面向大数据的多维粒矩阵关联分析及应用

吴 珺[1,2] 王春枝[1]
(湖北工业大学计算机学院 武汉 430068)[1]
(武汉理工大学交通物联网技术湖北省重点实验室 武汉 430070)[2]

摘 要 当前日益增长的大数据备受青睐,大数据的核心是数据分析。然而聚焦大数据的动态、多维特性,传统数据分析方法难以获取可靠且准确的分析结果,数据分析方法面临着重要的发展机遇和严峻的挑战。对动态大数据的多维关联性分析问题进行研究和探讨,以动态大数据为研究对象,以粒计算(Granular Computing,GrC)理论为研究基础,提出粒矩阵思想,研究构建面向动态大数据的粒矩阵方法,分析粒矩阵的逻辑约简运算,确定了基于粒矩阵的动态大数据多维关联性分析模型。本文旨在为高效利用动态大数据进行多维关联性分析和揭示数据隐含的客观规律提供科学依据,对大数据的可持续发展也具有重要意义。

关键词 大数据,粒计算,多维关联分析
中图法分类号 TP393 **文献标识码** A

Multiple Correlation Analysis and Application of Granular Matrix Based on Big Data

WU Jun[1,2] WANG Chun-zhi[1]
(School of Computer Science, Hubei University of Technology, Wuhan 430068, China)[1]
(Hubei Key Laboratory of Transportation Internet of Things, Wuhan University of Technology, Wuhan 430070, China)[2]

Abstract The ever increasing big data is acclaimed, and the key point of big data is data analysis. However focusing on the big data with dynamic and multiple-dimensional characteristic, it is difficult for traditional data analysis methods to obtain reliable and accurate analytical results. Therefore there is an important opportunity and a great challenge for the data analysis methods to be developed. It aims to make an important research and investigation of the multiple correlation analysis for dynamic big data. The research object is dynamic big data, and the research is based on the theory of granular computing (GrC). We got the theoretical thought of granular matrix completely new in this paper. It was expected to reveal the multiple correlation analysis for dynamic big data. On one hand the achievements of this research would provide a scientific basis for multiple correlation analysis and revelation of the objective law in big data area. On the other hand it is also an important implication for sustainable development of big data.

Keywords Big data, Granular computing, Multiple correlation analysis

1 引言

近年来,随着信息技术的创新发展,社交网络、电子商务、智慧地球等进入蓬勃的发展期,与此同时产生的结构化和非结构化的海量动态网络数据充斥着人们生活的各个角落,证明我们已经进入大数据时代。面对日益倍增的大数据,我们却正逐渐进入大数据发展的停滞期——学会了存储数据,但却不能从中提取价值。

目前大数据已经成为研究热点,科技发展和社会需求也推动大数据技术不断前进。以IBM的超级计算机Watson[1]为代表的大数据智能化机器产品,实现了对海量复杂自然语言的学习和处理。Watson在没有任何背景知识的条件下,通过阅读学习相当于2亿页用自然语言撰写的文档来获取知识,并分析挖掘多个知识间的关联性来确定解决问题的信息,最终完成击败人类智力竞赛冠军的壮举。此外Google公司也在研发构造具有独立思考能力可以进行深度学习的知识图谱[2],目前已掌握180亿个事实及其关联关系,以实现通过大数据的关联分析为用户提供更加智能的服务和密切的互动。面对大数据的海量、复杂多样、变化快的特性,对于大数据环境下的应用问题,很多传统的在普通数据上的机器学习算法已不再适用。通过上述成功的商业大数据智能化产品表明,研究大数据环境下的机器学习算法[3]成为学术界和产业界共

本文受国家自然科学基金(61602161),湖北省自然基金(2014CFB590),交通物联网技术湖北省重点实验室(2015III015-A03)资助。
吴 珺(1984—),女,博士,讲师,CCF会员,主要研究方向为粒计算及应用、智能方法与数据挖掘,E-mail:wujun@whut.edu.cn(通信作者);
王春枝(1963—),女,博士,教授,CCF会员,主要研究方向为计算机网络、智能方法及应用,E-mail:lavazza@Foxmail.com。

同关注的话题。学术界、工业界及全球各国的政府机构都已经密切关注大数据的数据挖掘和知识发现技术，如何获得大数据蕴含的有效信息并挖掘多维数据间的关联性等问题[4-9]成为研究的热点。

因此迫切需要发展适合大数据的数据挖掘技术，这也是计算机科学和人工智能[10-15]等相关学科研究的热点和难点问题。李志杰等[10]针对在大数据时代越来越多的领域中出现的对海量、高速数据进行实时处理的需求，介绍了大数据分析的动机与背景，集中展示了经典的和最新的在线学习方法与算法。张蕾等[11]使用了深度神经网络及其学习算法，该研究方法作为成功的大数据分析方法，已为学术界和工业界所熟知。与传统方法相比，深度学习方法主要基于数据驱动，能自动从数据中提取特征知识，对于分析非结构化、模式不明多变、跨领域的大数据具有显著优势。近期孟小峰[15]提出数据特征和现实需求都发生了变化，以大规模、多源异构、跨领域、跨媒体、跨语言、动态演化、普适化为主要特征的数据发挥着更重要的作用，相应的数据存储、分析和理解也面临着重大挑战。当下亟待解决的问题是如何利用数据的关联、交叉和融合实现大数据的价值最大化，由于认为解决该问题的关键在于数据的融合，因此提出了大数据融合的概念。

2 大数据的多维特性与分析研究

2.1 大数据的多维特性

大数据时代，传统的数据挖掘思想、规则挖掘环境和关联规则挖掘系统都面临着前所未有的挑战。传统的数据挖掘思想需要从静态规则转变到动态规则，从单目标分析扩展到多目标融合分析，从个体决策转化到群体决策。规则挖掘环境已经由静态确定型向动态不确定型转变；相应的关联规则挖掘系统也从集中式向分布式发展。动态大数据时代的推进使许多科研领域共享使用大规模的数据集成为可能，而且这类数据产生的速度远远超过了对它们进行人工分析的速度。

多维关联性分析是整个大数据处理流程的核心，因为大数据的价值产生于多维关联性分析的过程。从异构数据源抽取和集成的数据构成了数据分析的原始数据，根据不同应用的需求可以从这些数据中选择部分或全部进行分析。传统的分析技术如数据挖掘、机器学习、统计分析等在大数据时代需要做出调整，因为这些技术在大数据时代面临着一些新的挑战，主要有以下两个方面：

(1)大数据并不一定意味着数据价值的增加，相反往往意味着数据噪音的增加。在进行多维关联性分析之前必须进行数据清理、数据分类等预处理工作，但是预处理如此大量的数据对于机器硬件以及算法都是严峻的考验，因此需要引入新的理论方法，比如粒计算理论，它在处理复杂问题时的特性，在动态大数据的知识发现研究中能够表现出一定的优势。

(2)数据分析结果好坏的衡量。得到数据分析结果并不难，但是结果好坏的衡量却是大数据时代数据分析的新挑战，主要表现在以下 3 个方面：数据的动态更新、数据的多维融合、数据的关联性分析。传统方法进行数据分析时对整个数据的分布特点往往掌握得不太清楚，这会导致最后在设计衡量的方法以及指标时遇到诸多困难。

2.2 面向大数据的多维关联分析

关联性是一种反映事件之间的依赖关系或者关联程度的有效信息。现实世界中存在的各种事物间的关联规则其实就是从海量的、多源的、复杂的数据源中挖掘出数据间隐含的关联性。大数据关联性挖掘方法的主要工作就是从非结构化的数据源中寻找数据有序的组织形式并找到数据间隐藏的关联信息。

传统的关联规则算法都是基于频繁项目集的关联规则挖掘。最经典的关联规则挖掘算法是由 Agrawal 等[16]建立的用于事务数据库挖掘的项目集合空间理论的 Apriori 及其改进算法。Han Jiawei 等[17]提出的 FP_Growth 算法，是通过在内存中构造 FP_Tree，减少对原数据集的读取次数及候选频繁项目集的生成，从而克服了传统的 Apriori 算法需多次扫描数据集，生成大量候选频繁项目集而带来的执行效率较低的问题。

基于多维分布的关联规则挖掘是将数据库技术与关联规则挖掘方法相结合的一种新颖的方法。由于数据库系统的索引和查询处理机制可以与数据源本身的特性相结合，因此被用于进行关联分析。最早 Houtsma 等[18]通过运用 SQL 查询支持关联规则挖掘，提出了 SETM 算法，随后又出现了许多类似算法，实现了在数据库系统中提供关联规则挖掘的能力，这类关联规则挖掘的操作是一种表达能力很强的 SQL 操作，可以处理多维数据的聚集属性和项分层结构。Chen 等[19]结合数据仓库的特点，提出了分布的 OLAP 挖掘多层关联规则的框架。

随着人们对关联规则挖掘的重视，越来越多的人们对动态数据的关联规则挖掘进行学习和研究。例如数据库的更新就需要将动态性作为关联规则挖掘的一个新指标，因此数据的关联规则会随着时间的推移而发生改变。针对动态数据的关联规则挖掘方法，何清等[20]提出了一种大数据库的机器学习挖掘方法。沈斌等[21]在此方法上进一步改进获得了新的动态关联规则及其挖掘算法 EFP-growth，并通过实验分析证明了基于扩展 FP-树的 EFP-growth 算法具有较好的可理解性，适用于高密度海量动态数据的挖掘。针对云计算环境下的海量数据的关联规则挖掘算法，李玲娟等[22]提出了云计算环境下关联规则的并行挖掘，并实现了基于 MapReduce 编程模式的 Apriori 算法。杨勇等[23]根据传统的 FP-growth 算法提出了 Pruned FP-tree 算法，减少了传统算法挖掘的迭代过程，并且挖掘出精确的条件模式基，提高了后续建立和挖掘 FP-tree 的效率，它具有很好的规模增长性、可扩展性和良好的加速比，同时有效解决了海量数据处理时的内存瓶颈。

总体而言，目前关联规则挖掘方法得到了广泛的应用；然而对于大数据，由于数据的动态以及复杂多源等特性，传统的关联规则挖掘方法难以取得令人满意的结果，影响人们通过大数据获取信息，阻碍了信息社会的进步和发展。因此需要进一步针对动态大数据的特点，研究更加切实有效的关联规则挖掘方法。

3　粒计算

3.1　粒计算理论及应用

粒计算(Granular Computing，GrC)理论最早是由 T. Y. Lin 教授于 1997 年首次提出，旨在解决多层次问题，随后引起了众多学者的广泛关注，并成为人工智能领域的研究热点。人们面对复杂的、难以准确把握的问题时，由于能力有限，通常还不能采用系统的、准确的方法来获取问题的最优解，而是通过逐步尝试的方法达到有限的、合理的目标，也即采用由粗到细、不断求精的多粒度分析法，避免过于复杂的计算，从而获得较为满意的解，以使原来看似非多项式难解的问题得以解决[24]。人类智能的这种特点正是粒计算的基本思想。

粒计算的实质是通过选择合适的粒度，来降低问题求解的难度，从而找到一种较好的解决方案。人类智能公认的特点之一就是人们能从极不相同的粒度上观察和分析同一问题，不仅能在不同粒度的世界上进行问题的求解，而且能够很快地“从一个粒度世界跳到另一个粒度世界”。这种处理不同粒度世界的能力正是人类问题求解的强有力的表现。

在粒计算的理论方面已提出了一系列的粒度计算模型，如 Lin. T. Y 的领域系统粒度模型[25]、Zadeh 的模糊集词计算模型[26]、Pawlak 的粗糙集粒度模型[27]、张玲的商空间粒度模型[28]和王国胤的知识不确定性粒计算模型[29-30]。

粒计算理论被公认为是信息处理的一种新的概念和计算范式，凡是在求解和分析问题的过程中，应用了分组、分类、聚类和关联手段的一切理论与方法均属于粒计算的范畴[31-33]。面对诸如多维、非结构化等日益复杂的大数据，传统的分类方法已不能很好地满足需求。粒计算数据挖掘方法因其自身处理复杂问题时的特性，在动态大数据的知识发现的研究中能够表现出一定的优势。粒计算是一种看待客观世界的世界观和方法论。在现实世界中，粒度的思想是广泛存在的，信息粒是对现实世界不同层次信息的一定程度的抽象[34-35]。对粒度的思考是建立在人们对不同对象的相似性和不确定性的一个可容忍的限度之内的，其目的是使问题简单化、清晰化，从而有利于问题的多维关联性求解。在大数据时代，粒计算是一种针对高速发展的数据信息产生的新一代人工智能软计算方法，它具有特殊的数据属性定义及分析，能通过逻辑特性加工形成高速海量处理分析数据的能力。

3.2　粒度分析

Yao 从信息计算的角度给出粒计算的定义：粒计算是一个信息处理的典型方法。然而在实际的复杂问题的求解过程中，随着求解问题的不同，需要不同粒度的世界描述。

粒计算是信息处理的一种新的概念和计算范式，覆盖了所有有关粒度的理论、方法、技术和工具的研究。

粒计算中最基本的概念是粒、粒化和粒度。粒是指一些个体通过不分明关系、相似关系、邻近关系或功能关系等多种形式形成的块；粒化就是对粒进行操作；粒度本来是一个物理学概念范畴，指微粒大小的平均度量，粒计算中指信息粗细的平均度量，信息粒度表示对信息和知识细化的不同层次的度量，下面给出粒度的定义[36]。

定义 1　设给定论域 U 和 U 上的一个关系 $R:U\to P(U)=>U=\bigcup_{i\in X}G_I$，则称每一个 G_i 为一个粒子，$\{G_i\}_{i\in X}$ 是论域的一种粒度。

其中，$P(U)$表示论域 U 的冥集；R 可代表不可区分关系、相似关系、等价关系、模糊关系和一般的函数等。对 $\forall i,j\in X, i\neq j$ 时有 $G_i\cap G_j=\varnothing$，则称$\{G_i\}_{i\in X}$是论域的一种划分，记为$\{G_i\}_{i\in X}=[U]$；当 $\forall i,j\in X, i\neq j$ 时有 $G_i\cap G_j\neq\varnothing$，则称$\{G_i\}_{i\in X}$是论域的一种覆盖，记为$\{G_i\}_{i\in X}=\langle U\rangle$。

3.3　粒矩阵的定义及构建

粒矩阵是以粒计算理论为基础的，通过集合的思想刻画粒，实现对信息系统的表示和关系划分；通过构建粒矩阵实现规范管理动态大数据中的大量非结构化数据；它是进行更高层次粒矩阵逻辑约简运算的前提，粒矩阵构建的质量直接影响到后续粒矩阵各类逻辑约简运算的质量以及数据间的关联性分析。

粒矩阵是一种新的基于粒计算理论的数学模型。根据集合理论的思想，采用二进制表示法对数据进行粒表示，从而构建粒矩阵。

定义 2　设信息系统 $InfoSys=\langle U,A,V,f\rangle$，令 $R\subseteq A$，则属性集 R 划分为 $U/Ind(R)=\{X_1,X_2,\cdots,X_m\}$，$X_i$ 用一个长度为 $|U|$ 的二进制数来表示，即：

$$X_i=\{x_{1i},x_{2i},\cdots,x_{ni}\},x_{ki}=\begin{cases}1, & x_k\in X_i\\ 0, & x_k\notin X_i\end{cases}$$

定义 3　$|X_i|=\sum_{k=1}^{n}x_{ki}$，$|X_i|$表示等价类的大小，即等价类中不可分辨的元素的个数。令属性集 $R\subseteq A$，由 R 导出的等价类可以表示为 $X_{m\times n}=\begin{bmatrix}X_1\\X_2\\\cdots\\X_m\end{bmatrix}$。

定义 4　根据以上推导，粒矩阵可以为：

$$X_{m\times n}=\begin{bmatrix}x_{11}&x_{12}&\cdots&x_{1n}\\x_{21}&x_{22}&\cdots&x_{2n}\\\cdots&\cdots&\cdots&\cdots\\x_{m1}&x_{m2}&\cdots&x_{mn}\end{bmatrix}$$

定义 5　通过二进制信息粒，即 1 或者 0 组成的若干信息粒子来表示等价类；那么根据定义 4 得到的矩阵 $X_{m\times n}$，就是由以上导出的等价类构造而成的，其中包括若干 0 和 1 粒矩阵。

4　粒矩阵的多维关联分析与实验

研究基于粒矩阵的逻辑约简运算，通过 3.3 节粒矩阵的定义及构建章节中获得的粒矩阵判断知识粒化后的信息粒，从而更加直观地刻画表现面向动态大数据的粒矩阵结果。

本文提出基于粒矩阵的多维关联分析方法(Correlation Analysis based on Granular Matrix，CA-GM)，其具体的步骤如下：

Step1　针对大数据集中的若干个属性集合，设 A,B 为信息系统 $InfoSys=\langle U,A,V,f\rangle$ 中的属性集，并有 $A,B\subseteq R$，且 $A\neq B$。

Step2　通过粒矩阵相与 $\wedge$ 运算，设 $\mathrm{Min}(A_{l\times n}\wedge B_{p\times n})$ 为

粒矩阵相与运算约简后的粒矩阵。

Step3 在获得粒矩阵的基础上，定义并计算 $\mathrm{Min}(A_{l\times n} \wedge B_{p\times n})$，即为同时用知识 A,B 进行粒化后的最简化粒矩阵。

为进一步研究基于粒矩阵的多维关联分析方法，证明针对动态多维大数据的可用性和有效性，本文选取 Web 大数据进行实验分析。根据以下步骤分析，首先通过构造信息系统中的多个数据集进行多维关联分析实验。

本文实验的硬件环境是 CPU 为 i5，内存 4G 的高性能 PC 机，运行环境是 Win7 操作系统下的 Eclipse4.3.2。

相关实验步骤如下。

Input：Set of $InfoSys=\langle U,A,V,f\rangle$

Output：关联信息粒 ϕ

Step1 根据定义 1 构造粒计算集合属性 A 空间，可以得到 $U/Ind(A)=\{A_1,A_2,\cdots,A_l\}$，同理得到 $U/Ind(B)=\{B_1,B_2,\cdots,B_p\}$。

Step2 根据定义 2—定义 4 构造粒矩阵，分别得到大数据集合中的属性集合 A,B 的粒矩阵。

Step3 定义粒化相与计算：对于两个粒矩阵相，将其中一个粒矩阵中的每一行元素与另一个粒矩阵的所有行逐一进行相与运算。

Step4 粒矩阵相与运算表示知识 A,B 的等价类的包含关系。运算后的结果表示同时用知识 A,B 进行粒化后的等价类。

Step5 粒矩阵相与运算后的结果分析：

当 $|\sum a_{l\times n} \wedge b_{p\times n}|=0$，则表示该信息粒为 ϕ，表示该信息粒不包含任何信息，需要从粒矩阵中约简，从而得到关联信息粒 ϕ。

为验证基于粒矩阵的多维关联分析方法的可用性，分别针对不同数量级的信息系统数据集进行多维关联分析实验 1，其结果如表 1 所列，随着信息系统数据集规模的增大，本方法正常运行且处理时间正向增加。

表 1 实验 1 的结果

实验数据集	数据量/M	运行时间/s
InfoSys1	100	1.231
InfoSys2	500	3.673
InfoSys3	1200	8.267

将本文的研究方法与传统数据关联分析方法进行比较，取相同信息系统数据集进行实验 2，实验结果对比如图 1 所示。随着信息系统数据集的增大，本文的基于粒矩阵的多维关联分析方法在处理时间和关联分析结果上都优于传统方法。

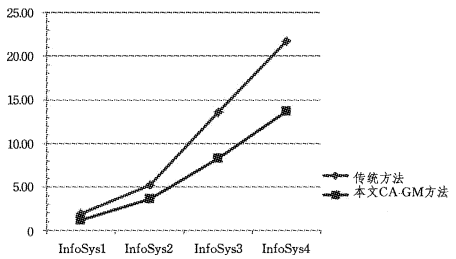

图 1 实验 2 对比分析结果图

结束语 面向动态大数据的粒矩阵多维关联分析方法与传统的关联性分析方法不同，首先要构建面向动态大数据的粒矩阵，得到粒矩阵；然后通过基于粒矩阵的逻辑约简运算，获得动态大数据的约简属性集；最后进行多维关联分析。本文的研究内容主要从构建面向动态大数据的粒矩阵出发，建立基于粒矩阵的数据粒化表示方法，重点解决基于粒矩阵的逻辑约简运算和属性约简多维关联分析算法等理论与技术上的难题。

参 考 文 献

[1] http://www.ibm.com.

[2] http://www.google.com.

[3] KUMAR A, NIU F, RÉ C. Hazy: making it easier to build and maintain big-data analytics[J]. Communications of the ACM, 2013, 56(3): 40-49.

[4] GUO P, WANG K, LUO A L, et al. Computational Intelligence for Big Data Analysis: Current Status and Future Prospect[J]. Journal of Software, 2015, 26(11): 3010-3025.

[5] BENNETT P, GILES L, HALEVY A, et al. Channeling the deluge: research challenges for big data and information systems [C]// Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management. ACM, 2013: 2537-2538.

[6] LU C W, HSIEH C M, CHANG C H, et al. An Improvement to Data Service in Cloud Computing with Content Sensitive Transaction Analysis and Adaptation[C]// 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops. 2013: 463-468.

[7] YANG Q. Big data, lifelong machine learning and transfer learning[C]// Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. ACM, 2013: 505-506.

[8] ORDONEZ C. Can we analyze big data inside a DBMS[C]// Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP. ACM, 2013: 85-92.

[9] NGUYEN D, VO B. Mining Class-Association Rules with Constraints[M] // Knowledge and Systems Engineering. Springer International Publishing, 2014: 307-318.

[10] 李志杰，李元香，王峰，等. 面向大数据分析的在线学习算法综述[J]. 计算机研究与发展，2015，52(8)：1707-1721.

[11] 张蕾，章毅. 大数据分析的无限深度神经网络方法[J]. 计算机研究与发展，2016，53(1)：68-79.

[12] 罗军舟，王兴伟，尹浩. 大数据驱动网络科学研究专题前言[J]. 计算机研究与发展，2015，52(4)：777-782.

[13] 陈世敏. 大数据分析与高速数据更新[J]. 计算机研究与发展，2015，52(2)：333-342.

[14] 王文剑，于剑，高阳. 面向大数据的人工智能技术专题前言[J]. 计算机研究与发展，2015，52(8)：1705-1706.

[15] 孟小峰，杜治娟. 大数据融合研究：问题与挑战[J]. 计算机研究与发展，2016，53(2)：231-246.

[16] AGRAWAL R, SRIKANT R. Fast Algorithms for Mining Association Rules in Large Databases[C]// Proceedings of the 20th International Conference on Very Large Data Bases. Morgan Kaufmann, 1994: 487-499.

（下转第 421 页）

(3)通过改进的词频统计结果和时间聚类中心的对比可以发现演化过程基本一致，因此综合两者之间的关键词变化情况，可以加强对舆情的内容演化分析。同时注意到改进的词频统计所得到的关键词比聚类所得到的要多，分析所得到的内容更加完整，更具有代表性，也更符合舆情的实际演化情况。在进行分类聚类时，可以考虑用改进的 TFIDF 所得到的词频统计结果指导聚类的初始中心和类别数。

结束语　本文通过对加权词频统计和时序聚类结果的分析初步完成了对女排夺冠话题内容演化的模拟分析，有效地实现了时间序列的聚类，能够在舆情分析预测上提供支撑依据。同时以时间为维度，考虑公众的情感倾向，以里约奥运会女排夺冠为例，得到舆情从话题本身往话题内在影响方面转移的演化趋势，提高了话题内容演化分析的准确性，从而能够在舆情发展上提供一定的指导。但本文需要在以下两个方面进一步深化研究：一是数据爬取的全面性和完整性；二是数据预处理中对噪声和无用数据的过滤。

参 考 文 献

[1] 刘毅. 网络舆情研究概论[M]. 天津人民出版社，2007.

[2] 侯万友. 群体性突发事件微博舆情演化分析[D]. 哈尔滨：哈尔滨工业大学，2013.

[3] 王雪梅，李晓峰，高巍巍. 一种改进的 K-Means 聚类算法的研究[J]. 计算机与数字工程，2013，41(11)：1717-1719.

[4] 刘慧婷，倪志伟. 基于 EMD 与 K-means 算法的时间序列聚类[J]. 模式识别与人工智能，2009，22(5)：803-808.

[5] 李深洛. 基于特征的时间序列聚类[D]. 桂林：广西师范大学，2014.

[6] 韩娜. 聚类算法在时间序列中的研究与应用[D]. 广州：广东工业大学，2011.

[7] 黄晓军，王博，包秀国. 基于层次分析法的微博文本特征权重计算方法[J]. 通信学报，2016，37(12)：50-55.

[8] 雷春，付业勤. 旅游网络舆情事件的时空分布与演化规律分析——以海南旅游热点事件为例[J]. 韶关学院学报，2014，35(1)：114-119.

(上接第 410 页)

[17] HAN J W，KOPERSKI K. Discovery of spatial association rules in geographic information databases[C]//Proceedings of the 4th International Symposium on Advances in Spatial Databases，Maine. USA，1995：47-66.

[18] HOUTSMA M，SWAMI A. Set-oriented mining for association rules in relational databases[C]//Proceedings of the Eleventh International Conference on Data Engineering. 1995：25-33.

[19] CHEN C，YAN X F，ZHU F D，et al. Graph OLAP：a multi-dimensional framework for graph data analysis[J]. Knowledge and Information Systems，2009，21(1)：41-63.

[20] 何清，李宁，罗文娟，等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能，2014，27(4)：327-337.

[21] 沈斌，姚敏. 一种新的动态关联规则及其挖掘算法[J]. 控制与决策，2009，24(9)：1310-1315.

[22] 李玲娟，张敏. 云计算环境下关联规则挖掘算法的研究[J]. 计算机技术与发展，2011，21(2)：43-46.

[23] 杨勇，王伟. 一种基于 MapReduce 的并行 FP-growth 算法[J]. 重庆邮电大学学报(自然科学版)，2013，25(5)：651-659.

[24] BOUKOUVALA F，DUBEY A，et al. Computational Approaches for Studying the Granular Dynamics of Continuous Blending Processes，2-Population Balance and Data-Based Methods[J]. Macromolecular Materials and Engineering，2013，297(1)：9-19.

[25] LIN T Y. Granular computing[M]//Rough Sets，Fuzzy Sets，Data Mining，and Granular Computing. Springer Berlin Heidelberg，2003：16-24.

[26] ZADEH L A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic[J]. Fuzzy Sets System，1997，90(2)：111-127.

[27] PAWLAK Z. Rough sets[J]. International Journal of Computer and Science，1982，11：341-356.

[28] 张铃，张钹. 模糊商空间理论(模糊粒度计算方法)[J]. 软件学报，2003，14(4)：770-776.

[29] 徐计，王国胤，于洪. 基于粒计算的大数据处理[J]. 计算机学报，2014，37(11)：1-22.

[30] 王国胤，张清华. 不同知识粒度下粗糙集的不确定性研究[J]. 计算机学报，2008，31(9)：1588-1598.

[31] YAO Y. Perspectives of granular computing[C]//2005 IEEE International Conference on Granular Computing. 2005：85-90.

[32] YAO Y Y. Granular computing：basic issues and possible solutions[C]//Proceedings of the Fifth Joint Conference on Information Sciences. 2000：186-189.

[32] 张清华，王国胤，胡军. 多粒度知识获取与不确定性度量[M]. 北京：科学出版社，2013.

[33] 苗夺谦，王国胤，刘清. 粒计算：过去、现在与展望[M]. 北京：科学出版社，2007.

[34] 王国胤，李德毅，姚一豫，等. 云模型与粒计算[M]. 北京：科学出版社，2012.

[35] 张钹，张铃. 粒计算未来发展方向探讨[J]. 重庆邮电大学学报(自然科学版)，2010(5)：538-540.

[36] 钟珞，吴珺. 粒度计算在数据仓库挖掘中的应用[J]. 华中师范大学学报(自然科学版)，2009，43(3)：392-395.