

# 基于因子分解机和隐马尔可夫的推荐算法

王晓耘<sup>1</sup>, 李 贤<sup>2</sup>, 袁 媛<sup>1</sup>

(1. 杭州电子科技大学 管理学院, 浙江 杭州 310018;

2. 杭州电子科技大学 计算机学院, 浙江 杭州 310018)

**摘 要:**因子分解机是一种基于矩阵分解的机器学习方法,通过在矩阵中引入上下文信息,构建含有上下文信息的矩阵,能够很好地进行用户评分预测。隐马尔可夫模型是一种含有隐含未知参数的统计模型,使用隐藏状态能够更好地符合实际情况。传统的推荐算法在进行推荐时通常并没有引入上下文信息,这通常会影响推荐算法的效果。鉴于上下文感知推荐算法通常能有效提高推荐精度,文中通过对推荐系统引入上下文信息并为用户添加用户隐藏兴趣状态,能够更精确地对用户进行推荐。为此,提出了一种结合因子分解机和隐马尔可夫模型的方法。在公开数据集上的验证结果表明,该方法相较于一些传统的推荐算法能够有效地提升推荐精度,并且在数据量增加的情况下也有较高的推荐精度。

**关键词:**上下文感知;因子分解机;隐马尔可夫模型;隐藏状态

中图分类号:TP391.3

文献标识码:A

文章编号:1673-629X(2019)06-0085-05

doi:10.3969/j.issn.1673-629X.2019.06.018

## A Recommendation Algorithm Based on Factorization Machines and Hidden Markov

WANG Xiao-yun<sup>1</sup>, LI Xian<sup>2</sup>, YUAN Yuan<sup>1</sup>

(1. School of Management, Hangzhou Dianzi University, Hangzhou 310018, China;

2. School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China)

**Abstract:** Factorization machine is a machine learning method based on matrix decomposition. By introducing context information into the matrix and constructing the matrix containing context information, it can predict the user rating well. Hidden Markov model is a statistical model with hidden parameters implied, using a hidden state can better meet the actual situation. Traditional recommendation algorithms usually do not introduce context information when making recommendations, which usually affects the effect of recommendation. In view of the fact that the context-aware recommender algorithm can effectively improve the precision, the users can be recommended more accurately by introducing context information into the recommendation system and adding users' hidden interest state. Therefore, we present a method of binding factorization machines and hidden Markov models. The verification results on the public data set show that the proposed method can effectively improve the recommendation accuracy compared with some traditional recommendation algorithms, and it also has higher recommendation accuracy when the amount of data increases.

**Key words:** context-aware; factorization machines; hidden Markov model; hidden state

## 0 引言

近年来,随着互联网技术的飞速发展,数据呈现出了爆炸式的增长。如何快速又有效地从海量数据中寻找用户所需要的信息成为了一个迫切的问题。为了应对海量的数据信息,搜索系统和推荐系统应运而生。对于那些有着明确目的的用户,信息搜索系统是一种非常有效的方法,但通常用户并不是非常清楚自己需要什么,在这种情况下使用推荐系统便是一个较好

的应对策略,现如今推荐系统的应用已经遍布网络的各个领域。协同过滤是一种常见且有效的推荐方法,协同过滤是通过寻找与当前用户最相似的用户,根据相似用户的历史购买记录来进行推荐的。此外还有基于内容的推荐方法和其他一些传统的推荐方法,这类传统的推荐方法起初往往在推荐过程中并不会引入上下文信息,这导致在推荐精度存在一些不足。在这种情况下,研究者们开始在推荐系统中引入上下文信息,

收稿日期:2018-07-02

修回日期:2018-11-08

网络出版时间:2019-03-06

基金项目:浙江省高校人文社科研究项目(GK140203204004/02)

作者简介:王晓耘(1956-),男,教授,研究方向为数据挖掘与商务智能;李 贤(1994-),男,硕士研究生,研究方向为数据挖掘与机器学习。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190306.0938.042.html>

通过结合上下文信息来对用户进行个性化推荐。这便出现了将传统的“用户-产品”二维的评分模型  $User \times Item \rightarrow Rating$  扩充为含有上下文信息的模型  $User \times Context \times Item \rightarrow Rating$ 。

文中主要基于上下文信息来对用户进行个性化推荐,提出了一种结合因子分解机<sup>[1]</sup>和隐马尔可夫模型的方法来进行个性化推荐,通过因子分解机进行初始预测,之后利用隐马尔可夫模型对用户的兴趣爱好进行预测,从而为用户进行个性化推荐。通过引入上下文信息来构建包含用户和产品的上下文信息矩阵,但该矩阵通常是稀疏矩阵,而因子分解机能够较好地解决稀疏矩阵进行学习,能够较好地解决实际情况中一些数据稀疏的问题,而拥有隐藏状态的隐马尔可夫模型则能够模拟以用户兴趣为隐藏状态的用户评分模型。在 MovieLens 数据集上的实验结果表明,该方法能够有效提高个性化推荐的准确度。

## 1 相关工作

### 1.1 推荐算法

目前推荐系统<sup>[2-4]</sup>可以分为两类:上下文感知推荐系统和非上下文感知推荐系统。它们之间的主要区别可以表述如下。对于一个用户集合  $U = \{u_1, u_2, \dots, u_m\}$  和一个产品集合  $I = \{i_1, i_2, \dots, i_n\}$ , 传统的推荐技术即非上下文感知推荐系统,是基于用户对产品的评分矩阵  $M \in R^{m \times n}$  来进行的,通过拟合矩阵特征模型来预测那些空白的数据,这类被称为非上下文感知的推荐。这种推荐模型能够适应很多的案例,但那些上下文信息,例如:时间、地点、用户情绪等<sup>[5]</sup>并未参与其中。这些信息能为推荐提供更多的维度。可以为上下文信息<sup>[6]</sup>定义一系列上下文的维度  $\{C_3, C_4, \dots, C_k\}$ , 每个上下文维度  $C_i$  是由一系列在某一情境下所捕获的属性集所组成的。事实上用户和产品也可以由上下文维度  $C_1, C_2$  来表示。

使用用户对产品的评分来进行个性化推荐,在部分情况下有着不错的效果,但在许多应用场景下并不能进行精确的推荐。上下文感知推荐系统是通过利用用户以及产品的额外信息来构建适应面更加广泛的推荐系统,以达到提高个性化推荐准确度的目的,具有重要的研究意义和应用价值。

通常用户的选择是受当前上下文信息影响的,包括时间、地点、人物、环境等等。Karatzoglou 等<sup>[7]</sup>提出了一种将用户、产品、上下文信息作为  $N$  维张量的多层推荐模型。对特征因子采用了高阶奇异值分解。然而,这种模型只能应用在一些特定的上下文信息之中,而且计算复杂度较高。Rendle 等<sup>[8]</sup>通过使用因子分解机(FM)来进行上下文感知推荐。因子分解机中可

以将大量的上下文信息转换到特征矩阵中。与多层推荐模型相比,因子分解机具有较短的计算时间,并且在预测结果和预测的质量上都要优于多层推荐模型。Hong 等<sup>[9]</sup>提出了一种 co-FM 方法来对用户的兴趣进行建模,并进行个性化推荐。Cheng 等<sup>[10]</sup>提出了一种具有损失函数的因子分解机模型,将一种特征选择算法和因子分解机进行合并。Wang 等<sup>[11]</sup>提出了一种随机分割的因子分解机模型来进行上下文感知推荐。在训练随机分割的因子分解机模型之前随机选择树需要手动进行,另外还具有非常高的计算复杂度。Yin<sup>[12-15]</sup>等采用主题模型进行兴趣点推荐,利用一个包含隐藏变量的中间层以及用户和产品空间的链接,但该模型面对大量信息时效果不佳。

### 1.2 隐马尔可夫模型

隐马尔可夫模型的主要特点是具有隐藏的状态序列,以及在观测序列的一般随机过程。目前,基于隐马尔可夫模型的改进包括对隐马尔可夫模型的改进和将隐马尔可夫模型与其他模型的结合。文献[16-17]将一阶隐马尔可夫模型改为二阶隐马尔可夫模型,这在实际应用中能更加符合对象特征,但算法的计算复杂度将会是一个重要的问题。Krishnalal<sup>[18]</sup>提出一种将隐马尔可夫模型和支持向量机相结合的方法,先使用隐马尔可夫模型进行预分类,再使用支持向量机进行二元分类。为解决小波系数之间的依赖问题,Crouse 等<sup>[19]</sup>提出一种小波域马尔可夫模型,该模型在分类、预测等方面也有不错的表现。

## 2 方法

### 2.1 因子分解机

因子分解机是一种基于矩阵分解的机器学习方法,对于稀疏数据有较好的学习效果。因子分解机<sup>[20]</sup>的 2 维模型可表示为:

$$y(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \quad (1)$$

在该模型中,模型参数  $\Theta = \{w_0, w_1, \dots, w_n, v_{1,1}, \dots, v_{n,k}\}$  为  $w_0 \in R, W \in R^n, V \in R^{n \times k}$ 。并且

$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (2)$$

$V$  中的行向量  $v_i$  表示第  $i$  个变量,并且行向量是  $k$  维的,  $k \in N_0^+$  是一个超参数,定义了因式分解的维数。而该模型需要在训练集中学习到参数  $w_0, W$  和  $V$ 。对于这些参数的学习和其他监督学习方法相似,可以通过在数据集  $S$  中最小化损失函数的和来进行。

$$OPT(S) := \arg \min_{\Theta} \sum_{(x,y) \in S} l(y(x | \Theta), y) \quad (3)$$

对于不同的目的,损失函数的选择也可以有所不同。

同,通常采用的是:

$$l^{LS}(y_1, y_2) := (y_1 - y_2)^2 \quad (4)$$

或者对于二分类问题 ( $y \in \{-1, 1\}$ ), 有:

$$l^C(y_1, y_2) := -\ln \sigma(y_1, y_2) \quad (5)$$

$$\text{其中, } \sigma(x) = \frac{1}{1 + e^{-x}}.$$

由于 FM 模型通常含有大量的模型参数  $\Theta$ , 因此容易出现过拟合。可以通过引入正则化进行处理。

$$\text{OPTREG}(S, \lambda) := \arg \min_{\Theta} \left( \sum_{(x, y) \in S} l(y(x | \Theta), y) + \sum_{\theta \in \Theta} \lambda_{\theta} \theta^2 \right) \quad (6)$$

其中,  $\lambda_{\theta} \in R^+$  是模型参数  $\theta$  的正则化系数。

对于上述优化问题可以采用 SGD 算法来求解, SGD 算法具有较低的计算复杂度和存储复杂度, 以及易实现等特点, 常用来优化因子分解机模型。参数的更新如下:

$$\mathbf{w}_0 \leftarrow \mathbf{w}_0 - \eta \left( \frac{\partial}{\partial \mathbf{w}_0} l(y(x | \Theta), y) + 2 \lambda^0 \mathbf{w}_0 \right) \quad (7)$$

$$\mathbf{w}_i \leftarrow \mathbf{w}_i - \eta \left( \frac{\partial}{\partial \mathbf{w}_i} l(y(x | \Theta), y) + 2 \lambda_{\pi(i)}^w \mathbf{w}_i \right) \quad (8)$$

$$\mathbf{v}_{i,f} \leftarrow \mathbf{v}_{i,f} - \eta \left( \frac{\partial}{\partial \mathbf{v}_{i,f}} l(y(x | \Theta), y) + 2 \lambda_{f, \pi(i)}^v \mathbf{v}_{i,f} \right) \quad (9)$$

$$\theta \leftarrow \theta - \eta \left( \frac{\partial}{\partial \theta} l(y(x), y) + 2 \lambda_{\theta} \theta \right) \quad (10)$$

其中,  $\eta \in R^+$  是学习速率, 也可理解为梯度下降的下降速度。

## 2.2 隐马尔可夫

一个离散的马尔可夫链是一系列具有马尔可夫性质的随机变量  $X_1, X_2, X_3 \dots$ 。也就是说下一状态的可能取值只依赖于当前状态, 而与历史状态无关。

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n)$$

一个马尔可夫模型可以表示为  $(S, \mathbf{P}, Q)$ , 其中  $S$  为  $x_i$  的所有可能取值所构成的可数集, 称为该马尔可夫链的状态空间;  $\mathbf{P} = [P_{ij}]_{n \times n}$  是状态转移矩阵,  $P_{ij}$  则表示在某一时刻处于状态  $i$  而下一时刻处于状态  $j$  的概率, 其中  $n$  表示所有的状态个数;  $Q = \{q_1, q_2, \dots, q_n\}$  是初始的各个状态的概率分布。

隐马尔可夫模型也是一种马尔可夫链, 不同的是它的状态转移过程是不可观察的, 定义如下:  $X = \{s_1, s_2, \dots, s_n\}$  表示一组状态集合, 其中  $n$  为状态数;  $O = \{v_1, v_2, \dots, v_m\}$  为一组可观察符号集合,  $m$  为观察值的数目; 状态转移概率为  $A = \{a_{ij}\}$ ,  $B = \{b_i(k)\}$  为状态  $i$  的观察概率分布, 表示状态  $i$  为对应观察值的概率;  $\pi = \{\pi_i\}$  表示初始化状态分布。至此, 一个隐马尔可

夫模型可以定义为:

$$\lambda = (X, O, \pi, A, B)$$

## 2.3 推荐方法

在推荐系统中, 数据稀疏是一个常见的问题, 由于因子分解机能够在稀疏数据中有较好的学习效果。因此, 可以使用因子分解机来进行处理。首先构建含有上下文信息的矩阵, 如图 1 所示。

	$U$				$I$				$L$				Rate
$x_1$	1	0	0	0	1	0	0	0	1	0	0	3	
$x_2$	1	0	0	0	0	1	0	0	0	1	0	5	
$x_3$	0	1	0	0	1	0	0	0	1	0	0	4	
$x_4$	0	0	1	0	0	0	1	0	0	0	1	5	
$x_5$	0	0	0	1	0	0	0	1	0	1	0	3	

图 1 含有上下文信息的矩阵

图 1 表示用户  $U = \{u_1, u_2, \dots, u_m\}$  对于产品  $I = \{i_1, i_2, \dots, i_n\}$  在条件  $L = \{l_1, l_2, \dots, l_k\}$  的情况下进行的评分, 其中  $L$  便是由上下文信息所组成的, 在  $L$  中可以添加想要加入的上下文信息, 当与该信息所描述的情况符合时由 1 表示, 否则为 0。例如, 第一行的  $x_1$  向量表示用户  $u_1$  在  $l_1$  的条件下对于产品  $i_1$  的评分为 3。在构建完含有上下文信息的矩阵后便可使用因子分解机来进行学习。

用户的兴趣是对用户进行个性化推荐的一个关键点, 如果根据用户的兴趣来进行推荐便可大大提高推荐的准确度。用户的兴趣通常来说是无法直接观察到的, 但是可以通过用户的历史记录来对用户的兴趣进行猜测。

通过引入隐马尔可夫模型来进行预测操作, 以用户的兴趣作为隐马尔可夫模型中的隐藏状态, 可以事先手动设置隐含状态集合  $X$ , 接着根据用户对产品的评分, 对每一个用户按照时间的顺序来构建一组可观察符号集合  $O$ , 之后便可以通过这组可观察的状态序列来对隐马尔可夫模型进行学习。具体的方法可以使用前向后向算法进行, 这里就不再进行介绍了。

文中采用的推荐方法的整体流程如图 2 所示。

首先对数据集进行随机划分, 按一定比例分为测试集和训练集。接着对训练数据集构建上下文信息矩阵, 再对上下文信息矩阵引入因子分解机来进行求解, 便可求得用户评分的初始预测值, 之后将初始预测值加入到上述可观察符号集  $O$  中, 结合用户的历史评分对隐马尔可夫模型进行学习, 最后再根据得到的隐马尔可夫模型求解用户对产品的最终预测值。得到最终预测值后便可根据最终预测评分的最高  $n$  位来进行推荐。

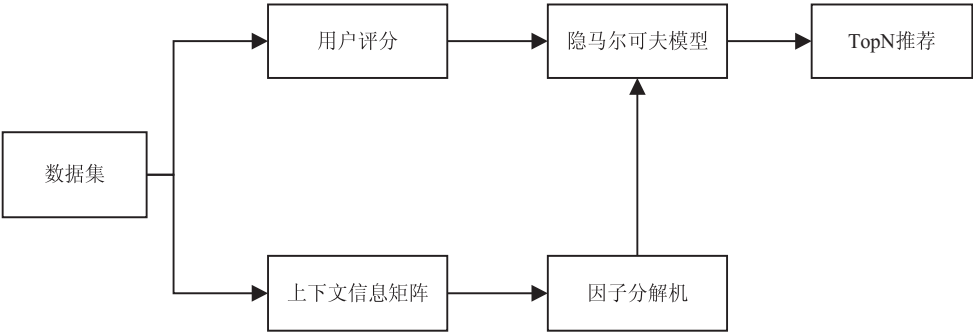


图 2 算法流程

3 实验

3.1 数据集

为验证文中方法的有效性,使用 MovieLens-1M 和 MovieLens-10M 数据集,其中 MovieLens-1M 数据集包含由 6 040 名用户对于 3 900 部电影进行的约一百万的评分,而 MovieLens-10M 数据集包含 71 567 名用户对 10 681 部电影进行的约一千万的评分。其中评分为 1 到 5 星,对于用户记录了用户性别、年龄、职业和邮编,其中对于年龄的划分分为:小于 18,18-24,25-34,34-44,45-49,50-55 以及大于 55。在职业上则进行了 20 种划分,包括作家、律师、无业游民、大学生等。而对于电影分成 18 种类型。

用户对电影的评分构成可观察的集合,再根据电影的类型作为用户兴趣集合。而对于上下文矩阵对用户考虑性别、职业和年龄属性。而对于电影则考虑电影类型。此外对用户的评分按照评分时间的顺序构建每个用户的评分序列。在数据上按照 80% 和 20% 进行划分,其中 20% 的测试集合为随机选取,剩余的 80% 则是训练集。

3.2 评估指标

推荐方法的评价指标有很多,但目前使用较多也是最直接的一类指标是准确度,比较典型的有平均绝对误差(mean absolute error,MAE)。

$$MAE = \frac{1}{|E^p|} \sum_{(i,j) \in E^p} |r_{ij} - \hat{r}_{ij}| \tag{11}$$

另外,均方根误差(root mean squared error, RMSE)也是一个使用较为频繁的指标。

$$RMSE = \sqrt{\frac{1}{|E^p|} \sum_{(i,j) \in E^p} (r_{ij} - \hat{r}_{ij})^2} \tag{12}$$

其中, $r_{ij}$  为用户  $i$  对于商品  $j$  的真实评分; $\hat{r}_{ij}$  则是预测值; $E^p$  为数据集。

3.3 实验结果

在对提出的方法(FM-HMM)进行实验时,选择了 SVD++<sup>[20]</sup> 和因子分解机(FM)两种方法分别对 MovieLens-1M 和 MovieLens-10M 两个数据集进行

对比实验,并选择了平均绝对误差和均方根误差作为评估标准。另外 SVD++是一种典型的非上下文感知推荐算法,而 FM 和 FM-HMM 则是基于上下文信息来进行推荐的。在构建含有上下文信息的矩阵时,对于用户选取了性别、年龄和职业作为上下文信息,在年龄和职业的划分上,采用了数据集所使用的划分方法分别将年龄划分为 7 段,而对职业则分为 21 种。对于电影,使用了电影类型来作为上下文信息,将电影分为了 18 种类型。而在对用户建立按照时间顺序的用户评分序列时,对于使用因子分解机初步预测得到的数据集,选择将这些预测值作为最新的评分来进行排列。

对于两种数据集得到的实验结果如表 1 和表 2 所示。分别查看可以明显地发现,上下文感知推荐在误差上明显要低于非上下文感知推荐方法。再对比 FM 模型和 FM-HMM 模型可以发现 FM-HMM 模型表现要略优于 FM 模型。再对比表 1 和表 2 的结果可以发现,当数据量增加时,三种方法在误差上都有相应的减少并且 FM-HMM 方法减少的量要略微大于 FM 方法。以上结果表明,引入上下文信息以及隐藏信息来进行推荐是可行的。

表 1 MovieLens-1M 数据结果

方法	RMSE	MAE
SVD++	1.006	0.837
FM	0.896	0.713
FM-HMM	0.875	0.691

表 2 MovieLens-10M 数据结果

方法	RMSE	MAE
SVD++	0.952	0.742
FM	0.854	0.661
FM-HMM	0.820	0.622

4 结束语

文中提出了一种应对上下文感知的推荐算法,通过结合因子分解机和隐马尔可夫模型来实现个性化推荐。该算法使用因子分解机进行初始预测,再将用户



的兴趣偏好作为隐藏属性来构建隐马尔可夫模型进行预测。通过与其他机器学习方法的对比来验证该方法的有效性,并采用 MovieLens 数据集进行实验分析。实验结果表明,引入因子分解机和隐马尔可夫模型能够有效地提升推荐结果的准确性。但是文中研究也存在一些不足之处,在对用户的兴趣偏好方面比较依赖具体的应用领域,而且兴趣的选择上也比较单一,今后将会尝试在兴趣的选择上进行更多元化的实现。此外,也会思考如何在大数据平台下进行实现。

#### 参考文献:

- [1] 王子豪,徐桂琼. 基于高阶偏差的因子分解机推荐算法[J]. 计算机应用研究,2017,34(2):339-342.
- [2] WANG Shaoqing, LI Cuiping, ZHAO Kankan, et al. Learning to context-aware recommend with hierarchical factorization machines[J]. Information Sciences, 2017, 409-410: 121-138.
- [3] 杨 博,赵鹏飞. 推荐算法综述[J]. 山西大学学报:自然科学版,2011,34(3):337-350.
- [4] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//ACM SIGKDD international conference on knowledge discovery and data mining. [s. l.]: ACM, 2008: 426-434.
- [5] 窦羚源,王新华,孙 克. 融合标签特征和时间上下文的协同过滤推荐算法[J]. 小型微型计算机系统,2016,37(1): 48-52.
- [6] 徐风琴,孟祥武,王立才. 基于移动用户上下文相似度的协同过滤推荐算法[J]. 电子与信息学报,2011,33(11): 2785-2789.
- [7] KARATZOGLOU A, AMATRIAIN X, BALTRUNAS L, et al. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering[C]//ACM conference on recommender systems. Barcelona, Spain: ACM, 2010: 79-86.
- [8] RENDLE S, GANTNER Z, FREUDENTHALER C, et al. Fast context-aware recommendations with factorization machines[C]//International ACM SIGIR conference on research and development in information retrieval. Beijing, China: ACM, 2011: 635-644.
- [9] HONG Liangjie, DOUMITH A S, DAVISON B D. Co-factorization machines: modeling user interests and predicting individual decisions in Twitter[C]//Proceedings of the sixth ACM international conference on web search and data mining. Rome, Italy: ACM, 2013: 557-566.
- [10] ZHANG T, ZHANG T, ZHANG T, et al. Gradient boosting factorization machines[C]//ACM conference on recommender systems. [s. l.]: ACM, 2014: 265-272.
- [11] WANG Shaoqing, DU Cuilan, ZHAO Kankan, et al. Random partition factorization machines for context-aware recommendations[M]//Web-Age information management. [s. l.]: Springer International Publishing, 2016: 219-230.
- [12] YIN Hongzhi, CUI Bin, HUANG Zi, et al. Joint modeling of users' interests and mobility patterns for point-of-interest recommendation[C]//Proceedings of the 23rd ACM international conference on multimedia. Brisbane, Australia: ACM, 2015: 819-822.
- [13] YIN Hongzhi, CUI Bin, ZHOU Xiaofang, et al. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation[J]. ACM Transactions on Information Systems, 2015, 35(2): 1631-1640.
- [14] YIN Hongzhi, ZHOU Xiaofang, CUI Bin, et al. Adapting to user interest drift for POI recommendation[J]. IEEE Transactions on Knowledge & Data Engineering, 2016, 28(10): 2566-2581.
- [15] YIN Hongzhi, ZHOU Xiaofang, SHAO Yingxia, et al. Joint modeling of user check-in behaviors for point-of-interest recommendation[C]//ACM international on conference on information and knowledge management. Melbourne, Australia: ACM, 2015: 1631-1640.
- [16] 韦 炜,全渝娟,卓奕涛,等. 基于多阶马尔可夫预测的个性化推荐算法[J]. 计算机工程,2015,41(11): 59-66.
- [17] KUNDU A, HE Y, BAHL P. Recognition of handwritten word: first and second order hidden Markov model based approach[J]. Pattern Recognition, 1989, 22(3): 283-297.
- [18] KRISHNALAL G, RENGARAJAN S B, SRINIVASAGAN K G. A new text mining approach based on HMM-SVM for web news classification[J]. International Journal of Computer Applications, 2010, 1(19): 98-104.
- [19] CROUSE M S, NOWAK R D, BARANIUK R G. Wavelet-based statistical signal processing using hidden Markov models[J]. IEEE Transactions on Signal Processing, 1998, 46(4): 886-902.
- [20] RENDLE S. Factorization machines with libFM[J]. ACM Transactions on Intelligent Systems & Technology, 2012, 3(3): 1-22.