

# 一种用于视频推荐的基于 LDA 的深度学习模型

程 涛, 崔宗敏, 喻 静

(九江学院 信息科学与技术学院, 江西 九江 332005)

**摘 要:**随着视频产业的发展,大量的视频已沉积在信息海洋中,为了缓解这种现象,越来越多的推荐算法开始应用于个性化视频推荐。然而目前的推荐算法都是以协同过滤为自主,都只注重通过评分矩阵提高捕捉用户与视频的低阶的交互,忽略了用户兴趣与视频属性的高阶关联。在这种背景下,文中通过 LDA 主题模型预测用户兴趣主题,引用干扰词典和关键词词典来提高 LDA 模型对视频文本聚类的准确率,然后利用最近提出的神经协同框架建模用户兴趣和视频属性的高阶关联,把 LDA 模型与深度学习相结合提出了一种新的模型 LIVR 用于视频推荐,最后在通过网络爬虫爬取的数据集上验证并对实验结果进行分析。结果表明,该模型的 Top-N 推荐准确率较常见的几个深度学习模型提高了约 1.9 个百分点。

**关键词:**LDA 主题模型;深度学习;隐式反馈;视频推荐;个性化推荐

中图分类号:TP391.5

文献标识码:A

文章编号:1673-629X(2020)08-0086-05

doi:10.3969/j.issn.1673-629X.2020.08.014

## A LDA-based Topic Attribute-aware in-Depth Learning Model for Video Recommendation

CHENG Tao, CUI Zong-min, YU Jing

(School of Information Science and Technology, Jiujiang University, Jiujiang 332005, China)

**Abstract:** With the development of video industry, a large amount of videos have been deposited in the information ocean. In order to alleviate this phenomenon, more and more recommendation algorithms have begun to be applied to personalized video recommendation. However, the current recommendation algorithms are all based on collaborative filtering. They only focus on improving the low-level interaction between users and videos through the scoring matrix, ignoring the high-order association between user interests and video attributes. In this context, we use the LDA topic model to predict user interest topics, cite interference dictionaries and key dictionaries to improve the accuracy of LDA model clustering of video texts, and then use the recently proposed neural collaboration framework to model high-order associations of user interests and video attributes. Combined with LDA model and deep learning, a new model LIVR is proposed for video recommendation. Finally, it is verified on the dataset crawled by the web crawler and the experimental results are analyzed. The results show that the Top-N recommendation accuracy of the model is about 1.9 percentage points higher than the common deep learning models.

**Key words:** LDA theme model; deep learning; implicit feedback; video recommendation; personalized recommendation

## 0 引 言

随着通讯技术与大数据产业的高速发展,信息资源在网络中以指数爆炸式增长的同时,信息载体携带的信息量也逐渐增大。其中作为人们常见的信息媒体之一,视频已成为人们消遣娱乐的主要方式。如今,抖音、B站、爱奇艺等视频业务蒸蒸日上,视频资源也日益增多,用户开始困于无法在短时间寻找到自己喜爱的视频,他们迫切需要一种既可以节约时间又可以向他们进行个性化视频信息推荐<sup>[1-2]</sup>的推荐系统。为此,视频推荐系统以其独特的解决视频信息过载<sup>[3]</sup>的

能力开始渐渐地走入了人们的生活,在近几年也受到了越来越多的重视。

传统的视频推荐方法主要包括协同过滤<sup>[4-5]</sup>、基于内容的推荐方法<sup>[6]</sup>和其他混合推荐方法<sup>[7-8]</sup>等,其中协同过滤是目前应用最为广泛的推荐算法,然而由于经典的协同过滤方法采用浅层模型,其无法学习到用户和项目的深层次特征<sup>[9]</sup>。比如在现实生活中,用户的兴趣和视频的属性总存在着大量交互关联,拥有“漫威”兴趣属性的用户喜欢看拥有“漫威”属性的视频。倘若能考虑到这些丰富的关联,将大大提高视频

收稿日期:2019-09-13

修回日期:2020-01-16

基金项目:国家自然科学基金(61762055)

作者简介:程 涛(1998-),男,研究方向为个性化推荐;崔宗敏,通讯作者,博士,副教授,硕导,研究方向为大数据计算。

推荐系统的性能,然而目前提出的方法<sup>[10-12]</sup>都是通过计算向量的距离,矩阵分解,文本分析等方法进行相关的视频推荐。它们都更多关注于通过改进距离计算和矩阵分解提高推荐效果,大多忽略了交互关联的潜在作用。

考虑类似属性的传统算法便是基于内容的推荐方法<sup>[13]</sup>,通过自动分析视频内容提取的视频视觉特征,提出了一种新的基于内容的推荐系统<sup>[14]</sup>。通过从候选视频的标题中提取的关键词与目标词进行比较,提出一种基于视频标题的视频推荐系统。这些方法主要通过用户观看过的视频属性为其推荐其他具有类似属性的视频,其依赖于人工设计特征,需要对视频属性特征进行有效的提取,是一种依赖于人工设计特征传统的浅层模型,因此其有效性和可扩展性非常有限。

作为人工智能的一个热潮,深度学习于近几年在图像处理、自然语言理解和语音识别<sup>[15-16]</sup>等领域取得了突破性进展,为个性化推荐的研究带来了新的机遇。深度学习常见的模型包括玻尔兹曼机、RNN、LSTM,以及<sup>[17]</sup>提出的MLP和GMF等通过将不同的数据映射到一个相同的隐空间,使其具备能够获得数据的统一表征的能力,因此,基于深度学习的融合传统推荐方法能够有效利用多源异构数据<sup>[18]</sup>对用户进行个性化推荐,缓解传统推荐系统中存在的数据稀疏和冷启动问题。当然,目前常用的深度学习模型也存在花费巨大和依赖大数据的问题,当需表征得数据过大时,会面临训练耗时延长等问题。

为了解决上述问题,文中结合传统的基于LDA模型<sup>[19]</sup>的视频推荐算法和当下火热的深度学习,充分利用用户兴趣和视频丰富属性间的关联,使用视频LDA主题模型对视频的丰富属性进行聚类,从而降低数据维度以减少深度学习的负担,然后基于近期提出的神经协同过滤框架<sup>[17]</sup>进行建模训练从而得到推荐结果。

## 1 LDA 模型

LDA (latent Dirichlet allocation) 是 2003 年由 David Blei 等提出的一种文档主题生成模型<sup>[20]</sup>,也称为一个三层贝叶斯概率模型,包含词、主题和文档三层结构。它是利用先验分布对数据进行似然估计并最终得到后验分布的一种方式,如式(1)所示:

$$P_{(\text{Word}|\text{Document})} = P_{(\text{Topic}|\text{Document})} * P_{(\text{Word}|\text{Topic})} \quad (1)$$

对每一篇文档,从主题分布中抽取一个主题,然后从上述被抽到的主题所对应的单词分布中抽取一个单词,这样重复上述过程直至遍历文档中的每一个单词。其生成方式为首先从 Dirichlet 分布  $\alpha$  中取样生成文档的主题分布  $\theta_m$ ,从主题的多项式分布  $\theta_m$  中取样生成文档  $m$  第  $n$  个词的主题  $Z_{m,n}$ ,然后从 Dirichlet 分布  $\beta$

中取样生成主题  $Z_{m,n}$  对应的词语分布  $\phi_{m,n}$ ,最后从词语的多项式分布  $\phi_{m,n}$  中采样最终生成词语  $W_{m,n}$ 。

## 2 LIVR 模型

众所周知,每个视频都有相当多的文档信息,比如介绍信息以及评论信息等,倘若将这些相关信息看作一个整体的文档,便可以通过 LDA 主题模型,获得每个电影文档的主题-概率分布,然后可以利用这些概率预测每个视频的主题属性,从而根据用户历史观看数据预测出用户感兴趣的主体。首先进行对数据的爬取,通过对这些数据的处理构建训练样本以及相关词典。对于爬取的视频信息文本文档,先进行 jieba 分词,并将分词结果作为语料进行 LDA 模型训练。然后通过视频 LDA 训练结果预测出用户的兴趣主题集和视频的属性主体集,再基于神经协同框架将用户/视频与他们相关主题属性进行建模,通过对两种模型的结合,便构成了 LIVR 模型。最后,只需用户的隐式反馈便可以训练出最优模型,从而对用户生成推荐列表。

### 2.1 视频 LDA 模型

传统的基于 LDA 模型的文本分析方法中,比如用于微博的文本分析方法,其常是单纯的进行语料爬取,然后将语料经过去停用词和分词后直接进行 LDA 训练得到文档-主题概率分布。考虑到所有视频的信息文档都具有视频媒体类的特征,传统方法可能会把《钢铁侠》和《生化危机》聚为一类即视频类,因为它们的信息文档中都存在大量的视频类的特征词,诸如“主演”“导演”“演员”等等。并且尽管 jieba 有着新词识别能力,但是一些关键的特征词诸如“美国队长”可能会被分成“美国”和“队长”,这样将严重降低分词的正确率。为了提高聚类效果,在传统的基于 LDA 模型的文本分析方法的基础上,引入了视频词典和干扰词典进行改进视频文档的聚类效果。

定义 1(视频词典):由所有视频的标题属性和爬取的输入法中的影视词汇构成的词典定义为关键词典,记为  $U_{dic}$ 。

定义 2(干扰词典):由所有会影响到聚类效果的词汇定义为干扰词,记为  $I_w \in I_{ws}$ ,  $I_w$  构成的词典定义为干扰词典,记为  $I_{ws}$ 。

定义 3(特征词):将视频整体文档  $D$  中具有实体意义的词定义为特征词,记为  $F_w$ ,所有  $F_w$  构成的集合定义为特征词集,记为  $F_{ws}$ 。

根据以上定义,将一个视频的整体文档中的标题属性和所有的关键属性诸如导演、演员、标签等直接作为这个视频的特征词  $F_w$ ,通过对视频的文本属性进行去干扰词和视频词典的验证,可以将每个视频的特征词词集描述为:

$$F_{ws} = (Jb - Jb \cap I_{ws}) \cup At \quad (2)$$

其中,  $Jb$  为导入视频词典后的 jieba 分词结果,  $At$  为相关关键属性集合。

## 2.2 构建主题兴趣集和主题属性集

通过视频 LDA 模型,可以预测出每个视频的视频-主题概率分布,进一步将其描述为一个矩阵  $\mathbf{VT}$ :

$$\mathbf{VT} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nk} \end{bmatrix} \quad (3)$$

将该视频的每个主题预测的概率作为该主题所占的属性主题权重,然后通过设置阈值  $s$ ,从视频的所有主题中挑选出概率大于  $s$  的主题构成这个视频的属性主题集  $ATS$  (Attribute-Tfs),若  $ATS$  为空则取前 3 个主题。

然后可以利用用户的观看历史构建的交互视频集  $IFVs$  和视频-主题矩阵  $\mathbf{VT}$  预测出他的主题兴趣集  $ITS$  (Interest-Tfs)。首先根据用户观看过的所有视频记录,利用式(4)计算出用户偏爱的兴趣主题权重。

$$W_i^u = \frac{\sum_{m=1}^n w_i^{IFVs_m}}{n} \quad (4)$$

其中,  $u$  代表用户  $u$ ,  $i$  代表第  $i$  个主题,  $n$  为用户  $u$  的交互视频集元素数量,  $m$  为交互视频集中第  $m$  个视频的编号。进一步用  $W_i^u$  表示用户  $u$  的第  $i$  个主题的权重,通过设置阈值  $s_2$ ,从所有主题中挑选出  $W_i^u$  大于  $s_2$  的主题构成该用户的主题兴趣集  $ITS$ ,若  $ITS$  为空则根据  $W_i^u$  取前 3 个主题。

## 2.3 生成推荐列表

利用神经协同过滤框架对用户/项目和视频 LDA 模型预测的  $ITS$  和  $ATS$  进行建模训练,考虑到要想建模用户/视频和它们的主题属性之间的成对关联,以及它的属性之间的所有嵌套关联,最常用的便是神经网络建模中的池操作,因此,决定使用文献[21]提出的 pairwise-pooling 操作进行建模。如式 5 建模用户  $u$  和他的兴趣主题集  $ITS$  的成对关系,其中  $u'$ ,  $ITS^{u'}$  为用户,用户兴趣主题集的  $K$  维嵌入量,  $|ITS^u|$  为兴趣主题集的长度,  $\odot$  表示两个逐元素相乘。同理,如式 6 建模视频  $v$  和他的属性主题集  $ATS$  的成对关系。

$$\text{pariwise}(u, \{ITS^u\}) = \sum_{i=1}^{|ITS^u|} u \odot ITS_i^u + \sum_{i=1}^{|ITS^u|} \sum_{i'=i+1}^{|ITS^u|} ITS_i^u \odot ITS_{i'}^{u'} \quad (5)$$

$$\text{pariwise}(v, \{ATS^v\}) = \sum_{i=1}^{|ATS^v|} v \odot ATS_i^v + \sum_{i=1}^{|ATS^v|} \sum_{i'=i+1}^{|ATS^v|} ATS_i^v \odot ATS_{i'}^{v'} \quad (6)$$

通过在神经网络的嵌入层上添加 pooling 层进行

pool-wise 操作,便可以利用全连接层捕获用户、项和属性之间的非线性和高阶关联。然后将  $Y_{u,v}$  的值作为一个标签,1 表示项目  $i$  和用户  $u$  相关,否则为 0。通过使用概率函数 Sigmoid,如式(7),作为激活函数作用在输出层,使得输出的  $\hat{Y}_{u,v}$  在 0~1 之间,代表项目  $i$  和用户  $u$  相关的可能性大小。

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (7)$$

通过二类交叉熵损失函数最小化  $\hat{Y}_{u,v}$  和其目标值  $Y_{u,v}$  之间的损失,并使用随机梯度下降(SGD)来进行训练优化。若  $W_l$ ,  $b_l$ ,  $\sigma_l$ ,  $E_l$  分别表示第  $l$  层隐层的权重矩阵,偏好向量,激活函数以及输出向量,将预测层形式上定义为式 8:

$$\begin{aligned} E_1 &= \sigma_1(W_1(\text{pairwise}(u, ITS^u) \odot \text{pairwise}(v, ATS^v)) + b_1) \\ E_2 &= \sigma_2 W_2 \sigma_1 + b_2 \\ &\dots \\ E_l &= \sigma_l(W_l \sigma_{l-1} + b_l) \\ \hat{Y}_{u,v} &= w^T * E_l \end{aligned} \quad (8)$$

其中,  $w$  为预测层的权重向量,  $E_l$  为隐藏层最后的输出向量,  $\hat{Y}_{u,v}$  为预测得分。

## 3 实验检测

### 3.1 实验数据与测试模型

爬取了豆瓣电影网中的 1 204 部电影的相关信息,966 个用户,96 255 条评分记录作为实验数据集。然后对爬取到的评论与简介信息进行传统 LDA 方法中的同义词处理,将中文常用的比如“黄渤”改为“黄勃”,“周星弛”改为“周星驰”,使用 python 的 gensim 库进行 LDA 模型训练,通过评估 perplexity 来确定最优主题个数为  $K=30$ ,通过预测样本的主题兴趣集和主题属性集构成神经网络的样本集。然后对样本集随机取 70% 作为训练集,30% 作为测试集,并在 python 中使用 Keras 深度学习框架进行神经网络的训练。最后通过对比传统深度学习模型 MLP(多层感知机)和 GMF(广义矩阵分解),对三种方法进行评估与分析。

### 3.2 实验性能评估指标

使用离线实验方法进行评测,最终结果取 3 次实验数据的平均值。评价指标选用准确率、召回率和  $F$  值。召回率描述有多少真正产生过行为的物品包含在最终的推荐列表中,计算公式为:

$$\text{Recall}_u = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (9)$$

准确率描述最终的推荐列表中包含多少真正产生



过行为的物品,计算公式为:

$$\text{Precision}_u = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |R(u)|} \quad (10)$$

$F$  值是准确率和召回率之间的调和平均值。用户  $u$  在测试集合上产生过行为的物品记为  $T(u)$ , 计算公式为:

$$F = \frac{2 * P * R}{P + R} \quad (11)$$

### 3.3 实验结果与分析

图1为在推荐视频个数  $K=10$ , LIVR 模型的 LDA 主题个数为 30 时,不同模型随迭代次数的增加  $F$  值的变化曲线。

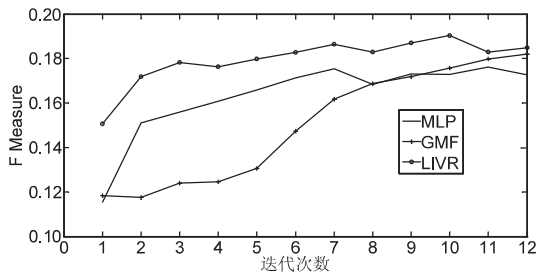


图1  $F$  值变化曲线

从图1可以看出,随着迭代次数的增加,LIVR 性能的改善速率虽然没 GMF 以及 MLP 快,但两种模型在前 10 次迭代过程中均远远落后 LIVR,且 LIVR 在第 10 次时最快迭代时达到最优,即使是 10 次迭代之后仍处于优势,其中可以看出,由于过拟合的原因,添加了隐层的 LIVR 和 MLP 在第 8 次迭代后的准确度和召回率都开始下降。关于过拟合问题,可以通过正则化<sup>[22]</sup>缓解,这里就不再累述。而 GMF 的准确度和召回率仍处于上升阶段,但是这并不能说明 MLP 和 GMF 的回归表现比 GMF 差,因为这里展示的只是具有 3 层隐层的 MLP 和 LIVR。

图2~图4分别表示各方法随着  $K$  的变化 precise、recall 和  $F$  值的变化。可以看出,在三种模型中,随着  $K$  的增大,LIVR 在三种评价指标下都明显处于上风,通过图中 MLP 与 GMF 的表现比较,MLP 较 GMF 的三种评价指标都低,但是这并不能说明 MLP 的比 GMF 的表现差,因为 MLP 模型可以通过添加更多的隐藏层来进一步改进,值得一提的是,LIVR 同样具有 MLP 的这个特点。

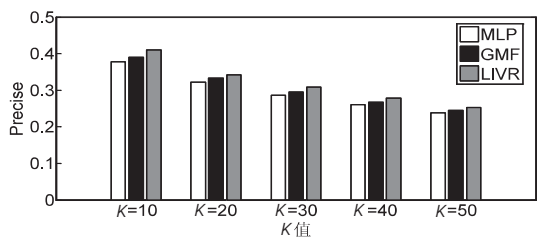


图2 precise 的比较

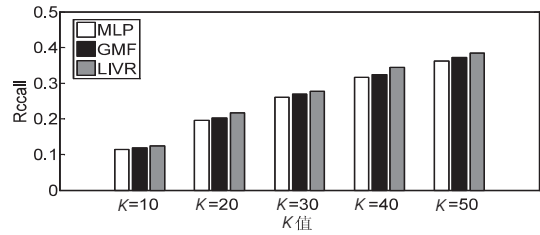


图3 recall 的比较

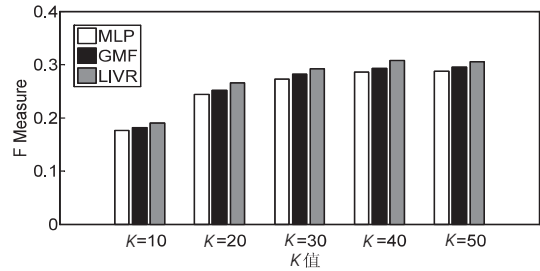


图4  $F$  值的比较

从三种方法的对比中可以看出 LIVR 表现最好,这都归功于隐藏层前使用了 poolingwise 操作,充分地为用户/视频与视频 LDA 主题模型的预测主题集的相关性建模。由图可以看出,与其他两种方法相比,LIVR 在训练阶段准确率和召回率上都有很大的提升,对比准确率和召回率之间的调和平均值,相比于 MLP 模型,LIVR 大约提升 1.9%,相比于 GMF,大约提升 1.1%。值得一提的是受干扰词的影响,LIVR 样本仍存在些许脏数据,然而样本的数据质量是可以通过人工标记进行进一步改善的,LIVR 效果也可以得到进一步的提升。

## 4 结束语

针对如何利用现实生活中用户和视频间丰富的属性关联提高推荐系统性能的问题,提出了一种基于 LDA 主题属性感知深度学习模型。通过构建的 LDA 视频模型对视频进行聚类,从而预测出用户与视频的相关主题属性,并利用 pooling-wise 操作改进的神经协同框架进行学习。最后通过实验验证,该模型在一定程度上在神经协同框架传统模型上提高了推荐质量。面对日益增多的视频文档信息,用户与视频数据量的急剧增加,首先考虑到模型的扩展性问题,当文本数据量达到一定水平时,模型聚类效果将渐渐降低,因此下一步的重点是解决系统的扩展性问题。

### 参考文献:

- [1] DU X, YIN H, CHEN L, et al. Personalized video recommendation using rich contents from videos[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 21: 31-27.
- [2] TRIPATHI A, ASHWIN T, GUDDETI R M R. EmoWare: a context-aware framework for personalized video recommen-

- ation using affective video sequences [J]. IEEE Access, 2019, 7: 51185–51200.
- [3] SAHU H, SHARMA N, GUPTA U. A new framework for collecting implicit user feedback for movie and video recommender system [M]//Recent trends in communication, computing, and electronics. Berlin: Springer, 2019: 399–408.
- [4] CHU P M, TSAI H R, LEE S J, et al. Improving collaborative filtering recommendation [C]//2018 3rd international conference on control, automation and artificial intelligence (CAAI 2018). Beijing: Atlantis Press, 2018.
- [5] YAO L, XU Z, ZHOU X, et al. Synergies between association rules and collaborative filtering in recommender system: an application to auto industry [M]//Data science and digital business. Berlin: Springer, 2019: 65–80.
- [6] LOPS P, JANNACH D, MUSTO C, et al. Trends in content-based recommendation [J]. User Modeling and User-Adapted Interaction, 2019, 32: 1–11.
- [7] SMITH J, DILLON W M J, FREEMAN J, et al. Towards a hybrid recommendation system for a sound library [C]//Joint proceedings of the ACM IUI 2019 workshops. Tokyo, Japan: Alan Said, 2019.
- [8] LOGESH R, SUBRAMANIASWAMY V. Exploring hybrid recommender systems for personalized travel applications [M]//Cognitive informatics and soft computing. Berlin: Springer, 2019: 535–544.
- [9] PERERA P, PATEL V M. Learning deep features for one-class classification [J]. IEEE Transactions on Image Processing, 2019, 34: 223–235.
- [10] JAIN S, PAWAR T, SHAH H, et al. Video recommendation system based on human interest [C]//2019 international conference on innovations in information and communication technology (ICICT). Greater Noida: IEEE, 2019: 1–4.
- [11] BARLASKAR U H, PALNITKAR A. Live video recommendation by an online system: U. S, 15/823256 [P]. 2019–05–30.
- [12] NILASHI M, IBRAHIM O, BAGHERIFARD K. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques [J]. Expert Systems with Applications, 2018, 92: 507–520.
- [13] DELDJOO Y, ELAHI M, CREMONESI P, et al. Content-based video recommendation system based on stylistic visual features [J]. Journal on Data Semantics, 2016, 5 (2): 99–113.
- [14] CHECKLEY G. Video recommendation based on video titles: U. S, 14/833496 [P]. 2017–03–02.
- [15] HATT M, PARMAR C, QI J, et al. Machine (deep) learning methods for image processing and radiomics [J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 2019, 3 (2): 104–108.
- [16] LU W, CHUNG F L, JIANG W, et al. A deep bayesian tensor-based system for video recommendation [J]. ACM Transactions on Information Systems, 2018, 37: 1–7.
- [17] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C]//Proceedings of the 26th international conference on world wide web. Cavtat, Croatia: IEEE, 2017: 173–182.
- [18] LIU W, ZHANG C, YU B, et al. A general multi-source data fusion framework [C]//Proceedings of the 2019 11th international conference on machine learning and computing. Zhuhai: ACM, 2019: 285–289.
- [19] CHAPLOT D S, SALAKHUTDINOV R. Knowledge-based word sense disambiguation using topic models [C]//Thirty-second AAAI conference on artificial intelligence. Louisiana: AAAI, 2018.
- [20] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993–1022.
- [21] WANG X, HE X, NIE L, et al. Item silk road: recommending items from information domains to social users [C]//Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval. Tokyo: ACM, 2017: 185–194.
- [22] BENNING M, BURGER M. Modern regularization methods for inverse problems [J]. Acta Numerica, 2018, 27: 1–111.