

doi:10.3772/j.issn.2095-915x.2016.05.005

基于社会信息重排序的图书检索方法

金建霖, 周芳, 张博文, 冯少辉, 晋赞霞, 殷绪成

(北京科技大学计算机科学与技术系 北京 100083)

摘要: 随着社交网络的兴起和发展, 互联网上出现了大量与商品有关的社会信息。如何利用这些社会信息结合商品元数据进行检索和推荐是信息检索领域中一个热门的研究问题。本文以社会图书检索为例, 提出了一种通用的信息检索方法来解决这一问题。首先, 通过分析原始图书数据集和图书的用户标签、用户评分和流行度等社会信息, 从图书中提取不同的社会特征构建特征矩阵; 然后分别计算图书在各种社会特征上的相似度, 并使用不同的策略对搜索引擎返回的排序结果进行重排序; 最后使用学习排序的方法进行重排结果融合, 得到最终的图书检索结果。在实验中, 使用该检索方法在 INEX Social Book Search 2015 和 2016 数据集上分别进行了训练和测试。结果表明, 相比现有的技术, 该检索方法能够有效提升图书检索的效果。

关键词: 社会图书检索, 社会信息, 社交网络, 重排序

中图分类号: TP391

A Book Retrieval Method Based on Social Information Reranking

JIN JianLin, ZHOU Fang, ZHANG BoWen, FENG ShaoHui, JIN ZanXia, YIN XuCheng

(Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: With the development of social networks, a great deal of social information arised on the internet. How to make full use of these social information and products metadata in search and recommendation is a

基金项目: 本文受国家自然科学基金“结合前馈和反馈机制的自然场景文本识别技术”(61473036)资助。

作者简介: 金建霖(1987-), 硕士研究生, 研究方向: 信息检索; 周芳(1972-), 博士, 副教授, 研究方向: 信息检索; 张博文(1992-), 博士研究生, 研究方向: 信息检索; 冯少辉(1990-), 硕士研究生, 研究方向: 信息检索; 晋赞霞(1992-), 硕士研究生, 研究方向: 信息检索; 殷绪成(1977-), 通讯作者, 博士, 教授, 博士生导师, 北京科技大学计算机与通信工程学院计算机科学与技术系模式识别技术创新实验室主任, 研究方向: 信息检索、模式识别, xuchengyin@ustb.edu.cn。

hot topic in the research field of information retrieval. In this paper, the authors took social book search as an example, proposed a general search-recommendation hybrid system for this problem. Firstly, by analyzing the original book data set and the social information of books, such as tags, rating and popularity, this study constructed the feature matrix by extract social features in books. Afterwards, we computed the similarity of books in different social features, and reranked the initial result of search engine with different strategies. Finally, we used learning-to-rank technique to combine a wide range of diverse reranking results. In this experiment, we trained and test the proposed system on the INEX Social Book Search 2015 and 2016 datasets with this method respectively. The result showed that our system can effectively improve the retrieval performance compared with other state-of-the-art systems.

Keywords: Social book search, social information, social network, rerank

1 引言

互联网的出现,使得信息技术飞速发展。人们发布、获取、分享信息的方式已经从根本上发生改变。但是互联网规模和覆盖面的迅速增长带来了信息超载的问题,过量信息同时呈现使得用户无法从中获取对自己有用的内容,信息的使用效率反而降低。搜索引擎作为一种主流的信息检索和过滤的手段,能够帮助用户过滤互联网上海量的信息,找到自己感兴趣的信息或关心的内容。

在以传统搜索引擎为核心的信息检索中,用户需要提供有关被检索对象的关键字。这些关键字通常由搜索对象的元数据构成,例如:分类、标题、品牌、材质等。这些信息描述了被检索对象自身的属性。可是在实际的情况中,用户由于知识面的限制并不一定能够给出合适的关键字,造成检索效果的降低。以图书检索为例,亚马逊、当当等图书电子商务网站能够提供给搜索引擎的只有图书的标题、作者、ISBN 号和出版社等最基本的信息。在这种情况下,如果用户不知道图书的相关信息,就很难检索到关心的书籍;同时,只有用户向搜索引擎输入足够相似的关键词,才能够得到想要的结果^[1]。

与此同时,社交网络在互联网上为人们提供了一个交际往来平台。互联网用户在相互交流时产生了大量的社会信息,例如用户在使用过某种商品后推荐给自己的好友,或者对商品的质量、内容、优缺点作出的评判和意见。这些社会信息可以使互联网用户从更多的渠道了解商品信息,而不只是从商品信息的发布者一个渠道进行了解。用户对商品信息也能够有更加客观的评价^[2]。

互联网上这些用户反馈的社会信息能够从更多的角度描述被检索的对象。使用户可以有更加多样化的方式来准确地描述检索需求,从用户的角度提高检索效果。弥补了检索过程中信息量不足的问题。这种融合社会信息的信息检索方法就叫做社会检索。如何利用社会信息结合商品元数据进行检索和推荐是信息检索领域中的研究热点之一^[3]。

本文以图书的检索和推荐为例,提出了一种基于社会信息的重排序检索算法。能够处理互联网上各种各样的图书数据资源,包括用户历史记录、用户收藏、以及社交网络中用户生成的图书描述、打分、评论等。通过从图书中提取多种社会信息,并对其进行表示和有效融合,使图书检索和推荐的效果得到提高。为用户提供更优秀的

图书搜索体验。

本文在第 2 节中，介绍了社会图书检索和相关的工作；第 3 节详细的描述了基于社会信息重排序的图书检索算法及其框架；第 4 节给出了对比实验和结果分析；最后进行总结。

2 社会图书检索和相关工作

图书作为一种知识信息存储和传播的介质长久以来一直被广泛的使用。现在，随着计算机和互联网技术的发展，越来越多的用户通过互联网查询自己喜欢的图书。社会图书检索的目的就是根据用户提出的查询，使用图书元数据和相关的社会信息，在指定的图书数据集中检索，为用户提供一个有序的图书推荐列表。

LibraryThing.com(LT) 是一个社会图书分类信息网站，同时也是一个社交平台。用户可以通过在网站的论坛中发布一个话题 (topic) 来获取其他用户推荐的图书。在 2011 年，社会图书检索开始作为 INEX 的一项比赛。INEX 社会图书检索

的组织者对这些话题和推荐的书籍进行整理和筛选，选出若干查询作为比赛任务 (查询的示例见表 1)。参赛者被要求根据话题发布者的需求从图书数据集中检索出相应的书籍^[4]。书籍信息来自 Amazon 的图书集，其中包含了大量的社会信息 (图书的示例见表 1)。由于传统的信息检索技术在社会图书检索中并不能取得较好的效果^[5,6]。如何利用这些社会信息提高图书推荐检索的效果，就成了一个值得研究的问题。

在以往的社会图书检索比赛和其他的相关研究中，一些技术方法表现出了较好的检索效果。Bellot 等人^[7,8] 等人分析了书籍的评论和打分，通过定义“可能性”对书籍进行打分和重排。但是，他们的实验结果并没有显示出社会信息在图书检索中的优势。Bogers 等人^[9] 研究了图书在作者、评论、评分方面的相似性，实验表明亚马逊图书集中的“相似图书”域在重排上能够取得较好的效果。Melanie 等人^[10] 研究了图书的价格、页数、用户评分等社会信息，并使用随机森林进行重排序融合，取得了较好的效果。在最新的研究中，

表 1 社会图书检索 XML 文档示例

话题	图书
<pre> <topics> <topic> <topicid>75275</topicid> <request>Do you like a specific series of books (not Twilight, those books r a little dumb, srry) that u recommend 2 others?</request> <group>What Are You Reading Now?</group> <group>Returned Peace Corps Volunteer Readers</group> <title>A series of good books</title> <examples> <work> <booktitle>Twilight</booktitle> <author>Stephenie Meyer</author> <workid>8384326</workid> </work> ... </topics> </pre>	<pre> <book> <isbn>0006174000</isbn> <title>The Endless Game</title> <manufacturer>Fontana Press</manufacturer> <publisher>Fontana Press</publisher> <dewey>823</dewey> <numberofpages>384</numberofpages> <reviews> <review> <authorid>A222LQEPE7O7BV</authorid> <date>2002-04-09</date> <summary>A Soild Effort</summary> <content>Not a bad little page-turner. I liked the story line ...</content> <rating>4</rating> <totalvotes>2</totalvotes> <helpfulvotes>2</helpfulvotes> </review> </reviews> <tags> <tag count="3">spy</tag> </tags> <similarproducts><similarproduct>1400066026</similarproduct></similarproducts> <browseNodes> <browseNode id="18">Mystery & Thrillers</browseNode> </browseNodes> </book> </pre>

Zhang 等人^[11-13]利用多元化的社会信息设计了一种整合了搜索和推荐的图书检索模型。虽然上述的研究为社会信息在社会图书检索中的应用提出了一些思路,但是这些研究都没有提出一种处理图书社会信息的通用方法。

3 基于社会信息的图书检索重排序算法

信息检索技术发展的目的是为用户提供更好的搜索体验。为此,本文提出了一种结合网络上丰富社会信息的图书检索方法,该检索方法使用了一些通用的图书社会信息,如:评分、标签、评论和相似图书等。整个检索系统主要包括三个模块。首先利用图书元数据进行检索,获得初排结果;第二步,使用图书的社会信息进行重排序,设计了一种基于内容过滤推荐的方法计算用户对书籍的偏好,获得图书的重排结果。最后,通过学习排序技术对不同的重排序结果加以整合,得到最终的检索结果。

3.1 基于数据集优化的初排序

3.1.1 图书数据预处理

在信息检索工作中,高质量索引的建立是提升检索效果的基础。因此在建立索引之前,需要对原始数据进行处理。首先,要去除 XML 文档中无法被信息检索使用的信息,如“image url”。其次,在 XML 文档中有若干域显示的是有用的

数值信息,例如 `<tag count=" 12" >crime</tag>` 表示有 12 位用户给该图书打上了“crime”的标签、`<dewey>810</dewey>` 表示该图书的杜威十进分类,这些信息在信息检索中与检索目标息息相关,但是无法直接被搜索引擎使用。实验中,根据 Bogers 使用的方法^[14]对 XML 文档进行处理,将数值信息替换为有意义的文本信息。

3.1.2 建立索引

Galago 是开源的搜索引擎,可以对数据建立索引并提供查询。为了提升检索效果,我们设计了两方案来建立索引。在第一种方案中,考虑到 XML 文档中的某些字段(例如: title、tag、content 和 summary 等)具有比其他字段更高的相关度和更明确的含义,所以去除其他的信息,只使用这些相关度较高的字段来建立索引。另一种策略是对 XML 文档进行扩展。通过研究数据集来源 Library Thing 论坛上的原始图书信息,发现很大一部分的书籍都缺少检索中用户关注程度较高的一些字段,如 content 和 summary。所以在这种策略中,需要参考互联网上其他网站的图书信息对 XML 文档中的这些字段进行扩展,然后再建立索引。在实验中,参考了豆瓣(douban.com)和 Lookupbyisbn(lookupbyisbn.com)两个网站上的信息。从豆瓣获取图书简介,从 Lookupbyisbn 获取了图书的描述信息。两个网站的图书信息都以 ISBN 作为标识。使用这些信息,对图书数据集中的约 60 万图书的 XML 文档进行了扩展(示例见表 2)。

表 2 扩展前后的书籍 XML 文档

扩展前	扩展后
<code><book></code>	<code><book></code>
<code><title>Mister Monday</title></code>	<code><title>Mister Monday</title></code>
<code><summary>So good, you can't put it down!</summary></code>	<code><summary>So good, you can't put it down!</summary></code>
<code><content>Now, I had...</content></code>	<code><content>Now, I had...</content></code>
<code><tag count=" 9" >children's literature</tag></code>	<code><tag count=" 9" >children's literature</tag></code>
...	<code><brief introduction>the content from the douban.com</ brief introduction ></code>
<code></book></code>	<code><description>the content from the lookupbyisbn.com</ description ></code>
	...
	<code></book></code>

3.1.3 建立活动图书集

通过分析用户简介。可以从中获取了一个活动图书集。这个集合中的图书最有可能被推荐给话题的发布者。

3.1.4 检索

使用 Galago 进行对图书索引进行检索，获得初排结果。Galago 使用的模型是目前广泛被采用的查询似然模型。这是一种融合了词项频率的概率检索模型。假设每个文档的先验概率是相同的，文档通过评价概率 $P(Q|D)$ 来排序，具体计算公式为：

$$S = \log P(Q|D) = \log \prod_{i=1}^n p(q_i|D) = \sum_{i=1}^n \log \frac{f_{q_i,D} + \mu \frac{c_{q_i}}{|C|}}{|D| + \mu} \quad (\text{公式 1})$$

公式中 $f_{q_i,D}$ ，表示图书中的特征 q_i 在文档 D 中现的次数， $\frac{c_{q_i}}{|C|}$ 表示特征 q_i 的数据集概率， μ 是 Dirichlet 平滑参数。

3.2 融合社会特征的重排序

在社会图书检索中，根据亚马逊图书集和相关的社会信息，设计了两类一共七种不同的重排策略(表3)。第一类基于相似度计算: 标签重排(T) 关注 <tag> 域，即读者为图书打的标签；相似图书重排(S) 关注 <similarproduct> 域，表示信息发布者选出的相似图书；Read by one 重排(R) 关注被相同读者看过的书籍。第二类重排基于话题发布者给出的样例图书的特征: 流行度重排(P) 关注图书的流行程度、推荐范例重排(E) 关注话题发起者列举的样例图书中的特征、读者量重排(N) 关注某一本书的阅读量、贝叶斯评分重排(B) 关注图书在读者中的评分特征。

3.2.1 重排特征提取

通过提取书籍的社会特征构建图书集的特征矩阵。首先分析图书数据集，根据图书描述文档的结构定义特征集合 F ，对于 F 中的每一个特征都根据其描述建立一个特征向量。例如，在社会图书检索中的标签 t 就是 F 中的一个特征。

表 3 重排序策略

重排序策略	描述
Tag(T)	用户标签
Similar product (S)	相似图书
Read by one (R)	被相同的读者阅读
Popularity (P)	图书流行度
Example recommended (E)	推荐示例图书
Reader number (N)	读者数量
RatingBayes (B)	用户对书籍的评分

如果假定总共只有两种不同的标签，其中两本书 i 和 j 的标签向量就可以表示为 $\vec{t}_i = [3,0]$ 以及 $\vec{t}_j = [1,5]$ 。表示有 3 个用户为书籍书 i 打上了第一个标签，而分别有 1 个和 5 个用户为书籍 j 打上了第一个标签和第二个标签。使用这种方法，如果 f 用表示某一特征，则所有书籍的特征矩阵可以表示为公式 2：

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix} \quad (\text{公式 2})$$

公式中 m 表示特征的种类数， n 表示数据集中书籍的数量。特征矩阵中的每一行都表示某一本书的特征向量。

3.2.2 重排得分计算

在重排阶段，通过计算书籍的重排得分，调整检索结果的排序。针对初排结果中的前 1000 项，使用上面的 7 种策略分别对初排结果进行重排，公式为：

$$S_{new} = \alpha \cdot S_{old} + (1 - \alpha)S_R \quad (\text{公式 3})$$

其中 S_{old} 为重排前的得分， S_{new} 为重排后的得分， S_R 为重排评分。 S_R 的计算方法根据重排策略的不同而有所区别。

标签重排 (T)

借鉴基于内容的推荐算法的思想，设计了一种基于用户标签相似度的重排序算法，提高初排序过程中排序靠后的相关结果，具体的计算方法为：

$$S_R(i) = \sum_{j=1}^N sim_{ij} \cdot S_{old}(j) (j \neq i) \quad (\text{公式 4})$$

sim_{ij} 表示 i 和 j 两本书之间的相似度。在特征矩阵 (公式 2) 中, 通过计算两个特征向量 $f_i=[f_{i1}, \dots, f_{im}]$ 和 $f_j=[f_{j1}, \dots, f_{jm}]$ 之间的相似度表示两本书 i 和 j 之间的相似度。在传统的基于内容的过滤模型中, 这个相似度通常使用余弦相似度:

$$sim_{ij}(f) = \cos \langle \vec{f}_i, \vec{f}_j \rangle = \frac{\vec{f}_i \cdot \vec{f}_j}{|\vec{f}_i| |\vec{f}_j|} \quad (\text{公式 5})$$

其中, $\vec{f}_i \cdot \vec{f}_j$ 表示两个向量的点积, $|\vec{f}_i|$ 表示向量的模。

相似图书重排 (S)

相似图书重排同样基于内容的推荐算法思想。 S_R 重排评分的计算方法与标签重排相同 (公式 10), 但是考虑了书籍之间相似性的传递性, 并且使用了定量的相似度表示:

$$sim_{ij}(S) = \begin{cases} 1, & i \text{ 是 } j \text{ 的相似图书,} \\ & \text{或者 } j \text{ 是 } i \text{ 的相似图书} \\ 0.5, & i \text{ 是 } j \text{ 的相似图书的相似图书,} \\ & \text{或者 } j \text{ 是 } i \text{ 的相似图书的相似图书} \\ 0, & \text{其他情况} \end{cases} \quad (\text{公式 6})$$

Read by one 重排 (R)

Read by one 重排基于这样一个假设: 两本书被同一个读者阅读过, 那么这两本书就是相关的。如果一本书的评分较高, 那么另一本书的评分也应该升高。

$$S_R(i) = \log(n) \cdot \frac{\sum_{r \in R_i} r}{n} \cdot S_{old}(i) \quad (\text{公式 7})$$

其中, R_i 表示用户对书籍的所有评分的集合, r 为用户评分。

贝叶斯评分重排 (B)

根据读者对某本书籍的评分进行重排, 提升有评分或者评分较高的书籍的排名, 降低没有用户评论过或者评分较低的书籍排名。计算公式为:

$$S_R(i) = \frac{1 + BA(i)}{1 + BA_{max}} \cdot S_{old}(i) \quad (\text{公式 8})$$

$BA(i)$ 表示书籍 i 的贝叶斯平均评分^[15], 其计算公式为:

$$BA(i) = \frac{\bar{n} \cdot \bar{m} + \sum_{r \in R_i} r}{n + \bar{n}} \quad (\text{公式 9})$$

其中, \bar{m} 是所有书籍的评价评分, n 为评分数量。

流行度重排 (P)

考虑到话题的发布者更有可能去选择那些流行程度较高的书籍, 所以根据一本书的流行程度, 对初排结果进行重排。流行度可以体现在销量、读者量、论坛热度和出版时间等方面。根据流行度计算重排评分的公式为:

$$S_R(i) = S_{old}(i) \cdot \frac{pop(i)}{\beta} \quad (\text{公式 10})$$

$pop(i)$ 表示图书的流行度, β 根据流行度的取值范围确定。

读者量重排 (N)

读者数量是一本书流行程度的重要体现, 所以单独根据读者量进行重排:

$$S_R(i) = S_{old}(i) \cdot \frac{reader_num(i)}{\beta} \quad (\text{公式 11})$$

$reader_num(i)$ 表示书籍 i 的读者数量, β 根据读者量的范围确定。

推荐范例重排 (E)

通过分析查询文档发现, 一些话题存在一个推荐范例图书集, 根据重排的图书是否属于推荐范例图书, 重排得分的计算公式为:

$$S_{new}(i) = S_{old}(i) + \alpha S_{old}(i) \quad (\text{公式 12})$$

如果图书 i 是样例图书 α 的取值由训练集数据得到, 如果不是则 $\alpha=0$ 。

3.3 基于学习排序的重排结果融合

在前面讨论的重排策略中, 某一重排策略只关注了书籍在某一方面的特征。有必要对重排的结果进行整合, 综合考虑书籍在不同方面的特征。按照传统的方法, 结果的整合与参数的设置通常是手动或者半自动的, 往往会消耗大量时间或者无法获得最优的解决方案。学习排序^[13] (Learning-To-Rank) 是一种监督学习的排序方

法,通过机器学习融合多种特征对数据进行排序。

学习排序的基本方法是:首先准备训练集数据,提取数据特征,并为其打上标签。其次,选择一种学习模型对训练数据进行学习。最后,使用学习到的模型对新的数据评分,根据评分完成新数据的排序。学习排序使用的算法有随机森林、RankNet 等。

综上所述,基于社会特征重排序的图书检索算法框架如图 1 所示:

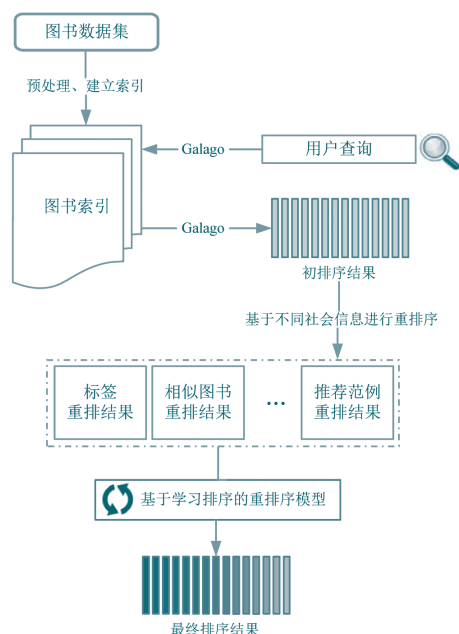


图 1 基于社会特征重排序的图书检索算法框架

首先对图书数据集进行处理,建立索引。对于一个用户的查询,首先使用搜索引擎结合图书内容、图书元数据(作者、简介、出版商等)和其他信息进行检索,得到图书的初排序结果。然后根据不同的社会信息使用不同的重排策略进行初排结果的重排序。重排的依据是社会信息的推荐模型。最后,通过学习排序技术对不同的重排序结果加以整合。通过监督学习得到一个融合不同知识的训练模型,并对新的数据集进行排序。最终,经过以上的步骤,获得图书检索的最终排序结果。

4 实验

4.1 实验数据

使用 CLEF Social Book Search (SBS) Lab Suggestion Track 比赛提供的数据集进行实验。该数据集使用了 LibraryThing 网站上的书籍推荐作为话题,使用 NDCG@10 进行检索结果的评价。SBS 比赛使用的图书集包含了亚马逊图书数据库中的约 280 万图书,该图书集不仅包含了基本的图书信息 (ISBN、标题、作者、简介等),还包含了亚马逊网站上与图书相关的大量社会信息(包括评论、评分、标签、相似图书等)。比赛标准答案中图书的相关度由人工标记的相关度决定,如果话题的发布者将其他用户推荐的图书添加到了自己的收藏目录中,就认为该图书拥有较高的相关度^[18]。在 2015 年和 2016 年的比赛中,赛方分别选出了 208 个和 120 个话题作为比赛任务。在实验中,使用 SBS2015 数据集进行训练,使用的测试集为 SBS2016。

4.2 实验设置

为了选出最有用的特征并确定最优的参数值,首先在 SBS2015 数据集上进行重排序模型的学习。训练的结果如表 4:

表 4 在 SBS2015 数据集上的训练结果

重排序模型	NDCG@10	Best
keywords	0.1291	-
keywords + active	0.1518	-
keywords + active + Similar product	0.1528	0.991
keywords + active + Example recommend	0.1581	0.384
keywords + active + Popularity	0.1538	0.09
keywords + active + Reader number	0.1709	0.706
keywords + active + Read by one	0.1548	0.997
keywords + 所有重排策略 + 样例书籍过滤	0.1972	-

如表4中所示，“keywords + 所有重排策略 + 样例书籍过滤”、“活动图书集”、“读者量”，在重排中的效果提升最大。因此，最终选定了6种重排序融合模型进行实验。

4.3 实验结果和分析

使用表5中的6种实验方案在SBS2015数据集上学习出模型，在SBS2016数据集上进行测试，结果表6所示。

表5 实验方案

实验方案	初排索引	重排模型
Run1	部分活动图书集	读者量 + 相似图书 + 流行度。
Run2	部分活动图书集	读者量重排。
Run3	部分活动图书集	Read by one 重排。
Run4	部分活动图书集	相似图书重排。
Run5	完全活动图书集	相似图书重排。
Run6	完全活动图书集	读者量 + 相似图书 + 流行度。

表6 不同的实验方案在SBS2016上的实验结果

实验方案	NDCG@10	MRR	MAP	R@1000
Run1	0.2157	0.5247	0.1253	0.3474
Run2	0.2047	0.4700	0.1177	0.3474
Run3	0.1989	0.4923	0.1157	0.3474
Run4	0.1935	0.4685	0.1106	0.3474
Run5	0.2009	0.4767	0.1128	0.3146
Run6	0.2030	0.4868	0.1144	0.3146
初排 + 去停用词	0.1265			

NDCG@10代表了前10个检索结果排序的准确性，将实验结果与INEX Social Book Search其他参赛者的NDCG@10结果进行对比（图2所示）。

实验结果表明，相比初排检索，基于社会信息的重排模型能够有效提升图书检索的精度。其中Run1方案融合了所有重排模型，在测试的120

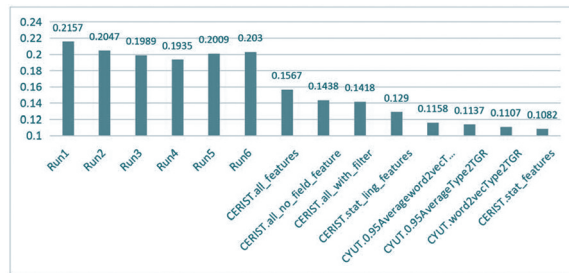


图2 六种实验方案与其他参赛者在SBS2016数据集
中的检索效果比较

个话题中，检索结果的NDCG@10达到了0.2157。在SBS2016中，相比其他参赛者的最优结果，基于社会信息的检索模型能够将检索结果的精度提高约30%和70%。

5 总结

本文提出了一种能够融合多种社会信息的查询推荐方法。首先对用户查询和图书数据集的特征进行分析，并使用Galago搜索引擎进行初排检索；然后使用特征矩阵来表示图书的社会特征，并通过不同的策略在初排的基础上进行重排序，将社会特征融合到了信息检索工作中，最后使用学习排序的方法融合重排序结果。经过实验和分析，基于社会信息重排序的图书检索方法能够有效提升图书检索的效果，而且能够在大多数的商品检索和推荐系统中使用。在未来的工作中，我们将对豆瓣等中文图书社区进行实验分析，设计新的商品检索和推荐系统来提高检索效果，并深入研究商品社会信息的获取和应用等问题。

参考文献

- [1] Deveaud R, Sanjuan E, Bellot P. Social Recommendation and External Resources for Book Search[J]. Lecture Notes in Computer Science, 2011, 7424:68-79.
- [2] Ozyar E, Gurdalli S. Social Network Analysis: a Powerful Strategy, also for the Information

Sciences[J]. *Journal of Information Science*, 2002, 28(6):441-453.

[3] Sinha R, Swearingen K. Comparing Recommendations Made by Online Systems and Friends[J]. In *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*, 2001.

[4] Koolen M, Bogers T, Gäde M, et al. Overview of the CLEF 2015 Social Book Search Lab[C]// *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Springer International Publishing, 2015:545-564.

[5] Koolen M, Kazai G, Kamps J, et al. Overview of the INEX 2011 Books and Social Search Track[J]. *Lecture Notes in Computer Science*, 2012, 7424:1-29.

[6] Koolen M, Kazai G, Kamps J, et al. Overview of the INEX 2012 Social Book Search Track[J]. *Clef Evaluation Labs & Workshop Online Working Notes*, 2012, 1180.

[7] Bonnefoy L, Deveaud R, Bellot P. Do Social Information Help Book Search?[EB/OL]. [2016-05-31]. <http://www.clef-initiative.eu/documents/71612/f1163b00-fde2-4f4e-b327-acaa04aae51b/>.

[8] Benkoussas C, Bellot P. Book Recommendation Based on Social Information[EB/OL]. [2016-07-01]. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-BenkoussasyEt2013.pdf>.

[9] Bogers T, Christensen K W, Larsen B. RSLIS at INEX 2011: Social Book Search Track[M]//

Focused Retrieval of Content and Structure. Springer Berlin Heidelberg, 2012:45-56.

[10] Imhof M, Badache I, Boughanem M. Multimodal Social Book Search[EB/OL]. [2016-04-30]. <http://ceur-ws.org/Vol-1391/7-CR.pdf>.

[11] Zhang B W, Yin X C, Cui X P, et al. Social Book Search Reranking with Generalized Content-Based Filtering[C]// *ACM International Conference on Conference on Information and Knowledge Management*, ACM, 2014:361-370.

[12] Yin X C, Zhang B W, Cui X P, et al. ISART: A Generic Framework for Searching Books with Social Information[J]. *Plos One*, 2016, 11(2):1-27.

[13] Zhang B W, Yin X C, Zhou F. A Generic Pseudo Relevance Feedback Framework with Heterogeneous Social Information[J]. *Information Sciences*, 2016.

[14] Bogers T, Larsen B. RSLIS at INEX 2013: Social Book Search Track[C]// *In INEX' 13 Workshop Pre-proceedings*. Springer, 2013:31-38.

[15] Koolen M, Huurdeman H C, Kamps J. Comparing Topic Representations for Social Book Search[EB/OL]. [2016-05-31]. <http://ceur-ws.org/Vol-1179/CLEF2013wn-INEX-KoolenEt2013.pdf>.

[16] Cao Z, Qin T, Liu T Y, et al. Learning to Rank: from Pairwise Approach to Listwise Approach[C]// *Machine Learning, Proceedings of the Twenty-Fourth International Conference*, 2007:129-136.