基于向心扩散加权 XML 模型的异构用户个性化模式匹配方法*

李树青　刘晓倩
(南京财经大学信息工程学院　南京 210046)

【摘要】介绍一种利用同文词语共现和引文词语共现分析实现的领域本体自动构建方法,该本体采用加权 XML 模型,利用概念联系中的权值设定可以有效地表达用户兴趣程度的差异,并利用基于向心扩散的扩散激活方法对用户兴趣特征及其联系提供更强的表达能力,以便于发现更有价值的潜在用户兴趣。进而介绍如何利用该本体按照"先打碎后重构"的策略将异构用户个性化模式转换为可以进行比较的一致模式,并对相关的异构用户个性化模式匹配方法做出详细说明。最后总结相关测试实验及其结果。

【关键词】领域本体　加权 XML　个性化　词语共现分析
【分类号】G202

The Matching Algorithm of Heterogeneous User Personalized Profile Based on Centripetal Spreading Weighted XML Model

Li Shuqing　Liu Xiaoqian
(College of Information Engineering, Nanjing University of Finance & Economics, Nanjing 210046, China)

【Abstract】This paper introduces an automatic construction method for domain Ontology implemented by words co - occurrence analysis in both document and citation. This Ontology adopts weighted XML model and uses weight in concepts and their relationship to express the difference of users' interest effectively, which can improve the ability of expressing users' interest and their relationship with centripetal weight spreading activation strategy in order to explore more valuable users' interest. Meantime, this paper also discusses how to use this Ontology to transform heterogeneous user personalized profile to consistent comparable model with the broken - and - reconstruction strategy, and how to match corresponding heterogeneous user personalized profile in detail. Finally, the result of correlative tests and experiments are concluded.

【Keywords】Domain Ontology　Weighted XML　Personalization　Words co - occurrence analysis

1　引　言

用户个性化模式是实现个性化信息推荐服务的基础,合理地选择和设计用户个性化模式对于提高信息推荐服务的有效性非常重要。目前,基于本体的用户个性化模式设计方法已经成为一种主流方式。笔者以前的研究采用了同构异值的加权个性化本体设计方法,利用本体中的权值差异来表征用户兴趣的差异[1]。这种方法简单易行,也便于人们理解。然而,它也存在很多问题,如不同的用户兴趣差异程度很大,同构异值的设计方法往往需

收稿日期:2012 - 04 - 24
收修改稿日期:2012 - 05 - 18
*本文系国家自然科学基金项目"基于通用加权 XML 模型的个性化用户兴趣本体研究"(项目编号:71103081)和江苏省高校自然科学研究面上资助项目"通用加权 XML 模型在便携式个性化用户兴趣本体中的表达方法研究"(项目编号:11KJB630001)的研究成果之一。

要在现有的用户个性化本体中引入大量的无关内容来进行补齐填充,从而形成一致的模式,因此带来大量的冗余计算,这一问题在用户个性化本体的权值扩散过程中表现尤为明显。异构设计方法可以给用户个性化本体带来一种新的解决思路,使其可以在结构上不受拘束地被扩展和自定义,以便更好地表达用户个性化兴趣特征。但其中的关键问题和难点并不在于此,如何实现对不同的异构用户个性化本体进行相似度比较,从而完成个性化信息推荐服务,才是亟需解决的重要问题。

造成用户个性化本体结构相异的因素主要有两个:本体节点语义不统一,不同用户所产生的个性化本体节点内容往往随意性很大,复杂的语言现象导致最终很难对不同的用户个性化本体进行比较和相似度判断;节点联系不统一,不同用户对组成自己兴趣特征的不同概念之间的关系,往往有不同的理解,比如有人采用层次型的组织方法,也有人采用网络型的组织方法,甚至还有人采用多维结构型的组织方法。即便是把概念联系限定在层次型结构中,相关类目和组织次序也会因人而异。

因此,要想解决异构用户个性化模式之间的相似度计算问题,就必须解决上述两个问题。

2　文献回顾

对于用户个性化模式而言,早期的表达方法往往采用关键词向量和类别向量[2]、概念集合[3]等方法,后来学者开始尝试使用本体来表达用户个性化模式[4,5]。如将传统的用户个性化模式中的关键词通过本体映射转换为概念词[6],完全使用本体来直接构建用户个性化模式[7],利用本体来构建细粒度用户个性化模式但却没有本体的自动构建方法[8]等。较新的研究提到了利用 WordNet 本体来实现“森林模型(Forest Model)”,并据此测度用户个性化模式之间的相似度,来实现社交网络中的朋友查询[9]。国内类似的研究使用本体来构建高校专家的集成信息,但是信息的收集方法却需要手工进行[10]。

对于这些异构用户个性化模式相似度的比较,相关方法的研究也由来已久。从总体上看,该问题属于信息异构问题。信息异构有三个层次,分别是句法(Syntax)、结构(Structure)和语义(Semantic)[11]。句法层次是其中最为简单的问题,主要原因在于数据格式的不一致,随着诸如 XML、RDF 和 OWL 等各种标准格式的推出,该问题已经逐渐得以解决。随之而来的主要是结构层次和语义层次的问题。其中,语义层次的异构问题仍没有取得较为明显的突破[12]。

借助本体工具和本体映射方法是一种解决信息异构问题的有效手段,在信息集成和信息转换领域中应用很广,但是面向于整个网络的通用本体并不存在,得到广泛使用的仍然以面向特定领域的领域本体为主。本文所采用的本体也是一种面向学术文献的领域本体。在使用不同本体时,本体映射方法是建立用户与服务之间联系的关键条件[13],它也是一种测度实体相似性的方法[14]。具体的本体映射方法有编辑距离匹配、语义匹配和结构匹配等,也有很多学者通过引入其他方法来解决,如人工智能技术等[15],还有学者尝试使用与协同推荐方法的结合来处理这一问题[16]。

然而,现有研究仍然存在很多问题:

(1)用户个性化本体表达能力亟需提高。为了实现这一目标,学者们引入了本体概念的加权表达方法和扩散激活(Spreading Activation)方法。

加权本体是一种有效的用户个性化模式解决方案。传统的个性化本体主要利用节点及其联系来表达各种语义关系,后续的研究发现,与兴趣权值的结合是一种有效的个性化本体构建方法[17]。它是一种“概念空间[18]”的发展和扩展,实践证明也是有效的[19]。它结合了本体方法和兴趣权值方法,有助于个性化本体的创建和进化。但现有的方法往往侧重于对概念加权,而忽略了对概念联系权值的考虑。

同时,在本体中引入语义网络中的扩散激活方法来增强对本体概念相似度的测度能力也是可行的[20]。扩散激活方法通过初始概念集合和相应的初始权值来寻找本体中的其他相关概念[21]。该方法可以较好地解决推荐系统中常见的冷启动问题,即使用现有的本体作为用户初始个性化模式,可以完成基本的概念扩展和关系识别[22]。复杂的方法还考虑了向上层概念的扩散或者对不同的概念联系采用不同的扩散方法等[23]。但是,相关的研究并不多见[24]。

在前期研究中,笔者也证实了引入加权表达和扩散激活方法的有效性,本文所采用的方法也沿用这一思路[25]。

(2)本体设计过于简单。由于缺乏有效的本体自动构建方法,同时诸如 WordNet 等传统人工组织方法效率低下,难以有效扩展,所以很多研究不得已在精确性和完备性上取得一定的折中,比如只使用用户最为关注的兴趣特征来提高精确度,当然在一定程度上以牺牲完备性为代价[26]。有学者只关注于 is - a 的基本联系[22],还有学者没有使用全部的本体成分,利用简化的结构设计了一种利用用户会话中最能反映用户兴趣特征的相关关键词组成图结构来表达用户个性化模式,而且采用了权值扩散的构造策略[27]。还有学者利用本体技术实现了个性化新闻推荐,但是仍然采用传统的向量模型来构建用户个性化模式[28]。虽然自动构建的本体在可读性方面逊于人工编辑的本体,但在可行性和易用性方面仍有优势。

本体自动构建需要解决如何从词语中识别概念等级及其联系,具体包括三种联系:属种关系(Generic Relationship)、实例关系(Instance Relationship)、整部关系(Whole - part Relationship)[29]。从现有的知识库中自动获取本体也是一种可行的方案,但是它需要知识库作为前提[30,31]。其他常见的方法有词语共现分析方法、句法模式识别方法和词语上下文分布相似度方法等。其中词语共现分析方法简单易行,属于内容分析方法的一种[32],目前被广泛地应用于聚类研究[33]、战略情报研究[34]、专利地图绘制[35]等方面,同时词语共现分析方法也有助于从特定领域中快速自动构建本体。但是它却难以区分上述三种等级关系,只是将其全部转换为上下级的相关联系[36]。不过,这一缺点对于处理用户个性化本体而言,问题并不明显[37]。

为了提高利用词语共现分析方法自动构建本体的有效性,有学者利用了包括分类相似性和内在概念相关度在内的多种相似度关系,实践证明具有可行性[38]。该方法的有效性还依赖于抽取上下文的调节参数、数据集合的规模与质量等[39]。本文采用了一种综合考虑同文词语共现和引文词语共现的方法,并据此实现了领域本体的自动构建。

(3)异构用户个性化本体的相似度计算方法需要探索。为了测度异构本体之间的相似度,必须要解决词语之间的语义相似度(Semantic Similarity)问题[40]。根据使用领域知识的类型不同,该方法可以分为基于分类结构的方法、基于概念的信息内容方法和基于上下文环境相关度的方法。最早的基于分类结构的方法往往侧重于测度两个概念之间的最短距离[41],复杂的考虑往往还会结合概念的所在层次,如认为概念所在层次越高相似度越小等。该方法的问题在于仅仅利用本体层次结构信息,因此本体信息的完备性、同质性和覆盖度都会影响测度的有效性[42]。而且这些方法往往忽略诸如共有上级节点等其他非最短路径信息[43],并且对不同路径的定量化区分也做得不足[44]。最早的基于概念的信息内容方法是根据最近共有上级概念的信息量[45],但是这种方法往往需要对概念进行迭代计算以寻找最近共有上级概念,而且一旦现有的分类层次体系发生变化,重新计算的需求将会导致可扩展性较差,同时它对分类体系的完备性要求更高[46]。基于上下文环境相关度的方法认为如果两个词语所在的上下文环境越相似,则两个词语的语义相关度越大。词语共现分析方法本身就是一种常见的基于上下文环境相关度的方法。本文所探讨的异构用户个性化本体相似度测度方法,是一种利用词语共现分析方法得到的加权本体,通过权值差异程度来间接测度概念内容和结构的差异。值得注意的是,该方法也可以为异构 XML 信息查询提供一种新的思路[47]。

3 设计思路

3.1 领域本体的构建

自动构建的领域本体必须在语义上具有丰富的特点,同时还能尽可能地表达出各种不同的概念联系。在学术文献这个特定的研究领域中,本文将文献的关键词作为构建领域本体概念的主要数据源,同时采用词语共现分析方法来识别概念联系,并引入下面两种设计策略:

(1)考虑同文词语共现和引文词语共现两种概念联系

传统的词语共现分析方法主要考虑同文共现关系,然而不同文献关键词不仅在一个时间点上具有相关性,而且还会在不同的时间点上具有相关性。利用引文之间的时序关系,可以更好地发掘出关键词的演变序列关系和时序相关性,这也充分印证了引文网络分析与其他方法的结合日益密切[48]。

从引文关系中发现引文词语共现对如图 1 所示。可以看出,关键词对(A1,A2)和(B1,B2)都是同文词

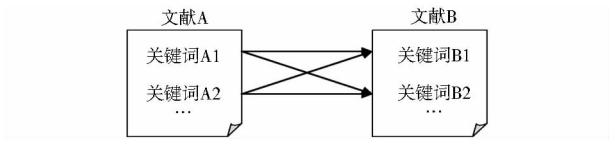

图 1　从引文关系中发现引文词语共现对

语共现关系,由于文献 A 引用了文献 B,所以还可以从中得到 4 个引文词语共现对,分别是(A1,B1)、(A1,B2)、(A2,B1)和(A2,B2)。

当然,对于不同含义的关键词共现对,应该予以不同的权重处理。对同文词语共现对赋予 1 的权值,而对引文词语共现对赋予 0.5 的权值。最终将所有的关键词共现对分组归类,并对权值进行累加,可以得到最终的带有权值的关键词共现对集合。这个权值在一定程度上可以揭示关键词共现对的有效性,权值越高,相应的关键词共现对越常见,实际语义相关性就越高,反之亦然。在实际运算中,可以考虑忽略那些权值较低的关键词共现对。最终可以对所有的权值规范化处理,如利用最大权值对所有的权值进行归一化处理,将权值限定在 0 到 1 之间。

这种关键词共现对具有两个特点:通过赋予不同的权重处理可以表达不同用户的兴趣特征;共现对为有向链接,如(A1,B1)表示 A1 的出现导致了 B1 的出现,本文称为关键词共现对链接。

(2)考虑网状的语义组织结构

该设计策略是将所有的共现对按照文档频率进行整序,将所有的共现对梳理成低文档频率词指向高文档频率词的次序,最终形成一个完整的网状领域本体组织结构。该结构的内核由一组具有较高文档频率的核心关键词组成,外部由多层较低文档频率的关键词组成,不同层次之间的关键词存在由外层指向内核的共现对链接关系。这种设计结构便于以后的权值扩散,权值可以由处于外层、专指性较强的一般关键词逐渐扩散到处于内核、概念含义更为广泛的核心关键词。

在实现上,该本体采用了加权 XML 模型,其中的每个概念对应一个 XML 节点,每个概念联系对应一个 XML 节点联系。同时,每个 XML 节点联系都被赋予对应的概念联系权值。部分结构如图 2 所示。

对应的加权 XML 模型信息如下所示:

```
<Root>
  <Nodes>
```

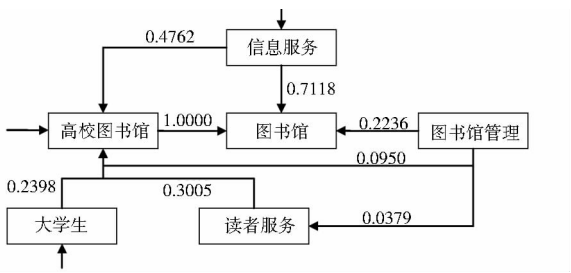

图 2　带有概念联系权值的领域本体部分示意图

```
    <图书馆 id=1 DF=16856/>
    <高校图书馆 id=2 DF=10937/>
    <信息服务 id=4 DF=4268/>
    <读者服务 id=9 DF=2169/>
    <图书馆管理 id=12 DF=1552/>
    <大学生 id=20 DF= 1171/>
    ...
  </Nodes>
  <Edges>
    <id=977473 sid=2 tid=1 weight=1.000/>
    <id=214531 sid=4 tid=1 weight=0.7118/>
    <id=143471 sid=12 tid=1 weight=0.2236 />
    <id=1150132 sid=4 tid=2 weight=0.4762/>
    <id=1972140 sid=9 tid=2 weight=0.3005/>
    <id=1588519 sid=20 tid=2 weight=0.2398/>
    <id=8831 sid=12 tid=2 weight=0.0950/>
    <id=1375122 sid=12 tid=9 weight=0.0379/>
    ...
  </Edges>
...
</Root>
```

其中,DF 表示词频,weight 表示权值,sid 和 tid 分别表示关键词共现对链接的起始关键词和终止关键词。

3.2　异构用户个性化模式的同质化处理方法

对于结构不同的各种用户个性化模式,必须通过将其映射到一个一致的参考本体(Reference Ontology)中,才能完成最终的相似度比较。为此,本文采用了“先打碎后重构”的策略。

如图 3 所示,用户个性化模式 a 和 b 的结构差异很大,甚至在节点内容上也完全不一致。将每个用户个性化模式打碎后,会各自得到一组由两两关键词组成的词语对,如模式 a 得到的词语对为(A,B)和(A,C),模式 b 得到的为(C,D)和(D,A)。接来下,对得到的这些词语对进行重构。首先是将每个词语对整理

成低频关键词指向高频关键词的次序,以便于后续将其纳入到领域本体结构中。假设 A、B、C 和 D 这 4 个关键词的文档频率大小序列为 A > B > C > D,则最终得到的 4 个序列,分别是(B,A)和(C,A)、(D,A)和(D,C),构成的最终个性化模式。

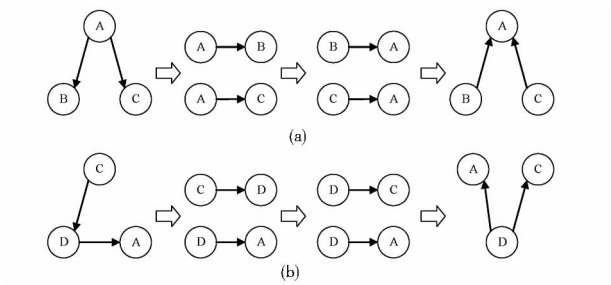

图 3　重构异构用户个性化模式实例

值得注意的是,由于不同的用户个性化模式会在每个关键词链接上具有反映不同兴趣差异的权值,而重构后的用户个性化模式只是重整了结构特征,对相关权值没有做调整。但是,重构后的用户个性化模式具有了和领域本体结构相似的链接指向关系,因此为后续的权值扩散提供了结构基础。

3.3　用户个性化模式在领域本体中的权值扩散方法

本文提出了一种沿着关键词共现对链接、由外层关键词向内核关键词进行权值扩散的迭代算法。具体的权值扩散计算如下:

$$weight(link(T_m,T_n)) = c\sum_{link(T_i,T_m)\in B(T_m)}\frac{weight(link(T_i,T_m))}{N_{T_i}}\times coef_k + (1-c) \quad (1)$$

其中,$link(T_m,T_n)$表示关键词 T_m 和 T_n 共现对链接,weight( )为共现对权值函数,$B(T_m)$表示以 T_m 为终点的所有关键词共现对链接,N_{T_n}表示以关键词 T_n 为起点的关键词共现对链接总数,coef 为衰减系数,由于权值扩散是采用由外层向内核的迭代扩散方式,因此在每次迭代中通过引入一个不断衰减的系数来控制权值扩散的强度,c 为保证权值迭代计算收敛的常量因子,设置为 0.8。

完整的具体算法伪代码如下:

```
// 衰减系数初始值
double coef = 0.8;
// 迭代运算
while(coef > 0.0001){
  // 取出现有用户个性化模式中的每个关键词共现对链接
    for each link_i in profile_k {
      // 取出当前关键词共现对链接的所有有效后续关键词共现对链接
      Collection links = getValidOutLink(link_i);
      // 循环处理以当前关键词共现对链接终点为起点的所有后续关键词共现对链接
      for each link_j in links {
        // 利用公式(1)得到扩散的权值
        getWeight(link_j);
        // 将其加入到现有的用户个性化模式中
        addToProfile(link_j);
      }
      // 控制衰减系数
    coef = coef/2;
}
```

值得说明一点,代码中引入的 getValidOutLink( ) 函数,主要是去除无效的较低权值的关键词共现对以降低计算量。在经过权值扩散的重构用户个性化模式中,利用权值的差别既可以表达共有关键词共现对链接的兴趣差异度,也能很好地测度模式中结构的差别。

3.4　用户个性化模式的相似度比较算法

本文设计的方法主要比较两两用户个性化模式的关键词共现对链接权值差异,并据此测度最终的相似度,计算公式如下:

$$similarity(profile_1,profile_2) = 1 - \frac{\sum_i |weight_1(link_i) - weight_2(link_i)|}{\sum_i weight_1(link_i) + \sum_i weight_2(link_i)} \quad (2)$$

其中,$weight_1(link_i)$表示一个用户个性化模式 $profile_1$ 中的关键词共现对链接,分子为两个用户个性化模式中所有关键词共现对链接权值差值的绝对值之和,分母为两个模式的所有关键词共现对链接权值之和,显然,该相似度最大为 1,最小为 0。

4　实验说明

4.1　数据准备

笔者对万方和 CSSCI 两大中文期刊数据库进行了文献数据获取,抽取了图情领域期刊文献共 59 种,其中核心期刊 35 种,时间跨度为 2000 到 2009 共 10 年,总共获得 202 759 篇有效文献,共计 232 987 个有效引文链接,限于数据集合的有限性,不包含不属于图情领域文献的引文链接。

4.2　构建图情领域本体的相关实验

实验获取的关键词总数为 113 626,通过同文共现

和引文共现获取的关键词共现对共计 2 059 083 个，其中按照低频关键词指向高频关键词的有序化整理，最终得到 982 402 个。通过可视化软件 Gephi，以关键词为节点、关键词共现对链接为边，最终得到的这个领域本体在宏观上表现为一个内核紧密、外层逐渐疏松的基本形态，如图 4 所示：

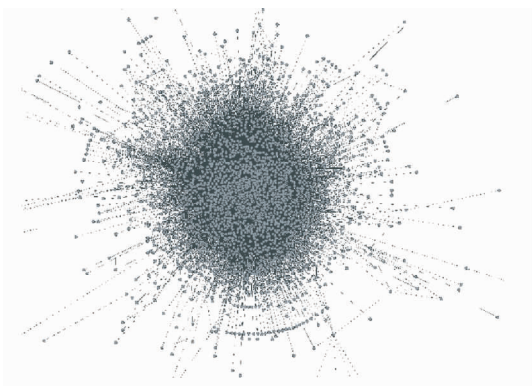

图 4　图情领域本体的宏观结构特征

其中，处于内核的、具有最高文档频率的关键词及其共现关系如图 5 所示：

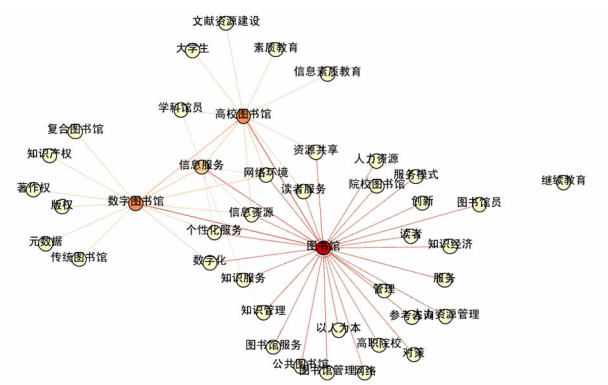

图 5　图情领域本体的核心关键词及其共现关系

图 5 中节点颜色的深浅表示了入度的大小，可以看出，“图书馆”、“数字图书馆”、“高校图书馆”等都是最为常见的核心关键词，也是最易于与其共现的关键词，链接颜色的深浅表示链接权值的大小。

处于外围的关键词共现对数量极大，权值也极小，实验对其进行了除去处理。本文设计了一个去除低权值关键词共现对的方法，即在图 6 中寻找权值变化曲线中的拐点，只保留处于拐点左侧的有效关键词共现对。具体方法描述如下：选择所有的以某一特定关键词为起点的关键词共现对链接，按照权值大小降序排列，然后由高权值链接开始，比较各个相邻的两两链接权值差值，如果连续有 10 次差值小于预设定的阈值 0.00046，则认为首次开始计数的关键词链接权值为截止阈值。以“个性化”为例，以它为起点的链接权值变化趋势如图 6 所示：

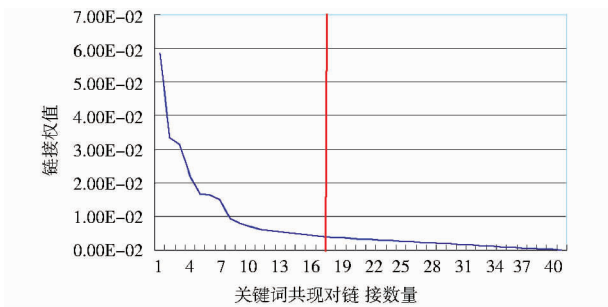

图 6　以“个性化”为起点的关键词链接权值由高到低的变化趋势

可以看出，大量的关键词共现对都具有较低权值，真正具有较高权值的链接数量并非很多。其中竖线标记的第 17 个关键词共现对权值（为 0.00415335）就是截止阈值。

4.3　用户个性化模式相似度计算的实验结果

为了便于测试分析，本文抽取了第一作者为南京地区的共计 5 438 篇图情领域的文献，作者共计 2 392 人，其中发文数量大于 10 篇的有 55 人。这些作者拥有足够多的文献数据，所以可以利用每位作者所发文献的关键词来表示他们各自的研究兴趣。

每个作者的初始用户个性化模式构造方法说明如下：抽取每个作者的每篇所发文献，抽取每篇文献中的所有关键词；将每篇文献中的关键词分成一组，按照文档频率升序排列，并自动构建出所有低频关键词指向高频关键词的共现对链接；对相同的关键词共现对链接进行累计，以出现次数作为权值，可以对每个作者，得到以一组相关联系的关键词共现对链接来表示的原始个性化模式。如抽取的南京大学苏新宁教授所发文献共计 11 篇，如表 1 所示。

从中自动构建的相应个性化模式如图 7 所示。

按照该方法，可以对所有的作者做相同的处理。可以看出，不同作者所对应的不同个性化模式在结构和权值上都具有较大的差异。

利用本文所述的权值扩散方法和相似度比较方法，对这55位作者进行了两两比较，以分析作者之间

表 1　作者发文信息

文献名称	期刊	年份	关键词
《中文社会科学引文索引》在科研及管理中的作用	图书情报工作	2003	中文社会科学引文索引;科学研究;科研管理;
企业知识管理研究与实践的进展	图书情报知识	2003	知识管理;企业管理;
引文索引数据质量控制研究	中国图书馆学报	2001	引文索引;数据质量控制;规范文档;
网络环境下竞争情报系统设计	情报理论与实践	2010	竞争情报;竞争情报系统;系统设计;网络;
网格环境下的个性化信息推荐服务模型研究	情报学报	2007	网格环境;个性化服务;信息推荐;网格中间件;
图书馆、情报与文献学研究热点与趋势分析(2000－2004)——基于 CSSCI 的分析	情报学报	2007	图书馆学;情报学;档案学;研究热点;研究趋势;CSSCI;
图书馆、情报与文献学学术影响力研究报告(2000－2004)——基于 CSSCI 的分析	情报学报	2006	图书馆学;情报学;档案学;分析评价;CSSCI;
视频信息索引技术研究进展	情报学报	2004	视频索引;MPEG 标准;视频信息自动处理;
中国社会科学引文索引设计	情报学报	2000	CSSCI;引文索引;系统设计;
超文本技术在全文检索系统中的实现	情报学报	2000	全文检索;超文本检索;动态超文本;
人文社会科学期刊评价指标体系研究	图书馆论坛	2006	期刊评价;评价指标;指标体系;人文社会科学;

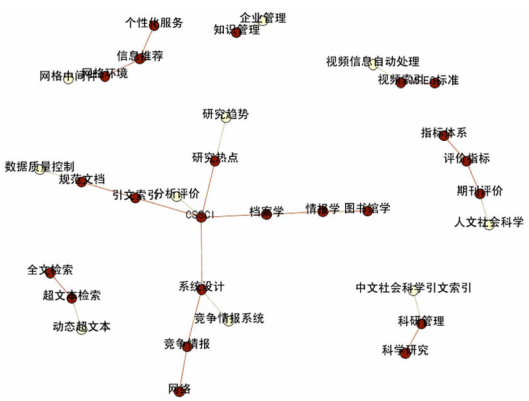

图 7　反映作者兴趣的原始个性化模式

的兴趣相似度,部分兴趣相似度较高的作者及其关系如图 8 所示:

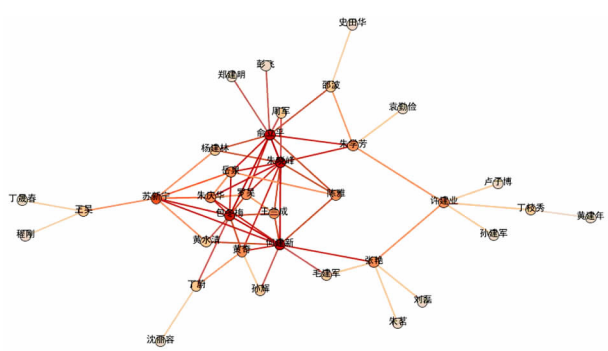

图 8　南京地区图情领域学者之间的兴趣相关度

其中,节点颜色越深,表示与该节点作者兴趣相近的作者数量越多,而链接颜色越深,则表示两者兴趣越相近。

5　结　语

本文所探讨的异构用户个性化模式相似度比较方法可以应用于各种特定领域的个性化信息推荐服务,同时也可以作为一种领域本体映射的解决方法。在研究实践中,笔者选择了学术文献这个特定的研究领域,并尝试将异构用户个性化本体引入到现有的个性化学术文献信息推荐系统中。从目前的使用效果来看,该方法实现了预期的目标。未来研究将重点放在对领域本体的设计改进上,如现有的主流方法主要考虑双词共现,而多词共现的现象也很值得关注,从中也可以发现更多的概念联系。另外,引文词语共现对充分反映了概念的时序演化关系,因此在现有的领域本体中通过引入时间维度来构建多维本体结构,可以实现更好的概念表达效果。

参考文献:

[1] 李树青. 基于加权 XML 数据模型的个性化本体研究[J]. 情报学报,2010,29(5):826－834. (Li Shuqing. Study of Personalized Ontology Based on Weighted XML Data Model[J]. *Journal of the China Society for Scientific and Technical Information*, 2010, 29(5):826－834.)

[2] Tamine－Lechani L, Boughanem M, Zemirli N. Personalized Document Ranking: Exploiting Evidence from Multiple User Interests for Profiling and Retrieval[J]. *Journal of Digital Information Management*, 2008, 6(5):354－361.

[3] Liu F, Yu C, Meng W Y. Personalized Web Search for Improving Retrieval Effectiveness[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2004, 16(1):28－40.

[4] Gauch S, Chaffee J, Pretschner A. Ontology－based Personalized Search and Browsing[J]. *Web Intelligence and Agent System*, 2003, 1(3－4):219－234.

[5] Sieg A, Mobasher B, Burke R. Web Search Personalization with Ontological User Profiles[C]. In: *Proceedings of the* 16*th ACM Confer-*

ence on Information and Knowledge Management(*CIKM*'07). New York, NY, USA:ACM,2007:525 – 534.

[6] Gabrilovich E, Markovitch S. Computing Semantic Relatedness Using Wikipedia – based Explicit Semantic Analysis[C]. In: *Proceedings of the* 20*th International Joint Conference on Artificial Intelligence*. 2007:12 – 18.

[7] Vallet D, Cantador I, Fernández M, et al. A Multi – purpose Ontology – based Approach for Personalized Content Filtering and Retrieval[C]. In: *Proceedings of the* 1*st International Workshop on Semantic Media Adaptation and Personalization*. 2006:19 – 24.

[8] 颜端武,刘明岩,许应楠. 基于领域本体的细粒度用户兴趣建模研究[J]. *情报学报*,2010,29(3):433 – 442. (Yan Duanwu, Liu Mingyan, Xu yingnan. Toward Fine – grained User Preference Modeling Based on Domain Ontology[J]. *Journal of the China Society for Scientific and Technical Information*,2010,29(3):433 – 442.)

[9] Bhattacharyya P, Garg A, Wu S F. Analysis of User Keyword Similarity in Online Social Networks[J]. *Social Network Analysis and Mining*,2011,1(3):143 – 158.

[10] 刘萍,叶燕. 基于本体的高校专家定位系统研究[J]. *情报学报*,2010,29(5):813 – 819. (Liu Ping, Ye Yan. An Ontology – based Experts Locator System Within Academia[J]. *Journal of the China Society for Scientific and Technical Information*,2010,29(5):813 – 819.)

[11] Stuckenschmidt H, Harmelen F. Information Sharing on the Semantic Web[M]. Springer,2005:3 – 4.

[12] Mao M, Peng Y F, Spring M. An Adaptive Ontology Mapping Approach with Neural Network Based Constraint Satisfaction[J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2010,8(1):14 – 25.

[13] Ehrig M. Ontology Alignment: Bridging the Semantic Gap(Semantic Web and Beyond)[M]. Springer,2006:1 – 2.

[14] Mao M. Ontology Mapping: An Information Retrieval and Interactive Activation Network Based Approach[C]. In: *Proceedings of the* 6*th International Semantic Web and the* 2*nd Asian Conference on Asian Semantic Web Conference*. 2007:931 – 935.

[15] Mao M, Peng Y F, Spring M. Ontology Mapping: As a Binary Classification Problem[C]. In: *Proceedings of the* 4*th International Conference on Semantics, Knowledge and Grid*. 2008:20 – 25.

[16] Koren Y, Park F. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model[C]. In: *Proceedings of the* 14*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA:ACM,2008:426 – 434.

[17] Haase P, Hotho A, Schmidt – Thieme L, et al. Collaborative and Usage – driven Evolution of Personal Ontologies[C]. In: *Proceedings of the* 2*nd European Conference on the Semantic Web: Research and Applications*. 2005:486 – 499.

[18] 邓珞华. 概念空间——定义、意义和局限[J]. *情报学报*,2003,22(4):393 – 397. (Deng Luohua. Concept Space——Its Definition, Significance and Limitation[J]. *Journal of the China Society for Scientific and Technical Information*,2003,22(4):393 – 397.)

[19] 吕刚,郑诚. 基于加权的本体相似度计算方法[J]. *计算机工程与设计*,2010,31(5):1093 – 1095. (Lv Gang, Zheng Cheng. Method of Ontology Similarity Calculation Based on Weighted[J]. *Computer Engineering and Design*,2010,31(5):1093 – 1095.)

[20] Tsatsaronis G, Vazirgiannis M, Androutsopoulos I. Word Sense Disambiguation with Spreading Activation Networks Generated from Thesauri[C]. In: *Proceedings of the* 20*th International Joint Conference on Artificial Intelligence*. 2007:1725 – 1730.

[21] Rocha C, Schwabe D, Aragao M P. A Hybrid Approach for Searching in the Semantic Web[C]. In: *Proceedings of the* 13*th International Conference on World Wide Web*. New York, NY, USA: ACM,2004:374 – 383.

[22] Middleton S E, Shadbolt N R, De Roure D C. Ontological User Profiling in Recommender Systems[J]. *ACM Transactions on Information Systems*,2004,22(1):54 – 88.

[23] Thiagarajan R, Manjunath G, Stumptner M. Computing Semantic Similarity Using Ontologies[R]. HP Labs Technical Report, HPL – 2008 – 87,2008.

[24] Sieg A, Mobasher B, Burke R. Improving the Effectiveness of Collaborative Recommendation with Ontology – based User Profiles [C]. In: *Proceedings of the* 1*st International Workshop on Information Heterogeneity and Fusion in Recommender Systems* (*HetRec*'10). New York, NY, USA:ACM,2010:39 – 46.

[25] 李树青,徐侠,钱钢,等. 基于震荡算法和领域本体的学术文献关键路径自动识别和可视化展示方法[J]. *情报学报*,2012,待发. (Li Shuqing, Xu Xia, Qian Gang, et al. An Automatic Recognition and Visualization Method of Main – path in Academic Documents Based on Vibration Algorithm and Domain Ontology[J]. *Journal of the China Society for Scientific and Technical Information*,2012, Accepted.)

[26] Zhou T, Kuscsik Z, Liu J G, et al. Solving the Apparent Diversity – accuracy Dilemma of Recommender Systems[J]. *Proceedings of the National Academy of Sciences*(*PNAS*),2010, 107(10): 4511 – 4515.

[27] Daoud M, Tamine – Lechani L, Boughanem M, et al. A Session Based Personalized Search Using an Ontological User Profile[C]. In: *Proceedings of the* 2009 *ACM Symposium on Applied Computing*. New York, NY, USA:ACM,2009:1732 – 1736.

[28] IJntema W, Goossen F, Frasincar F, et al. Ontology – based News Recommendation[C]. In: *Proceedings of the* 2010 *EDBT/ICDT Workshops*. New York, NY, USA:ACM,2010:16 – 23.

[29] National Information Standards Organization. Guidelines for the

Construction, Format, and Management of Monolingual Controlled Vocabularies[S]. Bethesda: NISO Press, 2005.

[30] 余传明,张小青. 从 Wikipedia 中获取本体:原理与方法研究[J]. *情报学报*,2011,30(3):244 - 252. (Yu Chuanming, Zhang Xiaoqing. Learning Ontology from Wikipedia :Principles and Methods[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(3): 244 - 252.)

[31] 薛建武,勾苗,吴拓. 基于 SKOS 的国防科学技术叙词表向本体的转换研究[J]. *情报学报*,2011,30(3):310 - 317. (Xue Jianwu, Gou Miao, Wu Tuo. The Transformation from Thesaurus of National Defense Science and Technology to Ontology Based on SKOS [J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(3): 310 - 317.)

[32] 冯璐,冷伏海. 共词分析方法理论进展[J]. *中国图书馆学报*, 2006,32(2):88 - 92. (Feng Lu, Leng Fuhai. Development of Theoretical Studies of Co - Word Analysis[J]. *Journal of Library Science in China*, 2006, 32(2): 88 - 92.)

[33] 李枫林,何洲芳. 基于关键词共现分析的检索结果聚类研究[J]. *情报学报*,2011,30(8):819 - 825. (Li Fenglin, He Zhoufang. Study on Clustering of Retrieval Results Based on Co - occurrence Analysis of Keywords[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(8): 819 - 825.)

[34] 杨颖,崔雷. 基于共词可视化的学科战略情报研究[J]. *情报学报*,2011,30(3):325 - 330. (Yang Ying, Cui Lei. Subject Strategic Information Research Based on Visualization of Co - Word Network[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(3): 325 - 330.)

[35] 王胜君,吴冲,张新颖,等. 基于共现分析的专利地图绘制及实证研究——一个政府信息重构的视角[J]. *情报学报*,2011,30(3):318 - 324. (Wang Shengjun, Wu Chong, Zhang Xinying, et al. Patent Map Drawing and Its Application of Based on Co - occurrence Analysis: Perspective of Government Information Restructuring[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(3): 318 - 324.)

[36] 杜慧平. 概念等级关系自动识别研究[J]. *情报学报*,2011,30(3):237 - 243. (Du Huiping. Automatic Extraction of Concept Hierarchical Relationships[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(3): 237 - 243.)

[37] 李树青. 基于引文关键词加权共现技术的图情学科领域本体自动构建方法研究[J]. *情报学报*,2012,31(4):371 - 380. (Li Shuqing. Research of Automatic Construction of Domain Ontology in Library and Information Science Based on Weighted Co - occurrence of Citation Keywords[J]. *Journal of the China Society for Scientific and Technical Information*, 2012, 31(4): 371 - 380.)

[38] 张学福. 基于词共现的可视化概念空间研究[J]. *情报学报*, 2008,27(2):205 - 211. (Zhang Xuefu. Research on Visualization Concept Space Based on Co - word Occurrence[J]. *Journal of the China Society for Scientific and Technical Information*, 2008, 27(2): 205 - 211.)

[39] Pedersen T, Pakhomov S V S, Patwardhan S, et al. Measures of Semantic Similarity and Relatedness in the Biomedical Domain[J]. *Journal of Biomedical Informatics*, 2007, 40(3): 288 - 299.

[40] 董慧,唐敏. 语义检索在 Web2.0 环境下的应用探讨[J]. *中国图书馆学报*,2011,37(2):115 - 119. (Dong Hui, Tang Min. Application of Semantic Search in the Web2.0 Environment[J]. *Journal of Library Science in China*, 2011, 37(2): 115 - 119.)

[41] Rada R, Mili H, Bichnell E, et al. Development and Application of a Metric on Semantic Nets[J]. *IEEE Transactions on System Management and Cybernetics*, 1989, 19(1): 17 - 30.

[42] Cimiano P. Ontology Learning and Population from Text: Algorithms, Evaluation and Applications[M]. Springer - Verlag, 2006: 44 - 45.

[43] Patwardhan S, Pedersen T. Using WordNet - based Context Vectors to Estimate the Semantic Relatedness of Concepts[C]. In: *Proceedings of the EACL* 2006, *Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*, Trento, Italy. 2006: 1 - 8.

[44] Bollegala D, Matsuo Y, Ishizuka M. WebSim: A Web - based Semantic Similarity Measure[C]. In: *Proceedings of the* 21*st Annual Conference of the Japanese Society for Artificial Intelligence* (*JSAI*2007), Miyazaki, Japan. 2007: 757 - 766.

[45] Resnik P. Using Information Content to Evaluate Semantic Similarity in a Taxonomy[C]. In: *Proceedings of the* 14*th International Joint Conference on Artificial Intelligence* (*IJCAI*' 95), Montreal, Canada. 1995: 448 - 453.

[46] Batet M, Sánchez D, Valls A. An Ontology - based Measure to Compute Semantic Similarity in Biomedicine[J]. *Journal of Biomedical Information*, 2011, 44(1): 118 - 125.

[47] 温馨,陈群,娄颖. 基于词项扩展的 XML 信息检索反馈技术[J]. *计算机工程*,2011,37(20):36 - 38. (Wen Xin, Chen Qun, Lou Ying. Feedback Technique for XML Information Retrieval Based on Term Expansion[J]. *Computer Engineering*, 2011, 37(20): 36 - 38.)

[48] 韩毅,张克菊,金碧辉. 引文网络分析的方法整合研究进展[J]. *中国图书馆学报*,2010,36(4):83 - 89. (Han Yi, Zhang Keju, Jin Bihui. Research Progress on Methodology Integration of Citation Network Analysis[J]. *Journal of Library Science in China*, 2010, 36(4): 83 - 89.)

(作者 E - mail: leeshuqing@163.com)