

数据仓库及其在城市规划决策 支持系统中的应用探讨^{*}

谢 榕

(武汉测绘科技大学城市建设学院, 武汉市珞喻路 129 号, 430079)

摘 要 在分析传统决策支持系统在城市规划决策应用中存在的问题的基础上, 初步提出一种基于数据仓库的城市规划决策支持系统的基本框架, 探讨了该系统建立中数据仓库的数据组织、数据挖掘、知识发现方法等关键技术问题, 并进一步阐述城市规划决策支持系统的建立方法, 最后以荆州市环境规划为例, 说明数据仓库在城市规划决策支持系统中的具体应用。

关键词 数据仓库; 数据挖掘; 知识发现; 城市规划决策支持系统

分类号 TU98

城市规划决策支持系统支持复杂空间问题的决策研究, 它将数据库、分析模型、决策者的知识、图形显示、列表报表及用户接口等集成在一个统一的系统中。为了支持这种研究, 一般认为它应具有决策支持系统的基本特点^[1]。同时, 为了处理空间问题, 系统还应具备以下特点^[1, 2]: ①提供多源空间数据的输入与存储功能; ②表达复杂空间关系与结构; ③提供空间分析技术; ④提供多种空间形式的输出功能。

但是这种决策支持系统也面临着其发展上的巨大障碍, 投入应用的成功实例并不多, 主要原因是: ①决策支持涉及大量历史数据和半结构化问题, 在传统数据库技术基础上建立城市规划决策支持系统, 只能提供辅助规划决策的数据级支持, 难以求解复杂的半结构化决策问题。②决策支持系统以集成数据为基础。现实中数据往往分散管理且大多分布于异构的数据平台, 数据不易集成。③决策支持系统的建立需要对数据、模型、知识和接口进行集成。数据库语言数值计算能力较低, 采用数据库管理技术建立的决策支持系统知识表达和知识综合能力较弱, 难以满足人们日益提高的决策要求。

90 年代初, 数据仓库技术的发展给以上问题的解决带来了新的契机。数据仓库来自各个数

据库的信息进行集成, 从事物的历史和发展的角度来组织和存储数据, 供用户进行数据分析, 并辅助决策支持^[3, 4], 成为决策支持的新型应用领域。由于数据仓库理论与技术的研究尚处于起步阶段, 将它用于规划决策支持还有许多特殊的问题亟待解决。本文着重探讨这一技术在城市规划决策支持系统中的应用及理论框架。

1 城市规划决策支持系统的基本结构框架

基于数据仓库的城市规划决策支持系统基本结构框架如图 1 所示, 系统的主要输入是数据库中的数据以及知识库中的知识和经验。数据仓库管理模块完成规划数据仓库的创建以及数据仓库中数据的综合、提取等各种操作, 负责整个系统的运转。数据挖掘工具用于完成实际决策问题中的各种查询、多维数据分析和数据开采。知识发现模块控制并管理知识发现过程, 将数据的输入和知识库中的信息用于驱动数据选择过程、知识发现引擎过程和发现的评价过程。人机交互模块通过自然语言处理和语义查询在规划部门和系统之间提供相互联系的集成界面。

图 1 中箭头方向为控制流。决策支持同数据

仓库管理是密切联系的, 用户发出决策请求命令后, 通过数据挖掘工具触发数据仓库管理模块从数据仓库中获取与任务相关的数据。被选择的数据经过知识发现模块中的知识发现引擎抽取算法处理, 生成辅助模式和关系。这些模式和关系被评价后, 其中一些被认为感兴趣的数据将提供给规划部门应用。有些发现还可能加入到知识库中, 以用于后继的知识发现和知识评价。

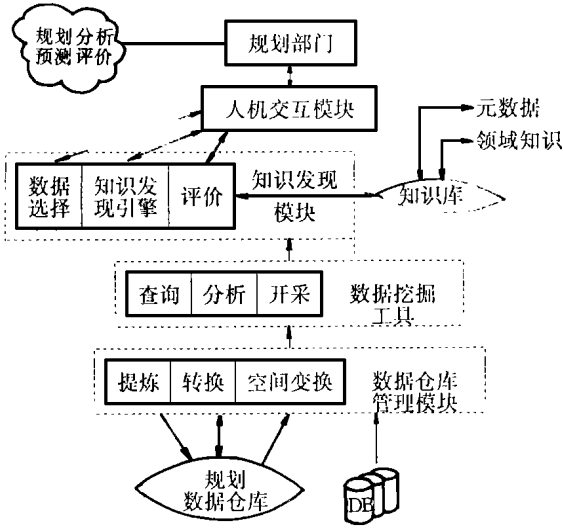


图 1 城市规划决策支持系统的基本结构框架

Fig. 1 Basic Framework of Urban Planning Decision Support System

数据仓库的数据存储根据对数据不同深度的处理, 分为 3 个不同的层次, 即历史数据层、综合数据层和专题数据层(如图 2 所示)。规划决策所需要的信息是通过基本数据所体现的整体趋向或随时间变化而表现出来的变化趋势, 必须对基本数据进行分类、析取、归纳、加工等处理才能得到这些信息。基本数据在时间控制机制下生成历史数据, 放入历史数据层, 用于城市动态趋势分析和模拟。在综合机制下对基本数据进行综合、提取生成综合数据, 放入综合数据层, 它包括社会、经济、人口、土地调查等各种统计分析数据、总规、分规、控制性详规等规划控制指标、评价计算结果、预测分析数据等。从传统数据库中获得的原始数据, 按照辅助决策的主题需求形成专题数据层。专题数据层存储对基本数据加工处理所形成的人口、资源、经济、环境等专用数据, 以及规划分析中所建立的数学模型、预测模型、评价模型中各类参数, 用于辅助城市规划与城市设计。

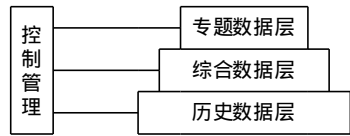


图 2 数据仓库的数据组织

Fig. 2 Data Organization in Data Warehouse

2 技术关键

基于以上提出的系统基本结构框架, 建立该系统的关键问题在于如何创建规划数据仓库, 如何从规划数据仓库中发现知识以及如何向用户解释和表达知识。

2.1 规划数据仓库的数据组织

城市规划与管理中涉及的数据主要来自规划部门内部具体的日常业务数据, 如基础地理信息、规划成果、建设用地、道路、综合管线、政策法规、人口、经济等, 一般常规数据库可对其进行处理, 这类数据又称为系统的基本数据。而城市规划决策支持分析需要大量数据, 这些数据可能分布于不同的应用中, 如土地利用、环境资源管理等; 或存储于不同的平台中, 如 Arc/Info、MapInfo 等。数据仓库为决策支持系统提供了可取的数据组织方式。数据仓库的建立并不取代传统的数据库应用, 它仍以基本数据为基础, 但是这些数据的增值与统一, 能最大限度地满足规划领域高层次决策分析的需要。

规划数据仓库涉及空间数据和属性数据。目前在大多数城市信息系统应用中, 空间数据存储于文件系统中, 属性数据由关系数据库来管理。空间数据和属性数据的分离管理必然造成系统查询与分析效率低下。规划数据仓库以一个多维数据库来实现。在规划数据仓库中, 空间数据和属性数据统一在一个数据仓库管理系统之下, 源数据必须经过提炼、转换和空间变换后方能进入数据仓库。通过提炼对数据项进行重构, 删去无用的信息, 检查数据的完整性和一致性, 并对数据仓库中的数据统一编码, 加上时间标志, 形成时间维数据, 同时将它们转换为统一的坐标系和相同的比例尺, 以形成空间维数据。空间维数据的组织是整个数据仓库建立的关键。当前关系数据库技术对规划数据仓库的支持能力不足, 因此可采用以下两种方法来支持数据仓库技术: (1) 对现有数据库作技术性改进, 引入面向对象技术, 建立面向对象数据库, 以适应规划数据仓库的要求。(2) 以当前关系数据库技术为基础, 按规划数据仓库的要求进行扩展和完善, 扩充数据类型, 并增加扩充

数据类型所需的操作和转换机制,使扩展的数据类型与关系数据库系统融合为一个整体。

2.2 数据挖掘的层次空间与知识发现

从信息源获取信息,并将它们组织集成到数据仓库,其目的在于对数据仓库中的数据进行分析和综合。数据挖掘是一种有效的工具,它从数据库中提取人们感兴趣的数据模式、数据的普遍关系及其潜在的、未知的数据特征^[5,6],即数据挖掘是在数据仓库的基础上进行知识发现的,因此这一技术在决策支持方面具有广泛的应用。

从应用深度上,将数据挖掘划分为3个层次空间(如图3所示):(1)数据空间。在该空间上利用现有数据库管理系统的查询检索功能,进行基于关键字的信息查询,实现联机事务处理(on-line transaction processing, OLTP)。例如,从城市用地分析中可获得某地区各类用地(如居住用地、工业用地等)的用地构成。(2)聚合空间。在该空间上利用聚集运算(Sum、Ave、Max、Min),结合多维分析和统计分析,实现在线分析处理(on-line analytical processing, OLAP),以提供决策参考的统计分析数据。例如,从城市人口分析中可获得某一地区、某一年龄段的文化程度分布。(3)影响空间。按照相似性的聚类、差异性的分类方法,发现关联性、相似时序、结构模式,从数据库中发现隐含的有用信息。这是在更深层次上的知识发现,是数据挖掘的实质性内涵。

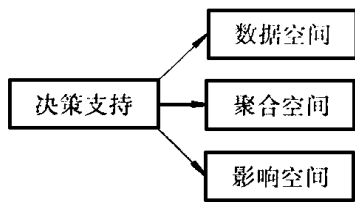


图3 数据挖掘的层次空间

Fig. 3 Level Space of Data Mining

在以上各个层次空间上进行数据挖掘反映了不同级别的查询请求,这种划分有利于知识的逐步提取,知识的提取过程即为决策支持过程。在传统决策支持系统中,知识库中的知识和规则由专家或程序人员建立,从外部输入,而数据挖掘是从系统内部自动获取知识的过程。同数据库管理系统查询检索的信息相比,数据挖掘的知识是隐含的、精练的和高水平的^[7]。

2.3 知识发现方法研究

从规划数据仓库中可发现的主要知识类型概括为:①地理几何知识,即有关目标的位置、形状、大小及结构等集合特征;②拓扑关系的知识,即空

间目标间相连、相邻、共生及包含等空间关系信息;③空间分布规律,即目标在地物空间的分布规律;④空间特征规则,即空间目标的几何与属性的对应关系;⑤区域差异规则,即区分不同类目标的特征;⑥区域演变规则,即目标依时间的变化规则。数据挖掘通常将这些知识表现为概念、规则、规律、模式、约束和可视化等形式,经过解释后可以直接在实际系统中应用,如城市人口分析、城市交通规划等辅助决策过程,或者提供给领域专家,修正专家已有的知识体系,也可以作为新的知识转存到应用系统的知识库中。发现的过程是使数据挖掘利用各种知识发现算法从规划数据仓库中发现、表达、更新和解释有关知识^[8]。

1) 基于关联规则的知识发现

数据关联是数据库中存在的一类重要的可被发现的知识^[9]。若两个或多个变量的取值之间存在某种规律性,则称为关联。数据间的关联通过关联规则表示,其形式为: $A_1 \wedge A_2 \wedge \dots \wedge A_i \rightarrow B_1 \wedge B_2 \wedge \dots \wedge B_j$ 。如果 B_1, B_2, \dots, B_j 出现,则 A_1, A_2, \dots, A_i 一定出现,这表明数据 A_1, A_2, \dots, A_i 和数据 B_1, B_2, \dots, B_j 之间存在某种联系。关联分析采用关联规则归纳技术找出目标在空间上的相连、相邻、共生和包含等空间拓扑关系^[10]。

2) 不确定性知识表示

人们在规划中对事物的判断、预测和决策是在问题领域的信息不完全、不精确或者模糊的条件下进行的。粗集理论作为一种智能数据决策分析工具,被研究并应用于这种不确定性的知识获取和知识表达中。它通过构造描述领域知识的概念集 $A = \{a_1, a_2, \dots, a_n\}$, $a_i \in A$, 由系统的事实库形成对象集 $E = \{e_1, e_2, \dots, e_m\}$, $e_j \in E$, 基于分类集 P 和对象集 E , 对 E 经过一定的操作产生核 $t^c(P, E)$ 和包络 $t^e(P, E)$, 从而形成该问题的一个粗集 $[t^c(P, E), t^e(P, E)]$, 构成不确定性区间,使用上限和下限两个量作为不确定性的测度。

3) 知识的更新和完善

神经网络通过对大量样本模式的学习,得到从 n 维输入向量空间到 m 维输出向量空间的非线性映射 F , 即 $F: R^n \rightarrow R^m$ 。利用神经网络输出结果,经专家认可后,将其作为新的样本实例存入系统中。因此,建立这种系统自学习模型,可以不断地从样本模式中学习专家用于城市决策的定性的、经验性的知识,从而保证系统不断地更新知识和获取新知识。

4) 知识的表达和解释

系统中最重要应用是用户能够理解所发现

的知识,这要求知识的展现不应局限于传统的数字或符号,而是更容易理解的方式,如表格、直方图、散点图或自然语言等。尤其在规划决策中更加注重时间和空间的过程及分布特征,这需要给规划分析、决策提供一个实时可视的集成环境。通过数据可视化表达方式对数据进行计算机交互和控制,以多维形式进行数据的表示和再表示、数据的可变视角显示和动态链接等,使规划人员交互地分析复杂的空间相互关系和动态变化趋势。

5) 预测建模

在数据仓库的全体数据记录上,建立待分类的样本集 U , 将待预测的样本 $u(u_1, u_2, \dots, u_n)$ ($u_1, u_2, \dots, u_n \in U$) 与数据仓库中已分类的模式作比较, 求出它们之间的贴近度, 即

$$(U, \text{Mode}_i) = (U \circ \text{Mode}_i + (1 - U \odot \text{Mode}_i)) / 2$$

式中, \circ 和 \odot 代表模糊运算中的内积和外积。根据择近原则, 即

$$(U, \text{Mode}_i) = \max((U, \text{Mode}_1), (U, \text{Mode}_2), \dots, (U, \text{Mode}_s))$$

判断该样本接近于哪一模式, 从这一模式的整体情况预测其结果。在规划中, 将城市发展水平的评价指标进行比较, 建立预测模型, 能确定综合因素影响下城市的最优发展目标。

另外, 统计 (statistics) 方法、模糊集 (fuzzy sets) 理论、决策树 (decision tree)、遗传算法 (genetic algorithm)、聚类 (cluster) 方法等也可用于数据挖掘和知识发现。

3 城市规划决策支持系统的建立

基于以上讨论, 建立城市规划决策支持系统的过程如下: (1) 分析决策需求, 确定决策主题, 描述和表示决策问题。(2) 确定数据来源, 将来自规划部门及相关部门的数据重新进行组织, 建立规划数据仓库。(3) 针对所要发现任务的所属类别, 设计或选择有效的数据挖掘算法并加以实现。(4) 调用数据挖掘功能, 从平凡的历史数据中提炼出综合数据, 并与最终用户交互、协同, 得到宏观性数据和趋势性知识。(5) 测试与评价所发现的知识, 对知识进行一致性、效用性处理。(6) 根据最终用户的要求, 建立适用于决策支持的集成界面和应用程序, 使用户能在决策支持中运用所发现的知识。

以上过程不是简单的线性流程, 而是一个学习、发现和修改的过程, 步骤之间包含了循环和反复, 这样可以对所发现的知识不断求精、深化, 并

使其易于理解。

4 应用实例

下面以荆州市环境规划为例, 具体说明数据仓库在城市规划决策支持系统中的应用。

4.1 系统目标

荆州市环境信息系统的目标是针对荆州市的环境状况和条件, 对该市环境现状作出综合评价, 为环境规划与环境发展目标提供科学的依据^[1]。

荆州市环保局除直接管理所辖三区的污染源外, 还接收所辖县市环保局的汇总统计信息, 同时定期向省局上报环境信息。现在其管理范围已扩大到四县三市三区, 管理业务更加繁重, 管理内容与范围也大为增加。已建成的环境信息系统形成了一套比较完整的环境指标体系和数据汇总渠道, 但由于数据量的增加, 高效率的管理成为十分迫切的需要。本次研究主要是在原有系统基础上进一步采用数据仓库技术建立环境规划决策支持系统, 在此系统的支持下开展环境规划决策将十分有利。

4.2 数据的组织

系统的基础数据由区域背景信息、环境信息与辅助信息组成。区域背景信息包括基础底图、行政区划图、土地利用图及规划图。环境信息包括污染源、环境监测、环境管理、环境统计等数据; 辅助信息包括人口、经济、水文与气象等数据。

针对系统目标, 将以上基础数据重新进行组织, 建立面向主题的、不同时间的、集成的空间数据仓库, 其设计采用 SQL Server 7.0 实现。具体包括以下内容:

(1) 支持环保各业务的专题数据, 如环境监测、排污收费管理、“三同时”管理、财务管理、污染源治理、环保设施管理等。

(2) 不同时期的历史数据 (目前时间区间为 5~10 年), 如大气、废气、降水酸度、河流水质、城市地下水、工业废水、城市饮用水、土壤、交通噪声、城市区域环境噪声、功能区噪声的日监测数据、污染源月排污状况、月排污费征收、污染事故记录等。

(3) 综合数据, 如环境资源、灾害污染、人口经济等生态环境因子、环境质量评价等。

4.3 系统功能

为了支持以上多种复杂的数据应用和综合性决策分析, 在该数据仓库的基础上建立规划决策支持系统, 系统总体结构如图 4 所示。系统功能

包括:

(1)比较分析。按照相似性聚类,发现关联性
及结构模式,对该市不同时段的环境现状进行对
比分析,从变化数量、变化的空间位置和分布上反
映环境变化趋势和水平。通过对荆州市 1993、
1994 年大气监测数据的统计分析,作出大气污染
因子 SO₂ 浓度等值线图。分析表明,SO₂ 浓度呈
现出由郊区向市区逐步增高的趋势。利用浓度等
值线分析,反映污染变化趋势,有利于快速找出该
地区污染的主要原因。

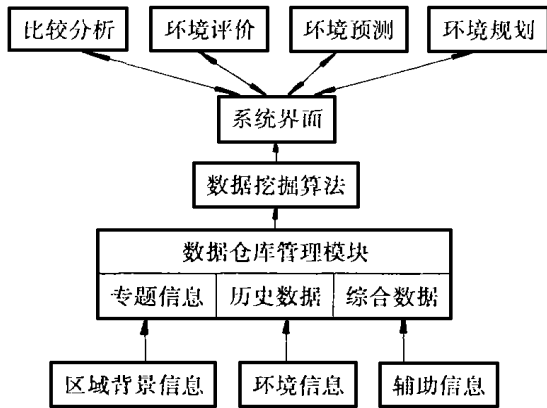


图 4 环境规划决策支持系统总体结构

Fig. 4 System Architecture of Environment Planning Decision Support System

(2)环境评价。通过数据挖掘综合、转化数据
库中资源、灾害、污染、人口、经济等规划数据,确
定对环境产生影响的主要因素,给出环境评价图
与相应结果的数据、表格。在对 1993 年荆州市环
境质量评价中,选择环境污染、人口状况中的 10
个评价因子(废水日排放量、废气日排放量、废渣
日排放量、SO₂、NO_x、水环境超标量、噪声、人口
密度、人均居住面积、人均绿化面积)建立评价模
型,以图表形式反映出该地区污染状况。

(3)环境预测。运用预测模型,预估若干年后
城市人口、用地、经济发展达到一定范围时对环
境的影响,以进行城市发展趋势预测分析。

(4)环境规划。通过可视化工具,对环境综合
治理规划实施方案进行模拟分析,给出环境质量、
功能区近期、远期规划图及优化结果,使规划人员
了解方案实施后城市的状况及发展水平。

4.4 技术特点

初步研究表明,在数据仓库的基础上利用数
据挖掘进行数据分析、开采,提供来自种类不同的
应用系统集成化数据和历史化数据,能为有关规
划部门进行全局范围战略决策和长期趋势分析提

供有效支持。另外,随着数据仓库、数据挖掘产品
的逐步完善,决策支持系统的应用会更加成熟。

5 结 论

城市规划决策支持系统的建立是一项复杂的
系统工程。研究表明,数据仓库技术融合了数据
库理论、统计学、数据可视化和人工智能技术等多
项研究领域,与数据挖掘的结合,在大量数据中发
现有价值的知识,是新的决策支持结构。基于这
一技术的城市规划决策支持系统的建立为规划人
员提供了强有力的支持工具,能有力地推动规划
决策的现代化进程。进一步的研究工作还包括规
划数据仓库的时空数据模型、矢栅一体化数据挖
掘的算法、城市规划决策支持系统的集成等。

参 考 文 献

- 1 Maguire D J, Goodchild M. Geographical Information Systems Principles and Applications. New York: John Wiley & Sons, Inc., 1991. 403~412
- 2 Leung Y. Intelligent Spatial Decision Support Systems. New York: Springer-Verlag, 1997. 1~10
- 3 Moriarty T. Modeling Data Warehouse. Database Programming and Design, 1996
- 4 胡 侃, 夏绍玮. 基于大型数据仓库的数据采掘. 软件学报, 1998, 9(1): 53~61
- 5 Anand S S, Scotney B W. Designing a Kernel for Data Mining. IEEE Expert Intelligent Systems and Their Applications, 1997(3, 4)
- 6 姚卿达, 黄晓春, 刘向民. 数据仓库和数据采掘应用研究. 计算机科学, 1996, 23(6): 63~65
- 7 邱凯昌, 李德仁, 李德毅. 空间数据挖掘和知识发现的框架. 武汉测绘科技大学学报, 1997, 22(4): 328~332
- 8 Matheus C, Chan P K, Piatetsky-Shapiro G. System for Knowledge Discovery in Database. IEEE Trans. on Knowledge and Data Eng., 1993, 5(6): 903~913
- 9 王清毅, 陈恩红, 蔡庆生. 知识发现的若干问题及应用研究. 计算机科学, 1997, 24(5): 73~77
- 10 Edwin M K, Raymond T N. Finding Aggregate Proximity Relationships and Commonalities in Spatial Data Mining. IEEE Trans. on Knowledge and Data Eng., 1996. 884~897
- 11 黄正东, 谢 榕. 荆州市环境信息系统. 测绘信息与工程, 1998(2): 16~21

谢 榕, 女, 31 岁, 副教授, 现主要从事城市信息系统和数据库研究。代表成果:《城市信息系统》。

E-mail: luoxin@public.wh.hb.cn

Data Warehouse and Its Application in Urban Planning Decision Support System

XIE Rong

(School of Urban Studies, WTUSM, 129 Luoyu Road, Wuhan, China, 430079)

Abstract Under the urgent requirements of information development, research on data warehouse is now raising and becomes a new international frontier research direction. Firstly, the paper analyzes some problems in traditional urban planning decision support system, that is: (1) This system can only provide with data support of aided planning decision, but without semi-structured decision problems if traditional database techniques are used. (2) Integrated data are the basis of decision support system, however, it is not easy to integrate data which distribute in varied heterogeneous data platform. (3) If the decision support system is established on the basis of database management system, the ability of spatial problem of knowledge expression and knowledge synthesis will be not very strong because of weak calculation ability of database. Secondly, the paper presents preliminarily system architecture of urban planning decision support system based on data warehouse. The system consists of database, planning data warehouse, database warehouse management module, data mining tool, knowledge base, knowledge discovery module and user interface. The key problems are how to create planning data warehouse, how to discover knowledge from data warehouse, and how to apply knowledge discovered to the practice. Thirdly, the paper discusses organization of planning data warehouse, data mining techniques and methods of knowledge discovery. Data in urban planning and management come from basic geographical information, planning results, and so on. The planning data warehouse can be divided into three levels: historical data level, synthesized data level and special data level. The historical data level is used as urban trend analysis and urban simulation. The synthesized data level includes varied statistical data, planning control index, evaluation results and forecast analytical data. The special data level stores data of population, resource, economics, environment, and some parameters in mathematical models, prediction models and evaluation models. The planning data warehouse can be built with multi-dimensional database. The key problem of establishment of whole data warehouse is the organization of spatial dimensional data. Data mining is just knowledge discovery on data warehouse. It can apply to decision support. It is advantageous for knowledge discovery gradually when the level spaces of data mining are established. The process of knowledge discovery is the process of decision support. Comparing with information from database management system, the knowledge from data warehouse should be hidden, refined and high standard. In data mining, knowledge can be regarded as concepts, rules, regulars, modules, restrains and visualization. After explanation, these knowledge can be applied into practical system. The paper also presents some methods of knowledge discovery. On the above discusses, a method of establishment of urban planning decision support system is proposed in the paper. Finally, taking an example of Jingzhou Environmental Information System, it presents system application under the support of urban planning decision support system.

Key words data warehouse; data mining; knowledge discovery; urban planning decision support system