

## Intelligent Application of Library based on big data

Wenyan HU    Zhengde BAO    Yawen TANG

School of Computer and Software, Jincheng College, Sichuan University, Chengdu, Sichuan, 611731

### Abstract

The mass of data in today's mobile network is disorganized before the data analysis, and the existence of big data technology is to process and store these data. Because the library has many sources and large quantities of data, the problem that the intelligent library must deal with urgently is how to process the virtual data and the real data, where to store the data, and how to establish a high efficiency. A safe and stable data resource management system. Based on big data's basic technology of data analysis and data mining, this paper briefly analyzes the intelligent application of library management system.

### Key Words

Cloud Computing, Big Data, Data Processing, Intelligent Library, Data Analysis

DOI:10.18686/jsjxt.v1i2.657

## 基于大数据的图书馆的智慧化应用

胡文彦    鲍正德    唐娅雯

四川大学锦城学院计算机与软件学院, 四川成, 611731

### 摘 要

在当今的移动网海量的数据在没经过数据分析之前是杂乱无章的, 大数据技术的存在则是为了处理和存储这些数据。图书馆因其数据来源途径多、数量大, 设智慧图书馆所刻不容缓需要处理的问题便是如何对这些虚拟数据及现实数据进行处理, 将数据存放在何处, 建立一个高效、安全且能够平稳运行的数据资源管理系统。本文基于大数据的数据分析与数据挖掘基础技术之上, 简要分析了图书馆管理系统的智能化应用。

### 关键字

云计算; 大数据; 数据处理; 智慧图书馆; 数据分析

### 1.引言

2012 年是与众不同的一年。作为大数据信息与技术开始指数型爆发式增长的元年, 短短数年间, 大数据就以迅雷不及掩耳之势火遍国内外各行各业, 从商圈到工业技术层再到政界, 各个阶层无一例外的接受了大数据科学技术的洗礼。从人文思想的改变到技术的进步, 大数据已经融入了我们的学习和工作。尤其是在高校之中, 传统的信息科学技术已经无法满足日益增长的数据分析处理, 因此使用大数据科学技术来建立图书馆管理信息管理系统是十分必要的。

### 2.大数据科学的发展

数据科学最初是与统计学联系在一起的, 因此数据科学在最初可以分为以下三个部分: 数据收集、数据建模与分析, 以及决策制定。数据科学通常人们可以理解为数据分析预测和数据挖掘。<sup>[1]</sup>IBM 公司的 SPSS 软件就是学习数据科学时运用最多的的软件。在大数据的驱动下, 图书馆的智慧化建设也将进行一个从内而外的革新, 以便以更好的智能模式服务于大众。

#### 2.1 大数据至商业智能的发展演变

大数据的产生同云计算, 都是被移动互联网、物联网还有各种各样的社交媒体所催生<sup>[2]</sup>。同时在大数据爆炸发展的过程中对于数据存储的需求使得数据库出现, 由此以后数据发展步入稳定状态, 数据也经由逐步的过

滤及清晰,从最开始的数据,到了信息、知识、智慧,形成当今的智能模式,分成阶段来说,可以说是从RDBMS到Hadoop,再到NewSQL。同时在商业智能的实际应用中,它们并不是独立的存在,他们往往是相互交融并行的。

## 2.2 大数据在当今社会的重要意义与运用

众所周知,大数据存在的意义不在于数据的存储,而是将有意义的数,也可称之为信息,进行系统的、专业的“加工”处理,使原本冗余复杂的海量数据“活”起来,实现自身价值的增值。

自21世纪提出大数据的概念后,发展至今,国内最值得一提的就是各行业的“互联网+”以及落实到实际生活中的“智慧+”项目了。大量的数据流就如同早高峰的车流,如何正确的引流和规划路线,是对于大数据的准确运用的关键所在。互联网与各行各业的结合,智慧城市、智慧医疗、智慧图书馆更是大数据发展史上的新的里程碑式成果。

## 3. 大数据技术使得图书馆从传统图书馆转变到智慧图书馆

以身边最为贴切的图书馆为例,传统的图书馆在当今云计算、大数据、物联网以及人工智能等技术高速发展的时代已经可以说是十分落后了,新型图书馆服务与管理模式的出现是大势所需。智慧图书馆是当今图书馆发展的终极目标,它以数字化、网络化、可视化以及最为重要的智能化技术为基础进行定义发展<sup>[3]</sup>。怎样运用云计算平台和技术对图书馆大数据进行分析处理,将现实中的图书馆搬上云端,使得读者在智能设备上完成图书馆的信息搜索与查询,获得优质的服务,是如今所迫切需要做的事。

### 3.1 智慧图书馆在云计算层面的分析

云计算这个新技术的运用,为图书馆的智慧化过程提供了一个强有力的支撑,它将数据、信息和智慧上升到了智能这个层面。对于图书馆的海量数据来说,可以大致分为内部数据和开放数据。这个时候就可以将图书馆中的内部数据及备份数据以私有云的形式存储在云平台上;相对应的,那些开放型的数据资源就可以以公有云的形式进行存储。公有云和私有云这两者相结合的存储方式能够很大程度上的提高图书馆的数据资源管

理的流通速度和灵活性。同时云计算的安全性特性也是为不同身份的用户提供了一个加密的保障。

### 3.2 智慧图书馆与大数据

爆炸式的数据量给读者带来了很大的困扰,智慧图书馆也就应运而生。建立智慧图书馆,第一步是要能够知道在这大量的图书馆数据资源中,哪些是读者需要的信息,即需要进行一个系统的需求分析。知道了具体需求,才能够明确目标,从海量的、模糊的、有噪声的繁杂数据中清洗出有效信息。

基于云计算平台的大数据挖掘技术也能够为图书馆带来很大的好处,比如说灵活的时间和空间弹性、强大的大批量数据处理能力以及较低的硬件成本等。大数据资源清洗、挖掘与管理是优化图书馆业务流程的重要步骤,也是提高图书馆数据处理效率的重要因素,为智慧图书馆系统建设提供了战略指导。

## 4. 智慧图书馆的智慧化体现

智慧化图书馆的发展顺应了人们日益增长的对于精神层面的需求,同时也促进了社会文明的进步。全球绝大多数的图书馆都在经历着从纸质与电子版的碰撞摩擦到互联高效的智能化过程,其中以下几个方面智能化最能体现出智能图书馆的精髓所在。

### 4.1 大数据智能化

需求分析是一切技术实施的根本,以大数据信息技术为基础的智慧图书馆正结合着图书馆、图书馆管理员、读者等多方面对于图书馆管理信息系统进行重塑。随着大数据技术处理的发展,图书馆的网络虚拟信息与物理层面进行融合,图书馆人机也进行了深度的交互,整体步入了全新的模式。智慧图书馆于智慧化技术,对于数据的整合、不同技术的融合、时间空间的服务力,也是取得了日新月异的进步。当然,这个进步也还只是智慧化进程中的冰山一角而已。

#### 4.1.1 流量统计

同一时段图书馆的读者出入馆统计、同一时段的借、还书统计。从而通过大数据技术实现读者的流量控制<sup>[4]</sup>。

#### 4.1.2 容量统计

图书馆当日最高、最低馆内人员容量统计。一方面

可以在图书馆人流量即将达到饱和时对将要入馆的人数进行预防控制措施;另一方面在图书馆人员容量达到极小值时进行架外书籍的入架、架上书籍的整理统计以及馆内设施的清洁维护等。

### 4.1.3 大数据的深度运用

可以将大数据技术与人工智能技术、可视化技术、物联网技术等进行融合运用。以第一点中的流量数据为例,大数据技术结合人工智能技术可以在入馆、出馆时对读者信息进行处理,例如该读者上次入馆有违规操作,这条信息会使用大数据技术进行处理,下次入馆时将会使用人工智能技术阻止入馆。通过大数据统计出馆内人员容量及分布情况,在运用可视化技术直观的通过热点图等方式进行表示,可以非常有效的保障图书馆的平稳运行。

## 4.2 移动客户端智能化

随着近年来微信公众号、手机 APP 的广泛使用,图书馆智能化也不再仅仅局限于 PC 端了。关于图书馆智能化管理信息系统实施的可行性来说,软件方面是基于数据库、SPSS 等软件实现。经济可行性方面来说,虽然一开始的智能化实施所需要的成本较高,但是后期取得的效益回报是巨大的。社会可行性就更不必说了,智慧化系统铺设是社会发展的的大势所需,也是当今社会高技术人才所必须要掌握的技能。以已推行出的图书馆智能化 APP 使用的高校图书馆为例,结合管理信息系统相关知识分析,图书馆智能化管理系统主要实现了以下几个功能:

### 4.2.1 座位预约功能

在入馆前对图书馆座位进行查询、选座修改或次日

预约。<sup>[5]</sup>座位预约的同时也包括了入座时间和离座时间的选择。这一功能的实现能够从根源上对于图书馆人流进行控制,也在很大程度上解决的图书馆占座的问题。

### 4.2.2 签到功能

这个功能的实现为人员入馆带来了很大的便利。区别于传统的刷卡入馆,APP 签到具有强大的灵活性,也解决了没有卡无法入馆的问题。

### 4.2.3 临时离座功能

当馆内读者需要短时间内离开座位,如小憩、就餐时,为了避免占座及临时离开座位被清理等问题,临时离座功能提供了一个很好的支撑。

### 4.2.4 释坐功能

该功能可以分为两方面实现,一方面是到点系统自动释坐,另一个就是手动释坐。当预约座位时间结束时,若没有进行操作,系统就会自动释坐<sup>[6]</sup>;当未到预约时间却离开时,为了避免资源浪费和非自愿占座,就可以手动在 APP 上进行释坐。

## 5.对图书馆数据的挖掘

研究图书馆数据之间的潜在联系,对图书馆的书目利用率的提高有重要的意义。关联分析大致来说包括数据准备、数据预处理、关联规则的深入挖掘、关联分析的具体规则、关联分析结果的深度分析以及提出专业性的建议。<sup>[7]</sup>

### 5.1 准备图书馆数据

为了更直观的展示大数据技术在图书馆智慧化过程中的运用,使用了 SPSS Modeler 软件对图书馆数据进行了一个初步的关联规则的挖掘。图 1 为初始数据。

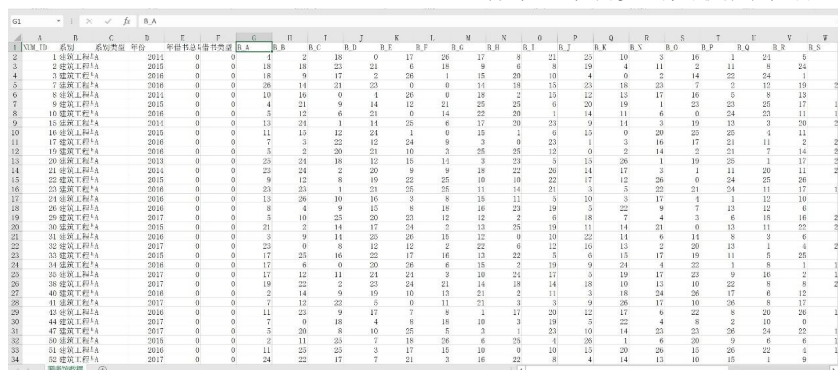


图 1 图书馆初始数据



### 5.2 对数据进行预处理

将数据进行预处理，并筛选出有效数据进行查看，如图 2、图 3 所示，分别为清洗过的图书馆数据的数据视图和变量视图。

图 2 数据视图

图 3 变量视图

### 5.3 数据的关联分析操作

这里通过 SPSS Modeler 软件来使用 Apriori 算法进行关联规则的数据分析与挖掘<sup>[8]</sup>。这里拿读者 ID 与借书类型 (A~Z) 之间做一个简单的关联分析，分析某学生借书种类之间的关系，以及在哪些数据之间已存在的关联之上为该学生推荐其他种类的书目。

首先将已清洗过的数据导入 Modeler，然后进行挖掘处理，具体流程如图 4 所示。

最终选择数据可以通过输出一个表 (图 5) 来进行查看。因数据量过大，在此只对一小部分的读者借书类型进行分析，再转当中可以输出一个网络图 (图 6) 进行查看。最终得到的关联规则如图 7 所示。

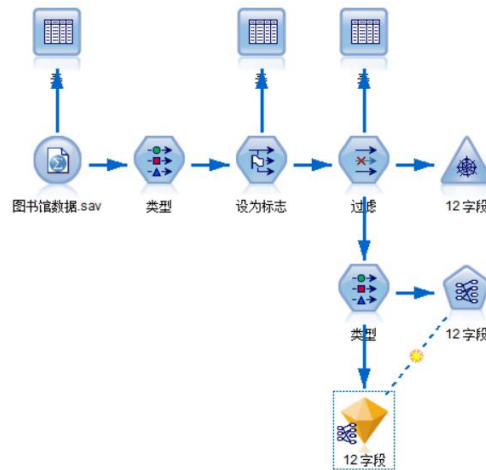


图 4 关联分析下 Apriori 算法具体流程

读者ID	系别	系列类型	A	B	C	D	E	F	G	H	I	J	K	N	O	P	Q	R	S	T	U	V	X	Z	年平均成绩	年级	年派卡消费
1	1	建筑工程与管理系 A	20	17	13	13	14	22	3	14	1	20	10	22	26	14	14	21	18	11	7	21	25	13	65.14	3	2100.00
2	2	建筑工程与管理系 A	19	24	20	17	9	4	20	4	18	9	15	13	0	10	19	17	23	4	7	18	15	3	62.16	3	720.00
3	3	建筑工程与管理系 A	2	1	2	13	18	26	23	5	0	3	20	19	23	24	15	8	7	7	4	18	7	4	0.00	4	0.00
4	7	建筑工程与管理系 A	7	15	13	2	23	20	19	14	6	7	14	26	10	10	24	24	17	9	4	2	21	50.00	4	0.00	
5	8	建筑工程与管理系 A	0	13	23	25	14	5	23	11	12	20	15	6	8	1	17	5	18	0	16	2	12	15	73.54	3	1890.00
6	9	建筑工程与管理系 A	20	4	11	1	4	16	0	3	1	26	18	19	10	8	6	16	14	19	23	17	10	2	14.45	4	1900.00
7	10	建筑工程与管理系 A	4	7	9	21	3	26	10	24	11	11	2	11	17	11	1	6	8	5	15	14	9	18	11.00	4	0.00
8	15	建筑工程与管理系 A	19	5	4	2	24	6	3	4	5	19	20	22	22	10	25	23	25	1	7	25	4	2	76.60	2	1720.00
9	16	建筑工程与管理系 A	0	5	8	5	26	12	9	16	26	2	21	16	3	15	9	20	4	1	12	5	25	23	13.36	4	1650.00
10	17	建筑工程与管理系 A	4	5	23	21	6	20	0	6	12	22	11	2	10	15	25	24	13	24	23	17	12	7	6.31	4	0.00
11	19	建筑工程与管理系 A	17	16	14	5	19	18	15	22	18	10	14	18	14	21	19	10	12	7	13	25	3	0	0.00	4	2510.00
12	20	建筑工程与管理系 A	10	4	6	19	16	8	22	25	5	6	1	10	17	13	21	2	26	22	0	20	3	1	72.27	3	0.00
13	21	建筑工程与管理系 A	23	26	21	10	23	19	26	10	10	5	3	20	0	23	26	6	3	0	15	21	8	17	0.00	4	1700.00
14	22	建筑工程与管理系 A	10	10	21	5	21	10	7	20	13	26	17	3	16	4	22	25	6	9	14	24	12	4	5.83	4	0.00
15	23	建筑工程与管理系 A	17	17	1	20	13	25	20	17	7	14	20	3	3	22	10	26	19	24	3	22	20	0	50.00	4	0.00
16	24	建筑工程与管理系 A	24	9	25	16	20	0	24	7	26	13	14	20	1	13	1	10	3	13	13	19	9	9	8.00	4	0.00
17	26	建筑工程与管理系 A	22	23	25	6	15	6	10	13	7	16	21	4	5	4	2	19	14	18	7	17	20	18	71.80	3	5073.00
18	29	建筑工程与管理系 A	13	15	12	4	19	2	19	16	14	20	1	12	14	10	10	23	22	22	17	16	23	12	0.00	4	200.00
19	30	建筑工程与管理系 A	23	3	3	3	4	6	17	13	23	6	14	1	1	18	11	21	9	11	7	26	19	5	70.96	3	2055.00
20	31	建筑工程与管理系 A	3	4	19	2	17	4	26	21	17	16	2	9	25	12	19	1	19	14	1	20	26	22	58.06	4	4927.00
21	32	建筑工程与管理系 A	11	17	18	16	4	3	17	9	2	2	2	16	21	1	5	17	1	5	16	3	9	4	0.00	4	100.00
22	33	建筑工程与管理系 A	17	26	5	1	17	18	25	15	26	0	9	1	5	9	9	11	22	17	8	9	11	77.17	2	1645.00	
23	34	建筑工程与管理系 A	19	21	14	7	22	21	14	9	0	8	26	22	15	16	0	9	5	14	6	25	15	13	78.20	2	1800.00
24	35	建筑工程与管理系 A	17	9	5	15	9	12	15	18	25	23	5	19	6	5	20	21	4	22	7	5	4	12	0.00	4	280.00
25	38	建筑工程与管理系 A	4	26	14	23	6	16	23	19	4	24	9	13	21	26	13	2	0	6	8	0	23	18	0.00	4	135.00
26	40	建筑工程与管理系 A	24	23	9	19	23	4	14	14	1	12	19	4	2	12	4	10	4	24	24	17	22	6	67.00	3	3820.00
27	41	建筑工程与管理系 A	1	25	19	8	16	19	15	8	18	7	20	2	20	22	8	6	18	20	23	6	26	5	0.00	4	0.00
28	43	建筑工程与管理系 A	16	19	12	25	25	12	25	10	25	1	6	20	23	6	5	6	0	23	11	9	6	18	60.25	3	5650.00
29	44	建筑工程与管理系 A	0	21	10	12	10	20	25	4	5	24	26	9	21	3	23	2	4	21	13	14	16	0	0.00	4	500.00
30	47	建筑工程与管理系 A	13	4	26	11	23	6	20	22	13	10	22	3	19	10	25	10	26	22	11	16	6	4	0.00	4	0.00
31	50	建筑工程与管理系 A	15	16	1	25	14	25	5	1	2	17	25	0	6	11	7	11	16	0	8	10	9	2	79.62	2	2080.00
32	51	建筑工程与管理系 A	19	19	1	17	11	3	7	23	9	23	19	11	25	26	14	9	20	23	9	3	5	24	67.31	3	4630.00
33	52	建筑工程与管理系 A	18	22	23	11	14	22	16	6	16	3	9	16	16	18	7	25	24	20	0	12	20	3	0.00	4	300.00
34	55	建筑工程与管理系 A	12	23	24	24	5	6	25	22	3	23	17	14	26	22	23	9	11	20	14	5	23	4	0.00	4	50.00
35	58	建筑工程与管理系 A	25	19	4	2	19	6	24	21	11	9	15	11	20	25	20	19	8	26	24	6	16	16	0.00	4	40.00
36	61	建筑工程与管理系 A	9	24	22	23	21	6	2	14	4	1	12	15	6	14	19	9	9	6	9	12	16	7	0.00	4	100.00
37	66	建筑工程与管理系 A	3	25	10	26	15	1	11	16	15	19	10	20	8	13	5	3	3	17	22	7	23	23	0.00	4	300.00
38	68	建筑工程与管理系 A	15	23	6	17	3	9	10	17	1	24	21	19	22	12	6	1	9	21	6	4	11	6	72.13	3	1350.00
39	71	建筑工程与管理系 A	23	11	21	3	26	3	22	4	12	26	25	13	21	1	14	8	24	11	21	20	9	19	0.00	4	0.00
40	74	建筑工程与管理系 A	14	4	9	22	13	1	4	12	4	12	1	10	18	10	23	12	2	23	14	26	13	14	0.00	4	0.00
41	77	建筑工程与管理系 A	2	23	11	7	20	4	14	24	7	3	24	13	10	12	24	24	7	16	23	2	8	17	0.00	4	100.00
42	80	建筑工程与管理系 A	14	11	13	0	3	19	8	23	19	19	23	7	2	13	3	6	18	20	4	22	15	11	0.00	4	100.00

图 5 最终数据表

### 5.4 对数据挖掘结果进行分析

由图 6 可以清晰的看到，在其中有很多粗细不一的线。线越粗，说明他们之间的联系关系就越强。所以我们可以

从右边的汇总数据中看出，A 类书籍和 B 类书籍直接的联系最强，这意味着在图书馆中，借了 A 类书籍的读者往往会借 B 类书籍的可能性最大，后面的数据以此类推。

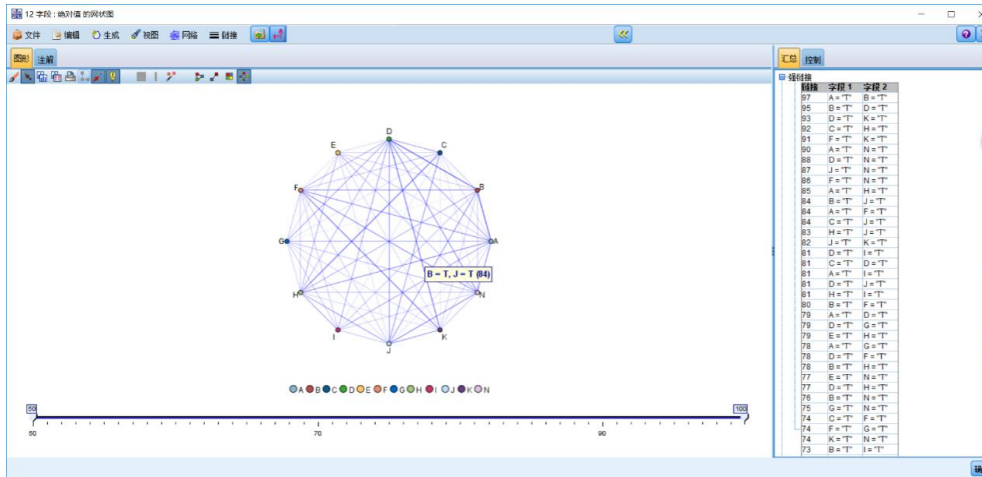


图 6 网络图

由图 7 的数据我们也可以很清晰的得出一些结论。以规则 ID1 为例，当前项为 I 和 N 时，后项为 A，用语言来解释就是当读者借阅了 I 类书籍和 N 类书籍时，他有很

大的可能性会借阅 A 类书籍，图中其他关联性数据同理。

后项	前项	规则 ID	支持度 %	置信度 %
A	I N	1	0.13	11.429
D	C G	2	0.121	12.308
C	G D	3	0.147	10.127
J	K A	4	0.128	10.145
K	J A	5	0.117	11.111
N	G B	6	0.11	11.864
B	I E G	7	0.002	100.0
G	I E B	8	0.002	100.0
E	I G B	9	0.004	50.0
I	E G B	10	0.007	25.0
B	I C K	11	0.002	100.0
K	I			

图 7 借书类型之间的关联关系

该数据挖掘反映出来的关联性为图书馆及读者带来了很大的好处。当读者借阅某种类型的书籍时，图书馆，或者说是智能化后的图书馆 APP 等就可以为该读者推荐与他借阅的书籍具有相关关联性的书籍，实现更加个性化的服务。

## 6.结束语

“智慧社会”是当今社会创新化建设的重要发展目

标之一。在以云计算、大数据等信息技术为基石的基础之上，智慧型建设还有很长的路要走，智慧图书馆作为智慧国家、智慧城市、智慧社会中重要的一员也要紧跟步伐快速发展，形成且完善自己的管理系统平台。在不断探索完善的过程之中，还有很多的难题需要进一步的研究。

## 参考文献

- [1]侯晓凌. 浅谈数据挖掘[J]. 科学之友: 中, 2009(4):140-141.
- [2]刘伟. 新兴 IT 技术背景下的信息管理与系统专业人才培养方案改革研究[J]. 信息技术与信息化, 2016(3):65-67.
- [3] 黄凯悦. 我国智慧图书馆的构建研究[D]. 2015.
- [4] 张红权. 试论门禁系统在高校图书馆安防系统中的作用[J]. 课程教育研究: 新教师教学, 2013(3).
- [5]张君一. 图书馆座位精细化管理系统[J]. 华人时刊旬刊, 2013(8).
- [6]胡华辉. 论屋面钢结构焊接变形控制措施[J]. 商品与质量, 2017(29).

- [7]黄超. 基于数据挖掘的移动图书馆研究与实现[D].
- [8]高洁. 基于数据挖掘技术的高校信息采集分析系统设计开发[D]. 电子科技大学.

### 作者简介

第一作者: 胡文彦 (1998-), 女, 汉, 安徽省合肥市, 本科, 四川大学锦城学院, 研究方向: 信息管理与信息系统、大数据应用与商业智能。

第二作者 (通讯作者): 鲍正德 (1989-), 男, 汉, 黑龙江省哈尔滨市, 研究生, 四川大学锦城学院, 研究方向: 电子商务。

第三作者: 唐娅雯 (1999-), 女, 汉, 四川省资阳市, 本科, 四川大学锦城学院, 研究方向: 信息管理、J2EE