

# 基于网格的聚类算法的虚拟物流企业伙伴选择方法研究<sup>\*</sup>

沈立新<sup>1</sup> 陈燕<sup>1</sup> 崔春雷<sup>2</sup> 王军<sup>1</sup> 李玖晖<sup>1</sup>

(大连大学管理学院 大连 116026)<sup>1</sup> (大连水产学院信息工程学院 大连 116023)<sup>2</sup>

**摘要** 虚拟物流企业已成为 21 世纪信息社会生产、物流的主流组织形式。准确地选择和确定合作伙伴是建立虚拟物流企业的核心。本文提出了应用基于网格的聚类算法来选择虚拟物流企业合作伙伴。该算法能对虚拟物流企业信息平台提供的庞大的高维的成员企业信息数据库进行快速有效的数据挖掘,从而在全球范围内寻找潜在的合作伙伴,以供虚拟物流企业核心企业作为下一步精选的根据。实例仿真说明了该算法的有效性。

**关键词** 虚拟物流企业,伙伴选择,网格,聚类算法

## Research on Virtual Logistics Enterprise Partner Selection Method of Grid-Based Clustering Algorithm

SHEN Li-Xin<sup>1</sup> CHEN Yan<sup>1</sup> CUI Chun-Lei<sup>2</sup> WANG Jun<sup>1</sup> LI Jiu-Hui<sup>1</sup>

(Management College, Dalian Maritime University, Dalian 116026)<sup>1</sup> (Information Engineering College, Dalian Fisheries University, Dalian 116023)<sup>2</sup>

**Abstract** Virtual logistics enterprise (VLE) is a mainstream organizational form for production and logistics in the information society of the 21<sup>st</sup> century. Correct selection of partners is one of the key links in establishing a VLE. A selection partner method based on grid of clustering algorithm is advanced in this paper, which can realize data mining in the bulkiness and high dimension information database of member enterprises provided by VLE information platform in order to find the latent partners in the whole globe and act as the basis for accurate selection of core enterprises of VLE. A simulating case will demonstrate the effectiveness of the selection method of the grid-based clustering algorithm.

**Keywords** Virtual logistics enterprise (VLE), Partner selection, Grid, Cluster algorithm

## 1 引言

第三方物流是目前物流行业运作的主要形式。而我国绝大部分都是中小型第三方物流企业,它们大多由原来的仓储企业、运输公司或货物代理类企业转型而来。由于先天功能的不足,加之物流系统复杂,涉及政府不同层次的多个管理部门、多个环节,物流企业经营管理模式跨行业的合并制度障碍凸现,因此他们在市场竞争中处于劣势。面对委托方提出的零库存、准时制、快速反应等日趋苛刻的要求,特别是由于经济全球化、网络化进程的加速以及当代信息技术的高速发展,我国的物流企业开始了对发展方向的思考。把虚拟企业技术引入到物流行业,建立不涉及资产合并,基于物流业务基础上的柔性的松散的虚拟物流企业既符合第三方物流本身具有的特点,又符合我国物流行业发展的现实情况。然而成功组建虚拟物流企业的前提是能否准确地选择灵活的、有竞争力的、能与核心企业相容的合作伙伴。

## 2 问题背景分析

对于虚拟企业伙伴的选择人们提出了多种模型,主要有:两阶段的伙伴选择过程模型<sup>[1]</sup>,文[2~4]对此方法进行了修正;以层次分析法与模糊数学为基础实现虚拟企业盟友选择模型<sup>[5]</sup>;基于活动的成本分析方法(ABC)<sup>[6]</sup>;基于遗传算法的

伙伴选择优化模型<sup>[7~9]</sup>,多阶段多指标的理想方案法<sup>[10]</sup>;时序多指标决策的灰色关联分析法<sup>[11]</sup>;投影决策方法<sup>[12]</sup>及案例推理法和神经网络方法等等。但针对虚拟物流企业(VLE)伙伴选择方法问题的研究却很少。

物流行业作业的最大特点就是物质实体在空间和时间上的流动。目前随着信息技术的飞速发展,打破了时间和空间对物流活动的限制,为国家间、企业间的经济关系发展提供了新的手段和条件。网络通信、数据仓库、数据挖掘、GIS、GPS 等技术在物流行业的应用,整个物流行业的运作环境、操作技术和行业规则也面临着新的巨大的变革。跨国跨地区作业日益频繁,所以虚拟物流企业(VLE)的核心企业在全世界范围内寻找合作伙伴的问题日益突出。而且基于物流行业的特点,在选择伙伴时采用的评价指标需增加诸如运输时间均值、运输时间方差、配送质量等等的指标。所以虚拟物流企业伙伴选择时所面临的问题,就是要对一个相对更庞大的高维的成员企业的信息数据库进行数据挖掘,从而在全球范围内找到潜在的合作伙伴,作为核心企业进行下一步精选的根据,而旧的伙伴选择方法已很难适应新时代虚拟物流企业伙伴选择的需要了。

数据挖掘是目前全球范围内重点投资研究的十大新技术之一,它引起了学术界和工业界的广泛关注。聚类分析是数据挖掘中重要的技术手段,是对数据进行分类的过程,同一类

<sup>\*</sup> 国家自然科学基金资助项目:多维空间数据集成环境下数据挖掘模型的研究,基金号:70540005。沈立新 博士生,研究方向:物流管理,数据仓库,数据挖掘,多 Agent 系统等。陈燕 教授,博士生导师,研究方向:数据挖掘,数据仓库,物流管理,多 Agent 系统等。崔春雷 博士生,讲师,研究方向:数据挖掘,数据仓库。王军 博士生,副教授,研究方向:交通运输规划与管理。李玖晖 博士生,副教授,研究方向:交通运输规划与管理。

的尽可能相似,不同类的尽可能相异<sup>[13]</sup>。它的主要算法有:(1)分裂法,典型算法有 K-Mean<sup>[14]</sup>, Fuzzy C-Means<sup>[15]</sup>, K-modes<sup>[16]</sup>和 K-prototypes<sup>[17]</sup>。算法给定参数 K, 给定一个初始分类方法, 然后通过迭代的方法改变分组, 使得每一次改进得到的分组方案都较前分组更好。(2)层次法, 这类算法将两个相似的对象合并到一个聚类中。这个过程时间花费大, 因为要对数据库中的数据进行两两比较。典型的大型数据库层次聚类算法有 BIRCH<sup>[18]</sup>, CURE<sup>[19]</sup>, CHAMELEON<sup>[20]</sup> 以及 ROCK<sup>[21]</sup>。(3)基于密度的算法, 是将数据的密集区域看做一个聚类, 可以得到任意形状的聚类。典型算法有 DBSCAN<sup>[22]</sup>, DENCLUE<sup>[23]</sup>和 OPTICS<sup>[24]</sup>。(4)基于网格的算法考虑单元格而非数据点, 它在分段数据空间网格结构上进行操作。因为这一特点, 基于网格的聚类算法计算效率更高。典型算法有 STING<sup>[25]</sup>, STING +<sup>[26]</sup>, WaveCluster<sup>[27]</sup>, CLIQ-UE<sup>[28]</sup>以及 GDILC<sup>[29]</sup>。(5)基于模型的算法是给聚类假定一个模型, 寻找能满足此模型的数据聚类。

由于虚拟物流企业(VLE)伙伴选择时要对一个高维数据

库进行聚类分析(本文所选实例中用于评价伙伴的指标就是 15 维, 为了探询评价指标与评估结果之间的关系, 我们还加了一维评估结果, 这样一共 16 维), 数据量大而且复杂, 聚类分析要求能够自动识别出包含有聚类的子空间。BIRCH, CLARANS 以及 DBSCAN 算法虽然也可用于解决高维数据库, 但是, 这些算法发现的是全维空间的聚类, 无法实现寻找原数据空间中包含聚类的子空间。CLIQUE 算法能够自动识别包含聚类的子空间, 以及能够寻找任意类型的聚类。所以本文给出的伙伴选择方法是基于 CLIQUE 聚类算法来实现的。它的优点是:(1)可自动发现存在聚类的最高维子空间, 而无需用户指定在哪一个子空间中对原数据表格进行聚类分析;(2)能够发现任意形状的聚类;(3)算法所需时间与要处理的总的记录数呈线性关系。

### 3 算法分析与实现

#### 3.1 算法框架(图 1)

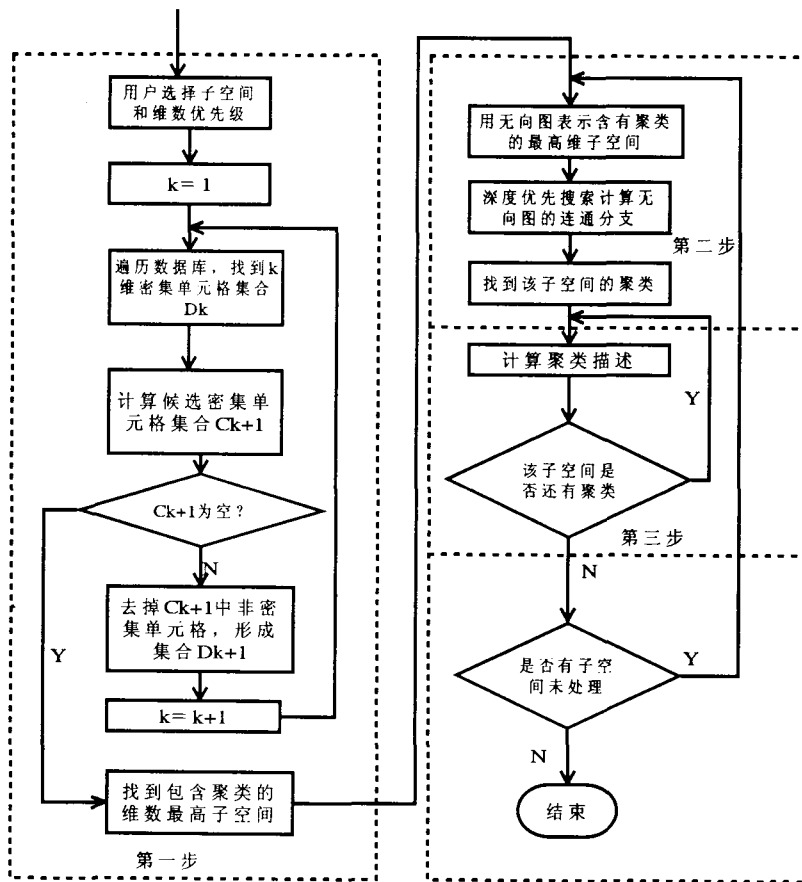


图 1

### 3.2 算法具体描述

3.2.1 寻找包含有聚类的子空间 本系统面对的是高维数据空间, 如果把所有的子空间都考虑, 这样的计算量和空间耗费都是很大的。所以采用自底向上的策略, 首先通过遍历一次数据库, 得到一维密集单元格。以此为基础, 得到 2 维候选密集单元格, 以此类推, 一直到无法产生更高维的密集单元格。设最高维是  $k$ , 对  $k-1$  维密集单元格集合  $D_{k-1}$  进行自连接, 连接条件是单元格的前  $k-2$  维相同, 生成  $k$  维候选密

集单元格集合  $C_k$ 。再遍历一次数据库, 把其中的非密集单元格滤掉, 得到了第  $k$  维密集单元格  $D_k$ 。这样作的依据是基于单调性原理<sup>[1]</sup>, 若  $k$  维单元格  $S$  是密集的, 则  $S$  是  $k-1$  维空间的投影的一部分, 则相应  $k-1$  维单元格也是密集的。

由于密集单元格的最高维是  $k$  维子空间, 因此其中产生过至多  $2^k$  个子空间。每增加一维子空间要遍历一次数据库来过滤掉非密集单元格, 设数据集大小为  $m$ , 则遍历数据库用  $mk$  的时间。所以此步的时间耗费最多为  $2^k + mk$ 。

尽管上面操作大大降低了处理单元格的数目,但是对于高维数据库而言,这个数目对于计算机而言仍然是很大的,因为随着子空间维数的增加,候选密集单元格的数目增加很快。对候选密集单元格作适当的裁减是必要的。此时我们采用MDL(最小描述长度)裁剪,按照某一模型对数据进行编码,目标是使得编码长度最短的方案,将覆盖率高的子空间保留在候选子空间中,其它的就被裁剪了。

假设所有的候选子空间为  $n$  个,裁剪点的位置  $1 < i < n$ ,首先要对这  $n$  个候选子空间排序,所需时间是  $O(n \log n)$ 。然后对这个排序队列进行两次扫描,第一次扫描计算  $\mu(i)$  和  $\mu_p(i)$ ,第二次扫描计算总编码长度。总时间耗费是  $O(n \log n + 2n) = O(n \log n)$ 。

3.2.2 寻找聚类 这一步是从包含有聚类的最高维子空间的密集单元格集合  $D$  中聚类。即要找到集合  $D$  的一个划分  $\{D_1, D_2, \dots, D_k\}$ ,要求  $D_i$  中的所有密集单元格都是相邻的,位于不同分类的单元格是不相邻的且子集互不相交。

聚类是由连通的密集单元格组成的连通分支。单元格的相邻关系可以用图论中的相关概念来表示。单元格可以用一个点来表示,若两个单元格相邻则在相应的点之间添加一条无向边。这样把算法第一步中产生的密集单元格集合  $D$  转换成了一个无向图,单元格的相邻关系转换成了图中结点的连接关系,由于图中没有环和多重边,因此构成了一个简单图。寻找子空间的聚类问题就转化为在无向图中寻找连通分支问题。寻找可以利用图的深度优先搜索算法来寻找图的连通分支。此步的时间耗费与最高维密集单元格的数据成线性关系。

3.2.3 生成聚类描述 这一步要进行的操作是对于前两步产生的密集单元格的每一个聚类  $C$ ,产生一个区域的集合  $R$ 。 $C$  是  $k$  维子空间中的一个密集单元格的集合, $R$  中的任一个成员都包含在  $C$  中,且  $C$  中的任一个单元格至少包含在  $R$  的一个成员中。

产生聚类描述问题已被证明是 NP-hard 的,即使在 2 维空间中的描述也是很难的<sup>[30,31]</sup>,最优覆盖是找到包含聚类的最小矩形。有一些计算二维覆盖的比较好的算法<sup>[17]</sup>,但并不适用于高维覆盖。对于一般覆盖,最好的算法由文<sup>[33,34]</sup>给出,但也只是适用于低维数据库中。考虑到本系统是高维数据库,算法分两步:1)贪心搜索策略尽可能搜索到一些极大

矩形覆盖;2)去掉多余矩形描述生成最小覆盖。

复杂度分析:第一步,得到每一个极大矩形区域  $R$ ,设  $|R|$  表示  $R$  中包含的密集单元格数目,贪心搜索会执行  $O(|R|)$  时间。设  $S$  是  $R$  所在的子空间, $k$  是  $S$  的维数, $n$  是  $S$  中密集单元格的个数。算法必须搜索  $R$  包含的每一个单元格来确认  $R$  是否是聚类的一部分。此外,它还必须访问  $R$  的每一个邻接单元格来确保  $R$  是极大的。邻接单元格的个数  $\leq 2k|R|$ 。因为每一个新的极大区域最少覆盖一个到目前为止没有包括的密集单元格,所以贪心算法至多会找到  $O(n)$  个新的区域,每个区域需要访问  $O(|R|) = O(n)$  个单元格,所以此处所需的时间小于  $O(n^2)$ 。假定子空间  $S$  中有  $n$  个密集单元且只包含一个聚类,界限为两个平行的超平面以及一个柱面。因为超平面不和任意一维平行,在这个超平面内,和聚类接触的单元格为  $O(n^{(k-1)/k})$  个,这些单元格都应该包含在极大区域中,每个区域的尺寸也是  $O(n^{(k-1)/k})$ ,因为每一个区域必须和另一个超平面相邻,所以贪心算法要访问  $O(n^{2(k-1)/k})$  个密集单元格。第二步,寻找最小覆盖。这一步比较好实现,首先需要区域进行排序,代价为  $O(n \log n)$ ,对每个区域  $R_i$  要访问  $|R_i|$  个单元格,总的时间为  $O(n^2)$ 。

综上所述,算法的三步总的时间复杂度为:

$$O(C^* + mk) + O(n \log n) + O(n) + O(n^2) = O(C^* + mk) + O(n^2)$$

$K$  为最高维密集子空间的维数, $m$  为数据库中记录的个数, $n$  为密集单元格的个数。

#### 4 算例分析

本文示例中把系统每一维分成  $\xi = 10$  个区间,一维子空间共有  $\xi \times k = 160$  个单元格,经过遍历数据库,得到的密集单元格为 81 个,一维密集单元格占有率约为 50%,而二维单元格本应有 24000 个子空间,但由一维密集单元格出发,得到的二维候选子空间为 3109 个,占有率约为 12.95%。到最后得到的 16 维密集单元格为 28 个,与实际的 16 维单元格  $10^{16}$  相比,这个数目就小得多了。由此看出,  $2^{16}$  的时间耗费大大降低了,基本上与数据量成线性关系。

下面表 1~表 3 分别就代表了源数据表格,以及表示结果的区间定义表格和聚类信息表格。

表 1 源数据表格

序号	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
1	150	15	5	400	2000	90	30	1.33	4	4	3	4	0.65	65	4	0.8
2	50	15	5	400	2000	90	30	1.67	4	4	4	4	0.65	65	4	0.88
3	150	15	3	200	1000	90	30	1.67	4	4	4	4	0.65	65	4	0.95
4	145	15	5	400	2000	90	30	1.33	4	4	3	4	0.65	65	4	0.85
5	130	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	1.0
6	130	10	1	50	200	100	15	0.67	3	3	3	2	0.85	70	5	0.40
7	130	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	1.0
8	130	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	1	0.7
9	130	10	1	50	200	70	35	1.67	5	5	5	5	0.85	70	5	0.75
10	130	10	1	50	200	100	15	1.67	5	3	3	5	0.85	70	5	0.80
11	150	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	0.95
12	165	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	0.12
13	160	10	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	0.52
14	130	30	1	50	200	100	15	1.67	5	5	5	5	0.85	70	5	0.70
15	130	10	5	50	200	100	15	1.67	5	5	5	5	0.85	70	5	0.66
16	130	10	3	50	200	100	35	1.67	5	5	5	5	0.85	70	5	0.76
17	170	25	8	2000	5000	60	60	0.67	2	2	2	2	0.95	95	1	0.05
18	135	25	8	2000	5000	60	60	0.67	2	2	2	2	0.10	8	1	0.05
19	135	15	8	50	500	99	15	0.67	2	2	2	2	0.90	85	4	0.05
20	155	15	4	200	2000	99	30	1.33	4	4	4	3	0.70	80	4	0.75

表 3  $\xi=10$ ,各个区间上下界

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
0	0	10	0	50	200	60	15	0	2	2	0	0	0.10	0.1	1	0.01
1	20	12.5	1	250	700	65	20	0.2	2.4	2.4	0.6	0.6	0.19	10.1	1.5	0.11
2	40	15	2	450	1200	70	25	0.4	2.8	2.8	1.2	1.2	0.28	20.1	2.0	0.21
3	60	17.5	3	650	1700	75	30	0.6	3.2	3.2	1.8	1.8	0.37	30.1	2.5	0.31
4	80	20	4	850	2200	80	35	0.8	3.6	3.6	2.4	2.4	0.46	40.1	3.0	0.41
5	100	22.5	5	1050	2700	85	40	1.0	4.0	4.0	3.0	3.0	0.55	50.1	3.5	0.51
6	120	25	6	1250	3200	90	45	1.2	4.4	4.4	3.6	3.6	0.64	60.1	4.0	0.61
7	140	27.5	7	1450	3700	95	50	1.4	4.8	4.8	4.2	4.2	0.73	70.1	4.5	0.71
8	160	30	8	1650	4200	100	55	1.6	5.2	5.2	4.8	4.8	0.82	80.1	5.0	0.81
9	180	32.5	9	1850	4700	105	60	1.8	5.6	5.6	5.4	5.4	0.91	90.1	5.5	0.91
10	200	35	10	2050	5200	110	65	2.0	6.0	6.0	6.0	6.0	1.0	100.1	6.0	1.01

表2 源数据表格中各项含义备注

A1:单位总成本(元/个)	A5:交货量期望值/个	A9:战略一致性(1-5)	A13 市场结构(Gjni 系数)
A2:运输时间均值/天	A6:对客户要求的反应率/%	A10:运输质量体系(1-5)	A14:市场规模/%
A3:运输时间均方差/天	A7:对客户要求反应时间/天	A11:供应体系质量(1-5)	A15:竞争环境(1-5)
A4:交货量最小值/个	A8:关键工序能力(CPk)	A12 制造质量保证(1-5)	A16:评价结果

表4 聚类结果

聚类号	A16	A15	A14	A13	A12	A9	A8	A5	A4	A1
0	6-9	8	6	8	8	7	8	0	0	6-8
0	9	6	6	6	6	5	8	4	0	7
0	7-8	6	6	6	6	5	6	9	1	7
1	0	0	8-9	8-9	3	0	3	9	9	8
1	0	6	8	8	3	0	3	0	0	6
2	1	8	6	8	6	7	8	0	0	7
3	3	8	6	8	3	2	3	0	0	6
4	5	8	6	8	8	7	8	0	0	8
5	3-4	2	0	0	8	0	8	0	3-4	7

我们共找到6个聚类,这里对其中的两个进行描述如下:

#### 聚类1描述:

$((120) \leq A1 < 180) \wedge (50 \leq A4 < 250) \wedge (200 \leq A5 < 700) \wedge (1.6 \leq A8 < 1.8) \wedge (4.8 \leq A9 < 5.2) \wedge (4.8 \leq A12 < 5.4) \wedge (0.82 \leq A13 < 0.91) \wedge (60.1 \leq A14 < 70.1) \wedge (5.0 \leq A15 < 5.5) \wedge (0.61 \leq A16 < 1.01) \vee ((140 \leq A1 < 160) \wedge (50 \leq A4 < 250) \wedge (2200 \leq A5 < 2270) \wedge (1.6 \leq A8 < 1.8) \wedge (4.0 \leq A9 < 4.4) \wedge (3.6 \leq A12 < 4.2) \wedge (0.64 \leq A13 < 0.73) \wedge (60.1 \leq A14 < 70.1) \wedge (4.0 \leq A15 < 4.5) \wedge (0.91 \leq A16 < 1.01)) \vee ((1140 \leq A1 < 160) \wedge (250 \leq A4 < 450) \wedge (4700 \leq A5 < 5200) \wedge (1.2 \leq A8 < 1.4) \wedge (4.0 \leq A9 < 4.4) \wedge (3.6 \leq A12 < 4.2) \wedge (0.64 \leq A13 < 0.73) \wedge (60.1 \leq A14 < 70.1) \wedge (4.0 \leq A15 < 4.5) \wedge (0.71 \leq A16 < 0.91))$

#### 聚类2描述:

$((160 \leq A1 < 180) \wedge (1850 \leq A4 < 2050) \wedge (4700 \leq A5 < 5200) \wedge (0.6 \leq A8 < 0.8) \wedge (2.0 \leq A9 < 2.4) \wedge (1.8 \leq A12 < 2.4) \wedge (0.82 \leq A13 < 1.00) \wedge (80.1 \leq A14 < 100.1) \wedge (1.0 \leq A15 < 1.5) \wedge (0.01 \leq A16 < 0.11)) \vee ((120 \leq A1 < 140) \wedge (50 \leq A4 < 250) \wedge (200 \leq A5 < 7000) \wedge (0.6 \leq A8 < 0.8) \wedge (2.0 \leq A9 < 2.4) \wedge (1.8 \leq A12 < 2.4) \wedge (0.52 \leq A13 < 0.91) \wedge (80.1 \leq A14 < 90) \wedge (4.0 \leq A15 < 4.5) \wedge (0.01 \leq A16 < 0.11))$

从上面聚类的结果我们可以发现 A8, A9, A12 和 A16 也偏低的伙伴企业会聚为一类;而 A1 和 A13 较高 A12、A16 偏低的一些伙伴企业会聚为一类。然后我们就可以根据不同 VLE 的核心企业的不同需要对他们进行了选择。

**结论** 本文在分析了国内外有关虚拟企业伙伴选择的各种方法和数据挖掘各种聚类算法基础上,针对虚拟物流企业的特点,综合已有的研究成果,提出了应用基于网格的聚类算法来选择虚拟物流企业伙伴的方法。该算法能够对信息平台提供的庞大的成员企业信息数据库进行较快速的数据挖掘,从而在全球范围内寻找潜在的合作伙伴,作为核心企业进行下一步精选的根据。实例仿真表明,该算法对于属于高维系统的虚拟物流企业伙伴选择问题具有较高的应用价值。

## 参考文献

- Talluri S, Baker R C. A quantitative frame work for designing efficient business process alliances[A]. Intl. Conf. on Engineering Management and Control(IEMC)[C]. 1996. 656~661
- 陆金伟, 达庆利. 一种选择虚拟企业伙伴的框架[J]. 管理工程学报, 2000, 14(4): 19~22
- 戴勇. 基于 AHP 的 DEA 分析基础上的虚拟物流企业联盟伙伴选择[J]. 系统工程, 2002, 20(3): 47~51
- 陈菊红, 汪应洛. 虚拟企业伙伴选择过程及方法研究[J]. 系统工程理论与实践, 2001, 21(7): 48~53
- 王丹, 杨晓春, 等. 基于模糊层次分析法首先实现虚拟企业中的伙伴选择[J]. 东北大学学报(自然科学版), 2000, 21(6): 607~609
- Roodhooft F, Konings J. Vendor selection and evaluation: An Activity-based Costing Approach[J]. European Journal of Operation Research, 1996, 96: 97~102
- 冯蔚东, 陈剑, 赵纯均. 基于遗传算法的动态联盟伙伴选择过程及优化模型[J]. 清华大学学报(自然科学版), 2000, 40(10)
- 沙磊, 徐晓飞, 李全龙, 谢炜. 基于遗传算法的动态联盟性能标准

优化方法[J]. 计算机集成制造系统, 2002, 8(6): 462~466

- Ip W H, Huang Min. Genetic algorithm solution for a risk-based partner selection problem in a virtual enterprise[J]. Computers & Operations Research. 2003 (30): 213~231
- 刘家学, 郑昌义. 多阶段多指标决策的理想方案法[J]. 系统工程理论与实践, 2001, 21(1): 61~64
- 刘家学. 时序多指标决策的灰色关联分析法[J]. 运筹与管理, 1997, 6(3): 6~10
- 王应明. 多指标决策与评价的新方法—投影法[J]. 系统工程与电子技术, 1999, 21(3): 1~4
- Han Jiawei, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000
- MacQueen J B. Some methods for classification and analysis of multivariate observations. In: Proc. of the 5th Berkeley Symposium on Mathematics Statistics and Probability, 1967. 281~297
- Dunn J C. A fuzzy relative of the ISODATA process and its use detecting compact well-separated clusters. Journal of Cybernetics, 1974(3): 2~57
- Huang Z. A fast clustering algorithm to cluster very large categorical data sets in data mining. In: SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997
- Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values, Data Mining Knowledge Discovery, 1998(2): 283~304
- Zhang T, Ramakrishnan R, Livny M. BIRCH: An efficient data clustering method for very large databases. In: Proc. of the ACM SIGMOD Conf. on Management of Data, Montreal, Canada, June 1996
- Guha S, Rastogi R, Shim K. CURE: An Efficient Clustering Algorithm for Large Data sets. In: the Proc. of the ACM SIGMOD Conf., 1998
- Zhao Y C, Song J. GDILC: a grid-based density-isoline clustering algorithm. In: Proc. of Intl. Conf. on Info-tech and Info-net, 2001, 3: 140~145
- Guha S, Rastogi R, Shim K. ROCK: a robust clustering algorithm for categorical attributes. In: Proc. of 15th Intl. Conf. on Data Engineering, 1999. 512~521
- Ester M, Kriegel H-P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial data sets with noise. In: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, Portland, OR, 1996. 226~231
- Hinneburg A, Keim D. An efficient approach to clustering in large multimedia data sets with noise. In: Proc. of the 4th Intl. Conf. Knowledge Discovery and Data Mining, 1998. 58~65
- Ankerst M, Breunig M, Kriegel H, Sander J. OPTICS: Ordering Points to Identify the Clustering Structure. In: Proc. ACM SIGMOD99 Int. Conf. on Management of Data
- Wang W, Yang J, Muntz R. STING: A Statistical Information Grid Approach to Spatial Data Mining. In: Proc. of the 23rd VLDB Conf. Athens, Greece, 1997
- Wang W, Yang J, Muntz R. STING+: an approach to active spatial data mining. In: Proc. of 15th Intl. Conf. on Data Engineering, 1999. 116~125
- Sheikholeslami G, Chatterjee S, Zhang A. Wave Cluster: a wavelet-based clustering approach for spatial data in very large data bases. VLDB J., 2000(8): 289~304
- Agrawal R, Gehrke J, Gunopulos D, et al. Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications. In: Proc. of the 1998 ACM-SIGMOD Conf. on the Management of Data, 1998. 94~105
- Karypis G, Han E H, Kumar V. CHAMELEON: A hierarchical clustering algorithm using dynamic modeling. Computer, 1999, 32(8): 68~75
- Masek W. Some NP-complete set covering problems; [ M. S. Thesis, MIT]. 1978
- Reckhow R A, Culberson J. Covering simple orthogonal polygon with a minimum number of orthogonally convex polygons. In: Proc. of the ACM 3rd Annual Computational Geometry Conf., 1987. 268~277
- Franzblau D. Performance guarantees on a sweep-line heuristic for covering rectilinear polygons with rectangles. SIAM J. Disc. Math, 1989(2): 307~321
- Feige U. A threshold of  $\ln n$  for approximating set cover. In: Proc. of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, 1996. 314~318
- Friedman J. Optimizing a noisy function of many variables with application to data mining. In: UW/MSR Summer Research Institute in Data Mining, July 1997