图书馆读者需求与服务匹配模型研究

林淑贞

(广州图书馆,广东 广州 510623)

[摘 要] 为提高图书馆服务的精准度和适用度,解决相关应用普遍存在覆盖度低、精度欠佳等问题,提出了新型的图书馆读者需求与服务匹配模型,详述了该模型的组成结构、数据处理流程以及关键算法;该模型从读者与服务两方面采集信息,并通过融合匹配算法将个性化的服务推荐给读者。实验证明,该算法具有良好的读者覆盖度、需求匹配精确度以及性价比。

[关键词] 图书馆服务;匹配;读者;大数据;信息融合

[中图分类号] G250 [文献标识码] A [文章编号] 2095-5197(2018)06-0106-06

Research on Library Reader and Service Matching Model

LIN Shu-zhen

(Guangzhou Library, Guangzhou 510623, China)

Abstract: In order to improve the accuracy and applicability of library services, and to solve the problems of low coverage and poor precision in related applications, this paper proposes a new library reader demand and service matching model, details the composition, data processing flow and key algorithms of the model. The model collects information from both readers and services, and recommends personalized services to readers through fusion matching algorithm. The experiment proves that the algorithm has good reader coverage, demand matching accuracy and cost performance.

Keywords: library service; match; reader; big data; information fusion

CLC number: G250

1 引言

读者大数据与精准画像技术是当前图书馆学界研究的重点和热点。其中,读者大数据的汇聚、提炼与应用是构建精准读者画像和实现图书馆个性化服务的关键所在,也成为图书馆服务领域关注的焦点[1]。随着阅读路径分析、云计算、深度学习等技术在图书馆领域的应用,已有一批基于读者大数据技术的图书馆读者与服务匹配模型及算法问世,其中较具特色的成果有:Zne-Jung Lee[2]等研究人员设计与实现了一个基于读者大数据的图书馆推荐模型,该模型通过对读者信息进行持续跟踪与融合,刻画出读者的阅读习惯,提高了个性化服务的读者满意度;Daniel Mican[3]等研究人员设计了基于读者社会媒体大数据分析的推荐系统,该系统通过读者关系分析,对读者需求进行了深入发掘,提高了推荐的准确度;Dharna Patel[4]等研究人员将云计算运用于读者大数据挖掘工作中,并据此设计了一款图书推荐系统,取得了较高的读者需求匹配度;Aravind Sesagiri Raamkumar[5]等研究人员对读者大数据与海量论

文之间的需求与匹配关系进行了分析,并开发了对应的科学文献服务系统,具有较高的读者满意度;Yifan Hu[6]等研究人员基于大数据技术开发了读者协同过滤推荐系统,极大地提高了读者荐读服务的准确度。Julien Verplanken[7]等研究人员设计了基于大数据技术的读者动态画像模型,并以此为依据开发了精度较高的推荐系统;Raymond J.Mooney[8]等研究人员应用自学习技术,构建了读者阅读成长模型,并将其应用于读者服务系统中,取得了很高的用户满意度。尽管上述成果具备一定的理论价值与实践意义,但从实际应用效果来看,普遍还存在着服务推荐精度不稳定、覆盖度较低、系统资源开销较大等问题。针对这些问题,本研究基于读者大数据融合技术,深入和全面地描绘读者画像,构建了较为完善的图书馆读者需求与服务匹配模型 LRSM(Library Reader and Service Matching)。该模型的结构、处理流程以及关键算法如下文所述。

2　模型结构与应用流程

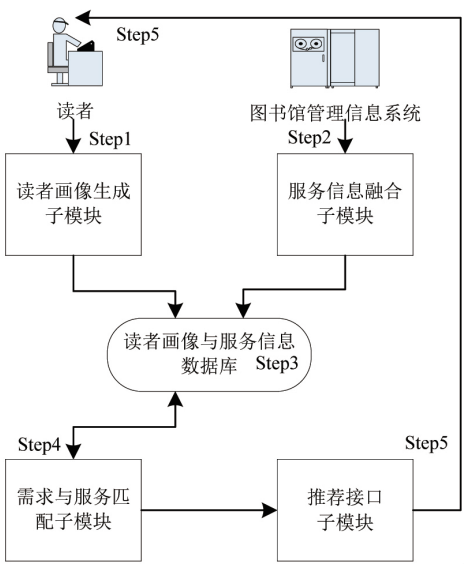

图1　模型结构与应用流程

2.1　模型结构

LRSM的总体特点是结构较为简单(全模型仅4个主要子模块),接口标准,可适用于各类图书馆的推荐服务运行;其结构特点是"紧内聚,松耦合",复杂的数据结构被封装在各个模块中,用户或第三方软件通过其接口就可以获取相关的推荐服务,使得服务的获取变得极为简单,而避免了过多繁复的配置和二次开发。如图1所示,LRSM模型主要嵌入在图书馆信息服务系统中,为读者提供高匹配度的个性化服务推荐。该模型主要有4个子模块。一是读者画像生成子模块。该模块主要从图书馆管理大数据系统中抽取读者相关信息,通过融合后,形成包含读者特征与需求信息的读者画像信息,并将这些信息存储在读者画像库中。LRSM模型通过该子模块,对读者的持续跟踪,不断丰富和细化这些画像信息,为需求与服务匹配工作提供读者的基础数据。二是服务信息融合子模块,该模块主要从图书馆各职能部门的服务器中获取服务与业务项目的特征信息,以及读者对于这些项目的反馈信息,最终形成图书馆的服务特征空间,为服务项目检索提供基础数据。三是需求与服务匹配子模块,该模块同时接收读者信息与服务信息作为检索依据,从读者画像库与服务资源库中选取匹配度较高的对应项目,推荐给读者。其中的需求与服务的融合匹配算法参见下文第3节。四是推荐接口子模块,该模块可以根据图书馆方或读者的具体要求,接收读者的需求报告,并向需求与服务匹配模块发出推荐申请,最终通过邮件、短信、微信等综合方式,向读者推荐图书馆的各项服务。

2.2　模型流程

LRSM模型对于读者需求与图书馆服务的信

息处理与匹配流程如下：

Step1：读者画像生成，即LRSM模型根据读者的注册信息，生成其静态画像属性；根据读者的借阅历史、服务使用记录、留言反馈等信息，生成动态的读者画像属性，并定时或实时地对其进行动态更新；最终，读者大数据将融合生成读者画像信息，存储在读者画像库中。

Step2：服务特征挖掘，即LRSM模型根据图书馆各服务职能部门提供的服务项目说明、读者意见反馈等信息，融合生成或定时更新数据库中的图书馆服务特征信息。

Step3：需求与服务预匹配，即为了提高二者的匹配速度，LRSM模型在系统空闲时，将对两类数据作预匹配，一方面自动提高近期访问频率较高的服务的权重；另一方面对近期访问图书馆的读者进行画像信息更新，对读者的潜在需求进行预测，并预先为其生成一部分高匹配度推荐服务列表。

Step4：需求与服务匹配，即一方面，当有读者进入图书馆管理信息系统时，LRSM模型将调用其读者画像作为检索依据，搜索匹配度较高的服务，形成推荐服务队列；另一方面，当有新服务上线，或旧服务更新时，LRSM模型将根据其服务特征，搜索匹配度较高的读者，形成推荐目标读者队列。

Step5：推荐实施与读者反馈，即LRSM模型根据读者订制或默认模式，将图书馆服务精准的推荐给目标读者，并收集读者的反馈，从而进一步细化和丰富读者画像信息以及服务特征信息。

3 需求与服务的融合匹配算法

本模型采用了基于读者大数据的需求与服务融合匹配算法。该算法的基本思路来自大数据传导模型，该模型为多层信息传导结构，其本质是一种信息能量传导模型的改进。总的来说，需求与服务融合匹配算法的核心（融合匹配度）可以有如下表示：

$$E(v,h)=-\sum_i a_i v_i-\sum_j b_j h_j-\sum_i\sum_j w_{ji}v_i h_j \tag{1}$$

在匹配度表达式（1）里，vi、hj是系统中匹配元素（读者需求与图书馆服务）的状态，而ai、bj则分别是它们的融合导向值，wji则是两类元素的匹配权重；该表达式中的具体求值计算方法如下：

$$h_j=\sum_i v_i w_{ij}+b_j \tag{2}$$

其中，$h_{io}(n)=[h_1\ h_2\ldots h_M]^T$，权值 $w_j(n)=[w_1\ w_2\ldots w_M]^T$，M是服务的个数。至此，可以通过下列公式求得融合匹配度：

$$\begin{aligned}E(v,h)=&-\sum_i a_i v(i)-\sum_j b_j[\sum_i w_{ji}v_i+b_j]\\&-\sum_i w_{1i}v_i[\sum_i w_{1i}v_i+b_1]-\sum_i w_{2i}v_i[\sum_i w_{2i}v_i+b_2]\\&-.....-\sum_i w_{Mi}v_i[\sum_i w_{Mi}v_i+b_M]\\&=-\sum_i a_i v_i-\sum_j b_j^2-[\sum_j b_j\sum_i w_{ji}v_i]\\&-\sum_i b_1 w_{1i}v_i-[\sum_i w_{1i}v_i]^2-\sum_i b_2 w_{2i}v_i[\sum_i w_{2i}v_i]^2\\&-.....-\sum_i b_M w_{Mi}v_i-[\sum_i w_{Mi}v_i]^2\end{aligned} \tag{3}$$

$$\begin{aligned}E(v,h)=&-\sum_i a_i v(i)-\sum_j b_j{}^2-2\sum_j b_j\sum_i w_j(i)v(i)\\&-\sum_j[\sum_i w_j(i)v(i)]^2\end{aligned} \tag{4}$$

为保证匹配度的收敛和最大化，应对（4）进行进一步的处理。首先为精确描述读者的需求，可以建立下列模型：

$$v^{(l)}(n)=s(n)+n^{(l)}(n) \tag{5}$$

在（5）式里，s(n)是读者画像的既往需求序列，而n(n)是系统中产生的信息影响，l为既有的读者需求标识，假设读者的画像模型中共有L个

样本,假设信息影响与读者需求彼此独立,有:$\sum_i s(i)n^{(l)}(i)=0$,且$\sum_i n^{(l)}(i)=0$。

至此,首先对公式(1)中的导向值ai进行求解;其中,ai可以表示为序列a(n),由于单个读者的需求实际上是有限的,因此可以将a(n)看作一个有限的能量传导型号,由于$\sum_{i=0} a^2(i)=K_1$,对其进行求解后,有:$a(n)=a_0(n)+\varepsilon\eta(n)$,带入上式(4)后,有下列融合匹配度表述:

$$E(v,h)=-\sum_i[a_0(i)+\varepsilon\eta(i)]v(i)-\sum_j b_j^{\ 2}-2\sum_j b_j\sum_i w_j(i)v(i)-\sum_j[\sum_i w_j(i)v(i)]^2+\lambda_1[\sum_i[a_0(i)+\varepsilon\eta(i)]^2-K_1] \tag{6}$$

$$\frac{\partial E(v,h)}{\partial\varepsilon}|_{\varepsilon=0}=-\sum_i\eta(i)v(i)+2\lambda_1\sum_i[(a_0(i)+\varepsilon\eta(i))\eta(i)]|_{\varepsilon=0}=-\sum_i[v(i)-2\lambda_1 a_0(i)]\eta(i)=0 \tag{7}$$

根据读者大数据融合模型中的匹配定义,$\eta(n)$通常是一个随机型号生成表达式,所以有:

$$-v(i)+2\lambda_1 a_0(i)=0 \tag{8}$$

$$a_0(i)=\frac{v(i)}{2\lambda_1} \tag{9}$$

公式(8)中的$v^{(1)}(n)$可以视为读者需求的不同表达,在L个需求时,有:

$$-v^{(1)}(i)+2\lambda_1 a_0(i)=0$$

$$-v^{(2)}(i)+2\lambda_1 a_0(i)=0$$

.....

$$-v^{(L)}(i)+2\lambda_1 a_0(i)=0 \tag{10}$$

求其总和,可以表达为:

$$-\sum_{j=1}^{L} v^{(j)}(i)+2\lambda_1 L a_0(i)=0 \tag{11}$$

进一步有:

$$a_0(i)=\frac{1}{2\lambda_1 L}\sum_j v^j(i)=\frac{1}{2\lambda_1 L}\sum_j[s(i)+n^{(j)}(i)]=\frac{s(n)}{2\lambda_1} \tag{12}$$

此时可得:

$$E(v,h)=-\frac{1}{2\lambda_1}\sum_i s(i)v(i)-\sum_i b(i)^2-2\sum_j b(j)\sum_i w_j(i)v(i)-[\sum_i w_j(i)v(i)]^2 \tag{13}$$

与上文类似,有$b(n)=b_0(n)+\varepsilon\eta(n)$,可得:

$$b_0(j)=\frac{L}{\lambda_2-1}\sum_i w_j(i)s(n)+\frac{1}{\lambda_2-1}\sum_i w_j(i)[n^{(1)}(n)+n^{(2)}(n)+.....+n^{(L)}(n)] \tag{14}$$

有:

$$E(v,h)=-k_1\sum_i s(i)v(i)-k_2[\sum_i w_i(i)s(i)]^2-2\sum_i w_1(i)s(i)\sum_i w_1(i)v(i)-[\sum_i w_j(i)v(i)]^2 \tag{15}$$

且有:$k_1=\frac{1}{2\lambda_1}$, $k_2=\frac{1}{(\lambda_2-1)^2}$。

最终,可以利用上式(15)反推求得w1(n),与上文相似,将$w_1(n)=w_{10}(n)+\varepsilon\eta(n)$引入,可得:

$$k_2[\sum_i[w_{10}(i)s(i)]\eta(i)-2\sum_i\eta(i)s(i)\sum_i w_{10}(i)v(i)-[\sum_i w_{10}(i)v(i)]\eta(i)=0 \tag{16}$$

$$w_1(n)=k_3 s(n)+n_L(n) \tag{17}$$

在(17)中,k3为设定值。至此,可以求得融合匹配度,当(1)中的匹配度较高时,进行推荐或个性化服务的效果较好。

4 实验与结果分析

LRSM模型在某图书馆信息服务系统中进行了测试,并与当前较为流行的读者辅助服务模型RSSM(Reader Supported Service Model)进行了

独立实验与对比。为了保证实验的公平公正，图书馆技术人员在两台服务器中分别部署了LRSM模型与RSSM模型作为后台，而两种模型基于各自的独立标注读者数据集，之后的服务推荐等信息处理任务均交由统一的界面完成。最终，两个实验读者组的数量分别为1 475人（LRSM）和1 461人（RSSM），人数差距符合统计学的差别分析要求；按照上述规范与要求，最终对两种模型的读者需求覆盖度、需求匹配精确度以及系统资源占用率等客观指标进行了为期30天的跟踪对比实验，并按照信息系统开发规范的要求，对两种模型进行了读者满意度方面的主观指标调查。最终的实验结果如图2所示：

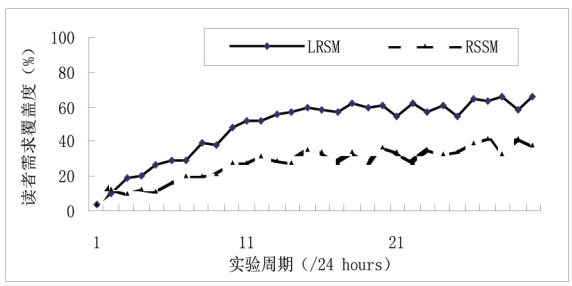

图2 两种模型的读者需求覆盖度对比

如图2所示：LRSM模型与RSSM模型在30天的实验过程中，均取得了良好的读者需求覆盖度。该指标指代图书馆个性化服务模型在一定的实验周期之中，为读者提供或者推荐的图书馆服务，在读者使用到的所有图书馆服务中所占的比例。如图2所示，尽管二者均具备良好的读者需求覆盖度，但从总体上看，LRSM模型的读者需求覆盖能力大大超过了RSSM模型。究其原因，主要是由于LRSM模型的读者需求挖掘效能更高，对读者新需求的发现更为灵敏。此外，从图2中也可以看出，LRSM模型的读者需求发现速度较快（曲线上升速度快），并在达到覆盖度稳定区后，长期保持较高的需求覆盖度。

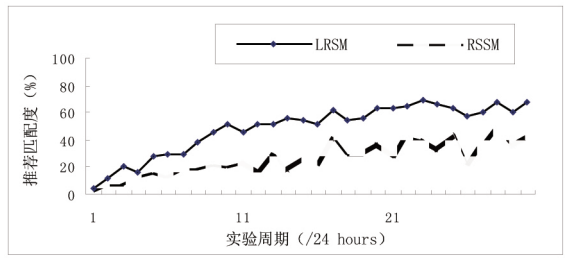

图3 两种模型的需求匹配精确度对比

如图3所示：LRSM模型与RSSM模型在30天的实验过程中，均取得了良好的需求匹配精确度。该指标指代图书馆个性化服务模型在一定的实验周期之中，两种模型提供或推荐给读者，并实际被采纳的服务，占到各自提供的服务数量的总比例。如图3所示，尽管二者均具备良好的读者需求匹配精确度，但从总体上看LRSM模型的需求匹配精确度超过了RSSM模型。究其原因，主要是由于LRSM模型的读者需求挖掘更为深入，对读者需求的刻画更为细致。此外，从图3中也可以看出，LRSM模型的需求精度上升速度较快，并在达到稳定区后，较长时间保持了匹配的高精确度。

实验完成后，将告知两组读者相关情况，并请他们为两个模型进行评价。表1是两种模型的读者主观评价情况（组内平均得分）对比。

表1 两种模型的读者主观评价对比

评价项目	LRSM模型得分	RSSM模型得分
总体满意度	81.03	53.28
推荐服务适用度	84.77	49.59
需求满足度	78.69	54.74
推荐便捷性	80.57	72.46
推荐服务全面性	83.14	59.45

最后，LRSM模型与RSSM模型在30天的实验过程中，均取得了良好的性价比；二者的CPU

占用率均未10%。在内存占用率方面,RSSM模型的峰值内存需求达到了150M,而LRSM的峰值内存需求仅为110M,体现了良好的性价比,参照目前主流的图书馆服务器配置(8G以上),该指标表明LRSM模型具有较高的系统可部署性和可扩容性。

5 结论

LRSM模型在图书馆服务上的应用体现出了其良好的应用价值,具有较高的读者需求覆盖度和需求匹配准确度。该模型的系统资源占用率较低,性价比突出,能够为读者提供个性化程度较高、需求满意度较高的图书馆服务。从目前的应用效果来看,该模型还需在以下几方面进一步扩展:首先,进一步丰富读者画像的内容,从而更全面地采集读者需求信息,深化和扩展图书馆服务的内容;其次,图书馆服务元数据挖潜,为需求—服务匹配提供更为精准和丰富的资源;最后,进一步优化需求—服务匹配算法,研究服务系统空闲期的游走需求采集模型。

[参考文献]

[1] 周强.利用Apache Mahout改善图书馆OPAC系统在大数据环境中用户体验的实践[J].图书馆研究,2015(3):91-94.

[2] LEE Z J,HU P K.A Recommender System for Library Based on Hadoop Ecosystem [J].Materials Science and Mechanical Engineering, 2017,5(3):213-221.

[3] MICAN D,MOCEAN L, TOMAI N.Building a Social Recommender System by Harvesting Social Relationships and Trust Scores between Users [J].Journal of Libraries, 2016,10(5):62-101.

[4] PATEL D, DANGRA J.A Book Recommendation System for Cloud Computing Framework [J].International Journal of Computer Applications, 2017,11(5):167-183.

[5] RAAMKUMAR A S, FOO S,PANG N.A Framework for Scientific Paper Retrieval and Recommender Systems [J]. Journal of Digital Libraries, 2016,8(12):34-55.

[6] HU Y F, KOREN Y,VOLINSKY C.Collaborative Filtering for Implicit Feedback Datasets for Libraries [J].Journal of Digital Information Processing, 2015,13(2):235-276.

[7] VERPLANKEN J.Implementation of a recommender system for the library of Ghent [J].Journal of Information Management, 2017,4(2):52-91.

[8] MOONEY R J,ROY L.Content-Based Book Recommending Using Learning for Text Categorization [J]. Journal of Digital Libraries, 2017,4(3):109-131.

[作者简介] 林淑贞(1975—),女,馆员,本科,研究方向:图书馆服务信息化等。

[收稿日期] 2018-05-24　(编发:王丽)