商务平台用户意图预测的时间感知分层自注意力网络

王森章^{1*}, 刘毅², 张家强², 尹成语²

¹中南大学计算机学院 长沙 中国 410083

²南京航空航天大学计算机科学与技术学院 南京 中国 211106

摘要 在推荐系统领域, 了解电商平台中在线用户的行为意图至关重要。目前的一些方法通常将用户与商品之间的交互历史数据视为有序的序列, 却忽视了不同交互行为之间的时间间隔信息。另外, 一个用户的在线行为可能不仅仅包含一种意图, 而是包含多种意图。例如, 当一位用户在浏览运动品类下的商品时, 其可能同时有购买足球和运动衫这两种商品的意图。但是现有的一些电商平台用户意图预测方法很难有效对用户-商品交互对时间间隔信息进行建模, 也难以捕捉用户多方面的购物意图。为了解决上述问题, 我们提出了一种时间感知分层自注意力网络模型 THSNet, 以更有效对电商平台的用户意图进行预测。具体而言, THSNet 模型采用一种分层注意力机制来有效地捕获用户-商品交互历史中的时间跨度信息以更有效建模用户的多种意图。THSNet 模型的注意力层分为两层, 底层的注意力层用于建模每个会话内部的用户-商品交互, 上层的注意力层学习不同会话之间的长期依赖关系。另外, 为了提高预测结果的鲁棒性和准确度, 我们采用 BERT 预训练的方法, 通过随机遮盖部分会话的特征表示, 构造了一个完形填空任务, 并将该任务与用户意图预测任务耦合成为多任务学习模型, 这种多任务预测方法有助于模型学到一个具有鲁棒性和双向性的会话特征表示。我们在两个真实数据集上对所提方法对有效性进行了验证。实验结果表明, 我们所提出的 THSNet 模型要明显优于目前最先进的方法。

关键词 意图预测; 商品推荐; 序列预测; 注意力机制; 深度学习

中图法分类号 TP399 **DOI号** 10.19363/J.cnki.cn10-1380/tn.2021.09.13

Time-Aware Hierarchical Self-Attention Networks for User Intent Prediction on E-Commerce Platforms

WANG Senzhang^{1*}, LIU Yi², ZHANG Jiaqiang², YIN Chengyu²

¹ School of Computer Science and Engineering, Central South University, Changsha 410083, China

² College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract Understanding the behavior intent of online users on E-commerce platforms is critically important in many recommender systems. Current approaches generally regard the behavior interactions between the users and the items as ordered sequences, which may largely ignore the time lag length between the behavior interactions. Meanwhile, instead of having only one intent, a user's online behavior on E-commerce platforms may have multiple intents. For example, when a user is browsing the sport equipment, she may want to buy a soccer and a sweatshirt simultaneously. It is difficult for existing approaches to both model the time lag length between the behavior interactions and capture the multi-facet user intents on E-commerce platforms. To address these issues, we propose a Time-Aware Hierarchical Self-attention Networks model named THSNet to more effectively predict the user intents on E-commerce platforms. Specifically, THSNet uses a novel hierarchical attention mechanism to effectively capture the time span length between user-item interactions and a user's multi-facet intents. The hierarchical attention mechanism contains two layers. The bottom attention layer focuses on capturing the user-item interaction within each session, and the upper layer attention aims to learn the long term dependencies among the sessions. In addition, to learn a more robust and bidirection session embedding, motivated by the pre-training method in BERT we propose to add a Cloze task which aims to predict the randomly masked session embeddings. The Cloze task is jointly conducted with the user intent prediction task under a multi-task learning framework. We conduct extensive experiments on two real-world datasets. The results show that the proposed THSNet outperforms multiple current state-of-the-art methods.

Key words intent prediction; product recommendation; sequence prediction; attention mechanism; deep learning

通讯作者: 王森章, 博士, 教授, Email: szwang@csu.edu.cn.

本课题得到中央高校基本科研业务经费“人工智能+”专项(No. NZ2020014)和广东省自然科学基金(No. 2021A1515012239)资助。

收稿日期: 2021-04-29; 修改日期: 2021-08-09; 定稿日期: 2021-08-10

1 引言

在电子商务平台中如何更好的向用户提供个性化服务,很大程度上取决于对用户意图^[1]的准确理解。例如,预测一名用户想要在亚马逊平台上购买什么商品,有助于提高商品推荐任务和销售预测任务的准确性。一名用户在电子商务平台中的行为可以看作是由一系列按时间顺序排列的会话组成。每次会话都是在给定的时间内发生的一组用户行为,会话内容是用户交互过的商品^[2]。其中,用户和商品之间的交互可以是用户对商品的一次点击,浏览或购买。来自一个用户的两个连续会话之间通常具有一定的时间间隔。它们可能发生在同一天中,也可能间隔几天、几周或几个月。一个会话通常反映了一名用户的一个特定意图,比如听某种风格的音乐或者找某件合适的商品^[3]。在电商领域,一名用户在一个会话中通常有一个明确的购物需求。如图 1 所示是用户 1 在亚马逊平台上的浏览历史,该浏览历史是由一系列会话组成的,从会话 $session_1$ 到会话 $session_{i+1}$ 。在第一个会话中,用户 1 与一组不同类型的耳机产生交互,该用户隐含的目的或行为意图是购买其中的某个耳机。然后经过了一段时间后,用户 1 想要购买一双鞋子,于是又建立了一个新的会话,在新会话中用户和一组鞋子产生交互。从历史会话中捕捉用户的偏好以及预测用户未来的意图,是服务提供商用来改善用户体验的一种行之有效的办法。

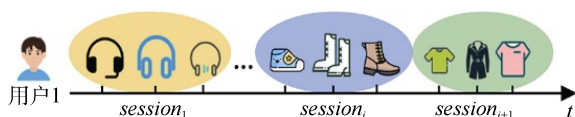


图 1 用户 1 在亚马逊平台中浏览历史示例

Figure 1 A toy example of a user's browsing history containing three sessions on Amazon

虽然在基于会话的推荐任务和序列预测任务等方面已经有了大量的相关工作,但由于以下局限性,导致现有方法在此研究问题上的性能表现仍然不尽如人意。首先,受到 BERT 模型(Bidirectional Encoder Representation from Transformers)^[4]在文本理解领域中相关研究的启发,一些模型将自注意力模型直接应用于用户会话中的用户意图预测任务^[5-6]。但这些模型只考虑了会话序列的顺序信息,而忽略了一个会话中用户行为之间的时间间隔信息,即一名用户在较短的时间间隔内的行为比在较长时间间隔内的

行为具有更强的相关性。其次,用户的意图具有复杂性,这是因为用户在一个会话中可能有多种意图^[2],而之前的一些相关工作提出的方法无法有效地捕获用户的多种意图。例如,当一名用户正在浏览运动品类的商品时,该用户可能同时具有购买一个足球的意图和购买一件运动衫的意图。第三,包含用户行为的会话数据与其他类型的序列数据(如文本和时间序列等)之间存在的一个显著区别是,用户行为的会话数据通常不仅在同一个会话中具有短期依赖关系,并且在不同的会话之间具有长期依赖关系^[7-8]。现有的一些相关工作也很难同时捕获这种长期和短期依赖关系。

为了有效解决上述问题,本文提出了一个名为 THSNet 的时间感知分层自注意力网络模型,以更精准的预测电商平台中在线用户的购物意图。具体说来,THSNet 模型首先建立了一个会话级别的时间感知 LSTM 层,并通过该层学习每个会话的特征表示。受文献[9]的启发,我们将时间间隔信息输入标准的 LSTM 单元,用来处理每个会话中不规律的时间间隔。接下来,为了捕捉用户的多种意图,我们专门设计了一个多意图自注意力模块,该模块利用一个多头自注意力机制来捕捉用户的多种意图。此外,我们还设计了一个掩码矩阵用以确保时间流动的单向性,这意味着当前的用户行为只能受到其之前行为的影响,而不会受到未来未发生的行为影响。最后,我们设计了一个分层自注意力机制,以同时捕获用户在线行为的短期和长期依赖关系,其中包含在为会话内部的用户行为建模的会话内自注意力机制,以及为多个会话之间的序列依赖关系建模的会话间自注意力机制。为了使会话表示学习具有更好的鲁棒性和双向性,我们还把训练完型填空模型作为一项辅助任务。具体来说,我们随机地遮盖了部分会话的特征表示,然后通过对它们的左右上下文信息进行联合调节,实现对被遮盖部分的预测。通过对被遮盖的会话特征表示进行预测与在多任务学习框架下对用户意图的预测相结合,可以进一步提升用户意图预测的性能。

我们的主要贡献总结如下:

- 据我们所知,目前还没有针对用户在电商平台上的商品购买意图进行量化预测的工作,本论文首次研究了基于用户在线会话的用户对商品购买意图的量化预测问题。我们提出了一个时间感知分层自注意力模型用来预测电商平台中的用户购买意图。

- 为了捕获在一次会话中某位用户的多种购物意图, 我们设计了一个多意图自注意力模块。并且, 为了同时捕获短期和长期依赖, 我们也提出了会话内和会话间两种层次的分层自注意力机制。
- 为了提升预测模型的鲁棒性和准确率, 我们提出了一个多任务学习框架, 将基于掩码的会话表示学习和用户购买意图量化预测进行联合学习。
- 在两个真实数据集上的广泛实验也证明了, 与五种基线模型相比, 我们的方法是有效的。此外, 我们还进行了相应的消融实验, 用来分析我们模型中的关键组成部分以及每个部分对应的贡献。

本论文的组织结构如下: 第 2 章会对相关工作进行介绍; 第 3 章给出所研究问题对正式定义; 第 4 章详细介绍时间感知分层自注意力机制模型; 第 5 章介绍实验及结果分析; 第 6 章对整篇论文进行总结并对未来工作进行展望。

2 相关工作

2.1 协同过滤推荐系统

由于互联网的不断发展, 网络中的信息呈现出爆炸式的增长, 大量的商品、音乐、视频等资源和信息可以提供给用户选择, 信息过载的问题非常严重, 推荐系统是解决这一问题的有效方式。协同过滤算法(Collaborative Filtering, CF)是非常经典的一种推荐算法, 其主要思想是在大量用户的行为数据中获取集体智慧并用于推荐, 大体可以分为基于用户的协同过滤^[10], 基于物品的协同过滤^[11]和基于模型的协同过滤^[12]。2003 年, Linden 等人^[13]提出了商品到商品的协同过滤算法, 即把一名用户所购买和评分的商品匹配到相似的商品, 并将所有相似的商品组合, 最后放入推荐列表中, 因为与电商平台中在线的顾客数量和产品目录中的商品数量无关, 所以能够产生高质量的实时推荐。2017 年, He 等人^[14]提出了基于神经网络的协同过滤模型 NCF(Neural collaborative filtering), 将矩阵分解模型中的内积操作替换为神经网络, 提升了特征的交叉能力。此类算法已经在京东、亚马逊等国内外主流的电商平台中得到了广泛部署。

2.2 基于图的推荐系统

随着人们研究的不断深入, 基于图的推荐算法也开始崭露头角^[15]。在电商推荐系统中, 可以将用户视为一个集合, 商品视为另一个集合, 用户会对商品产生交互行为, 这种交互行为刚好可以用二分图进行表示^[16]。于是给用户的推荐任务就可以转化成为度量用户顶点与没有连线的商品顶点之间的相关性, 相关性越高的商品在推荐列表中的位置就越靠前。2012 年, Chen 等人^[17]在传统的用户-项目二部图的基础上, 把用户的查询行为也纳入考虑范围, 建立了“用户-查询-项目”三部图, 由于从用户顶点到项目顶点可随机游走的线路数量增加, 因此在缓解数据的稀疏性问题上取得了良好的效果。传统的协同过滤推荐算法是对用户和项目之间的交互进行建模, 一般会带来系统冷启动的问题, 所以 Wang 等人^[18]于 2019 年提出了 KGCN 模型, 将知识图谱引入推荐系统, 并在知识图谱中利用 GCN 技术挖掘更丰富的辅助信息用于推荐, 有效地解决了冷启动的问题。

2.3 序列模型

序列模型是一种被设计用来处理输入数据具有顺序依赖性的模型。例如, 用来预测下一个单词或字母^[19]。又或者从之前观看过的视频列表信息中预测下一个视频^[20]。在序列模型中, CNNs 和 RNNs 是两种重要的主流结构。因为 RNN 的隐藏层可以从神经网络上一时刻的状态进行学习, 所以成为序列模型建模的默认选择。由于有些数据序列太长, 在训练中会出现梯度消失的问题, 因此 Hochreiter 和 Schmidhuber 在 1997 年提出了长短期记忆神经网络(LSTM)^[21], Cho 等人^[22]在 2014 年提出了门控循环单元(GRU), 这两种流行的变种 RNN 比较好地解决了在 RNN 训练中遇到的梯度消失问题。另外, CNN 也被 LeCun、Oord 等人用于对长期历史上的序列数据进行建模^[23-24], Javidani 等人^[25]于 2018 年将三维视频数据作为输入, 然后分解成为空间上的二维数据和时间上的一维数据, 并使用 1D-CNN 沿着时间维度学习时间特征, 最后达到视频分类的目的。Bai 等人^[26]于 2018 年提出了时间卷积网络(TCN), 并且表明了卷积结构和残差连接在建模序列任务上优于循环神经网络, 例如机器翻译任务。与 LSTM 相比, 尽管 TCN 在并行性和内存开销方面具有优势, 但 TCN 没有利用时间间隔信息, 并不适合用于解决我们所研究的问题。因此我们的工作利用时间感知 LSTM 来处理不规则的时间间隔信息, 这对于我们的预测任务来说无疑是大有裨益的。

2.4 序列感知推荐系统

序列感知推荐系统的目的是根据用户的序列历史信息来推荐与之相关的项目。不同于经典的推荐任务, 序列推荐任务更注重时间和上下文信息, 而在许多协同过滤(CF)方法^[27]中却很少考虑到这些信息的利用。序列推荐系统的早期工作利用了马尔可夫链和马尔可夫决策过程。例如, Steffen 等人^[28]通过矩阵分解的方式来模拟用户的兴趣偏好, 并结合马尔可夫链根据用户的近期行为来预测下一个时间点的用户行为; Guy 等人^[29]认为将推荐问题视为序列优化问题更加合适, 而且由于马尔可夫决策过程能够考虑到每次推荐带来的长期影响, 因此马尔可夫决策过程能为推荐系统提供一个较好的模型。然后, 许多基于 RNN 的方法由于能够有效地对序列性质进行建模, 因而引起了越来越多的关注, 在文献[30]中, 作者采用 LSTM 构建了一个自回归模型, 用于捕捉用户和项目的动态性; 在文献[31]中, 作者利用 GRU 来建模用户历史的时间序列, 在推荐模型中考虑了时序关系。最近, 随着一些基于自注意力的方法在 NLP 领域取得成功, 这些方法也被应用在了序列推荐任务中, 而且同样也取得了良好的效果。例如 SASRec^[5]采用自注意力机制对用户的历史行为信息建模, 提取其中更有价值的信息, 最后将这些信息分别与所有物品的特征做内积, 按照相关性大小排序后进行推荐; BERT4Rec^[6]则将 NLP 领域中的 BERT 模型用于推荐系统, 将用户的历史序列视为词序列, 利用 BERT 模型的双向注意力对用户行为序列进行建模。此外, 文献[32]模型中提出了一种多视

角学习机制, 分别为静态特征、动态特征以及静态动态交叉特征构建三个视图, 在每个视图中采取自注意力的方法提取重要信息, 通过静态特征与动态特征之间的特征交互, 保留了更多的信息, 有利于提高预测的准确性。与一般的会话推荐任务不同, 我们的工作主要是定量地预测用户在会话中对特定商品购买意图的强烈程度。

3 问题描述

令 U 为用户的集合, V 是商品的集合。对于每个用户 $u \in U$, 我们有一个会话序列 $S^u = [s_1^u, s_2^u, \dots, s_N^u]$, 其中 s_i^u 是用户 u 之前按照时间顺序与商品进行交互的商品集合, 记作 $s_m^u = [i_1^{s_m^u}, \dots, i_k^{s_m^u}, \dots | i_k^{s_m^u} \in V]$ 。一名用户和一件商品之间的交互可以是用户的一次点击、购买或者浏览某件商品。给定用户的历史会话序列 S^u 和用户的身份信息 e_u (例如年龄、性别、教育程度等), 我们的目标是预测用户在下一会话 s_{N+1}^u 中对每件商品购买意图的强度。购买意图强度在 $[0, 1]$ 范围内变化, 其中 0 表示用户没有购买该商品的意图, 1 表示用户对该商品非常感兴趣并将购买该商品, 从 0 到 1 表示用户的购买意图由弱到强。

4 时间感知分层自注意力机制

如图 2 所示, 我们提出的 THSNet 模型包含了以下四个层级。图的底部呈现的是模型的输入层, 输入

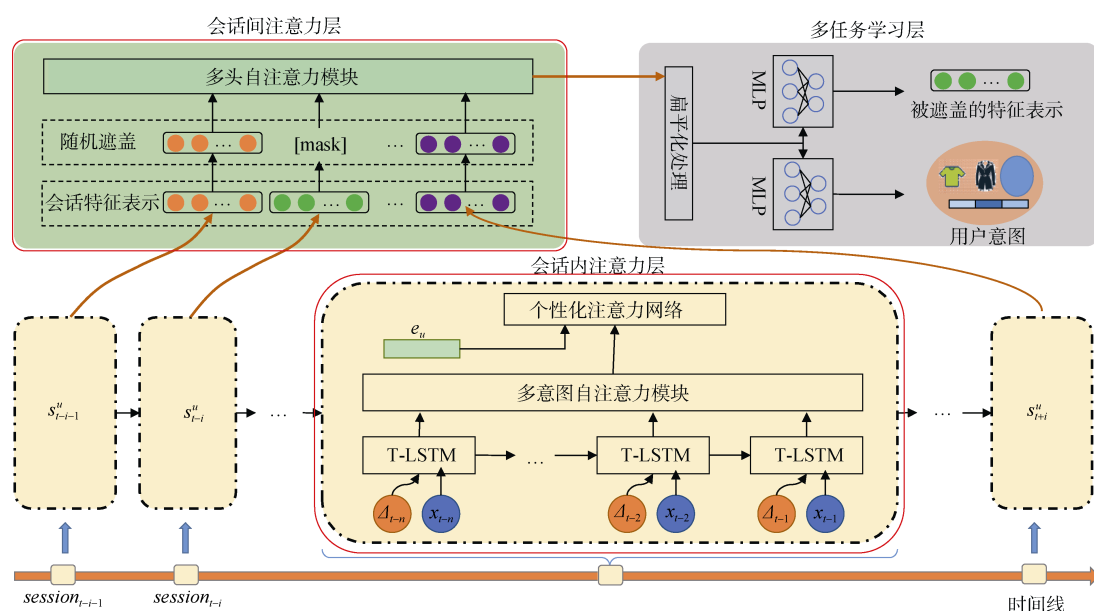


图 2 THSNet 模型结构, 红色方框内是模型创新的部分

Figure 2 The model architecture of THSNet, the contribution of the model is highlighted in red rectangle

的数据是用户的会话序列。第二层是内注意力层, 目的是通过时间感知 LSTM 单元来学习商品在每次会话中的特征表示, 并利用多意图自注意力模块来捕获每次会话中的短期依赖关系。接下来我们提出了一个个性化的注意力网络, 用来确保会话的特征表示是个性化的。这么做的目的是由于在很多情况下, 尽管不同的用户可能拥有相似度较高的会话, 但此时他们会话对应的特征表示却不尽相同。例如, 两个用户 A 和 B 可能有相似的历史交互行为, 但是由于 A 和 B 在个人喜好、年龄层次等方面的不同, 使得两人对未来的交互商品产生不同的选择。第三层采用了一个多头自注意力机制, 目的是为了捕获会话序列之间的长期依赖关系。最后, THSNet 模型通过结合完型填空任务(也称为“掩码语言模型^[4]”)和用户的意图预测任务进行训练, 其中, 完型填空任务是指随机遮盖部分输入的会话表示, 然后预测这些被遮盖的会话表示。完形填空任务有利于模型学习到具有良好鲁棒性和双向性的会话特征表示, 为进一步提高用户意图预测的准确性提供帮助。

4.1 会话内注意力层

4.1.1 会话级时间感知 LSTM

因为 LSTM 可以有效地捕获序列数据的顺序信息, 所以 LSTM 已经被广泛应用于各类序列建模问题中。但是 LSTM 没有考虑输入序列之间的时间间隔信息, 而这些时间间隔信息对于建模用户的序列行为又非常重要。例如, 如图 3 所示, 如果两名用户有相同的历史交互序列信息, 其中的一名用户是在 1 个小时内完成了这些交互, 而另外一名用户只在 10 分钟内就完成了这些交互。在这种情况下, 由于以下原因, 他们的行为意图可能截然不同。首先, 短时间内的两个动作往往比较长时间间隔的两个动作具有更强的相关性。其次, 行为的频率可能反映了用户的偏好强度, 即较高的交互频率可能意味着用户对所交互的商品有更浓厚的兴趣。因此, 当对用户行为进行建模的时候, 交互动作之间的时间间隔信息不容忽视。

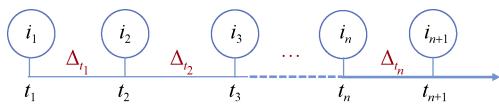


图 3 具有时间间隔信息的一次会话

Figure 3 A session with time intervals

用户与商品间交互的时间间隔信息有助于捕获序列数据的时间特征。受 Zhu 等人^[9]的启发, 我们提

出了一个会话级时间感知 LSTM 模型来捕获每个会话中不规则的时间间隔信息。具体来说, 我们引入时间间隔变量 Δ_k 并重新设计了遗忘门。会话级时间感知 LSTM 的公式如下:

$$\begin{aligned} i_t &= \sigma(W_i^1 x_t + W_i^2 h_{t-1} + b_i) \\ f_t &= \sigma(W_f^1 x_t + W_f^2 h_{t-1} + W_f^3 \Delta_t + b_f) \\ o_t &= \sigma(W_o^1 x_t + W_o^2 h_{t-1} + b_o) \\ c_t &= f_t * c_{t-1} + i_t * \tanh(W_c^1 x_t + W_c^2 h_{t-1} + b_c) \\ h_t &= o_t * \tanh(c_t) \end{aligned}$$

其中, i, f, o 分别表示 t 时刻网络的输入门, 遗忘门和输出门。 c 是细胞激活向量。 x 和 h 分别表示输入特征向量和隐藏层输出。 W 是权重参数, b 是对应的偏置, Δ_t 是 x_{t-1} 和 x_t 之间的时间间隔。在介绍完该模型如何学习会话表示后, 我们接下来介绍如何将隐藏层输出 h_t 传递到多意图自注意力模块。

4.1.2 多意图自注意力模块

在推荐任务中, 多头自注意力模型^[33]已经初露锋芒, 有着不凡的表现^[6]。自注意力模型将输入序列本身作为查询、键和值向量, 具有学习输入序列本身的结构和特点。自注意力模型的输出可以通过给输入序列的每个向量分配不同的权重分数, 自然地将各种向量汇聚成一个整体的表示。由于这一特点, 我们利用自注意力结构来捕获每个会话中的短期依赖。此外, 多头机制同样也适用于用户意图建模, 因为在一个会话中, 用户可能有多方面的意图。因此, 普通的注意力网络可能不足以捕获用户多方面的意图。

然而, 传统的多头自注意力模型不能直接地应用于我们的用户意图预测问题中。这是因为在一个句子中一个单词可以受它前后单词影响, 而在用户意图预测任务中, 项目之间的相互作用是单向的, 也就是说, 相互交互的项目只受之前交互过的项目的影响。为了解决这一问题, 我们设计了一个适合于本研究问题的多意图自注意力模块, 在该模块中增加了一个精心设计的方向掩码矩阵 M , 以保护序列中的时间信息, 确保时间信息只能单向流动。

$$M_{ij} = \begin{cases} 0, & i < j \\ -\infty, & \text{otherwise} \end{cases}$$

当 $i < j$ 时 $M_{ij} = 0$ 表示用户对商品 j 的购买意图只会受到之前交互过的商品 i 的影响, 否则 $M_{ij} = -\infty$ 表示用户对商品 j 的购买意图不会受到商品 i 的影响。

多意图自注意力模块的细节给出如下:

$$\begin{aligned}
Q_i &= W_i^Q X \\
K_i &= W_i^K X \\
V_i &= W_i^V X \\
head_i &= \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}} + M\right) V \\
Z &= \text{Multihead}(X) \\
&= W^O \text{concat}(head_1, \dots, head_h)
\end{aligned}$$

其中, $X=[h_1, \dots, h_n]$ 是会话级时间感知 LSTM 的隐藏层输出, $Q_i, K_i, V_i \in \mathbb{R}^{n \times d_k}$ 分别是查询、键和值向量。每个 $head_i \in \mathbb{R}^{n \times d_k}$ 代表一个单独的头注意力, $W^O \in \mathbb{R}^{hd_k \times d}$ 表示输出线性转换的权重矩阵, h 表示头的数量, d_k 是缩放因子, 且 $d_k = \frac{1}{h}d$ 。这样, 最后的输出 $Z \in \mathbb{R}^{n \times d}$ 就能捕获每个用户会话中的会话级信息。

4.1.3 个性化注意力网络

对于不同的用户, 即使他们的交互项目是相似的, 由于用户个体的偏好, 他们可能也有不同的会话意图。因此, 我们把用户身份信息的特征表示也纳入多意图自注意力模块中, 以建模个性化的会话表示。该模型的数学表示形式如下所示。

$$\begin{aligned}
\alpha_k &= \frac{\exp(e_u Z^T)}{\sum_{k=1}^t \exp(e_u Z^T)} \\
\phi_t^u &= \sum_{k=1}^t \alpha_k Z
\end{aligned}$$

其中, e_u 是用户的特征表示, s_t^u 是个性化注意力网络最终的输出, 代表了会话中用户 u 的短期行为。

4.2 会话间注意力层

如图 2 左上部分所示, 为了捕获不同会话之间的长期依赖关系, 我们在会话内注意力结构之后使用了多头自注意力机制^[33]。与从单词序列中理解句子的含义非常类似, 我们从会话序列中预测用户的意图。会话间注意力层将来自于个性化注意力网络的不同会话表示作为输入, 然后把该结构的输出传递到多任务层中。为了学习到具有双向性的会话表示, 我们随机遮盖了输入会话的部分信息, 然后在接下来的训练任务中预测这些被遮盖的会话特征表示。如图 4 所示。

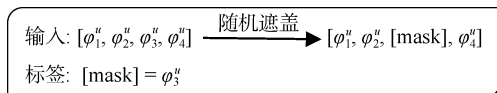


图 4 掩码语言模型示例

Figure 4 An example of masked Language Model

通过随机遮盖部分会话的特征表示, 可以获得鲁棒性更强的会话表示。另外, 通过上下文会话特征表示来预测被遮盖的特征表示也有助于学到一个具有双向性的向量。具体如下:

$$\begin{aligned}
\tilde{\Phi} &= \text{mask}(\Phi) \\
\Psi &= \text{Multihead}(\tilde{\Phi})
\end{aligned}$$

其中 $\Phi=[\phi_1^u, \phi_2^u, \dots, \phi_n^u]$ 是来自于个性化注意力网络的所有会话特征表示, $\tilde{\Phi}$ 是被随机遮盖的会话特征表示。 Ψ 是以 $\tilde{\Phi}$ 作为输入的多头注意力结构的输出。

4.3 多任务学习层

多任务学习已成功应用于机器学习的各种应用场景, 包括自然语言处理^[1-2]、语音识别^[3]、计算机视觉^[4]和推荐系统^[5]。通过学习相关任务之间共享特征表示, 多任务学习可以使模型捕获更多有用的特征表示, 丰富单任务学习的特征表示形式、从而提升特征学习的鲁棒性, 提高学习模型在多个学习任务之间的泛化性能。同时, 受到工作^[6]的启发, 多任务学习在电子商务场景中更有利于更好地学习用户的通用特征, 以及导致不同购买行为的特有特征, 因此可以更有效地预测用户行为。因此, 本论文也采用多任务学习的方法提升用户购买意图预测的准确率。在此次工作中, 我们提出了使用多任务学习框架来同时预测被遮盖的会话特征表示和用户意图。这样, THSNet 模型不仅能学到一个鲁棒性更强的表示以及提高模型的泛化能力, 而且可以进一步提升用户购买意图预测的准确性。如图 2 的右上区域所示, 第一项任务是预测在会话间注意力层中被遮盖的会话特征表示的原始会话特征表示, 第二项任务是预测在下一个会话中的用户购买意图。

我们首先把会话间注意力层的输出进行扁平化处理, 即把多维的输入数据转化为一维数据, 然后送入两个分支。在每个分支中, MLP 的作用是学习并获得最终的输出。遮盖的会话特征表示预测任务 and 用户意图预测任务的整体损失函数定义如下所示:

$$\begin{aligned}
L_\Phi &= -\frac{1}{|\tilde{\Phi}|} \sum_{\phi_m^* \in \tilde{\Phi}}^N \|\phi_m - \phi_m^*\|_2 \\
L_t &= -\frac{1}{N} \sum_{(x, y) \in D}^N \|y - p(x)\|_2
\end{aligned}$$

其中, $\tilde{\Phi}$ 是所有会话 Φ 中被随机遮盖的会话特征表示, ϕ_m 是真实的会话特征表示, ϕ_m^* 是被遮盖的会话特征表示。 D 是大小为 N 的训练集, x 是网络的输入, y 是标签, $y=0$ 表示用户没有购买该商品, $y=1$ 表示用户购买了该商品。 $p(x)$ 代表预测的用户购买某种商品

的意图强弱。全局损失由这两个损失相加求和而得, λ 是缩放参数。

$$L_{global} = L_i + \lambda L_\phi$$

整个 THSNet 算法的流程如下所示:

THSNet 算法

输入: 用户集合 U , 商品集合 V , 对于每个用户 $u \in U$, 都有会话序列 $S^u = [s_1^u, s_2^u, \dots, s_N^u]$, 其中 $s_m^u = [i_1^{s_m^u}, \dots, i_k^{s_m^u}, \dots | i_k^{s_m^u} \in V]$, 用户信息特征编码为 e_u

输出: 用户 u 在下一次会话中 s_{N+1}^u 中对每件商品购买意图的强度 $p(x)$

1: for 用户 u 的会话序列的一个会话 s_i^u do

2: 此会话 s_m^u 中商品集合为 $s_m^u = [i_1^{s_m^u}, \dots, i_k^{s_m^u}, \dots | i_k^{s_m^u} \in V]$, 用户与商品间交互的时间间隔信息为 $A_i = [A_1, \dots, A_{|s_i^u|}]$

3: 根据章节 4.1.1 可得:

$$T_LSTM_OUT = T_LSTM(s_m^u, A_i)$$

4: 根据章节 4.1.2 可得此会话中的会话级信息: $Z_u = Multihead(T_LSTM_OUT)$

5: 根据章节 4.1.3 可得融合用户信息后的会话表示: $\phi_i^u = ATT_NET(Z_u, e_u)$

6: return 会话特征表示: $\Phi = [\phi_1^u, \phi_2^u, \dots, \phi_N^u]$

7: 随机遮盖后的会话特征表示为 $\tilde{\Phi} = mask(\Phi)$, 被遮盖的表示为 ϕ_m

8: 根据章节 4.2 可得捕获了不同会话间依赖关系的表示: $\Psi = Multihead(\tilde{\Phi})$

9: 根据章节 4.3 可得多任务预测结果, 其中遮盖的会话表示预测结果为 $\phi_m^* = MLP_1(\Psi)$, 下一次会话中 s_{N+1}^u 中对每件商品购买意图的强度 $p(x) = MLP_2(\Psi)$

10: 最终的全局训练误差为: $L_{global} = -\frac{1}{N} \sum_{(x,y) \in D} \|y - p(x)\|_2 - \frac{1}{|\tilde{\Phi}|} \sum_{\phi_m^* \in \tilde{\Phi}} \|\phi_m - \phi_m^*\|_2$

5 实验

5.1 数据集

我们使用 LastFM 数据集和 JingDong 数据集作为我们实验评估的数据。这两个数据集的统计情况见表 1。

表 1 数据集的统计情况

Table 1 Statistics of the two datasets

| 数据集 | #用户数量 | #项目数量 | #交互数量 |
|----------|-------|--------|---------|
| LastFM | 1000 | 1000 | 1293103 |
| JingDong | 10692 | 303150 | 1198735 |

• **LastFM** 数据集: 这是一个基于用户的听歌序列进行音乐推荐的公共数据集 (<http://snap.stanford.edu/jodie/lastfm.csv>), 其中包含了 1 个月内有哪个用户听了哪些歌曲的信息。我们挑选了 1000 名用户和 1000 首最常听的歌曲, 并获得了 1293103 次交互的历史数据。在这个数据集中, 交互没有特征。

• **JingDong** 数据集: 本数据来自于中国最大电商平台之一的京东网, 其中包含了京东网的用户在线行为记录数据。用户每条在线行为记录数据包含了用户 ID、商品 ID、交互时间、交互行为(点击、购买、加购、收藏等)、商品属性信息等。我们采集了从 2020 年 3 月到 2020 年 4 月, 10692 名用户和 303150 个商品之间的交互数据, 并对这些数据进行数据清洗、异常点去除等预处理, 最终得到 1198735 次交互数据。然后我们根据用户 ID, 将同一个用户的所有商品交互行为记录信息进行合并整合。

这两个数据集都包含了用户与项目之间交互的时间戳或具体日期。在数据处理中, 我们减去序列中最小的时间戳, 让所有用户的时间戳从零开始。之后, 我们对交互记录按用户进行分组, 每个用户的所有交互记录分为一组, 并根据时间戳对每组的交互记录数据进行排序, 为每个用户建立交互序列。为了确保数据集的质量, 我们过滤掉冷启动用户以及少于 20 次交互的商品。然后把数据集分别划分为训练集、验证集和测试集。前 18 天的数据用于训练, 第 19 到 23 天共 5 天的数据用于验证, 剩下的 7 天数据用于测试。

5.2 对比方法和评估标准

我们实验中用于对比的算法包括:

- **LSTM**^[21] 是 RNN 模型体系中的一种重要方法。这里我们简单的记录了商品的序列, 并且删除了时间信息。
- **Time-LSTM**^[9] 是一种新的 LSTM 变种, 它为 LSTM 配备了时间门用来给时间间隔信息建模。

- **LatentCross**^[34]是一个基于 RNN 的推荐系统框架, 它通过对上下文特征进行编码, 将上下文数据信息引入 RNN 中。
- **Jodie**^[35]是一个耦合的循环神经网络模型, 用来学习用户和商品的动态特征表示。
- **LightGCN**^[36]是一种最新的基于协同过滤的推荐方法。它简化了 GCN 的设计, 使其更加简洁和适合于推荐。

为了说明我们提出的会话级时间感知 LSTM、多意图自注意力模型以及多任务学习的有效性, 我们进一步做了将 THSNet 模型和下面三种模型变体进行比较的消融实验。

- **THSNet-1** 删除了会话级时间感知 LSTM 模块, 并用一个传统的 LSTM 代替。只将用户-商品的历史交互序列输入到 LSTM 模型中, 而忽略了两个交互动作之间的时间间隔信息。
- **THSNet-2** 删除了多意图自注意力模块中的掩码矩阵 M 。我们用传统的多头注意力机制取而代之, 直接用该机制预测会话的特征表示。
- **THSNet-3** 删除了多任务学习模块。不采用随机遮盖会话表示的形式进行预训练, 直接进行用户意图的预测输出。

我们采用平均倒数秩(MRR)和 Recall@K 作为评估标准。MRR 用于评估能生成对查询样本响应的列表并按概率排序的任何过程。对于样本查询 Q 的结果的平均倒数秩可定义为: $MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$, 其

中 rank_i 是对于第 i 次查询第一个相关实例的排序位置。**Recall@K** 则衡量了实际检索的相关项目总数的比例。

5.3 实现细节

我们用 Pytorch 实现了 THSNet 模型并且在验证集上对超参进行了微调。批处理的大小为 128, 学习率是 0.001, 模型使用 Adam Optimizer 进行训练, 所有数据集的权重衰减系数为 0.01。我们考虑了隐藏层维度 d 的变换范围是 {16, 32, 64, 128}, 以及 heads 的数量 h 变化范围是 {2, 4, 8, 16}。所有的实验都是在一张 RTX-3090 的显卡上完成的。

5.4 结果

表 2 展示了我们的模型和对比模型的实验结果。我们观察到, THSNet 模型在这两个数据集上的性能明显优于所有的对比模型。在对比模型中,

Light-GCN 是基于协同过滤的方法, 它无法捕获用户历史行为在时间维度上面的时间信息, 因此这个方法不适合对时间信息有依赖关系的用户购买意图预测任务, 并且它的性能也是所有方法中最差的。尽管 LSTM 没有考虑两次连续的用户与商品交互的时间间隔信息, 但却利用了商品的序列顺序信息, 因此, LSTM 比 LightGCN 表现得更好。Time-LSTM 作为 LSTM 的一种变体, 考虑了时间信息而不是顺序信息, 在 LastFM 数据集上比 LSTM 提高了 16.9%, 在 JingDong 数据集上比 LSTM 提高了 12.2%。因为 LatentCross 将上下文信息引入到了 RNN 中, 所以比 LSTM 和 Time-LSTM 表现更好。Jodie 模型把用户-商品之间交互的动态和顺序依赖关系纳入考虑之中, 因此比其余对比模型表现更好。然而, 与 Jodie 相比, THSNet 模型在两个数据集上将性能分别提高了 4.6%和 6.8%。这主要是因为它捕捉到了用户-商品交互的时间间隔信息, 并且在一个会话中建模了多方面的用户意图。

表 2 实验结果, “*”表示在 $p=0.01$ 时 THSNet 相较于对比方法的性能提升是显著的。

Table 2 Experiment results. “*” indicates the improvement of the THSNet over the baseline is significant at the level of 0.01.

| 模型 | LastFM | | JingDong | |
|-----------------|---------------|---------------|---------------|---------------|
| | MRR | Recall@10 | MRR | Recall@10 |
| LSTM | 0.060* | 0.118* | 0.039* | 0.057* |
| Time-LSTM | 0.069* | 0.138* | 0.047* | 0.064* |
| LatentCross | 0.146* | 0.223* | 0.065* | 0.099* |
| Jodie | 0.195* | 0.305* | 0.080* | 0.131* |
| LightGCN | 0.037* | 0.072* | 0.009* | 0.013* |
| THSNet-1 | 0.194* | 0.305* | 0.082* | 0.135* |
| THSNet-2 | 0.198* | 0.316* | 0.085* | 0.138* |
| THSNet-2 | 0.192* | 0.313* | 0.081* | 0.136* |
| THSNet | 0.202* | 0.319* | 0.087* | 0.140* |

在表 2 中, 同样也表现了 THSNet 模型在这两个数据集上优于所有的变体, 这意味着我们提出的会话级时间感知 LSTM 和多意图自注意力机制对于性能的提高有很大帮助。THSNet-1 的结果表明, 去除两个交互动作之间的时间间隔信息可能会导致 THSNet 的性能急剧下降。没有时间间隔信息, 模型只考虑了序列的顺序信息, 这对为顺序用户行为建模来说并不适合。虽然与 THSNet-1 相比, THSNet-2 的提升很小, 但是 THSNet-2 的结果仍然说明, 在会话中考虑时间的单向性有利于用户意图预测。

为了更直观地展现预测性能,我们在图 5 中给了两个案例研究,其中绿色方框代表了一个女孩的两次历史会话,她在两个会话中分别与一些不同类型的卷棒和衣服进行了互动。接下来,给出一个包含不同类型的鞋子的新会话(红色方框),我们的模型预测出了她对每件商品的购买意图,如红色方框所示。我们可以很清楚地看出,这个女孩最喜欢第四双鞋子,而购买第二双鞋子的意图强度是最弱的。第三双鞋是最接

近第四双的,因为它们都是高跟鞋。这可能就是为什么这个女孩对购买第三双鞋子有这么高的购买意图的原因。相似地,图 5 中的另一个案例是一位男士近三次的会话。其中可以发现,在前两次会话中,他分别对足球和足球鞋进行了相关点击,可以说明此用户对足球类商品兴趣较大。在第三次的会话中,我们的模型预测出此人在 T 袖类别下每件商品的购买意图,可以发现其对偏运动偏足球风的商品更为喜爱。

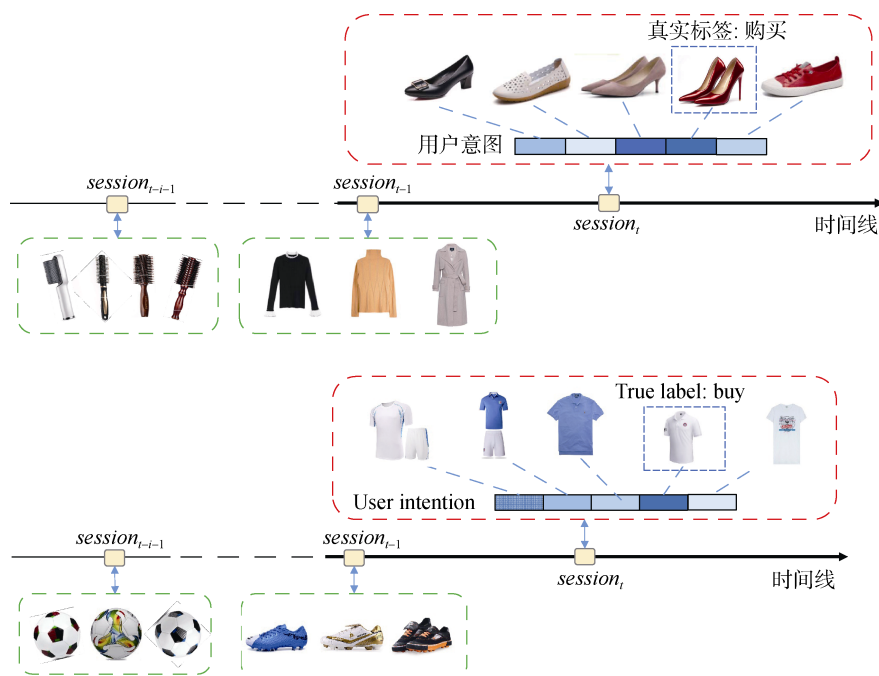


图 5 基于用户浏览历史的意图预测案例

Figure 5 Case studies of user intent prediction based on a user's browsing history

另外,我们对参数的敏感度进行了分析。图 6 展现了在两个数据集上, Recall@10 在特征表示向量维度不同和注意力机制 heads 数量不同情况下的表现。

我们接下来研究了参数 heads 的数量 h 对于 THSNet 模型的影响,其中 h 从 2 到 16 进行变化,同

时保持特征表示的维度 d 不变。结果如图 6(b)所示。我们可以观察到当 $h=8$ 时, THSNet 模型达到最佳性能。当 $h=2$ 时, THSNet 模型的性能最差。 $h=2$ 的情况与其他情况之间的巨大差距体现了多头在建模用户多方面意图时的有效性。我们还研究了参数 λ 的影响。研究结果见图 7。结果表明,模型的性能随参数

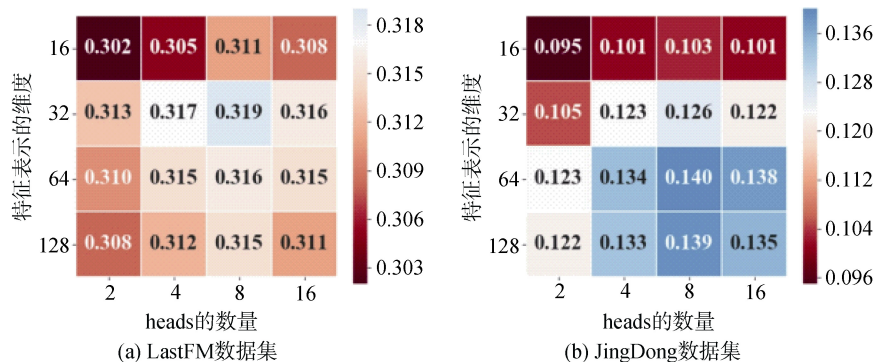
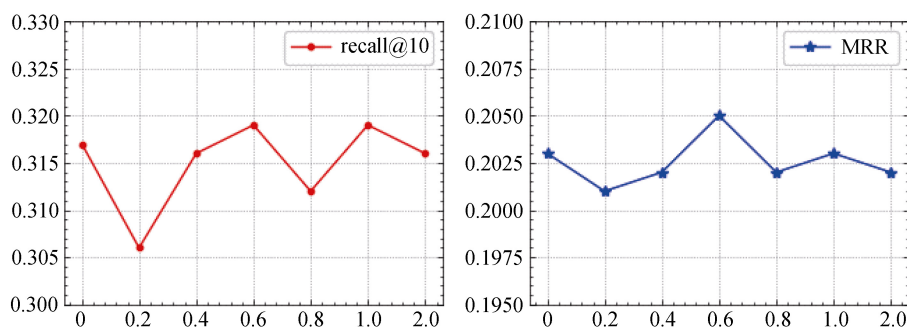


图 6 THSNet 分别在两个数据集上的 Recall@10

Figure 6 The Recall@10 of THSNet over the two datasets

图 7 不同 λ 对实验结果的影响Figure 7 Result with different λ

λ 值的变化而波动, 并没有出现显著的规律, 而对于给定的数据集, $\lambda = 0.6$ 是一个合适的数值。

6 结论

在本文中, 我们提出了一个时间感知分层自注意力模型 THSNet, 用来量化预测电商平台中在线用户的购物意图。THSNet 模型克服了现有工作的局限性, 即忽略了交互动作之间的时间间隔信息和难以捕获用户的多方面意图。在两个真实的数据集上, 与多个最近提出的对比模型相比, 我们提出的方法有了较为明显的性能提升。

参考文献

- [1] Fan S H, Zhu J X, Han X T, et al. Metapath-Guided Heterogeneous Graph Neural Network for Intent Recommendation[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 2478-2486.
- [2] Lv F Y, Jin T W, Yu C L, et al. SDM: Sequential Deep Matching Model for Online Large-Scale Recommender System[C]. *The 28th ACM International Conference on Information and Knowledge Management*, 2019: 2635-2643.
- [3] Quadana M, Karatzoglou A, Hidasi B, et al. Personalizing Session-Based Recommendations with Hierarchical Recurrent Neural Networks[C]. *The Eleventh ACM Conference on Recommender Systems*, 2017: 130-137.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. *NAACL-HLT*, 2019: 4171-4186.
- [5] Kang W C, McAuley J. Self-Attentive Sequential Recommendation[C]. *2018 IEEE International Conference on Data Mining*, 2018: 197-206.
- [6] Sun F, Liu J, Wu J, et al. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer[C]. *The 28th ACM International Conference on Information and Knowledge Management*, 2019: 1441-1450.
- [7] Guo L, Hua L F, Jia R F, et al. Buying or Browsing? : Predicting Real-Time Purchasing Intent Using Attention-Based Deep Network with Multiple Behavior[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 1984-1992.
- [8] Song Y, Elkahky A M, He X D. Multi-Rate Deep Learning for Temporal Recommendation[C]. *The 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2016: 909-912.
- [9] Yu Zhu, Hao Li, Yikang Liao, et al. What to do next: Modeling user behaviors by time-Istm[C]. *IJCAI*, 2017:3602-3608.
- [10] Bellogin A, Parapar J. Using Graph Partitioning Techniques for Neighbour Selection In User-Based Collaborative Filtering[C]. *The sixth ACM Conference on Recommender Systems*, 2012: 213-216.
- [11] Sarwar B, Karypis G, Konstan J, et al. Item-Based Collaborative Filtering Recommendation Algorithms[C]. *The tenth International Conference on World Wide Web - WWW'01*, 2001: 285-295.
- [12] Loepp B, Ziegler J. Towards Interactive Recommending In Model-Based Collaborative Filtering Systems[C]. *The 13th ACM Conference on Recommender Systems*, 2019: 546-547.
- [13] Linden G, Smith B, York J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering[J]. *IEEE Internet Computing*, 2003, 7(1): 76-80.
- [14] He X N, Liao L Z, Zhang H W, et al. Neural Collaborative Filtering[C]. *The 26th International Conference on World Wide Web*, 2017: 173-182.
- [15] Xiang L, Yuan Q, Zhao S W, et al. Temporal Recommendation on Graphs via Long- and Short-Term Preference Fusion[C]. *The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010: 723-732.
- [16] Haveliwala T H. Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2003, 15(4): 784-796.
- [17] Chen B S, Wang J D, Huang Q H, et al. Personalized Video Rec-

- ommendation through Tripartite Graph Propagation[C]. *The 20th ACM International Conference on Multimedia - MM '12*, 2012: 1133-1136.
- [18] Wang H W, Zhao M, Xie X, et al. Knowledge Graph Convolutional Networks for Recommender Systems[C]. *WWW '19: The World Wide Web Conference*, 2019: 3307-3313.
- [19] Barman P P, Boruah A. A RNN Based Approach for next Word Prediction In Assamese Phonetic Transcription[J]. *Procedia Computer Science*, 2018, 143: 117-123.
- [20] Hosseini M, Maida A S, Hosseini M, et al. Inception LSTM for Next-Frame Video Prediction (Student Abstract)[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(10): 13809-13810.
- [21] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [22] Cho K, van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation[C]. *The 2014 Conference on Empirical Methods in Natural Language Processing*, 2014: 1724-1734.
- [23] Bengio Y, LeCun Y. Convolutional Networks for Images, Speech, and Time-Series[J]. *The Handbook of Brain Theory and Neural Networks*, 1995, 3361(March): 1-14.
- [24] Aaron van den Oord, Sander Dieleman, Heiga Zen, et al. Wavenet: A Generative Model for Raw Audio. 2016: arXiv preprint arXiv:1609.03499.
- [25] Ali Javidani and Ahmad Mahmoudi-Aznavah. Learning Representative Temporal Features for Action Recognition. 2018: arXiv preprint arXiv:1802.06724.
- [26] Bai S J, Kolter J Z, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[EB/OL]. 2018: arXiv: 1803.01271[cs.LG]. <https://arxiv.org/abs/1803.01271>.
- [27] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems[J]. *Computer*, 2009, 42(8): 30-37.
- [28] Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing Personalized Markov Chains for Next-Basket Recommendation[C]. *The 19th International Conference on World Wide Web*, 2010: 811-820.
- [29] Shani G, Braffman R I, Heckerman D. An MDP-Based Recommender System [J]. *Journal of Machine Learning Research*, 2005, 6(1):1265-1295.
- [30] Wu C Y, Ahmed A, Beutel A, et al. Recurrent Recommender Networks[C]. *The Tenth ACM International Conference on Web Search and Data Mining*, 2017: 495-503.
- [31] Zhou G R, Mou N, Fan Y, et al. Deep Interest Evolution Network for Click-through Rate Prediction[J]. *The AAAI Conference on Artificial Intelligence*, 2019, 33: 5941-5948.
- [32] Chen T, Yin H Z, Hung Nguyen Q V, et al. Sequence-Aware Factorization Machines for Temporal Predictive Analytics[C]. *2020 IEEE 36th International Conference on Data Engineering*, 2020: 1405-1416.
- [33] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all You Need[C]. *NeurIPS*, 2017: 5998-6008.
- [34] Beutel A, Covington P, Jain S, et al. Latent Cross: Making Use of Context In Recurrent Recommender Systems[C]. *The Eleventh ACM International Conference on Web Search and Data Mining*, 2018: 46-54.
- [35] Kumar S, Zhang X K, Leskovec J. Predicting Dynamic Embedding Trajectory In Temporal Interaction Networks[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 1269-1278.
- [36] He X N, Deng K, Wang X, et al. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation[C]. *The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020: 639-648.



王森章 于 2016 年在北京航空航天大学计算机软件技术与理论专业获得博士学位。现任中南大学计算机学院特聘教授, 研究领域为数据挖掘、人工智能, 研究兴趣包括: 大数据分析, 城市计算。Email: szwang@csu.edu.cn



刘毅 于 2018 年在南京航空航天大学信息与计算科学专业获得理学学士学位。现在南京航空航天大学电子信息专业攻读硕士学位。研究领域为深度学习、数据挖掘。Email: liuyi96@nuaa.edu.cn



张家强 于 2020 年在南京晓庄学院信息与计算科学专业获得理学学士学位。现在南京航空航天大学电子信息专业攻读硕士学位。研究领域为深度学习、时空数据挖掘。Email: zhangjq@nuaa.edu.cn



尹成语 于 2018 年在重庆大学信息安全专业获得学士学位。现在南京航空航天大学计算机技术专业攻读硕士学位。研究领域为深度学习、数据挖掘。研究兴趣包括用户行为分析、推荐系统。Email: chen-gyuyin@nuaa.edu.cn