改进贝叶斯的语义推送算法设计

朱 睿,冯锡炜,窦予梓,高天铸,马 蕾,吴衍兵
(辽宁石油化工大学 计算机与通信工程学院,辽宁 抚顺 113001)

摘 要:教育信息语义本体构建是通过语义本体构建方式去设计教育信息本体库。本体间逻辑关系表示方法,是构建出有逻辑结构的教育信息集合的过程。实现教育信息的半结构化数据归类,对不同时间采集的归类数据在规定好的模型中进行计算—词汇频度分析模型。词汇频度分析模型运用逆概率的贝叶斯思想,经过对传统贝叶斯算法与语义本体性质相结合,使 MapReduce 善于处理半结构化数据;经过对语义本体构建的教育信息数据结合词汇频度分析模型进行计算,获得教育信息本体的推荐能力值 E_i ;通过对不同本体 E_i 值进行排序,获得了推荐信息的顺序;根据推荐权重进行信息的推送工作,同时根据 JS 指数,经过比较基于词汇频度分析模型与目录结构推送算法的分析结果得出:词汇频度分析模型优于基于目录结构推送算法。

关键词:语义本体;信息推送;词汇频度分析模型;教育信息化
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2020)03-0104-07
doi:10.3969/j.issn.1673-629X.2020.03.020

Design of Semantic Push Algorithm Based on Bayesian

ZHU Rui,FENG Xi-wei,DOU Yu-zi,GAO Tian-zhu,MA Lei,WU Yan-bing
(School of Computer and Communication Engineering,Liaoning Shihua University,Fushun 113001,China)

Abstract:The construction of educational information semantic ontology is to design educational information ontology database through semantic ontology construction. The representation method of logical relation between ontologies is the process of constructing the set of educational information with logical structure. The semi-structured data classification of educational information is realized,and the classification data collected at different time are calculated in a well-defined model—lexical frequency analysis model. The Bayesian idea of inverse probability is introduced in the lexical frequency analysis model. The combining of the traditional Bayesian algorithm with the semantic ontology property makes MapReduce deal with semi-structured data well. After calculating the educational information data based on semantic ontology and vocabulary frequency analysis model,the recommendation ability value (E_i) of educational information ontology is obtained. By sorting different ontology E_i values,the order of recommendation information is obtained. The information is pushed by the recommendation weight. According to JS index,by comparing the analysis results based on lexical frequency analysis model and directory structure push algorithm,it is concluded that the lexical frequency analysis model is superior to the push algorithm based on directory structure.

Key words:semantic ontology;information push;lexical frequency analysis model;educational informatization

0 引 言

教育信息化越来越受到教育研究者的关注,随着各类学科的电子化,人们访问这类网站所产生的浏览数据量越来越大。通过大数据技术,对这些浏览数据进行分析后,可以根据每个用户群体不同的浏览数据习惯进行相关教育方面的信息推送[1-3]。

专业化教育资源本体库的建立对于教学信息资源的推送有着不寻常的实践价值[4-5]。在 2017 年教育部发布了《基础教育教学资源元数据》系列教育行业标准通知,里面包括了《基础教育教学资源元数据 信息模型》、《基础教育教学资源元数据 XML 绑定》及《基础教育教学资源元数据 实践指南》,这些标准对于建立相关教育信息化本体有着非常重要的意义。

文中利用 Protégé,以计算机组成原理这一课程内容为本体设计数据来源,进行本体设计。基于百度指数中关于计算机组成原理的各项搜索数据,基于贝叶

收稿日期:2019-03-03 **修回日期**:2019-07-05 **网络出版时间**:2019-11-07
基金项目:辽宁省自然科学基金(20180550130)
作者简介:朱 睿(1993-),男,硕士研究生,研究方向为数据挖掘;冯锡炜,博士,教授,研究方向为分布式计算。
网络出版地址:http://kns.cnki.net/kcms/detail/61.1450.TP.20191107.0910.034.html

斯建立词汇频度分析模型,将百度指数中的搜索指数结合词汇频度分析模型进行计算,计算后的各个不同本体的词汇频度分析数据按照数值的从大到小进行推送。

1　教育信息化本体构建

1.1　教育信息化

教育信息化具有两层含义,一个在教育目标中加入信息素养,另一层指在教学与科研中加入信息技术手段,注重教育信息资源的探究与使用[6]。文中主要对后者进行阐述。在信息技术手段上利用大数据、语义分析及用户粘性等信息技术对教育工作者常进行浏览的网页记录进行分析,进而进行推送[7-8]。

1.2　教育资源本体

教育资源本体用来容纳和规范教育信息,根据实际的需求,将本体的属性分为数据属性(Data Property)和对象属性(Object Property)。数据属性定义域是本体的类,值域是数据类型。对象属性是表示所有个体之间的关系属性[9]。

数据属性:为使网络上分布的教育资源库有统一的语义标注标准,通过对《基础教育教学资源元数据》的每一个元数据项进行分析,然后整理出了数据属性。部分数据属性的定义与说明如表 1 所示[10-11]。

表 1　数据属性定义及说明

数据属性	定义域	说明
catalog	Resource	所属标识方案或编目方案的名称或指示符。一种命名方案
entry	Resource	在标识或编目方案中用于标识此学习对象的标识符一个与名域相关的字符串
title	Resource	所描述的教学资源的名称
proper title	Resource	对教学资源内容的揭示具有关键意义的主要名称
alternative title	Resource	正式标题之外的其他名称或替代写法
language	Resource	资源知识内容所使用的语言类型
description	Resource	以文本方式对资源内容的简介
keyword	Resource	用以描述资源主要内容的关键词语
code	Resource	描述该资源与教材内容框架的对应关系
special subject	Resource	在实施教育教学的过程中,专门研究或讨论的题目
coverage	Resource	资源所涉及的时间、文化和地理区域。资源内容的范围和广度,覆盖主要包括空间位置、时间段
format	Resource	资源在技术上的数据类型;该数据元素用于确定资源所需的运行软件
technical	Resource	该类别描述了资源的技术要求及其相关特征
size	Resource	数字化资源的大小,用十进制数字“0”到“9”表示,单位是字节(每字节 8 位),不是兆字节等;该数据元素表明了资源的实际大小,如果资源经过压缩,则该数据元素的值是未压缩时的大小
requirement	Resource	使用资源所需要的技术条件,如:硬件、软件、网络等
role	Resource	贡献的类型(注:至少应该描述资源的作者)
location	Resource	用于表明如何获取资源的字符串。它可能是一个位置(如:URL),或解析出位置的一种方法(如:URI)。最可取的位置优先列出
educational	Resource	该类别描述了资源在基础教育和教学方面的一些关键特征
learning resource type	Resource	描述该资源的一般范畴、功能、种属或聚类层次,越主要的类型越先列出
learning mode	Resource	该资源所适用的学习行为,体现学生在自主性、探究性和合作性方面的基本特征
Applicabili	Resource	该资源所适应的范围

对象属性:根据教育元数据进行教育资源领域的本体构建。主要对象属性是教育信息的对象属性[12]。教育资源之间存在丰富的语义关系,通过语义关系建立本体属性,利用这些属性进行本体推理和查询,作为教育资源语义搜索的基础[6]。

教育资源间属性关系,可根据教育信息的特点,对教育信息间关系进行分析抽象,得到表 2 所示的对象属性及对应公理。

其中对象属性的公理,为从离散数学当中借鉴过来的三种关系性质,分别是 Transitive(传递性)、Asymmetric(非对称性)和 Reflexive(自反性),具体对象属性及对应公理如表 2 所示[13]。

表 2 对象属性及对应公理

对象属性	数据间意义	定义域/值域	公理
contain	包含	Conept/Concept	Transitive/Asymmetric/Reflexive
belongTo	属于	Conept/Concept	Transitive/Asymmetric/Reflexive
isneighborOf	相邻	Conept/Concept	Transitive/Asymmetric/Reflexive
isparallelOf	平行	Conept/Concept	Transitive/Asymmetric
isprecursorOf	前驱	Conept/Concept	Transitive/Asymmetric/Reflexive
isrearOf	后继	Conept/Concept	Transitive/Asymmetric/Reflexive
isreferenceOf	参照	Conept/Concept	Asymmetric
issynonymyOf	同义	Conept/Concept	Transitive/Asymmetric

1.3 教育资源本体

利用 Protégé 进行计算机组成原理这一课程体系及相关知识的本体构建。层级关系采用目前本科计算机类学生教学常用的《计算机组成原理》中对计算机组成的分类方式作为分类标准,主题上分四个大块,分别是概论、计算机系统的硬件结构、中央处理器、控制单元。采用树状方式进行存储,深度为 4 层。图 1 和图 2 分别是在 Protégé 进行本体构建的结构图和可视化界面图。Protégé 会生成对应的 owl 及 xml 文件,可以方便在 Hadoop 中进行相关处理工作。

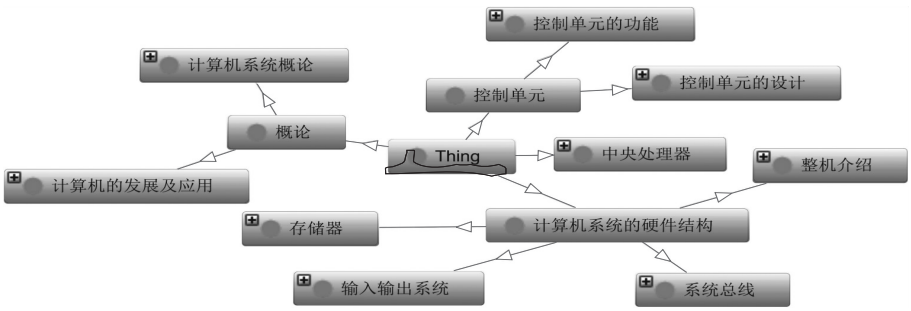

图 1 Protégé 本体之间结构关系简图

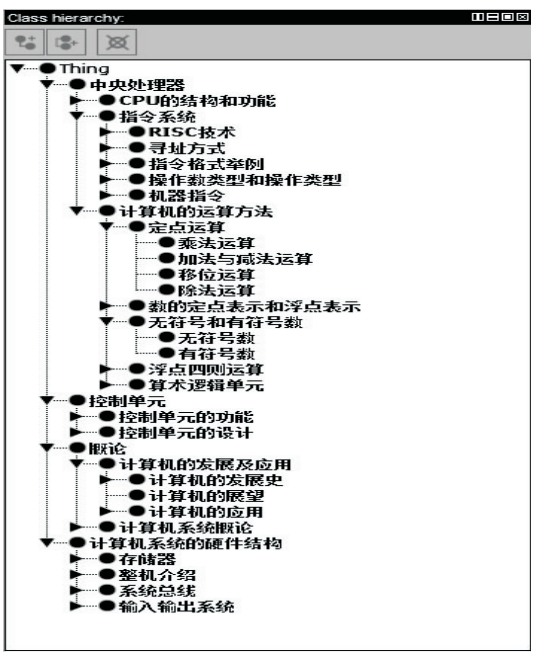

图 2 Protégé 本体之间可视化界面部分展开

2 教育信息化本体构建

教育信息之间的语义关系可以制定丰富的自定义推理规则[14-15]。这里假设 a、b 为教育信息,p、q 表示属性,p 具有传递性,p 和 q 互逆:

传递性规则:(? a p ? b)(? b p ? c)->(? a p ? c)

如果教育信息 a 和 b 之间具有属性 p,教育信息 b 和 c 之间也具有属性 p,属性 p 具有传递性,则可以推理得到教育信息 a 与 c 之间也具有属性 p。

互逆规则:(? a p ? b)->(? b q ? a)

如果教育信息 a 与 b 之间具有属性 p,由于属性 p 和 q 互逆,则可以推理得到教育信息 b 和教育信息 a 之间具有属性 q。

这里以计算机组成原理中的知识点为例,利用表 2 对象属性及对应公理中对象属性结合传递性或互逆规则,用 JSJZC 表示计算机组成原理的知识点作,在表 3 中写出为推理规则。

表 3 教育信息本体间逻辑

推理规则	描述
(? a JSJZC:contain ? b)(? b JSJZC:contain ? c)->(? a JSJZC:contain? c)	如果教学信息 a 和 b 是包含关系,教学信息 b 和 c 是包含关系,则教学信息 a 和 c 也是包含关系

续表 3

推理规则	描述
(? a JSJZC:isbelongTo ? b)(? b JSJZC:isbelongTo ? c)->(? a JSJZC: isbelongTo? c)	如果教学信息 a 和 b 是属于关系,教学信息 b 和 c 是属于关系,则教学信息 a 和 c 也是属于关系
(? a JSJZC: issynonymyOf ? b)->(? b JSJZC: issynonymyOf? a)	如果教学信息 a 和 b 是同义关系,则教学信息 b 和 a 是同义关系
(? a JSJZC: isparallelOf ? b)(? b JSJZC: isparallelOf ? c)->(? a JSJZC: isparallelOf ? c)	如果教学信息 a 和 b 是平行关系,教学信息 b 和 c 是平行关系,则教学信息 a 和 c 也是平行关系
(? a JSJZC: isneighborOf ? b)->(? b JSJZC: isneighborOf? a)	如果教学信息 a 和 b 是相邻关系,则教学信息 b 和 a 是相邻关系
(? a JSJZC: isprecursorOf ? b)->(? b JSJZC: isrearOf? a)	如果教学信息 a 和 b 是前驱关系,则教学信息 b 和 a 是后继关系
(? a JSJZC: isrearOf ? b)->(? b JSJZC: isprecursorOf? a)	如果教学信息 a 和 b 是后继关系,则教学信息 b 和 a 是前驱关系
(? a JSJZC: isreferenceOf ? b)(? b JSJZC: isreferenceOf ? c)->(? a JSJZC: isreferenceOf? c)	如果教学信息 a 和 b 是参照关系,教学信息 b 和 c 是参照关系,则教学信息 a 和 c 也是参照关系

属性约束,OWL 使用属性约束来描述那些特定类的属性条件,属性条件的基数约束如表 4 所示[11]。

表 4　属性条件约束规则

约束	解释
Owl:minCardinality	至少有 N 个属性
Owl:maxCardinality	至多有 N 个属性
Owl:Cardinality	恰好为 N 个属性

3　词汇频度分析模型

本体构建只是将零散的教育信息进行半结构化的数据构建过程,而词汇频度分析模型是将这类数据进行处理的模型。Hadoop 作为一个分布式计算基本框架,在对大数据进行分布式计算的过程中,需要对数据进行整理和规划,而作为 Apache 公司推出的 MapReduce 可以在大数据以及半非结构化的概况下进行数据处理[16-17]。教育信息数据具有半非结构化,需要通过本体构建的方式构建起一个相对的结构体系,所以通过对 MapReduce 和 Hadoop 进行配合,进行相关的数据计算,能更好地对数据进行处理。

而词汇频度分析模型 MapReduce 对教育信息资源进行管理,词汇频度分析模型的处理和表示是分类器构建的一个重要过程[18]。词汇频度分析研究的是对教育信息资源进行推送的相关算法,在前面已经基于本体进行个元数据的分类及结构构建工作,但只有结构无法进行相应的推送工作,因为对于元数据来说,每个元数据在推送过程中都具有相同的推送价值[19-20]。为了更好的进行相关信息资源的推送,文中在基于语义构建元数据的基础上加入了基于改良后的贝叶斯概率统计计算公式。贝叶斯概率统计计算公式相较于传统的频数概率统计方式有所不同,其概率统计会保留不确定性[7]。

$$P(A \mid X) = \frac{P(X \mid A)P(A)}{P(X)} \tag{1}$$

这与推送内容的目标用户对于推送内容的不确定性恰好吻合,而传统的贝叶斯公式如式(1)所示,其中 $P(A)$ 代表 A 发生的概率,其概率值在[0,1], X 代表在 A 之后进行测试的实验[7]。这个公式代表的含义是在已知 $P(A)$ (在推送中最开始的 $P(A)$ 可来自该行业专家的初始定值或小范围内的问卷调查赋值初始概率)的情况下,每次新的变化会让概率在[0,1]之间不停的变化。当中需要对每个教育本体进行附加属性,通过这些附加属性进行词汇频度分析模型的构造。文中采用词汇频度分析模型来对各个标题进行赋值,从而在进行推送的过程中可以更加准确地进行相关信息的推送工作[21]。

$$W = \{w_1, w_2, \cdots, w_n\} \tag{2}$$

$$w_i = \{\text{name}, \text{depth}\}, i \in [1, n] \tag{3}$$

式(2)中的 W 代表本体库,式(3)中的 w_i 为本体库中的本体,每个本体 w_i 含本体名称和在本体库中的本体层数,规定根节点(在文中是计算机组成)层数为 1,其中下角标 i 代表每个本体的标号, n 代表本体库中最大本体数目。

$$h_{ij} = \{h_{11}, h_{12}, \cdots, h_{1m}, h_{21}, \cdots, h_{2m}, \cdots, h_{f1}, \cdots, h_{fm}\} \tag{4}$$

式(4)中 h_{ij} 是各个本体词汇在不同日期下的热度值,其中 t 代表日期,最大日期值为 f , j 代表所对应本体的标号。w_i 通过记录的字段 name 与 h_{ij} 在代表本体进行互相映射。

$$P(w_i) = \frac{1}{\sum_{j=0}^{n} \text{dep}(w_i, w_j)} \tag{5}$$

式(5)为预先处理数据,根据已构建的本体库,其存在层级关系,层级越低,其概括越大。而层级越高,其内容越细。计算在本体库中与 w_i 具有较强连接度的本体数据的比例关系,进而得出与整体的关系。$P(w_i)$ 代表的是每个本体与整体的连接概括关系,而

dep(w_i , w_j)表示两个本体间的层级的距离,如果两个本体间越相近,其 dep( )值越小, $P(w_i)$ 越大,其本体 w_i 与其他本体的链接越紧密,在推送的时候更应该连带进行推送。此部分对应传统贝叶斯公式中的 $P(A)$ 部分。

$$P(h_{tj}) = \frac{h_{tj}}{\sum_{T=1}^{f} h_{Tj}} \tag{6}$$

式(6)为计算在对应本体的热度值,该热度值的来源为各大搜索引擎的热词搜索数据(文中采用的是百度指数中的相关数据)。这里设每日该本体对应的热词比例公式为 $P(h_{tj})$, h_{tj} 和 h_{Tj} 分别为获取当前词汇的热度数和不同天数下的该本体的热度数。$P(h_{tj})$ 值越高,代表 h_{tj} 在用户搜索中占有较重要的意义, h_{tj} 所对应的本体 w_i 的推送排名应该上升。$P(h_{tj})$ 值越低,代表 h_{tj} 在推荐中应该进行排名下滑。

$$P(w_i \mid h_{tj}) = \frac{P(h_{tj} \mid w_i)P(w_i)}{P(h_{tj})} \tag{7}$$

通过组合式(5)及式(6),可以得到简单的基于贝叶斯模型,如式(7)所示。但贝叶斯在进行统计概率的情况下,其容易受到单次数据较大波动导致统计概率发生较大的起伏,所以,文中在结合本体与贝叶斯统计概率公式的同时加入频数统计概率。

$$P(w_i, h_{tj}) = \frac{P(h_{tj}) + \alpha}{\sum_{T=1}^{f} P(h_{Tj}) + \beta P(w_i)} \tag{8}$$

式(8)为当有相应的用户搜索数据后,计算用户搜索频度值。式(8)中的 α 值为加权自定义值,默认情况下为 0。而 β 为本体加权变量,默认情况下为 1。如果需要特殊优先推送,增加该本体的 β 值或者调节 α 值,增加 β 值可以对本体进行正向加权,让 $P(w_i, h_{tj})$ 增加,而调节 α 即避免当前该信息过新无人查看 $P(h_{tj}) = 0$ 这种情况。

$$E_i = \frac{P(w_i)P(w_i, h_{tj})}{P(h_{tj})} \tag{9}$$

式(9)得到每个本体 W_i 所对应的粘性能量值 E_i ,推送系统根据 E_i 值进行相关内容的推送。E_i 值越大,其推送排名越靠前; E_i 值越小,其推送排名越靠后。

4　语义推送结果分析

基于逆概率的贝叶斯算法设计出的词汇频度分析模型和通过语义本体构建规则构建起来的半结构化教育信息本体结构,进行用户粘性模型教育信息推荐系统[22-23]的设计工作。通过在百度指数上的相关数据结合专业书籍及相关专业老师的意见,计算 $P(w_i)$ 及相关公式的结果 E_i ,得到计算机组成原理排名前 10 的词汇。教育信息推荐系统推送结果如表 5 所示。

表 5　教育信息推荐系统推送结果

序号	推送内容	E_i /热度
1	计算机组成	477.85
2	中央处理器	270.93
3	冯·诺依曼	256.55
4	i/o	170.71
5	存储器	162.72
6	机器码	150.49
7	risc	119.77
8	整机	74.24
9	指令周期	54.44
10	中断系统	18.41

图 3 是用 MATLAB 生成的教育信息推荐系统推送结果展示图,以搜索计算机组成一词例推送出的相关信息,点的大小代表访问量,距离的远近代表相关性层级的远近。图 4 是 MATLAB 生成的按书目录一级标题推送结果。从图 3 和图 4 对比中可以大致看出,基于按一级目录进行推送的结果在大多情况下不如教育信息推荐系统的推送结果,教育信息推荐系统的推送结果具有信息量大,相关信息多的特点。

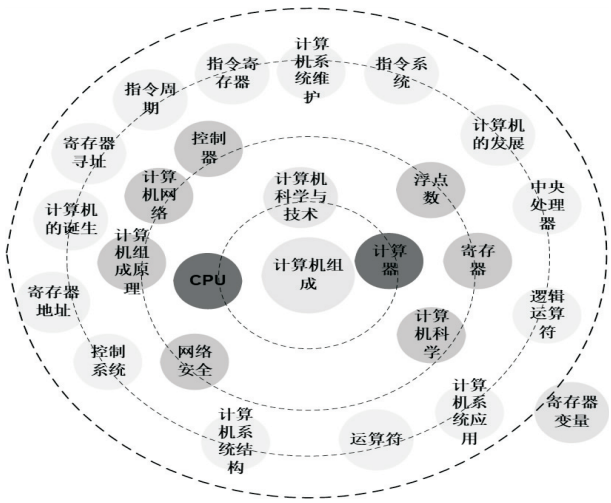

图 3　教育信息推荐系统推送结果展示图

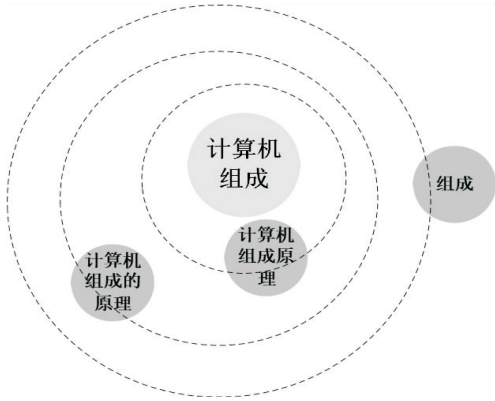

图 4　按书目录一级标题推送结果图

对于表 5 当中的教育信息推荐系统的推送结果,选取了 100 名相关计算机专业的学生,通过给他们推送基于词汇频度分析模型及按书目录一级标题排列进行推荐可靠度打分,让其判断需要程度的排序,得出如图 5 所示的百名用户满意度记录。从图中可以大致看出,基于按一级目录进行推送的结果在百名用户中大多情况下不如教育信息推荐系统的推送结果。

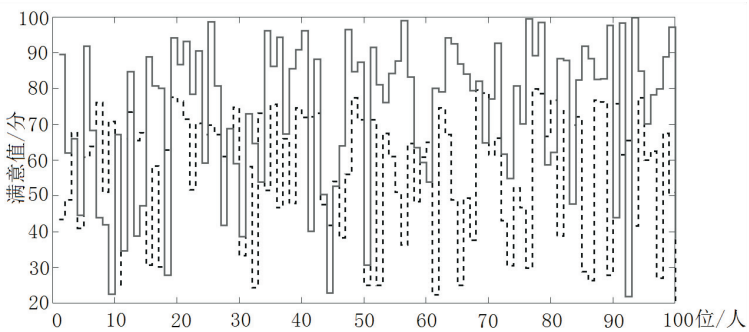

图 5　百名用户满意值记录

数值判断方面,利用 Jaccard Index(简称 JS 指数)进行用户对推送结果排序的符合程度计算。式(10)为 JS 指数计算方式,其中 A 为推送结果,B 为用户希望推送结果。$J(A,B)$ 为 JS 指数计算结果,当 JS 指数大于 0.70 时为优秀,大于 0.50 时为良好,低于 0.25 时,该系统不利于进行推送。

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|}=\frac{|A\cap B|}{|A|+|B|-|A\cap B|} \tag{10}$$

将表 5 当中的信息推荐系统表和按一级目录排列的结果同时让 100 名自愿用户(计算机专业学生)评判是否符合心理推送预期。并且利用式(10)进行计算。

根据图 5 中百名用户满意值记录,进行平均值计算,结果比较如表 6 所示。从表中可以看出,利用词汇频度分析模型结合语义本体分析后的推送系统 JS 平均指数达到了 0.73,达到了良好的标准,而根据一级目录进行推荐的推荐系统 JS 平均指数达到了 0.57,明显比基于用户粘性模型及语义本体分析后的 JS 平均指数低。

表 6　各类推荐算法比较表

推送方法	JS 平均指数计算结果
按一级目录推送	0.57
词汇频度分析模型推送	0.73

对于表 5 当中的教育信息推荐系统的推送结果,从多名自愿用户(计算机专业学生)的学生中选出 100 个计算机专业常见词汇,通过测试推送基于词汇频度分析模型及按书目录一级标题排列进行打分,能推送出准确的结果为 1,未能推送出结果的为 0,未能推送出准确结果但能推送出其泛词(相同或相关的词汇)的结果为 0.5。图 6 是 100 词汇测试结果记录图,其中实线代表教育信息推荐系统推送,虚线代表按照一级目录推送。

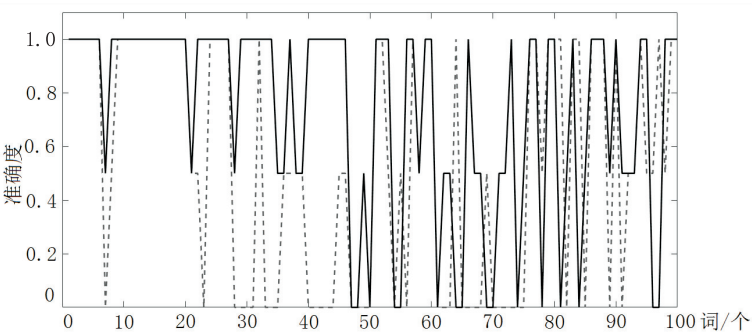

图 6　100 词汇测试结果

根据图 6,将图中数据进行推荐度计算(推荐结果累加总分/词汇总数),结果比较如表 7 所示。从表中可以看出,利用词汇频度分析模型结合语义本体分析后的推荐度分数达到了 0.73,达到了良好的标准,而根据一级目录进行推荐的推荐系统推荐度数仅仅达到了 0.535,显而易见,教育信息推荐系统的推送结果的准确性要远远高于按一级目录推送结果的准确性。

表 7　推荐算法比较

推送方法	推荐度	判定结果
按一级目录推送	0.535	差
词汇频度分析模型推送	0.73	良好

5 结束语

文中利用语义本体对教育信息进行本体构建,利用贝叶斯及频度统计概率的方式对构建的教育信息本体进行概率上的计算,得到每个本体的推送概率 E_i,根据 E_i 值的大小进行教育本体信息的推送工作。对推送的结果进行满意度判断,并且进行统计后,利用 JS 指数对该推送结果进行分析。

为了使推送的内容更加准确,从算法的实用性和健壮性出发,在教育信息研究领域当中应用改进贝叶斯算法设计的词汇频度分析模型,其推送结果的准确性和适应性优于基于目录结构推送算法,能够更加精确地对所服务的人群进行相应数据的推送工作。

参考文献:

[1] CHI N, JIN Y, HSIEH S. Developing base domain ontology from a reference collection to aid information retrieval[J]. Automation in Construction, 2019, 100: 180-189.

[2] NING Huansheng, SHI Feifei, ZHU Tao, et al. A novel ontology consistent with acknowledged standards in smart homes [J]. Computer Networks, 2019, 148: 101-107.

[3] LAJEVARDI A M, AMINI M. A semantic-based correlation approach for detecting hybrid and low-level APTs[J]. Future Generation Computer Systems, 2019, 96: 64-88.

[4] 刘媛媛. 基于本体的教育资源检索系统研究[J]. 电脑知识与技术, 2017, 13(35): 1-2.

[5] VELOUDIS S, PARASKAKIS I, PETSOS C, et al. Achieving security-by-design through ontology-driven attribute-based access control in cloud environments[J]. Future Generation Computer Systems, 2019, 93: 373-391.

[6] BOUIHI B, BAHAJ M. An UML to OWL based approach for extracting Moodle's ontology for social network analysis [J]. Procedia Computer Science, 2019, 148: 313-322.

[7] MODI K J, GARG S. A QoS-based approach for cloud-service matchmaking, selection and composition using the Semantic Web[J]. Journal of Systems and Information Technology, 2019, 21(1): 63-89.

[8] 巩利艳, 孙 力. 本体概述及其在教育领域的研究现状[J]. 中国教育技术装备, 2017(18): 9-13.

[9] DUARI S, BHATNAGAR V. sCAKE: semantic connectivity aware keyword extraction[J]. Information Sciences, 2019, 477: 100-117.

[10] 代晓宇. 基于本体的教学资源语义检索应用研究[D]. 哈尔滨: 哈尔滨工程大学, 2012.

[11] 王 燕, 孙秀英. 基于语义标注的文本聚类算法研究[J]. 科学技术与工程, 2012, 12(35): 9706-9709.

[12] 王 刚, 杨 波, 杨明杰. 云计算环境下分布式语义文本自适应分类方法[J]. 科学技术与工程, 2018, 18(7): 208-212.

[13] ANDREW S. Discrete mathematics by example paperback [M]. New York: McGraw-Hill Inc, 2001.

[14] 冯 瑶, 冯锡炜. 基于本体的教育资源推理查询原型系统设计与实现[J]. 计算机应用与软件, 2016, 33(10): 14-18.

[15] AHMAD A, CUOMO S, WU W, et al. Intelligent algorithms and standards for interoperability in Internet of Things[J]. Future Generation Computer Systems, 2019, 92: 1187-1191.

[16] 杨 鼎, 阳爱民. 一种基于情感词典和朴素贝叶斯的中文文本情感分类方法[J]. 计算机应用研究, 2010, 27(10): 3737-3739.

[17] UNGER L, FISHER A V. Rapid, experience-related changes in the organization of children's semantic knowledge[J]. Journal of Experimental Child Psychology, 2019, 179: 1-22.

[18] SAVAGE J, ROSENBLUETH D A, MATAMOROS M, et al. Semantic reasoning in service robots using expert systems [J]. Robotics and Autonomous Systems, 2019, 114: 77-92.

[19] CAMERON D P. Bayesian methods for hackers probabilistic programming and Bayesian inference[M]. New Jersey: Addison-Wesley Professional, 2015.

[20] 王 刚, 姜 山. 基于贝叶斯分类实现数字图书馆主动推送服务[J]. 中华医学图书情报杂志, 2011, 20(12): 54-56.

[21] 邓左祥, 涂 芳. 一种有效的多关系贝叶斯分类算法[J]. 微电子学与计算机, 2017, 34(7): 123-127.

[22] 王文相. 贝叶斯公式在数据挖掘中的应用[J]. 数学学习与研究: 教研版, 2017(13): 139.

[23] ALHAKBANI N, HASSAN M M, YKHLEF M, et al. An efficient event matching system for semantic smart data in the Internet of Things (IoT) environment[J]. Future Generation Computer Systems, 2019, 95: 163-174.