深 圳 大 学

本 科 毕 业 论 文 (设计)

题目： 基于密度的空间文本聚类

姓名： 彭婷婷

专业： 计算机科学与技术

学院： 计算机与软件学院

学号： 2015300091

指导教师： 吴定明

职称： 讲师

2019 年　月　日

深圳大学本科毕业论文（设计）诚信声明

本人郑重声明：所呈交的毕业论文（设计），题目《个人搜索引擎的实现》是本人在指导教师的指导下，独立进行研究工作所取得的成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式注明。除此之外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。本人完全意识到本声明的法律结果。

毕业论文（设计）作者签名：

日期：　　年　月　日

目 录

融合内容信息的单类协同过滤推荐算法研究

计算机与软件学院计算机科学与技术专业　徐留成

学号：2012080173

【摘要】对于基于隐式反馈的个性化推荐算法而言，pairwise learning 是一个非常重要的技术手段。pairwise learning 算法通常基于这样一个假设：对一个用户而言，相比于未选择过的物品往往会更感兴趣于已选择过的物品。这种假设在推荐算法的学习过程中会衍生出大量的 training pairs。而为了应对大规模的数据集，我们所研究的推荐算法往往都是基于均匀采样的随机梯度下降方法进行求解。不过，这种采取均匀采样的策略经常会导致算法收敛非常缓慢。在本文中首先讨论了均匀采样策略导致收敛缓慢的原因，并研究了通过在已有的 BPR 推荐框架中融合内容信息改进采样策略并最终提高推荐效果的方法。实验证明，相比于均匀采样策略，通过融合内容信息的适应性采样策略的确能够有助于提高推荐效果。

【关键词】推荐系统; 协同过滤; 适应性采样

1　推荐系统概述

1.1　主要符号表

表2中列举了大部分在本文中使用的符号及其意义。

1.2　推荐系统纵览

自从 20 世纪 90 年代中期第一篇关于协同过滤 (Collaborative Filtering) 的研究文章[?] 出现以后，推荐系统就开始成为了一个重要且有趣的研究主题。协同过滤通过收集推荐系统中相似用户的偏好进行推荐，而生成近邻用户 (neighbourhood formation) 是协同过滤中非常重要的一个方面[??]. 近邻用户生成的目的是为每个用户找到一些相似的用户群或其最近邻，然后基于有着相似偏好的近邻用户推荐产品或服务[??]。这里的近邻 (neighbourhood) 是指那些对于我们将要为之提供推荐建议的用户所感兴趣的物品有过相似交互行为的其他用户。在这里，我们把需要为之提供推荐的用户成为目标用户，那么通过比较目标用户与其近邻评分，就可以做出最终的推荐[??]。当缺乏用户评分数据的时候，协同过滤就会遇到所谓的稀疏性问题，这将导致推荐效果变得很差。因此，在推荐系统中预防稀疏性问题非常重要。为此一个很重要的途径便是从隐式反馈 (比如用户的购买行为，上线时间，历史浏览记录) 数据中提取用户的偏好信息来降低协同过滤对于用户评分数据的依赖，同时提高推荐效果[??]。隐式反馈数据能够通过对于用户行为的观测提供更多的信息来降低评分数据不充分的影响[??]。另一方面，协同过滤推荐技术的用户画像 (user profile) 通过用户对于物品的评分得以构建。为了降低协同过滤对于评分数据的依赖，用户行为 (user activity) 也已经成为研究调查的一个重要关注点，也就是说通过挖掘用户偏好的经验性知识来构建更加精确的用户画像 (user profile)[????]。

1.3　典型推荐算法概述

推荐系统通过识别用户的需求与偏好为其推荐合适的产品或服务。目前国内外关于推荐系统的研究下已衍生了很多推荐算法，这些推荐算法通常可以分为三类：基于内容的推荐 (Content-based recommendations), 协同过滤 (Collaboratibe Filtering) 和混合型 (Hybrid approches) 推荐.

1.3.1　基于内容的推荐系统

基于内容信息的方法[???] 来学习个体的隐式表达 (latent representation) 并缓解冷启动 (cold start) 问题。比如，在 FM[?] 中各种属性信息被放到特征矩阵中，然后通过对于评分数据回归分析相关属性。

基于内容的推荐系统从用户与物品的 content profile 之间的相似度出发进行推荐。他们从研究推荐系统中个体的内容信息角度进行分析。通常这类方法利用个体的内容信息，比如物品属性，用户文本，或照片的像素点，主要利用探索启发式 (heuristics) 的方法。在[???] 中, 他们使用诸如 cosine similarity 的方式来衡量相似度，然后推荐在内容上与用户过去所喜欢的相类似的物品。在[?] 中，基于物品内容信息并由用户标注的标签: “相关 (relevant)” 或者是 “不相关 (irrelevant)”，作者学习了一个贝叶斯分类器来对没有标注的物品进行分类。近来，也有很多社交媒体 (social media) 相关的

表 1: 主要符号表

常用符号	意义
s	user number
t	item number
u	user
v	item
u_m	the specified user m
v_i	the specified item i
v_j	the specified item j
b_i	item bias
r_{ui}	real rating of user u on item i
$\hat{r}_{ui}$	predicted rating of user u on item i
$\hat{r}_{uj}$	predicted rating of user u on item j
e_i	entity, e.g., user u or item v
T	iteration number in the algorithm
$k \in \mathbb{R}$	number of lentent dimensions
$r(j)$	the ranking place of the item v_j
$\mathcal{P}$	(user, item) pairs in training data
$\mathcal{P}^{te}$	(user item) pairs in test data
$\mathcal{U}$	the whole user set
$\mathcal{I}$	the whole item set
$\mathcal{I}_u^{re}$	recommended items for user u
$\mathcal{I}_u^{te}$	selected items by user u in test data
$\mathcal{I}_u^{tr}$	selected items by user u in training data
$\mathcal{I}_{u_m}^{+}$	the set of items selected by the user u_m
$U \in \mathbb{R}^{s \times k}$	user-specific lantent matrix
$V \in \mathbb{R}^{t \times k}$	item-specific lantent matrix
$U_{u\cdot} \in \mathbb{R}^{1 \times k}$	user-specific latent feature vector
$V_{v\cdot} \in \mathbb{R}^{1 \times k}$	item-specific lantent feature vector
$Y^e = [y_1^e, y_2^e, y_3^e, \cdots]$	latent representation of entities
$y_i^e \in \mathbb{R}^{1 \times k}$	the latent vector of entity e_i
$\mathcal{C} = \{c_1, c_2, \cdots, c_k\}$	categories
$D_S := \{(m.i, j) \mid v_i \in \mathcal{I}_{u_m}^{+} \wedge v_j \in \mathcal{I} \setminus \mathcal{I}_{u_m}^{+}\}$	the set of all pairwise preference

推荐系统关注 content-based 推荐方法并对其进行了很多研究。比如，在[? ?] 中通过基于可视性的内容相似度考虑它的最近邻标签, 然后来为目标图像推荐标签。[?] 提出了一个在线视频的推荐系统，而该系统则利用了在用户与视频间点击数据的多模态的内容关联度。

但是，这些基于内容的推荐方法大都具有以下局限性: 第一, 它们必须有足够的信息构建一个分类器, 并且显然会被推荐物品的特征所局限; 第二, 它们推荐的物品, 在内容上往往与用户已经有过评分行为的物品很相似, 显然这就会导致了较低的推荐多样性。

1.3.2 基于协同过滤的推荐系统

协同过滤 (Collaborative Filtering) 方法通过挖掘用户的评分历史来预测用户的偏好。它们并不需要内容信息 (content information)，并且能够发现一些基于内容的推荐方法所不能发现的一些有趣的联系。通常来说，协同过滤基于这样一个基本的设想：相似的用户对于相似的物品有着相似的行为[? ?]。这里的“相似”并不同于 content-based 方法中的内容相似度 (content similarity), 它指的是相似的评分偏好 (similar rating preference)。

协同过滤方法可大致分为两类：memory-based methods, model-based methods。memory-based 方法[? ? ? ?] 通常通过搜寻相似的用户或商品去进行推荐。而其相似度则是经由评分历史计算而得。memory-based 方法也可进一步的被分为 user-based 和 item-based 两类方法。通过与当前用户有着相似偏好的其他用户进行推荐即为 user-based, 通过推荐与当前用户喜欢过的物品所相似的物品即为 item-based。不过，当缺乏用户评分数据的时候，协同过滤就会遇到叫做稀疏性的一个问题，这将很容易导致推荐效果变得很差。因此，在推荐系统常常需要应对稀疏性这一大难题。应对稀疏性问题一个重要的途径便是从隐式反馈 (implicit feedback)(比如用户的购买行为，上线时间，历史浏览记录) 数据中提取用户的偏好信息来降低协同过滤对于用户评分数据的依赖，当然这往往同时也能够提高推荐效果[? ?]。另外, 相对于显式反馈，隐式反馈的数据更易采得也更丰富。隐式反馈能够通过对于用户行为的观测提供更多的信息来降低评分数据不充分的影响[? ?]。这时其实也就是变成我们所谓的单类协同过滤 (One-class Collaborative Filtering) 问题。

OCCF 问题的最典型特征是仅能够观测到正向采样 (positive examples), 比如用户的点击行为, 浏览行为，同时数据分类往往非常不均衡, 比如用户点击过物品可能只是占到整个物品集合的很小一部分。我们把用户未有过交互行为的物品，比如未点击过的物品，叫做 negative examples. 那么如何从大量未有过交互行为的物品集合中针对 negative examples 进行采样与建模是很多问题的关键所在。在前人的一些工作中，有几种直观的策略来处理这个问题。其实一个最常见的做法是将所有缺失的数据视作 negative examples, 显然这将导致推荐结果具有偏差，因为很多缺失数据很多可能是 positive examples。另一种做法是所缺失的数据是做未知的，这将导致协同过滤模型仅利用了 positive examples。近来的一些研究中，一些关于 OCCF 的研究人员将重点放到了对于 negative examples 的建模上 [19,34,35,44]。他们的一个基本的想法是将所缺失的数据视作是 negative, 但是给出了将其视作 negative 的一个概率权重。不过，他们当中的部分做法仅仅是通过简单地观测历史反馈的概率属性来区分 negative examples。比如，[19,34], 他们计算了每个用户给多少物品评过分，每个物品被多少用户评过分，由此来计算一个权重。进一步的说，他们认为如果一个用户浏览过的物品越多，那么他没有浏览过的物品便更大可能是 negative 类型；如果一个物品被越少的用户浏览过，那么这个物品相关缺失数据便更小可能是 negative, 这种做法仍然是略显粗糙。

协作型方法[? ? ?] 通过处理大量的用户与物品间的交互信息，比如隐式反馈和显式的评分 (也

叫作协同信息)。这些方法不同于 memory-based 方法，model-based 方法采用机器学习与概率统计的技术从已有的用户评分去学习一个模型，再将模型应用到推荐中。其中包括有隐语义模型 (latent semantic models), 图模型 (graphical models), 贝叶斯模型 (Bayesian models), 聚类模型 (clustering models). 在众多的 model-based 方法中，低秩矩阵分解 (low-rank Matrix Factorization) 由于在可扩展性与精确度方面的优势已经获得了许多研究者的关注。其实分解的方法在个性化的推荐系统中很常见。他们可以被用来处理推荐系统中收集的各种信息，比如隐式反馈[? ?], 物品属性[? ?], 用户画像[?] 和社交信息[?]。其中矩阵分解基于用户的偏好可以被一小部分因子表示，通过从 user-item rating matrix 来学习 user 与 item 一个低秩隐含因子，然后利用它们去预测未被观测到的 ratings。

矩阵分解[?] 及其一些扩展方法[? ? ?] 是用来处理协同信息的非常典型的分解方法, 它通过分解协同信息并试图在一个共享的隐式空间学习用户与物品的隐式表达。比如，隐式矩阵分解[?] 通过为每个 user-item pair 计算一个适应性的信任权重来扩展基础的 BPR 处理隐式反馈。尽管通过扩展 BPR 能够应对隐式反馈问题，但是由于在隐式反馈数据集中普遍存在的数据倾斜 (data skew) 问题 (正反馈数量常常不到总数的 1%), 他们很容易陷入过拟合问题。为了缓解数据倾斜与推荐系统的隐式反馈学习，Bayesian Personalized Ranking (BPR)[?] 和它的一些扩展方法[? ? ?] 被提出，其所基于的假设为：相比于未选择的物品用户更感兴趣已经选择的物品。这样假设会产生大量的训练数据，因此对应的学习算法通常基于均匀采样用户物品对的随机梯度下降。但是不同的训练采样可能会对参数学习产生不同的影响，均匀采样策略往往会产生大量低效的训练采样并导致收敛变得缓慢。尤其是当物品数量很大和物品的流行度有着长尾分布 (long distribution)[?] 的时候，均匀采样策略将会导致极其缓慢的收敛。因此，BPR 的作者 Rendle 进一步研究了长尾效应并利用它提出了非均匀的物品采样器[?]。对于给定的一个用户，他们计划挑选出那些在某一领域很流行并且尚未被该用户选择过的物品来构成训练对。理论上，这种采样方式很耗时，因为它将物品的隐式因子当做物品流行度的指示器并且需要在每轮迭代的每个区域对物品进行重新排序。为了考虑运行效率，Rendle 不得不减少重新排序的时间来妥协推荐性能。另一方面，为了获得一个通用的加速 BPR 学习的方案,[?] 尝试根据一个在两个不同未选择过的物品上的偏好差别来选取那些富含信息的训练对。但是，由于真实世界的数据集里物品数往往极其庞大, 这种策略不得不在计算偏好差别上花费大量的时间。因此,[? ?] 都陷入了平衡算法效率与性能表现的两难境地。在本课题中所研究的采样策略在效率与性能两方面都表现了很好的效果，并且有潜力加速 BPR 的学习。

传统的协同过滤对于评分预测问题往往能够取得很好的效果，比如 Netflix 的电影推荐。但是，它受制于一个众所周知的问题: 冷启动，当一个新的物品或用户进入系统时由于几乎无法获得任何评分记录, 在此种情况下推荐效果往往很不理想。为了缓解推荐系统中的冷启动问题，Map-BPR[?] 扩展了 BPR 框架，他们学习了一个将内容信息空间映射到隐式空间的一个映射关系。然后，Map——BPR 利用学习到了这个映射学习那些缺乏协同信息的新个体的隐式因子。不过，Map-BPR 将隐式因子的学习分割为两个不相关的部分。这会导致在隐式反馈数据集中的个体的隐式因子仅仅指示协同属性而不会显示内容属性。为了获得更可信的隐式因子，在本课题的研究方法在同一个学习过程中研究了通过协同信息与内容信息学习个体的隐式因子。

1.3.3 混合式推荐系统

混合方法尝试将基于内容与协同过滤的推荐方法结合起来应对它们的局限性。[?] 通过将基于内容与协同过滤的预测结果进行线性组合设计了一个混合推荐模型。[?] 提出从概率混合的角度将协同过滤与基于内容的推荐方法进行统一。近来也有很多工作都重点关注了社交媒体推荐 (social media

recommendation)，而他们中的大部分都采用了混合方法，在挖掘社交媒体内容的同时考虑了用户的历史行为来获得更高的推荐准确度。[?] 为在线社交网络中的视频推荐 (video recommendation) 设计了一个组合式的社交内容推荐框架. 他们的方法通过利用社交网络信息 (social network information) 与内容信息 (content information), 提出一个 user-conetnt matrix 填充冷启动中的 user-video 条目。[?] 研究利用了集成学习 (ensemble learning) 方法，在音乐推荐中将基于物品协同过滤结果与基于内容方法的结果进行融合。

1.4　推荐系统评价指标

所谓评价指标主要包括“技术评价指标”和“业务评价指标”。技术评价指标包括诸如 RMSE[1]、MAE[2]、NDCG[3]、MAP[4]、Recall、Precision 等，业务评价指标如成交转化率、用户点击率等。[?] 也介绍了推荐系统中的很多评测指标。这些评测指标可用于评价推荐系统各方面的性能。，它们包括用户满意度、预测准确度、覆盖率、多样性、实时性、健壮性等等。其中有些可以通过计算来定量衡量，有些则只能定性描述，有些可以通过离线实验计算，有些需要通过用户调查获得，还有些只能在线评测。这里主要介绍在技术评价指标中, 评分预测与 TopN 推荐的预测准确度定义。

1.4.1　评分预测

图 1: 用户评分

很多提供推荐服务的网站都有一个让用户给物品打分的功能。那么，如果知道了用户对物品的历史评分，就可以从中习得用户的兴趣模型，并预测该用户在将来看到一个他没有评过分的物品时，会给这个物品评多少分。预测用户对物品评分的行为称为评分预测。

评分预测的预测准确度一般通过 RMSE 和 MAE 计算。对于测试集中的一个用户 u 和物品 i，令 r_{ui} 是用户 u 对物品 i 的实际评分，而 $\hat{r}_{ui}$ 是推荐算法给出的预测评分，那么 RMSE 的定义为：

$$RMSE = \frac{\sqrt{\sum_{(u,i)\in\mathcal{P}^{te}} \left(r_{ui} - \hat{r}_{ui}\right)^2}}{|\mathcal{P}^{te}|}$$

[1]RMSE: **R**oot **M**ean **S**quared **E**rror, 均方根误差
[2]MAE: **M**ean **A**bsolute **E**rror, 平均绝对误差
[3]NDCG: **N**ormalized **D**iscounted **C**umulative **G**ain
[4]MAP: **M**ean **A**verage **P**recision, 平均准确率

MAE 采用绝对值计算预测误差，它的定义为：

$$MAE = \frac{\sum_{(u,i)\in\mathcal{P}^{te}} |r_{ui} - \hat{r}_{ui}|}{|\mathcal{P}^{te}|}$$

关于 RMSE 和 MAE 这两个指标的优缺点，Netflix 认为 RMSE 加大了对预测不准的用户物品评分的惩罚 (平方项的惩罚)，因而对系统的评测更加苛刻。研究表明，如果评分系统是基于整数建立的（即用户给的评分都是整数），那么对预测结果取整会降低 MAE 的误差。

1.4.2　TopN 推荐

图 2: TopN 推荐

网站在提供推荐服务时，一般是给用户一个个性化的推荐列表，例如购物网站上的热门推荐，这种推荐叫做 TopN 推荐。在现实场景下，TopN 推荐也是更常见的一种推荐形式。

TopN 推荐的预测准确率一般通过准确率 (precision)/召回率 (recall) 衡量。对于用户 u, 推荐列表 $\mathcal{I}_u^{re}$ 的准确率定义为：

$$Precision_u = \frac{|\mathcal{I}_u^{re} \cap \mathcal{I}_u^{te}|}{|\mathcal{I}_u^{re}|}$$

其召回率定义为:

$$Recall_u = \frac{|\mathcal{I}_u^{re} \cap \mathcal{I}_u^{te}|}{|\mathcal{I}_u^{te}|}$$

1.5　本章小结

本章首先对推荐系统进行了概括性的介绍，然后主要从典型推荐算法与推荐系统的评价指标两方面对推荐系统的整个框架形成了一个粗略的认识。

2　引言

2.1　研究背景及意义

互联网的出现和普及给用户带来了大量的信息, 满足了用户在信息时代对各种信息的需求, 但随着 Internet 的迅速发展而带来的网络上信息量的巨幅增长, 使得用户在面对大量信息时无法快速从中获得对自己真正有用的那部分信息。换言之, 在这种情况下人们对信息的使用效率反而降低了, 这就是所谓的信息过载 (information overload) 问题. 的确如此, 面对信息的汪洋大海, 人们往往感到无所适从, 信息过载已经成为一个不容忽视的问题.

目前, 应对信息过载的办法之一便是以搜索引擎为代表的信息检索系统, 比如国外的 Google[5]、国内的 Baidu[6]等, 它们在帮助用户从巨大的网络资源中获取信息方面发挥着极其重要的作用. 但对于使用搜索引擎的用户而言, 在使用同一个关键字搜索信息时, 在一段时间内所得到的结果都是相同的. 另一方面来看, 信息及其传播是多样化的, 而用户对信息的需求是多元化和个性化的, 那么通过以搜索引擎为代表的信息检索系统获得的结果显然不能满足用户的个性化需求, 它们仍然无法很好地解决信息过载问题.

面对信息过载, 另外一个非常有潜力的办法是个性化的推荐系统, 它是根据用户的信息需求、兴趣等, 将用户所感兴趣的信息、产品、服务等推荐给用户的个性化信息推荐系统. 和搜索引擎相比, 推荐系统通过研究用户的历史行为与兴趣偏好, 进行个性化考量, 由系统发现用户的兴趣点, 从而引导用户发现自己的信息需求. 一个优秀的推荐系统不仅能为用户提供个性化的服务, 还能和用户之间建立密切关系, 让用户对其推荐产生依赖. 个性化推荐系统现已广泛应用于很多领域, 其中最典型并具有良好的发展和应用前景的领域就是电子商务领域. 目前, 几乎所有大型的电子商务系统, 如 Amazon, eBay, 京东, 当当网上书店等, 都不同程度地使用了各种形式的推荐系统。同时学术界对推荐系统的研究热度一直很高, 逐步形成了一个独立的研究领域.

Internet 为人们提供了极其丰富的信息资源, 在这些海量、异构的 Web 信息资源中蕴含着具有巨大潜在价值的知识。根据用户访问的历史记录以及各种服务或商品之间的相关信息可以构建用户的兴趣模型, 从而凭借该用户的兴趣模型对繁杂的信息进行过滤, 然后向用户推荐其可能感兴趣的服务或商品。事实上, 推荐系统已经成为目前解决信息过载最有效的工具之一。

2.2　本文主要工作

本文从推荐系统的概述展开, 讨论了在推荐系统的学习算法中随机梯度下降方式中采用均匀采样策略而导致收敛缓慢的一些原因, 并通过融合内容信息改进了均匀采样策略–适应性采样策略, 然后将适应性采样策略放入已有的推荐算法框架中, 加快原有推荐算法的学习。

2.3　论文组织结构

本论文共分为七章，内容如下:

第一章为引言, 主要介绍了本论文的研究背景、意义, 主要工作及论文的组织结构.

[5] https://www.google.com/
[6] https://www.baidu.com/

第二章为推荐系统概览，并分类介绍了包括了基于内容、基于系统过滤与混合型推荐算法的一些典型的推荐学习算法。

第三章为预备工作，首先简要回顾了 Bayesian Personalized Ranking(BPR) 推荐算法，并对其局限性进行了一些探讨。

第四章为适应性采样策略，主要研究了通过融合内容信息提出了适应性采样策略改进已有的均匀采样策略。

第五章为整体的算法框架, 将适应性采样策略融入已有的 BPR 推荐模型。

第六章为实验论证，主要内容为在适应性采样策略下的推荐算法的实验表现。

第七章为结论与展望，首先简要总结了本文的一些工作，并对接下来进一步的研究工作做了展望。

致谢

首先衷心地感谢潘微科老师。在本科生涯最后的一年多里，不仅是现时的学业与学术，更是对于未来的发展给予了我很多指导与帮助。本次毕业设计，从选题到论文撰写，给予了我很多宝贵的意见。他渊博的学识、严谨的治学态度及认真负责的工作态度都使我受到鼓舞和熏陶。在此向潘微科老师表示崇高的敬意和衷心的感谢，他的言传身教将使我终生受益。

感谢 key 哥哥与在 453 认识的朋友们，与你们的交流大概就是我对计算机启蒙的开始. 如果不是有幸与你们相识，这一路走来必是要曲折地多.

感谢 Thuthesis 及其作者薛瑞尼. 最终虽未使用 Thuthesis 模板，但是此间对其研习所得对我顺利使用 LaTeX 完成论文撰写仍然起了很大作用.

感谢一直关心我的父母与兄长。远游在外，感谢还有你们牵挂.

感谢自己熬过了那段难捱的日子。从学习画画到广播电视再到计算机科学，在如今看来似曾是做了诸多无用功，不过幸而没有因为短时的平庸迷茫而消磨掉满心的戾气.

前路漫漫，不冀求大步流星，唯盼能步步坚实.

Research on Content-Aware Collaborative Filtering

【Abstract】 Pairwise learning algorithms are a vital technique for personalized ranking with implicit feedback. They usually assume that each user is more interested in items which have been selected by the user than remaining ones. This pairwise assumption usually derives massive training pairs. To deal with such large-scale training data, the learning algorithms are usually based on stochastic gradient descent with uniformly drawn pairs. However, the uniformly sampling strategy often results in slow convergence. In this paper, we first uncover the reasons of slow convergence. Then, we associate contents of entities with characteristics of dataset to develop an adaptive item sampler for drawing informative training data. In this end, to devise a robust personalized ranking method, we accordingly embed our sampler into Bayesian Personalized Ranking (BPR) framework, and further propose a Content-aware and Adaptive Bayesian Personalized Ranking (CA-BPR) method, which can model both contents and implicit feedbacks in a unified learning process. The experimental results show that our adaptive item sampler can indeed improve recommendation performance.

【Keywords】 Recommendation System; Collaborative Filtering; Adaptive Sampling

指导教师: 潘微科