

基于知识增强的细粒度个性化新闻推荐用户建模

熊晓波¹, 方文涛^{2†}

(1. 湖南省高速公路集团有限公司, 湖南 长沙 410153; 2. 中联重科股份有限公司, 湖南 长沙 410013)

摘要:传统的新闻推荐用户建模方法难以深入解析新闻的复杂语义和用户的真实需求,为此提出了一种知识增强的新闻建模方法,通过实体表示层、上下文嵌入层和注意力聚合层获取新闻文档表示。在此基础上提出了一种基于知识增强文档的细粒度用户建模方法,利用长文档建模技术将知识增强的新闻文档串联成长文档,通过捕获文档间的词级交互行为得到细粒度用户表示;通过捕获文档内的实体交互行为得到粗粒度用户表示,粗粒度的用户表示和细粒度的用户表示聚合得到最终的用户表示。实验结果显示,提出的新闻建模方法在 AUC 和 NDCG@10 指标上优于基线模型,基于此方法的用户建模方法在 AUC 上至少提升 2.51%,在 NDCG@10 上至少提升 4.75%。

关键词:新闻建模;用户建模;知识图谱;细粒度;个性化新闻推荐

中图分类号:TP391

文献标识码:A

User Modeling for Fine-Grained Personalized News Recommendation Based on Knowledge Enhancement

XIONG Xiaobo¹, FANG Wentao^{2†}

(1. Hunan Expressway Group Co., Ltd., Changsha, Hunan 410153, China;
2. Zoomlion Heavy Industry Science & Technology Co., Ltd., Changsha, Hunan 410013, China)

Abstract: Traditional user modeling methods for news recommendation struggle to deeply analyze the complex semantics of news and the genuine needs of users. To address this issue, this paper first proposes a knowledge-enhanced news modeling approach, which obtains news document representations through an entity representation layer, a context embedding layer, and an attention aggregation layer. Based on this, a fine-grained user modeling method based on knowledge-enhanced documents is proposed, utilizing long document modeling techniques to concatenate knowledge-enhanced news documents into a long document. Fine-grained user representations are obtained by capturing word-level interactions between documents, while coarse-grained user representations are derived from capturing entity interactions within documents. The final user representation is aggregated from both coarse-grained and fine-grained user representations. Experimental results show that the proposed news modeling method outperforms baseline models in terms of AUC and NDCG@10 metrics, with the user modeling method based on this approach achieving at least a 2.51% improvement in AUC and at least a 4.75% improvement in NDCG@10.

Key words: news modeling; user modeling; knowledge graph; fine-grained; personalized news recommendation

在当今互联网环境中,网络新闻发布平台吸引了大量用户,实现了新闻传播活动的实时监控和管

理。然而,面对海量新闻内容,用户难以逐一浏览寻找感兴趣的新闻,因此个性化新闻推荐技术应运

而生。该技术旨在探索用户隐性兴趣和需求,提供定制化的新闻推荐服务,减轻信息过载压力,提升用户阅读体验^[1]。

在构建个性化新闻推荐系统时,精准预测用户对新闻内容的偏好至关重要。传统方法通常将用户点击的新闻单独编码成新闻嵌入,再聚合形成用户嵌入,忽略了新闻间和新闻内部实体交互的重要信息。同时,知识图谱揭示了实体间深厚的语义联系,为推荐系统带来了可能的辅助信息。因此,本文首先提出了一种知识增强的新闻建模方法,并在此基础上提出了一种基于知识增强文档的细粒度用户建模方法。

知识增强的新闻建模方法包含实体表示层、上下文嵌入层和注意力聚合层。实体表示层通过 TransE^[2] 知识图谱嵌入方法得到新闻中的每个实体嵌入,并利用知识图注意力网络(Knowledge Graph Attention Network, KGAT)^[3] 的思想,通过动态调整邻居实体的重要性权重,更准确地捕捉实体间的复杂关系。上下文嵌入层加入了实体的位置、频率和类别信息来进一步丰富实体信息,弥补了交互信息的稀疏和缺失。注意力聚合层通过注意力机制对实体进行加权,将加权后的实体表示嵌入原始的文档向量中并经过 LAMA 编码器^[4] 的处理得到知识增强的文档表示。基于知识增强文档的细粒度用户建模方法采用 Longformer^[5] 模型将新闻文档拼接成长文档,通过稀疏注意力机制,在降低计算复杂度的同时全面捕捉文档间的细粒度实体交互行为,再通过多头自注意力处理得到细粒度用户表示,除此之外,通过 Transformer^[6] 和多头自注意力捕获新闻内的实体交互得到粗粒度的用户表示,细粒度的用户表示和粗粒度的用户表示进行聚合得到最终的用户表示。

实验表明,本文提出的知识增强的新闻建模方法和基于知识增强文档的细粒度用户建模方法相比传统方法在新闻推荐的 AUC 和 NDCG@10 指标上均有所提升,验证了各组件和建模框架的有效性和必要性。

1 相关工作

在目前的新闻推荐系统研究中,用户建模技术主要划分为两大类别:基于特征的用户建模和基于深度学习的用户建模。

基于特征的用户建模方法通常使用人工特征来表示用户。例如,Goossen 等^[7] 采集了用户以往

访问过的新闻的 CF-IDF(Collaborative Filtering-Inverse Document Frequency)特征来表示用户兴趣。Chu 等^[8] 通过将用户分为不同的年龄段和性别,探索了这些人口统计特征是如何影响用户对特定内容的偏好的,以此来建模用户的兴趣。Yi 等^[9] 使用停留时间作为点击新闻的权重进行用户建模。

近年来,大量基于深度学习的个性化新闻推荐算法被广泛应用于用户建模中,避免了大规模的特征工程。KRED 利用注意力机制,识别并加权用户历史点击新闻中的重要实体信息,从而生成一个综合反映用户偏好的用户表示^[10]。DKN 利用候选感知的注意力机制,对每一条用户点击过的新闻分配一个权重,如果某个点击新闻与候选新闻内容更相关,那么这个点击新闻在学习用户表示时就会被赋予更高的权重^[11]。EBNR 通过门控循环单元(Gate Recurrent Unit, GRU)从用户浏览新闻的表示中学习用户的表示^[12]。

2 知识增强的新闻建模

本节提出了一种知识增强的新闻建模方法,构建了如图 1 所示的模型,该模型包括三层:实体表示层、上下文嵌入层和注意力聚合层。首先给定一个初始的文档向量 DV,该文档向量经过实体表示层,通过 TransE 知识图谱嵌入方法和知识图注意力网络得到新闻文档的实体向量。将实体传入上下文嵌入层中,增加实体的位置、频率和类别信息,增强实体的上下文表示能力。将结合了上下文信息的实体传入注意力聚合层后,利用注意力机制给实体分配不同的权重,并将加权后的实体和原始的文档向量进行聚合,最后经过 LAMA 编码器处理得到知识增强的文档向量 KDV。

2.1 实体表示层

在实体表示层中,每个实体和关系的嵌入向量 e 是通过 TransE 模型学习得到的,维度为 d 。为了全面地表示一个实体,不仅需要考虑实体自身的嵌入向量,还需要考虑与其直接相关联的其他实体信息。于是,利用知识图注意力网络,在生成实体表示时考虑了邻居实体的信息。对于一个头实体 h ,其实体表示 e_{N_h} 是通过公式(1)计算的:

$$e_{N_h} = \text{ReLU}(\mathbf{W}_0(e_h \oplus \sum_{(h,r,t) \in N_h} \alpha(h,r,t) e_t)) \quad (1)$$

$\oplus$ 表示向量的拼接操作, e_h 和 e_t 分别表示头实体和尾实体通过 TransE 学习到的嵌入向量, N_h 表示三元组的集合,注意力权重 $\alpha(h,r,t)$ 由两

层全连接神经网络计算得到。

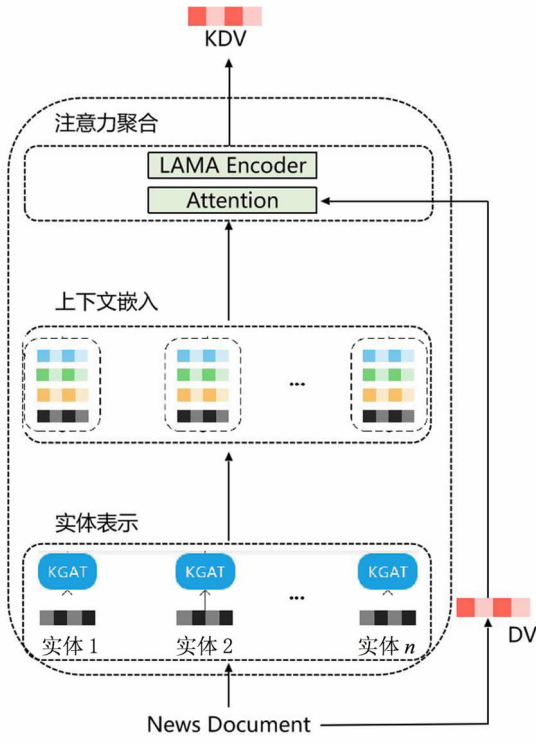


图 1 知识增强的新闻建模框架

注意力机制对不同的实体关系对赋予不同的权重,使模型专注于对头实体语义表示最有贡献的尾实体,首先通过 ReLU 得到 $\alpha_0(h, r, t)$, 如公式(2)所示:

$$\alpha_0(h, r, t) = w_2 \text{ReLU} (W_1 (e_h \oplus e_r \oplus e_t) + b_1) + b_2 \quad (2)$$

再通过 Softmax 函数归一化处理得到 $\alpha(h, r, t)$, 如公式(3)所示:

$$\alpha(h, r, t) = \frac{\exp(\alpha_0(h, r, t))}{\sum_{(h, r', t') \in N_h} \exp(\alpha_0(h, r', t'))} \quad (3)$$

模型中需要训练的参数包括 W_0, W_1, w_2, b_1, b_2 。在进行注意力权重归一化时,Softmax 函数确保所有相关联实体关系对的注意力权重总和为 1,从而使模型能够在考虑所有相关三元组时自适应地分配关注点。

2.2 上下文嵌入层

通过上下文嵌入层编码实体在文档中的上下文信息来精简输入,提高系统的处理效率,并极大地提高新闻文章中实体重要性和相关性的识别能力。

1)位置编码。位置编码反映了实体在文档中的位置,对用户来讲,标题中出现的实体可能比正文中出现的实体更重要。为了对实体位置进行编

码,引入了一个位置偏差向量,记为 $B_{p_i}^{(1)}$, 其中 $p_i \in \{1, 2\}$ 表示实体 i 出现在标题或正文的位置类型, $B_1^{(1)}$ 代表实体位于标题中, $B_2^{(1)}$ 代表实体位于正文中。

2)频率编码。实体在文档中出现的频率是另一个重要的指标,它可以反映实体与文章主题的相关性以及重要性。为了量化这一指标,引入了一个编码矩阵 $B^{(2)}$, 用于将实体的出现频率 f_i 转化为一个嵌入向量,记为 $B_{f_i}^{(2)}$ 。这一编码过程考虑了实体频率的离散特性,并允许模型通过累积实体出现的次数来学习实体的重要性。为了避免频率的异常值影响模型的稳定性, f_i 的上限设定为 20,即超过该阈值的实体频率都被认为是相同的。

3)类别编码。实体的类别信息可以为模型提供关于实体内容的直接线索。为了将类别信息集成到实体表示中,设计了一个类别编码矩阵 $B^{(3)}$ 。每个实体根据其类别 t_i 被赋予一个类别编码向量 $B_{t_i}^{(3)}$, 从而使得模型能够辨识并利用这些类别信息,以增强新闻内容的理解。

在上述三种上下文编码策略的基础上,每个实体 h 的最终上下文嵌入向量 e_{T_h} 是通过将这些编码向量与实体表示层生成的实体 e_{N_h} 进行元素级加法构建的,如公式(4)所示:

$$e_{T_h} = e_{N_h} + B_{p_h}^{(1)} + B_{f_h}^{(2)} + B_{t_h}^{(3)} \quad (4)$$

公式中 $B_{p_h}^{(1)}$ 、 $B_{f_h}^{(2)}$ 和 $B_{t_h}^{(3)}$ 分别代表实体 h 的位置编码、频率编码和类别编码向量。这种元素级加法保证了不同上下文特征的独立性,并允许模型在不同特征之间进行有效的信息融合。

2.3 注意力聚合层

为了在模型中准确捕捉并表示实体的动态重要性,使用了注意力机制,用来捕捉并加权实体间的相对重要性。原始的文档向量 v_d 为 Query, 上下文嵌入层生成的实体 e_{T_h} 用作 Key 和 Value。可以通过公式(5)和公式(6)计算得出注意力权重 $\beta(h, v)$:

$$\beta_0(h, v) = w_2 \text{ReLU} (W_1 (e_{T_h} \oplus v_d) + b_1) + b_2 \quad (5)$$

$$\beta(h, v) = \frac{\exp(\beta_0(h, v))}{\sum_{t \in \epsilon_v} \exp(\beta_0(t, v))} \quad (6)$$

ϵ_v 代表文档 v 的实体集。上下文嵌入层生成的实体表示 e_{T_h} 经过注意力权重 $\beta(h, v)$ 加权后,得到实体的加权表示 e_{O_h} , 如公式(7)所示:

$$e_{O_h} = \sum_{h \in \epsilon_v} \beta(h, v) e_{T_h} \quad (7)$$

最后,加权实体表示与原始文档向量 v_d 结合,

通过一个全连接的前馈网络进行进一步的转换,生成知识感知的文档向量 v_i , 其计算过程如公式(8)所示:

$$v_i = \text{Tanh}(w_3(e_{o_h} \oplus v_d) + b_3) \quad (8)$$

v_i 是经过注意力机制和前馈网络转换后得到的文档表示,它集成了实体的重要性信息和文档的上下文信息,为后续的新闻推荐任务提供了丰富的语义表示。之后进一步利用 LAMA 编码器对 v_i 进行处理,得到最终的新闻嵌入 v_k ,即知识感知的文档向量 KDV。

3 基于知识增强文档的细粒度用户建模

受 FUM^[13] 的启发,本节提出了一种新的用户建模方法,设计了如图 2 所示的模型。其中的新闻编码器(News Encoder)为图 1 所示的新闻建模框架,用来生成知识增强的新闻文档向量。首先,将用户点击的历史新闻文档经过新闻编码器处理后生成知识增强的新闻文档表示。对于粗粒度用户建模部分,输入新闻文档向量表示,通过 Transformer 和多头自注意力机制得到用户表示。在细粒度用户建模部分,首先通过 Longformer 进行长文档建模,再通过多头自注意力机制生成用户表示。最终将粗粒度用户表示和细粒度用户表示进行聚合,得到最终的用户表示。

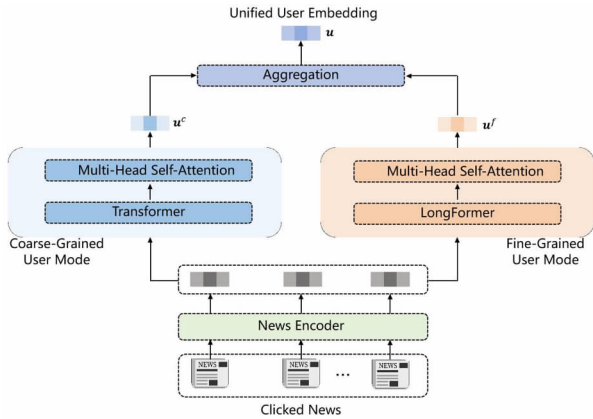


图 2 基于知识增强文档的细粒度用户建模框架

3.1 细粒度用户建模

为了在长文档中建模用户的细粒度行为交互,本文采用 Longformer 模型。Longformer 首先将全局上下文总结为一个嵌入 q , 然后基于与全局上下文的相关性转换每个 token 的嵌入,如公式(9)(10)(11)所示:

$$q = \text{Att}(q_1, \dots, q_L), q_i = W_q h_i \quad (9)$$

$$k = \text{Att}(q * k_1, \dots, q * k_L), k_i = W_k h_i \quad (10)$$

$$\hat{h}_i = W_o(k * v_i), v_i = W_v h_i \quad (11)$$

其中 h_i 和 $\hat{h}_i$ 表示行为嵌入序列中第 i 个 token 的输入和输出, $*$ 表示元素间的乘积, $\text{Att}(\cdot)$ 表示注意力汇聚网络, W_q, W_k, W_v 和 W_o 表示可训练的参数。然后,将不同注意力头的输出串联起来,为第 i 个 token 构建统一的上下文表示 g_i , 然后采用一个注意力网络学习每条点击新闻的嵌入,如公式(12)所示:

$$f_i = \text{Att}(g_{(i-1)kl+1}, g_{(i-1)kl+2}, \dots, g_{ikl}) \quad (12)$$

其中 f_i 代表第 i 条点击新闻。

最后,将它们聚合以构建用户嵌入,得到细粒度的用户表示 u^f , 如公式(13)所示:

$$u^f = \text{Att}(f_1, \dots, f_m) \quad (13)$$

通过这种方式,可以有效并准确地从词级细粒度行为交互中建模用户兴趣。

3.2 粗粒度用户建模

为了捕获用户兴趣的全貌,从宏观层面提炼用户整体的阅读习惯和用户偏好,在用户建模框架中还设计了一个粗粒度用户建模块件。首先利用新闻编码器处理用户点击的历史新闻,得到知识增强的新闻文档嵌入向量。这些嵌入向量被输入一个 Transformer 网络中,通过自注意力机制学习各个新闻嵌入之间的复杂关系。为了进一步捕捉用户的兴趣模式,引入了多头自注意力机制,每个注意力头独立学习序列中不同方面的特征,然后将它们整合起来,形成一个全面的用户表示。通过以上步骤最终得到粗粒度的用户表示 u^c 。

3.3 模型结合

通过结合粗粒度的用户表示 u^c 和细粒度用户表示 u^f , 形成了一个统一的用户表示 u , 用公式(14)表示如下:

$$u = u^c + u^f \quad (14)$$

其中, u^f 来自细粒度用户模型,反映了用户兴趣的微观特征, u^c 来自粗粒度用户模型,体现了用户兴趣的宏观特征。

4 实验及结果分析

4.1 数据集

实验所用的数据集与 KRED 中所用的数据集一致。该数据集收集了 2019 年 1 月 15 日至 1 月 28 日期间用户的点击历史和映像日志(在新闻推荐系统中,当给用户展示一组新闻时,就形成了一个映像)。其中第一周的数据用作模型的训练和验证,第二周的数据用作测试。此外还收集了训练日

期前两周的用户点击历史和映像日志,以便进行更为精准的用户建模。在此过程中,删除了那些用户历史点击新闻数量不足 5 篇的数据。经过筛选,最终得到的数据集(包括训练集、验证集及测试集)共计涵盖了 665034 位用户、24542 篇新闻文章以及 1590092 个映像日志,每篇新闻文章包含大约 701 个词。在实体链接的过程中,剔除了置信度小于 0.9 的实体数据,然后利用 KRED 提出的技术对知识图谱进行处理。经过处理之后,每篇新闻包含 24 个实体,这些实体在知识图谱中的一跳邻域总共覆盖了 3314628 个实体、1006 种不同的关系类型以及 71727874 个知识三元组。

4.2 实验设置

实验在云服务器上进行,实验环境及软硬件配置信息如表 1 所示。

表 1 实验环境及软硬件配置

实验环境/硬件	描述
编程语言	Python 3.10
深度学习框架	PyTorch 2.1.0
运行环境	Ubuntu 22.04
处理器型号	16 vCPU Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz(120GB)
显卡型号	NVIDIA GeForce RTX 4090(24GB)

模型训练的相关参数设置如表 2 所示。

表 2 模型训练参数设置

参数名称	数值
学习率	0.001
L2 正则化系数	1×10^{-5}
神经网络维数	128
实体嵌入维数	90
批大小	128

实验采用 AUC 和 NDCG@10 来评估模型的性能,并选择了 FM^[14]、Wide&Deep^[15]、NAML^[16]、DKN^[11]、STCKA^[17]、KRED^[10] 和 ERNIE^[18] 共 7 种方法作为实验对比的基线方法。

4.3 实验结果

为了评估模型的效果,利用了注意力机制对用户历史点击文档向量进行处理,以获取每篇文章的注意力得分。随后将注意力得分与文档向量进行加权融合,从而得到用户向量。通过将用户向量和新闻向量输入预测函数中,计算得到预测分数,以此来评估推荐效果的准确性和可行性。此计算过程如公式(15)所示:

$$\hat{y}(i,j) = g(\boldsymbol{u}_i \oplus \boldsymbol{v}_j)$$
 (15)

其中 $\boldsymbol{u}_i$ 是用户向量; $\boldsymbol{v}_j$ 是新闻文档向量; g 是预测函数; $\hat{y}(i,j)$ 是预测得分。

本文提出的模型与其他的基线模型在 AUC 和 NDCG@10 指标上的具体表现如表 3 所示。

表 3 本文提出的模型和基线模型在新闻推荐任务上的对比实验结果

模型	AUC	NDCG@10
FM	0.6569	0.3363
Wide&Deep	0.6635	0.3401
NAML	0.6740	0.3458
DKN	0.6677	0.3407
STCKA	0.6726	0.3460
ERNIE	0.6811	0.3524
KRED	0.6855	0.3632
知识增强的新闻建模+ 传统用户建模	0.6870	0.3688
知识增强的新闻建模+ 细粒度用户建模	0.7106	0.3817

实验结果表明:

(1)知识增强的新闻建模+传统用户建模优于 KRED,表明通过低秩分解来优化自注意力机制,将低秩分解的计算效率和自注意力机制的长距离依赖捕捉能力进行结合,能够进一步提升文档向量的语义表示。

(2)知识增强的新闻建模+传统用户建模优于其他基线模型,表明该模型基于知识图谱设计的三层结构的有效性,最终生成的知识增强的文档表示可以提升新闻推荐的性能。与 ERNIE、DKN 和 STCKA 等模型相比,该模型更深入地利用了知识图谱的结构化信息,使得它在理解复杂的用户兴趣和新闻内容方面表现更为出色。

(3)知识增强的新闻建模+细粒度用户建模在所有对比的模型中性能最优。在 AUC 和 NDCG@10 两个评价指标上都取得了最高的分数,分别为 0.7106和 0.3817。与其他模型相比,知识增强的新闻建模+细粒度用户建模在 AUC 上的性能至少提升了 2.51%,在 NDCG@10 上至少提升了 4.75%。

4.4 消融实验

知识增强的新闻建模框架由三层(实体表示层、上下文嵌入层、注意力聚合层)构成,为了验证每一层的有效性和必要性,通过去除每一层进行实验,观察结果是否会受影响,有助于确定模型各部分对整体性能的贡献,实验结果如表 4 所示。

表 4 知识增强的新闻建模+传统用户建模
消融实验结果

模型	AUC		NDCG@10	
完整模型	0.6870		0.3688	
去除实体表示层	0.6834	↓ 0.52%	0.3663	↓ 0.68%
去除上下文嵌入层	0.6838	↓ 0.47%	0.3624	↓ 1.74%
去除注意力聚合层	0.6822	↓ 0.70%	0.3651	↓ 1.00%

由结果可知,去除知识增强的新闻建模框架的每一层都会降低模型的性能,这表明模型每一层都是有效且必要的。

知识增强的新闻建模+细粒度用户建模消融实验结果如表 5 所示。

表 5 知识增强的新闻建模+细粒度用户
建模消融实验结果

模型	AUC		NDCG@10	
细粒度+粗粒度	0.7106		0.3817	
去除细粒度	0.7030	↓ 1.07%	0.3754	↓ 1.65%
去除粗粒度	0.6994	↓ 1.58%	0.3696	↓ 3.17%

结果表明,去除细粒度用户建模组件或粗粒度用户建模组件后,推荐性能均有所下降,并且仅保留粗粒度用户建模组件的模型表现优于仅保留细粒度用户建模组件的模型,这可能是因为细粒度用户建模组件无法有效利用新闻内部的交互。

5 结 论

新闻建模和用户建模是新闻推荐系统的核心组成部分,但在传统的新闻推荐系统中,新闻文档建模和用户建模均存在局限性。本文提出了一种知识增强的新闻建模方法,它由实体表示层、上下文嵌入层和注意力聚合层三个核心部分组成。这种通过知识图谱和注意力机制相结合的方法,不仅显著提升了新闻推荐的准确度和个性化程度,也提升了新闻推荐的可解释性。在用户建模方面,本文提出了一种基于知识增强文档的细粒度用户建模方法,更准确地反映用户的兴趣多样性和复杂性。实验表明,本文提出的新闻建模和用户建模方法,相比于传统的新闻推荐模型有更好的表现。并且通过消融实验,也验证了新闻建模中的实体表示层、上下文嵌入层和注意力聚合层以及用户建模中的粗粒度和细粒度建模组件在整个模型中的有效性和必要性。未来将从多模态数据融合、实时性与动态更新、跨域推荐技术、隐私保护与推荐公正性、

解释性和透明度几个方面继续展开研究。

参考文献

[1] WU C, WU F, HUANG Y, et al. Personalized news recommendation: methods and challenges[J]. ACM Transactions on Information Systems, 2021, 41: 1–50.

[2] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[J]. Advances in Neural Information Processing Systems, 2013: 2787–2795.

[3] WANG X, HE X, CAO Y, et al. Kgat: knowledge graph attention network for recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 950–958.

[4] MEHTA S, RANGWALA H, RAMAKRISHNAN N. Low rank factorization for compact multi-head self-attention[J]. ArXiv Preprint ArXiv: 1912.00835, 2019.

[5] BELTAGY I, PETERS ME, COHAN A. Longformer: the long-document transformer [J]. ArXiv Preprint ArXiv: 2004.05150, 2020.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017: 6000–6010.

[7] GOOSSEN F, IJNTEMA W, FRASINCAR F, et al. News personalization using the CF-IDF semantic recommender [C]//Proceedings of the International Conference on Web Intelligence, Mining and Semantics. 2011: 1–12.

[8] CHU W, PARK S T. Personalized recommendation on dynamic content using predictive bilinear models[C]//Proceedings of the 18th International Conference on World Wide Web. 2009: 691–700.

[9] YI X, HONG L, ZHONG E, et al. Beyond clicks: dwell time for personalization[C]//Proceedings of the 8th ACM Conference on Recommender Systems. 2014: 113–120.

[10] LIU D, LIAN J, WANG S, et al. KRED: Knowledge-aware document representation for news recommendations [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020: 200–209.

[11] WANG H, ZHANG F, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation[C]//Proceedings of the 2018 World Wide Web Conference. 2018: 1835–1844.

[12] OKURA S, TAGAMI Y, ONO S, et al. Embedding-based news recommendation for millions of users[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 1933–1942.

[13] QI T, WU F, WU C, et al. FUM: fine-grained and fast user modeling for news recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022: 1974–1978.

[14] RENDLE S. Factorization machines[C]//2010 IEEE International Conference on Data Mining. IEEE, 2010: 995–1000.

[15] CHENG H T, KOC L, HARMSSEN J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. 2016: 7–10.

[16] WU C, WU F, AN M, et al. Neural news recommendation with attentive multi-view learning[J]. ArXiv Preprint ArXiv: 1907.05576, 2019.

[17] CHEN J, HU Y, LIU J, et al. Deep short text classification with knowledge powered attention[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 6252–6259.

[18] ZHANG Z, HAN X, LIU Z, et al. ERNIE: Enhanced language representation with informative entities [J]. ArXiv Preprint ArXiv: 1905.07129, 2019.