

第七届中国 Rough 集与软计算、第一届中国
Web智能、第一届中国粒计算联合会议

相容粒度空间模型 及其应用

史忠植 郑征

shizz@ics.ict.ac.cn

中国科学院计算技术研究所



中国科学院
计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY

报告提纲

✓ 引言

- 相容粒度空间模型
- 基于相容粒度空间模型的信息分类
- 基于相容粒集的双层决策模型
- 相容粒度空间的其它应用
- 结束语

粒度计算理论

- 粒度计算从广义上来说是一种看待客观世界的世界观和方法论。
- 粒度计算的基本思想就是使用**粒**而不是**对象**为计算单元，使用粒、粒集以及粒间关系进行计算或问题求解。

粒度计算的必要性

● 从哲学的角度看

Yager和Filev指出“人类已经形成了世界就是一个粒度的观点”以及“人们观察、度量、定义和推理的实体都是粒度”。信息粒是一种抽象，它如同数学中的“点”、“线”、“面”一样，在人类的思维和活动中占有重要地位。

● 从人工智能的角度看

张钹院士指出“人类智能的公认特点，就是人们能从极不相同的粒度上观察和分析同一问题。人们不仅能在不同粒度的世界上进行问题求解，而且能够很快地从一个粒度世界跳到另一个粒度的世界，往返自如，毫无困难。这种处理不同世界的的能力，正是人类问题求解的强有力的表现”。

粒度计算的必要性

● 从优化论的角度来看

粒度计算的理论与方法在观念上突破了传统优化思想的束缚，不再以数学上的精确解为目标，即：需要的是很好地理解和刻画一个问题，而不是沉溺于那些用处不大的细节信息上。粒度计算的方法不要求目标函数和约束函数的连续性与凸性，甚至有时连解析表达式都不要，而且对计算中数据的不确定性也有很强地适应能力，计算速度也快，这些优点使粒度计算具有更广泛地应用前景，所以，粒度计算理论的研究对推动优化领域的发展极其重要。

粒度计算的必要性

● 从问题求解的角度看

用粒度计算的观点来分析解决问题显得尤为重要，这样就不用局限于具体对象的细节。除此之外，将复杂问题划分为一系列更容易管理和更小的子任务，可以降低全局计算代价。

● 从应用技术的角度看

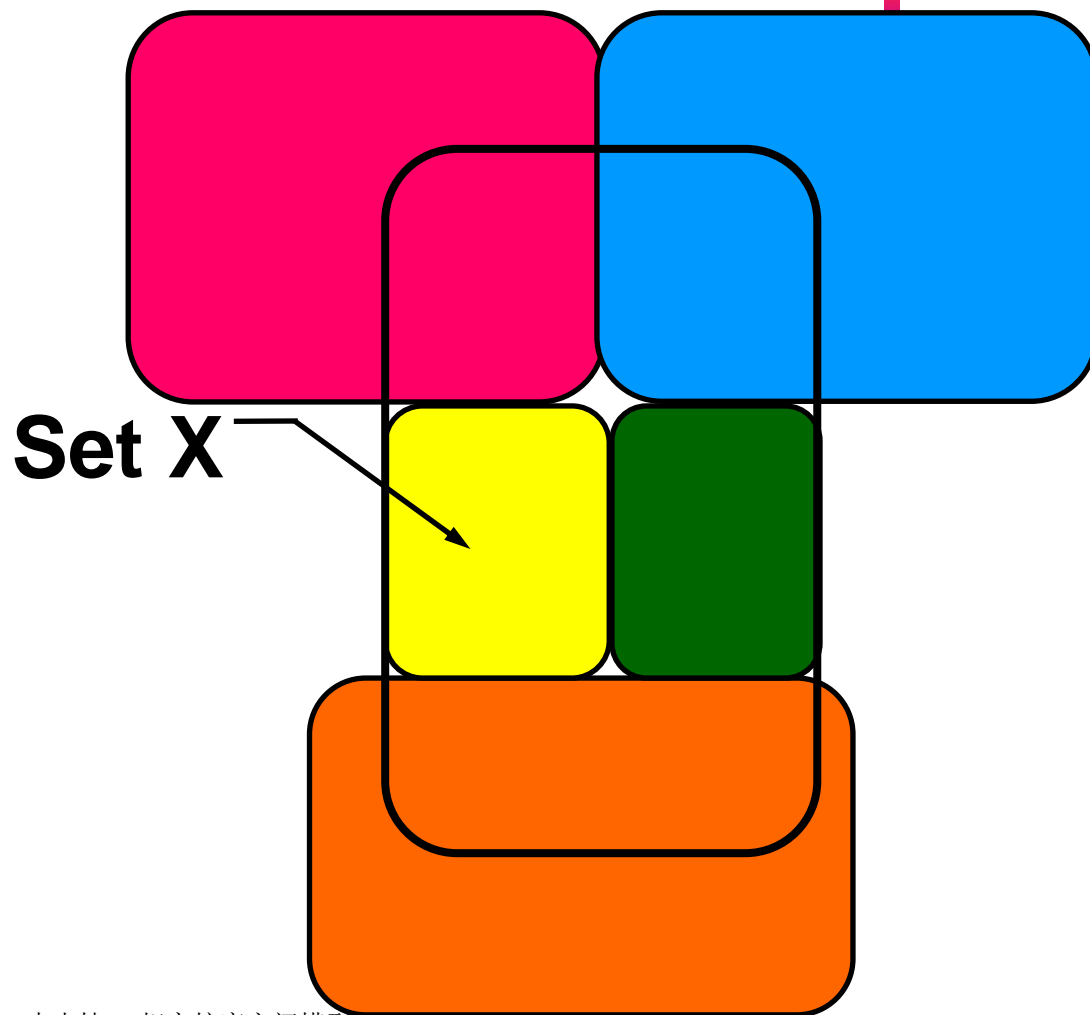
图像处理、语音与字符识别等，是计算机多媒体的核心技术。这些信息处理质量的好坏直接依赖于分割的方法和技术，而粒度计算的研究或许能够解决这一问题。

粒度计算的基本问题

- 两大问题
 - 粒的构造：处理粒的形成、表示和解释
 - 使用粒的计算：处理在问题求解中粒的运用
- 两个方面
 - 从语义上：侧重于对粒的解释，如为什么两个对象会在同一个粒之中，为什么不同的粒会相关。
 - 从算法上：如何进行粒化和如何进行基于粒的计算。对粒的分解与合并方法的研究，是构建任何粒度体系结构的本质要求。

粗糙集理论

- 粒：等价类，子集
- 粒的计算：粒之间的近似



商空间理论

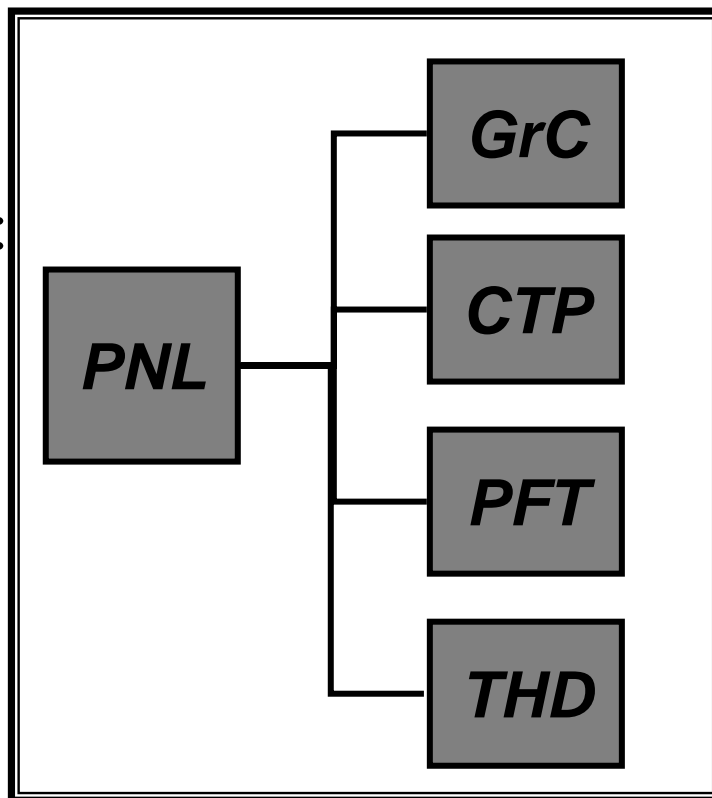
- ▶ 粒：等价类，子集，粒之间具有拓扑关系
- ▶ 粒的计算：合成、分解



词计算理论

CW

- ▶ 粒：词
- ▶ 粒的计算：
模糊数学



Granular Computing

*Computational Theory
of Perceptions*

Protoform Theory

*Theory of Hierarchical
definability*

**CW = Granular Computing + Generalized-Constraint-Based
Semantics of Natural Languages**

粒度计算的进展

- T. Y. Lin在粒度计算的结构、表示和应用方面进行了一系列的研究
- Y. Y. Yao等人对粒度计算的基本问题、基本方法进行了一系列研究；
- Witold Pedrycz在模糊数学相关的粒度计算领域做了一系列的研究；
- A. Skowron等在近似空间的基础上提出了粗糙神经计算及相关信息粒系统；

粒度计算的进展

- 总的来说，在粒度计算这个新兴的研究领域，上述各个方面的研究都还处于一种理论上的探索阶段，许多方面都不成熟，但这些思想为我们进一步研究打下了基础。

报告提纲

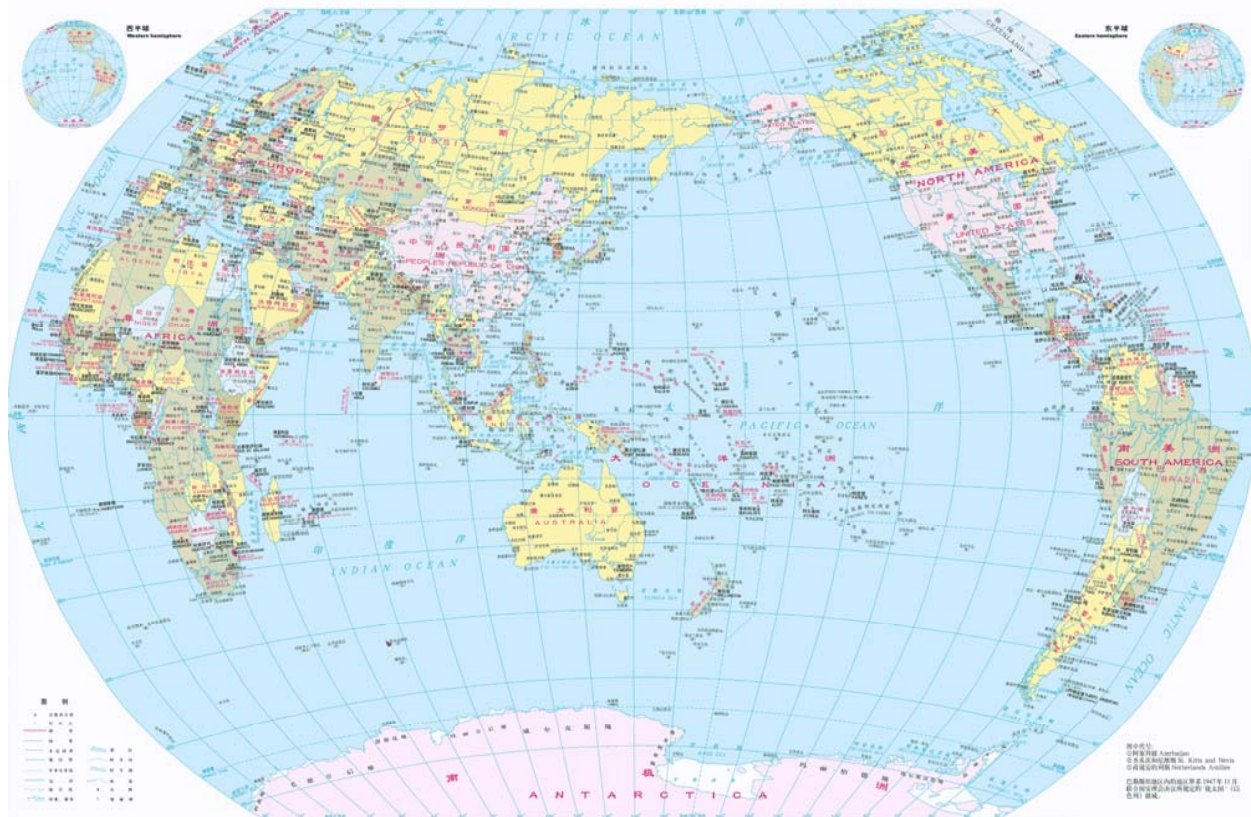
- ✓ 引言
- ✓ 相容粒度空间模型
 - 基于相容粒度空间模型的信息分类
 - 基于相容粒集的双层决策模型
 - 相容粒度空间的其它应用
 - 结束语

基本思路



- 人类不仅具有从极不相同的粒度上观察和分析同一问题的能力，更具有根据实际问题建立合适的粒度空间的能力。

世界地图



基本思路

- 人们在认识和解决问题时能够很快的从一个粒跳到另一个粒，或者从一个粒度层次转到另一个粒度层次，这需要粒度空间上的结构。
- 我们认为粒度计算的功能不仅仅在于对问题的简化和近似化，更在于以粒为单位通过粒和粒之间的关系进行计算在某些问题的解决中不可替代的作用。

粒度模型

- 基于近似和相容关系的粒度模型
 - 近似空间
 - 变精度粗糙集模型
 - 相容空间
- 层次和嵌套模型
 - 由嵌套等价关系序列引导的嵌套粗糙集近似
 - 由层次结构引导的层次粗糙集近似
 - 由邻域系统引导的层次粗糙集近似

问 题

- 粒的定义：子集，没有内涵，无法区分粒和类
- 粒的元素：粒的元素为基本对象，不能为粒
- 粒的嵌套层次结构简单
- 粒的功能是用于描述和近似，而对于问题求解作用不大（明显）

相容粒度空间模型

- 用一个三元组来描述相容粒 $G=(IG, EG, FG)$
 - IG : 相容粒 G 的内涵，用向量表示
 - EG : 相容粒 G 的外延，用向量的集合表示
 - FG : 内涵和外延之间的转换函数

描述了相容粒在特定环境下表现的知识，并表示在一个特定任务下相容粒中所有元素的一般性特征、规则、共同性等

包含这个粒所涵盖的对象或粒

定义了粒的内涵和外延之间的转换，可以用函数、规则、算法等形式来描述

相容粒

- 粒的定义需要考虑三个基本属性：
 - 反映粒中元素交互作用的内部属性；(**IG**)
 - 揭示一个粒与其它粒交互作用的外部属性；(**EG**)
 - 表示一个粒在特殊环境下存在的上下文属性。(**IG, FG**)

相容粒之间的关系

- 内涵关系 $G_1 R G_2 = IG_1 R' IG_2$
- 外延关系 $G_1 R G_2 = EG_1 R'' EG_2$
- 复合关系

$$G_1 R G_2 = (IG_1, EG_1) R'' (IG_2, EG_2)$$

相容粒的合成和分解技术

- 合成：小粒合成大粒，依赖于相容粒之间的关系
 - $EG = EG_1 \cup \dots \cup EG_n$
 - $EG = \{x / (x, a) \in tr \wedge x \in (EG_1 \cup \dots \cup EG_n)\}$
 - $EG = \{G / (IG, ig) \in tr \wedge IG \in \{IG_1\} \cup \dots \cup \{IG_n\}\}$
 - $EG = \{G / (IG, ig) \in tr_1 \wedge IG \in \{IG_1\} \cup \dots \cup \{IG_n\}\} \cup \{x / (x, a) \in tr_2 \wedge x \in (EG_1 \cup \dots \cup EG_n)\}$
- 分解：大粒分成小粒
 - $EG_i = \{x / (x, a) \in tr' \wedge x \in EG\}$

相容粒度空间模型

● 四元组(OS , TR , FG , NTC)

- OS 表示对象集
- TR 表示一个相容关系
- FG 表示相容粒转换函数
- NTC 表示一个嵌套相容覆盖

对象集系统由在相容粒度空间中处理和粒化的对象组成，它也可以看成是一个对象域

相容关系系统是一个参数化的关系结构，它由一组相容关系组成，包括一个粒度空间所基于的关系和参数

嵌套相容覆盖系统是一个参数化的粒度结构，其中定义了不同层次的粒和基于对象系统和相容关系系统的参数化过程。它定义了：

- 粒之间、粒集之间、对象之间以及粒和对象之间的关系；
- 粒的合成和分解。

对象集

- 对象本身也具有层次结构。例如，在图象处理中，对象存在像素、图像、视频流这样的层次结构。
- 对象和粒：不同之处在于对象是实际存在的数据和知识，而相容粒是根据具体任务背景通过粒化对象得到的知识结构，即可以认为是虚拟的知识或依赖于任务和上下文的知识。

对象集

定义（对象）： O_0 ，称为原始对象，表示为 R^n 中的一个向量，其中 R 是实数集。 O_1 ，称为一阶子集对象，表示为由原始对象构成的集合。一般的说， O_{k+1} ，称为 $k+1$ 层子集对象，表示为 k 阶子集对象 O_k 构成的一个集合。

基于上面的定义，一个对象集系统可以形式化为：

$$OS = \{ \bigcup_P \{ O_{o_r} \} \} \cup \dots \cup \{ \bigcup_P \{ O_{o_r} \} \} \cup \dots$$

对象集

例如，在图像处理中， $\mathbf{O}_0 = (x, y, R, G, B)$ 表示一个像素，其中 x, y 表示该像素的坐标值， R, G, B 表示该像素的RGB颜色值。 \mathbf{O}_1 表示一幅图像，并且

$$\mathbf{O}_1 = \begin{bmatrix} \mathbf{O}_{0_{11}} & \mathbf{O}_{0_{12}} & \cdots & \mathbf{O}_{0_{1m}} \\ \mathbf{O}_{0_{21}} & \mathbf{O}_{0_{22}} & \cdots & \mathbf{O}_{0_{2m}} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{O}_{0_{n1}} & \mathbf{O}_{0_{n2}} & \cdots & \mathbf{O}_{0_{nm}} \end{bmatrix}$$

其中是 $\mathbf{O}_{0_{ij}}$ 一个像素， (i, j) 表示该像素在 \mathbf{O}_1 中的位置； \mathbf{O}_2 表示一段视频流。

相容关系

相容关系系统由一组相容关系组成，它可以被形式化为：

$$TR = \cup tr_{(cp, \omega, DIS, D)}$$

相容命题 cp ，权重向量 ω ，距离函数向量 DIS 和半径向量 D 是相容关系系统的四个关键要素。许多空间形状可以由相容关系系统来描述，例如：

$$cp_1(\alpha, \beta | D) = dis(\alpha, \beta | \omega) \leq d = \omega_i (\alpha_i - \beta_i)^2 \leq d$$

$$cp_2(\alpha, \beta | D) = ((\omega_i | \alpha_i - \beta_i|) \leq d_i)$$

其中当 α 或者 β 固定时， cp_1 描述一个球， cp_2 描述一个n维立方体。

嵌套相容覆盖系统的构建

- 自顶而下构建嵌套相容覆盖系统的算法

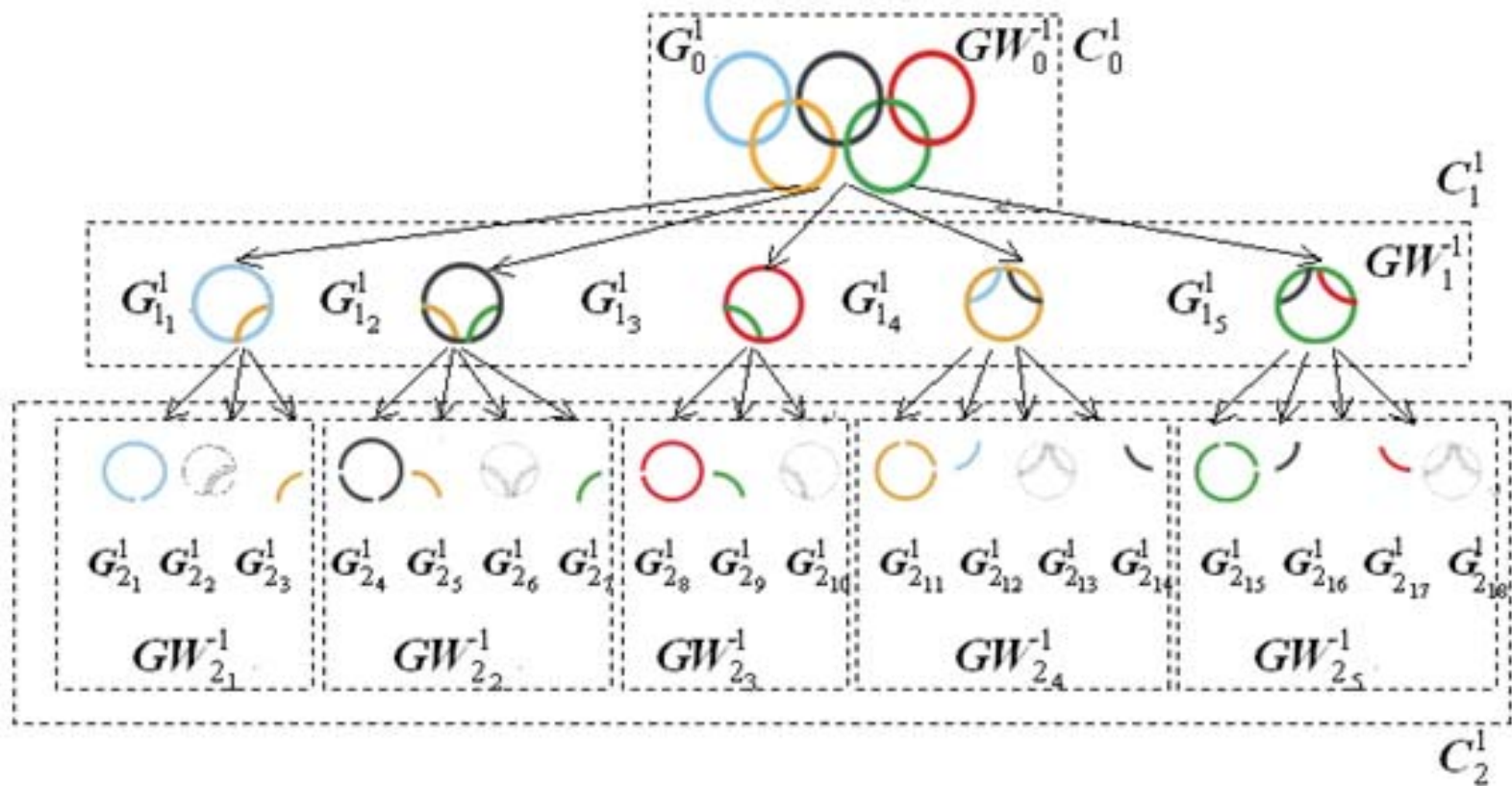
$$EG_{i+1}^1(\eta_{i+1}^1 | tr_{i+1}^1) = \{x | (x, \eta_{i+1}^1) \in tr_{i+1}^1(\varphi_{i+1}, \sigma_{i+1}, DIS_{i+1}, D_{i+1}) \wedge x \in EG_i^1\},$$

- 自底而上构建嵌套相容覆盖系统的算法

$$EG_{i+1}^1(\eta_{i+1}^1 | tr_{i+1}^1) = \{x | (x, \eta_{i+1}^1) \in tr_{i+1}^1 \wedge (x \in O_1)\},$$

$$EG_{i+1}^1(\eta_{i+1}^1 | tr_{i+1}^1) = \{G_i^1 | (IG_i^1, \eta_{i+1}^1) \in tr_{i+1}^1 \wedge G_i^1 \text{ 为一个 } i \text{ 层相容粒}\},$$

相容覆盖



相容粒度空间模型的主要特点

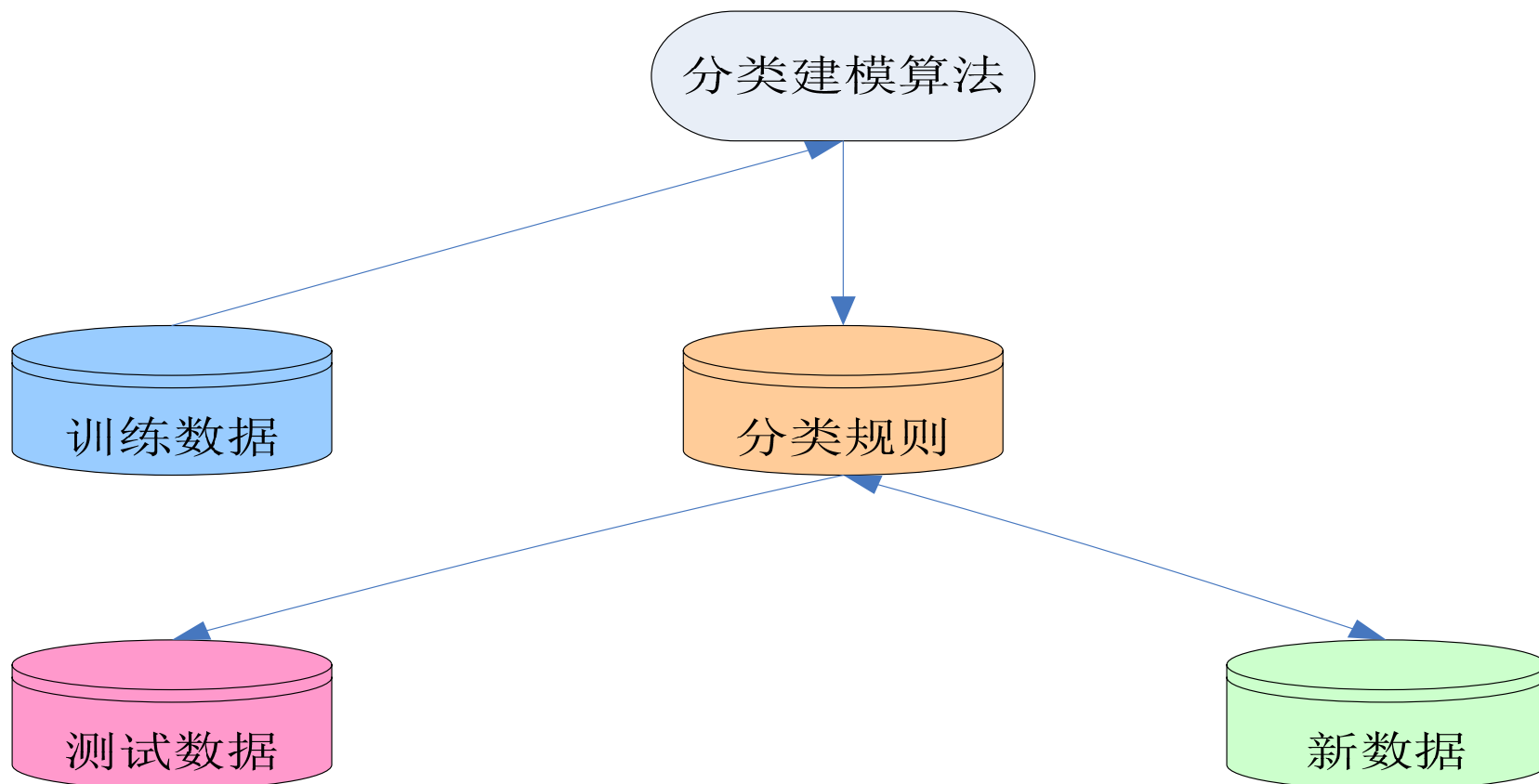
- 功能特点
 - 粒度计算的功能不仅仅在于对问题的简化和近似化，更在于以粒为单位通过粒之间的关系进行计算在某些问题的解决中不可替代的作用。
- 建模所基于关系的特点
 - 相容关系
- 粒的定义的特点
- 粒度空间结构的特点
 - 通过定义粒的三种关系：内涵关系、外延关系和复合关系，以及粒度空间的层次和嵌套结构实现了这种粒之间和粒度层次之间交互跳跃的能力。

模型	哲学基础	粒的定义	粒的性质	粒的层次	关系	在模型上	研究的对象	研究重点
模糊词计算理论	人类在进行思考、判断、推理时主要是用语言进行的，而语言本身就是“粒度”。	词。	模糊的粒。	没有明确的层次概念。	模糊关系。	粒是用“词”来表示，而描述“词”的有效的方法就是模糊集理论。	主要研究如何描述由词界定的不同粒度的对象，它更擅长描述由形容词、副词表达的不同粒度的概念。	主要是讨论粒度的表示问题，他们认为人类是用语言进行各种思考和推理的，不同的词就表示不同的粒度。
粗糙集理论	人的知识就是一种分类的能力。	子集。	清晰的粒。	没有明确的层次概念。	等价关系。	粒用子集来表示，不同粒度的概念可以用不同大小的子集来表示，所有这些表示可以用等价关系来描述。	通过元素的不同属性值，来描述元素之间的关系，并用元素按不同属性进行的分类来表示不同的概念粒度。	研究粒度的表示、刻划和粒度与概念之间的依存关系；其论域是简单的点集，元素之间没有拓扑关系。
商空间理论	人类智能的公认特点，就是人们能从极不相同的粒度上观察和分析同一问题。	子集。	清晰的粒。	有。	等价关系。	除了用子集来表示粒，更引入了拓扑结构，并通过粒度世界的合成与分解来解决问题。	通过元素的不同属性值，来描述元素之间的关系。除此之外，还引入元素之间的关系 T （用拓扑来描述）。	研究不同粒度世界之间的互相转换、互相依存的关系，是描述空间关系学的理论；在论域元素之间存在有拓扑关系的情况下进行研究的。
相容粒度空间理论	人类具有根据特定环境对相关数据和知识泛化或者特化成不同程度、不同大小的粒的能力，以及进一步根据这些粒和粒之间的关系进行求解的能力。	内涵、外延和转换函数。	清晰的粒。	有。	相容关系。	粒用相容粒来表示，一个粒不仅仅是对象的集合，更是这些对象的抽象。模型包含四个组成部分，对象系统，相容关系系统，转换函数和嵌套覆盖系统。	除了元素的属性（特征），元素之间的关系，还引入和区分了粒的内涵关系、外延关系和复合关系，以及依据此类关系构建的嵌套覆盖结构。	研究在具体任务背景下，对现有的知识和数据进行粒度化的理论和方法，以及在构建的粒度空间上根据粒、粒度层次以及粒度空间之间的往返交互实现问题的求解。

报告提纲

- ✓ 引言
- ✓ 相容粒度空间模型
- ✓ 基于相容粒度空间模型的信息分类
 - 基于相容粒集的双层决策模型
 - 相容粒度空间的其它应用
 - 结束语

相容粒度空间模型 在信息分类中的应用



信息分类中 相容粒度空间模型的建立

- 粒的定义：
 - 内涵：决策规则的向量形式
 - 外延：原始对象的集合
- 粒之间的关系：
 - 内涵关系，即决策规则之间根据所覆盖对象的情况定义的包含关系。
- 对象系统：
 - 一阶对象：原始对象
 - 二阶对象：决策表

信息分类中 相容粒度空间模型的建立

- 相容关系系统:

$$\text{dis}(\alpha, \beta | \omega) = \sum_{i=1}^{n-1} \omega_i (\alpha_i \oplus \beta_i),$$

其中 $\alpha = (\alpha_1, \dots, \alpha_n)$, $\beta = (\beta_1, \dots, \beta_n)$, 且

$$\alpha_i \oplus \beta_i = \begin{cases} 0, & \text{if } |\alpha_i - \beta_i| \leq r \\ 1, & \text{else} \end{cases}.$$

- 嵌套相容覆盖系统的
 ▶ 自顶而下的构建算法
 ◆ 粒的外延

$$EG_{i+1}^1(\eta_{i+1}^1 | \tau_{i+1}^1) = \{x | (x, \eta_{i+1}^1) \in \tau_{i+1}^1(\omega_{i+1}^1, \text{DIS}, D) \wedge x \in EG_i^1\};$$

基于相容粒度空间的信息分类建模算法

- 基于相容粒度空间的信息分类建模算法**TGM**

Step 1: 初始化;

Step 2: 计算1层粒;

Step 3: 递归地, 从*i*层父粒中分解生成*i+1*层相容粒;

Step 4: 最终, 我们得到嵌套相容覆盖系统为:

$$NTC = \{NTC_1\}。$$

相容粒度格

假设 $\langle GS, \leq \rangle$ 是一个偏序集, 对于任意两个相容粒 $G_i^k \in GS$ 和 $G_j^p \in GS$ ($G_i^k \neq G_j^p$), 满足下列条件的两个相容粒 G_l^m 和 G_e^f 称为 G_i^k 和 G_j^p 的上界相容粒和下界相容粒:

- 1) $G_i^k \leq G_l^m$ 和 $G_j^p \leq G_l^m$;
- 2) $G_e^f \leq G_i^k$ 和 $G_e^f \leq G_j^p$;

相容粒度格

假设 $\langle GS, \leq \rangle$ 是一个偏序集，对于任意两个相容粒 $G_i^{\neq} \in GS$ 和 $G_j^{\neq} \in GS$

($G_i^{\neq} \neq G_j^{\neq}$)，有两个且仅有两个相容粒 G_i^m 和 G_j^f 满足：

- 1) G_i^m 和 G_j^f 分别为 G_i^{\neq} 和 G_j^{\neq} 的上界相容粒和下界相容粒
- 2) 不存在另外一个相容粒 G_a^{\neq} 满足 $G_i^{\neq} \leq G_a^{\neq} \leq G_i^m$ 和 $G_j^{\neq} \leq G_a^{\neq} \leq G_i^m$ ；
- 3) 不存在另外一个相容粒 G_c^{\neq} 满足 $G_j^{\neq} \leq G_c^{\neq} \leq G_i^f$ 和 $G_j^{\neq} \leq G_c^{\neq} \leq G_j^f$ 。

此时 $\langle GS, \leq \rangle$ 称为是一个相容粒度格。

这里， G_i^m 称为 G_i^{\neq} 和 G_j^{\neq} 的最小上界相容粒， G_j^f 称为 G_i^{\neq} 和 G_j^{\neq} 的最大下界相容

基于相容粒度空间的信息分类算法



Why?

一个决策表

a	b	c	d
1	1	3	1
3	2	3	3
3	3	2	2
2	1	1	2
1	3	1	1
3	3	2	2



规则 1: $a=1 \Rightarrow d=1$

规则 2: $b=2 \Rightarrow d=3$

规则 3: $c=2 \Rightarrow d=2$

规则 4: $a=2 \Rightarrow d=2$



规则 1: $a=1 \Rightarrow d=1$

规则 3: $c=2 \Rightarrow d=2$



a	b	c
1	1	2



分类规则

规则 1.1: $a=1 \wedge b=1 \Rightarrow d=1$

规则 1.2: $a=1 \wedge b=3 \Rightarrow d=1$

规则 1.3: $a=1 \wedge c=3 \Rightarrow d=1$

规则 1.4: $a=1 \wedge c=1 \Rightarrow d=1$

规则 3.1: $a=3 \wedge c=2 \Rightarrow d=2$

规则 3.2: $b=3 \wedge c=2 \Rightarrow d=2$



a	b	c
1	1	2



规则 1.1: $a=1 \wedge b=1 \Rightarrow d=1$

Why基于相容粒度空间模型?

- 通过层次的分解可以节省复杂度
- 层次的分解利用内涵关系

Why自顶向下的构建?

- 复杂度低

基于相容粒度格的信息分类算法

TGLC

Step 1: 选择在相容粒度格中所有满足下列条件的相容粒构成的相容粒集为初始分类粒集 GS : 和 G_0^1 直接连接或者在和相容粒 G_0^1 的连接路径上只有冲突相容粒的相容粒。

Step 2: 对于待分类的每个原始对象 O_{0_i} , 循环下列步骤

Step 2.1: 用 GS 中粒的内涵所表示的决策规则进行分类;

Step 2.2: 如果只有一个匹配的一致相容粒, 则跳到 Step 2.8;

Step 2.3: 如果有多个匹配的一致相容粒而且这些相容粒的内涵所表示的决策规则的决策相同, 则跳到 Step 2.8;

Step 2.4: 如果这些匹配的一致相容粒集不存在子粒, 则存储 GS 中和 O_{0_i} 匹配的一致相容粒集到 S , 否则跳到 Step 2.6;

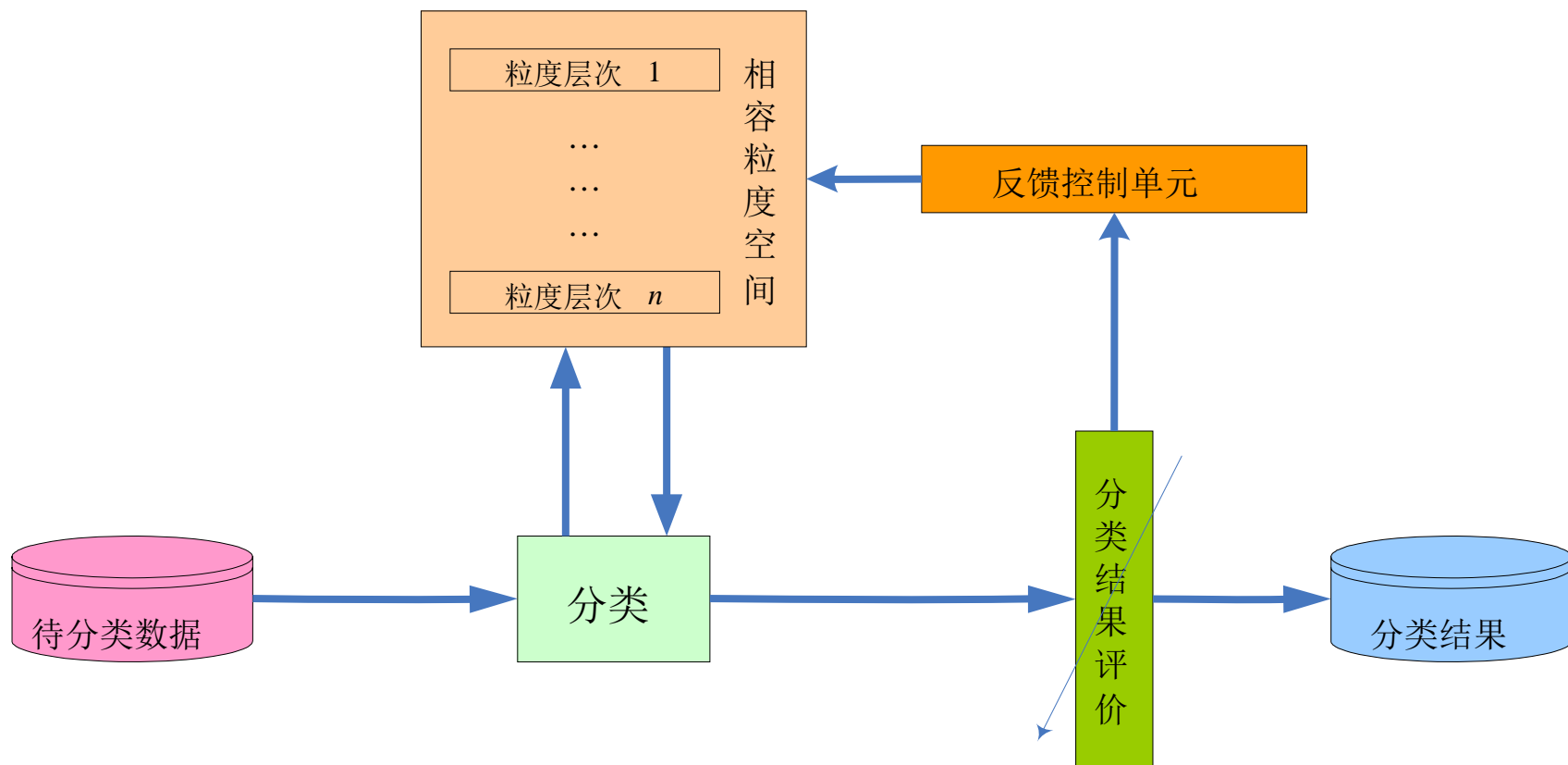
Step 2.5: 根据 S 中每个相容粒的外延以及原始对象集的统计特性度量相容粒 (例如可信度, 支持度等), 选择度量值最大或者最小的相容粒的决策为 O_{0_i} 的分类结果, 跳到 Step 2 进行下一个对象的分类;

Step 2.6: 设这些匹配的一致相容粒中所在层次最大为 i 层, 则选择所有以这些匹配的相容粒的第 $i+1$ 层一致子孙相容粒构成的相容粒集为分类粒集 GS' ;

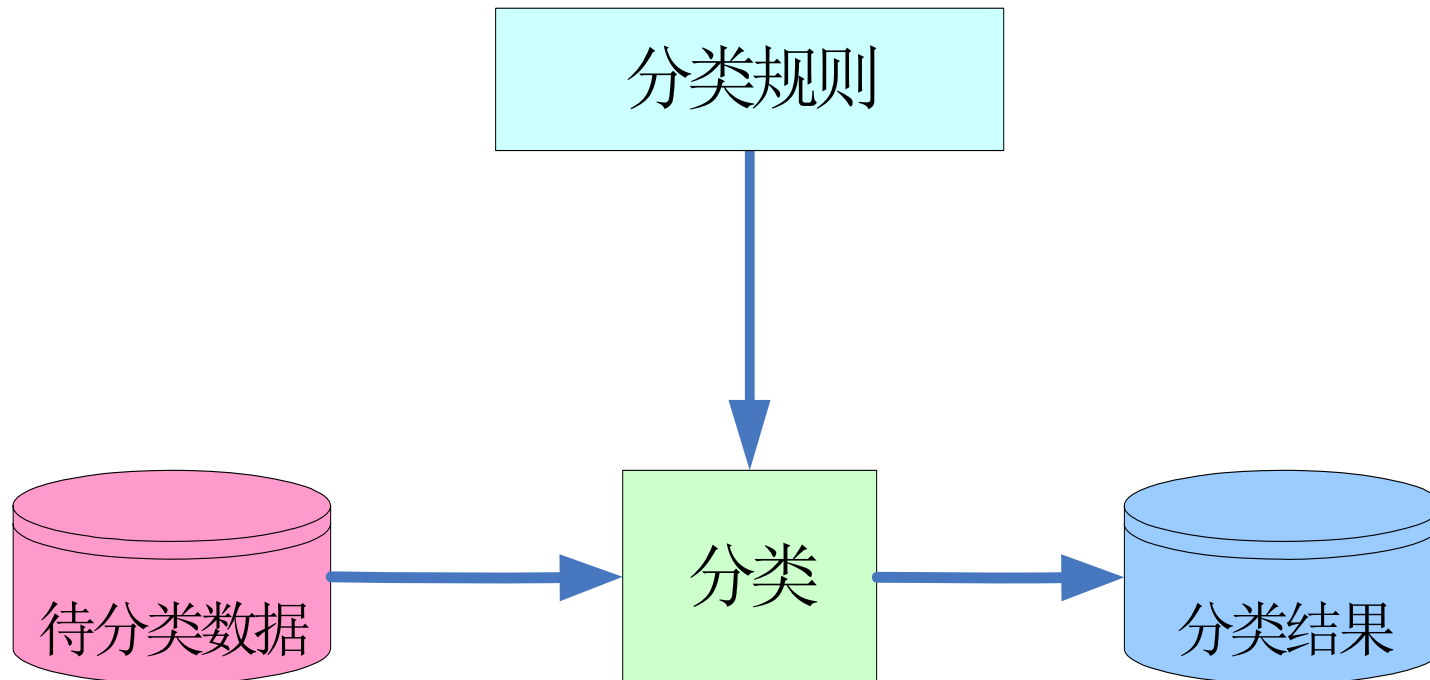
Step 2.7: 用 GS' 中相容粒的内涵所表示的决策规则进行分类, 跳到 Step 2.2;

Step 2.8: 原始对象 O_{0_i} 的分类结果为匹配相容粒的内涵所表示决策规则的决策, 跳到 Step 2 进行下一个对象的分类。

基于相容粒度空间模型分类的基本流程



传统分类流程



简化策略

- Why简化?

- 时间复杂度高 $O(e_{max}mn^2)$ (通常为 $O(m^2n^2)$)

- 简化策略

- 策略1: 在通过**TGM**构建相容粒度空间模型时, 一旦构建了一个一致相容粒后, 就可以不再进行分解。

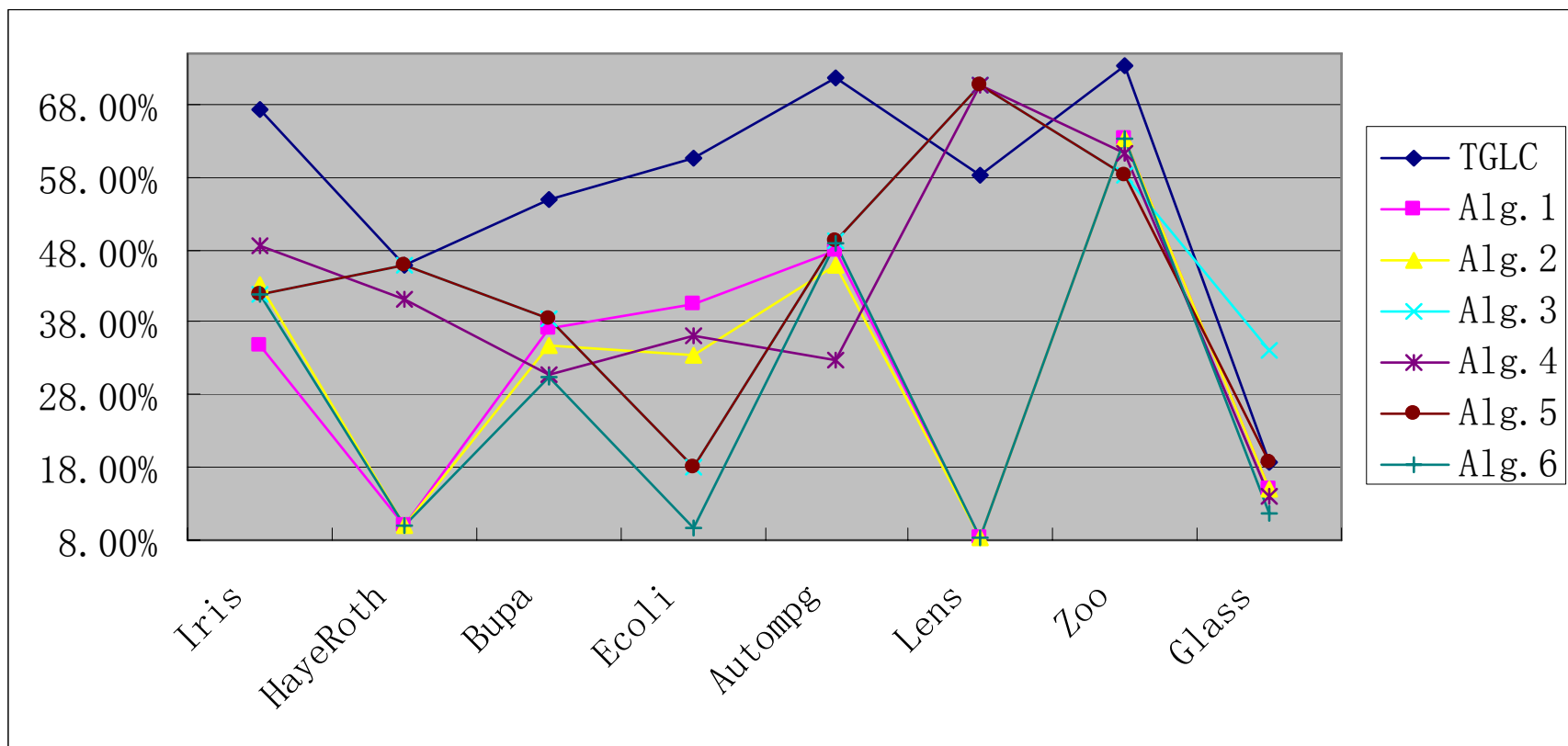
- 策略2: 通过限制相容粒度空间的相容粒层数来降低复杂度。

- 策略3: 去掉相容粒度空间中的冲突相容粒。

- 策略4: 每个相容粒的外延中不存储对象, 而是存储对象在原始数据表中的位置。

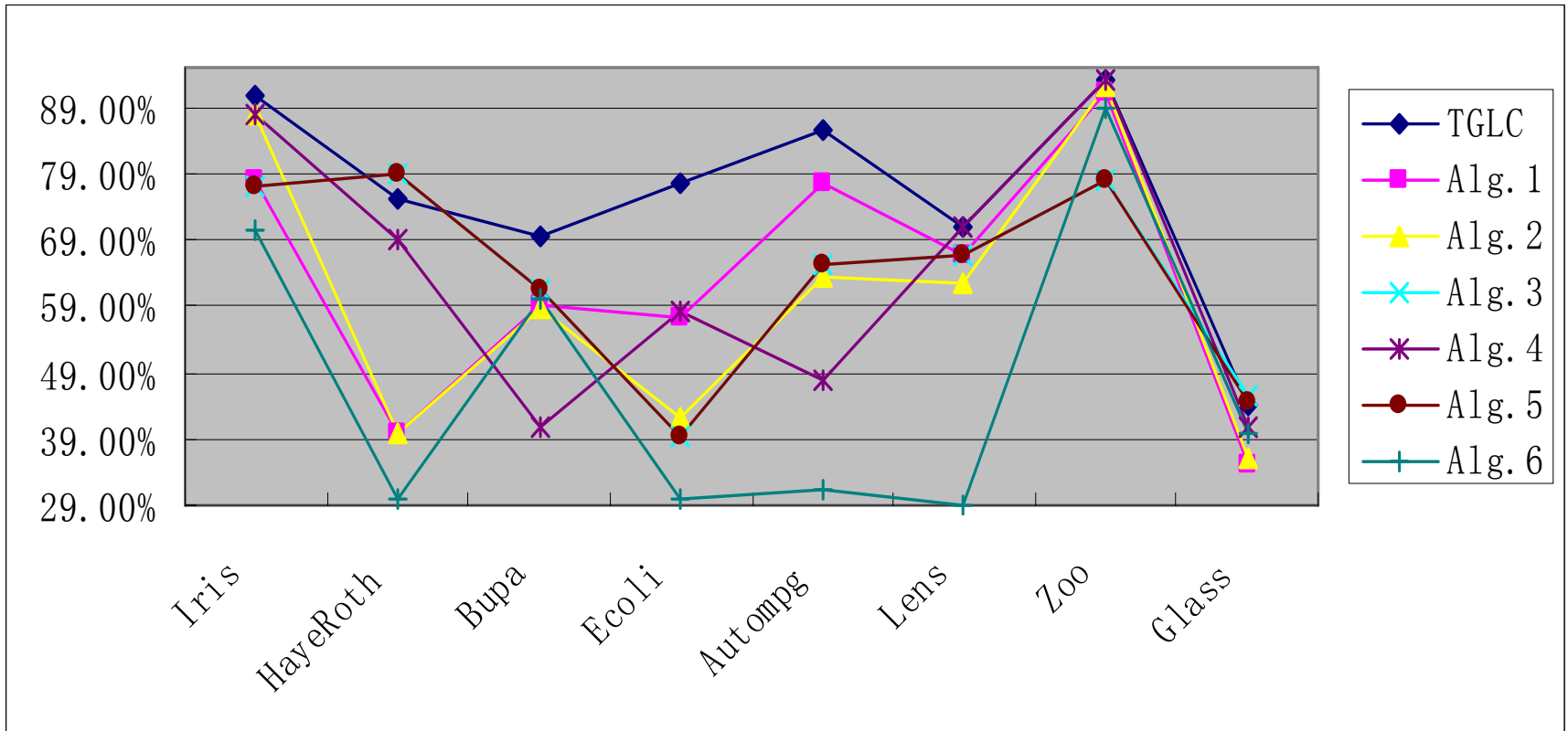
实验结果

—TGLC和几种算法的比较($\alpha = 10$)



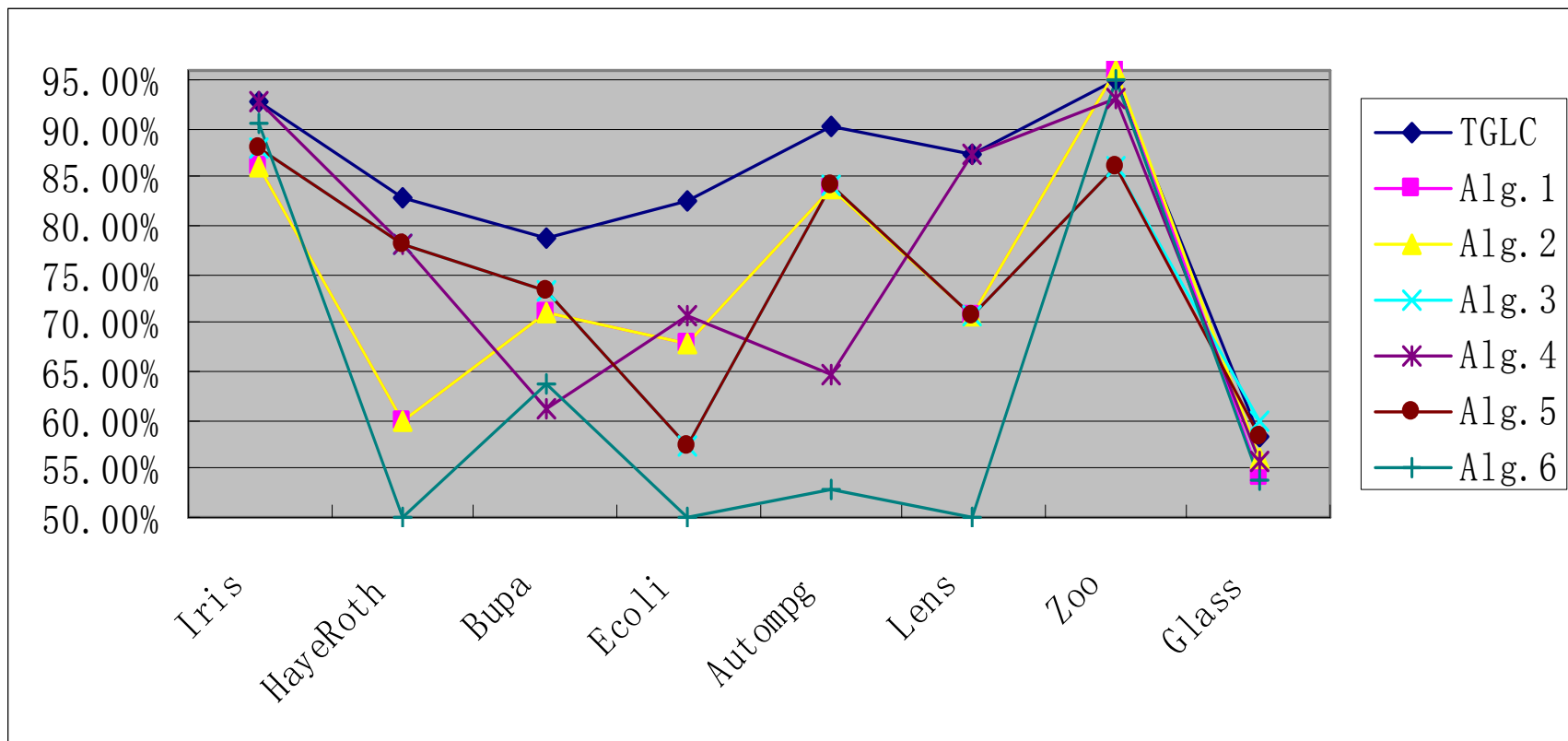
实验结果

—TGLC和几种算法的比较($\alpha = 30$)



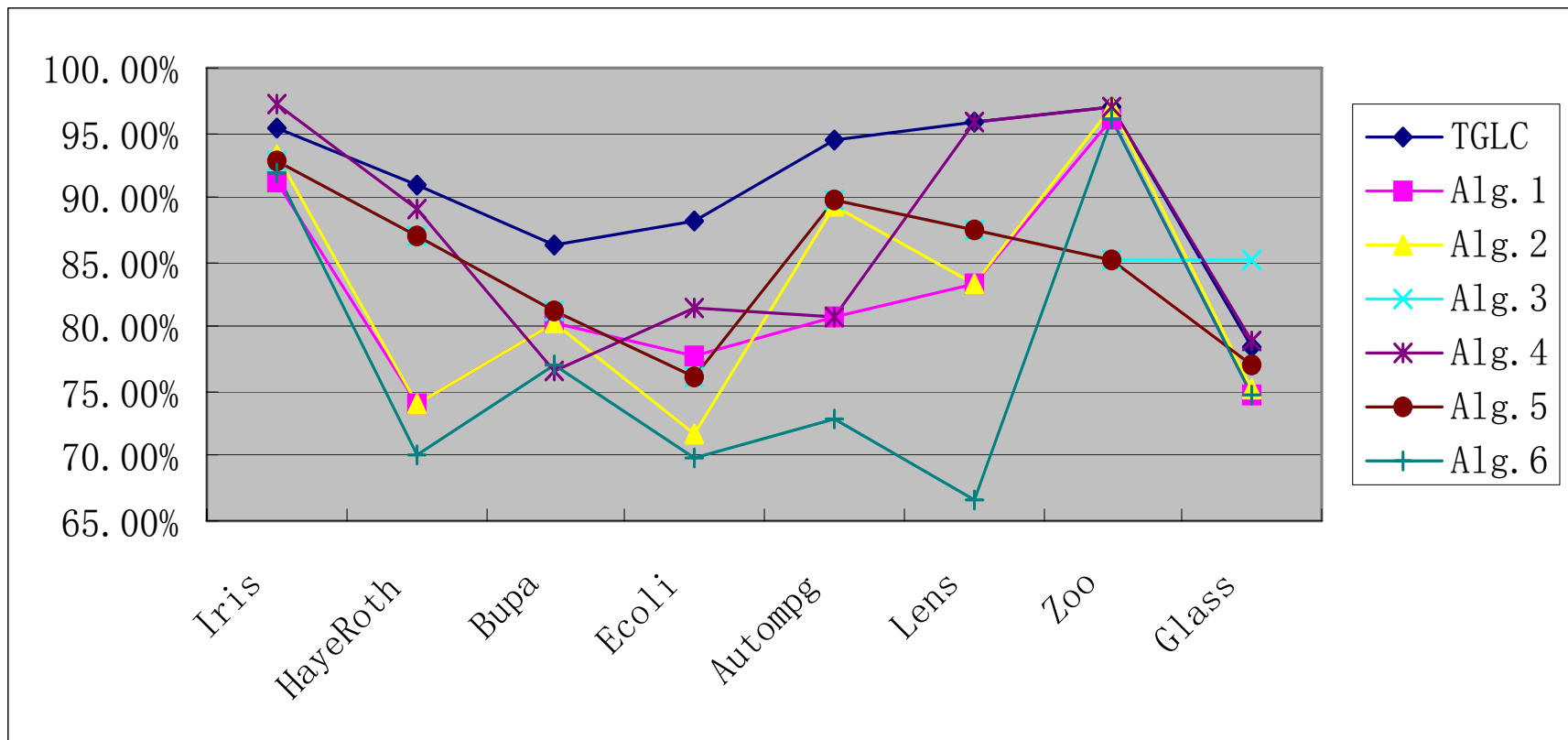
实验结果

—TGLC和几种算法的比较($\alpha = 50$)



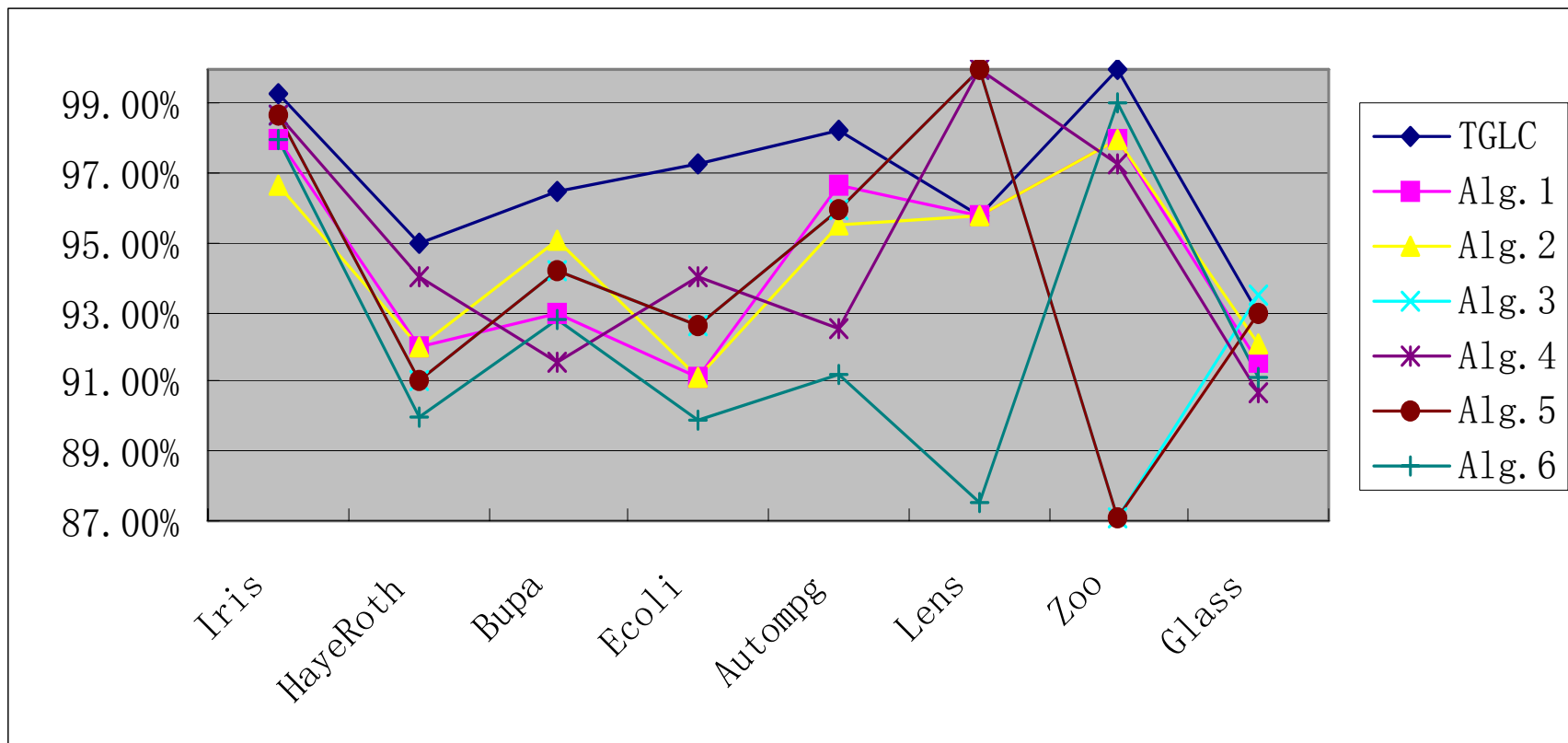
实验结果

—TGLC和几种算法的比较($\alpha = 70$)

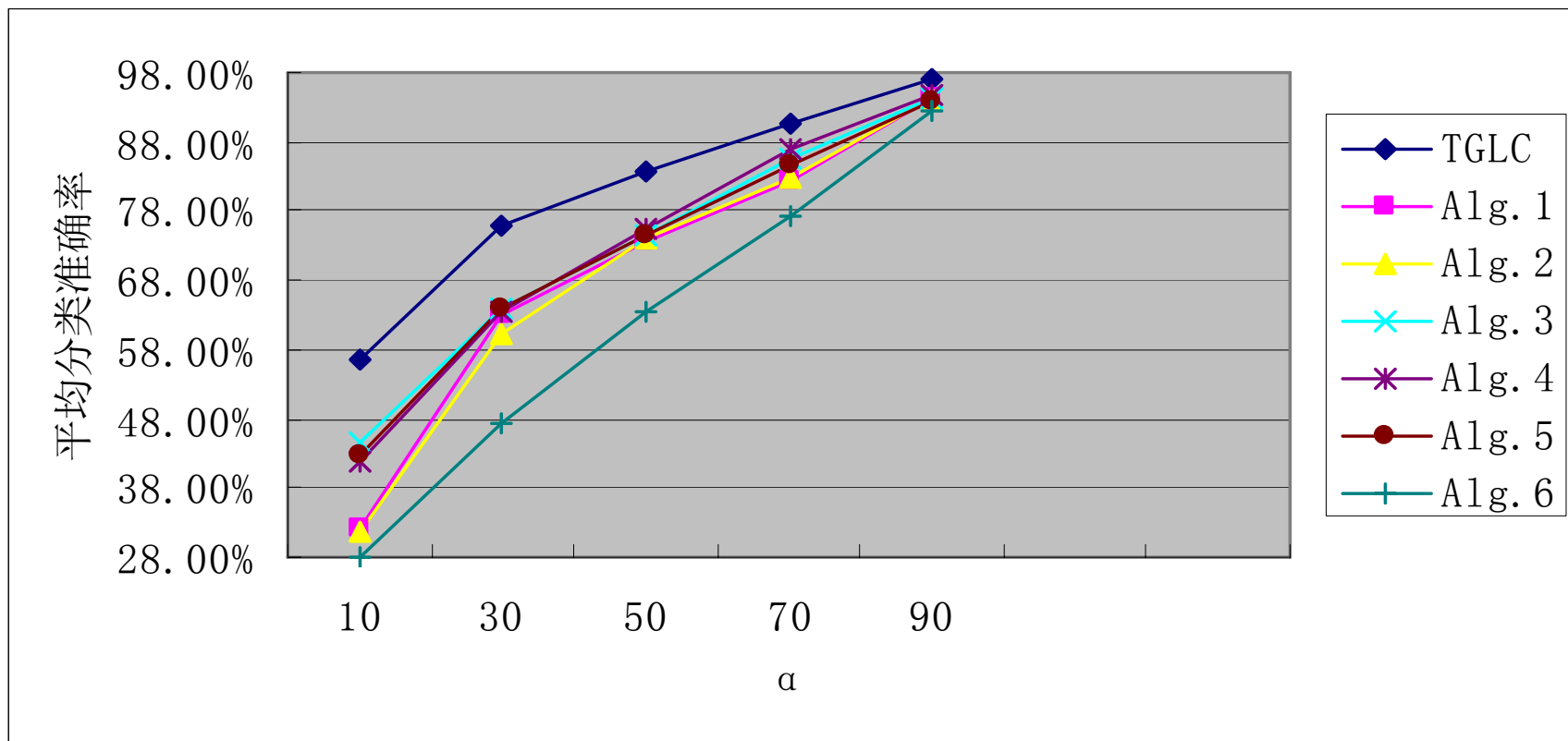


实验结果

—TGLC和几种算法的比较($\alpha = 90$)



鲁棒性比较



TGLC算法

- 优点

- 分类准确率提高
- 数据鲁棒性提高

- 缺点

- 时间复杂度高
- 错误识别率较高
- 未识别率较低

报告提纲

- ✓ 引言
- ✓ 相容粒度空间模型
- ✓ 基于相容粒度空间模型的信息分类
- ✓ **基于相容粒集的双层决策模型**
- 相容粒度空间的其它应用
- 结束语

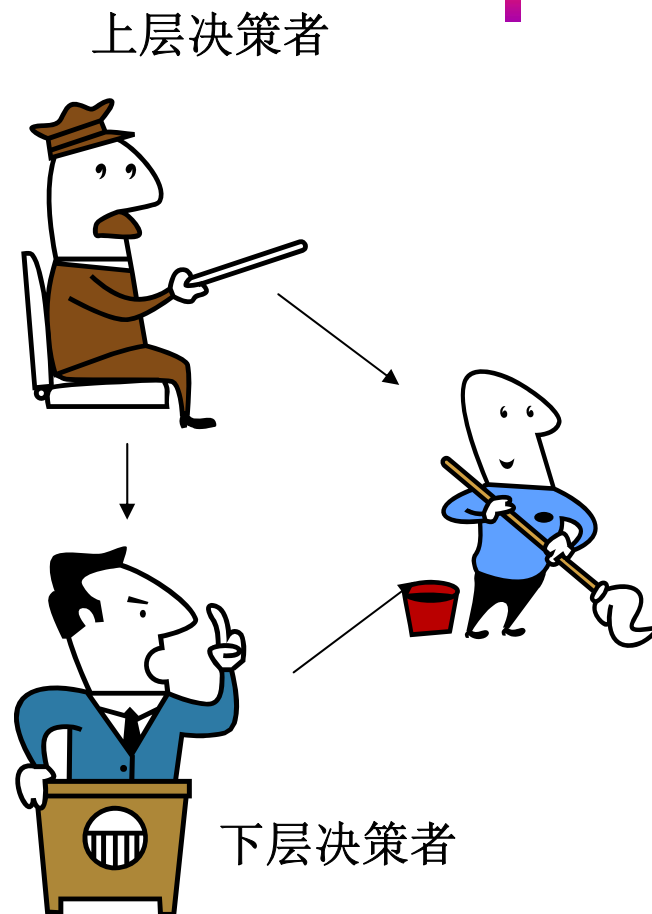
基于相容粒集的双层决策模型

● 研究动机

- 现代系统越来越复杂，知识的表示越来越多样，并且信息系统中常常包含着海量、不完整、模糊及不精确的数据及对象，使得传统的数学优化方法显得无能为力。
- 粒计算的方法不要求目标函数和约束函数的连续性与凸性，甚至有时连解析表达式都不要，而且对计算中数据的不确定性也有很强的适应能力。

什么是双层决策？

- 随着人类社会的发展，实际问题规模越来越大，结构越来越复杂，涉及到对问题做出决策的人越来越多，而且这些决策者各自处于不同的层次上。
- 一般地，高级决策机构（者）自上而下的对下一级若干决策机构（者）行使某种控制、引导权，而下一级决策机构（者）在这一前提下，亦可以在其管理范围内行使一定的决策权，虽然这种决策权比较起来处于从属地位。



双层决策问题的经典模型

$$\min_x F(x, y)$$

$$\text{s.t. } G(x, y) \geq 0$$

$$\min_y f(x, y)$$

$$\text{s.t. } g(x, y) \geq 0,$$

其中 x 和 y 分别为上层决策者和下层决策者的决策变量， F 和 f 分别为上层决策者和下层决策者的目标函数， G 和 g 分别为上层决策者和下层决策者的约束函数。

经典双层决策模型存在的问题

- 建模困难。很多问题很难描述成为线性或者非线性的函数；
- 求解困难。许多非线性函数构建的双层决策模型，甚至一些线性函数构建的双层决策模型都难以解决。

基于决策规则集的双层决策问题

- 理论上来说，通过枚举一个域内所有可能状态，一个决策问题的约束可以转化为一个决策表，称为约束决策表。用同样的方法，目标函数也可以转化为一个决策表，称为目标决策表。也就是说一个双层决策问题可以被转化为一组决策表，其中决策变量表示为决策表中的对象。

基于决策规则集的双层决策模型

$$\min_x f_L(x \oplus y)$$

$$\text{s.t. } g_L(x \oplus y) \geq 0$$

$$\min_y f_F(x \oplus y)$$

$$\text{s.t. } g_F(x \oplus y) \geq 0$$

其中 x 和 y 是待决策的对象。 f_L 和 g_L 分别是上层决策的目标决策规则集函数和约束决策规则集函数， f_F 和 g_F 分别是下层决策的目标决策规则集函数和约束决策规则集函数。 F_L 、 G_L 、 F_F 和 G_F 分别是上面四个决策规则集函数对应的决策规则集。

基于决策规则集的双层决策建模算法

输入： 一个双层决策问题，其中包括上层决策和下层决策的目标和约束；

输出： 一个基于相容粒度集的双层决策模型；

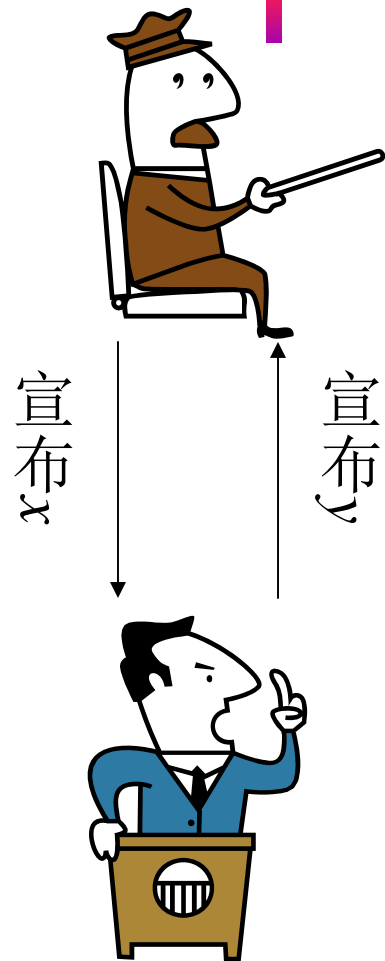
Step 1： 把该问题用一组决策表来表示；

Step 2： 根据规则抽取算法从每个决策表中抽取决策规则集；

Step 3： 根据生成的上层决策目标决策规则集，上层决策约束决策规则集，下层决策目标决策规则集和下层决策约束决策规则集，重新组合为前面公式所表示的基于决策规则集的双层决策模型。

双层决策的决策机制

- 上级决策者首先宣布它的决策 x ，这一决策将影响下级决策问题的约束和目标函数，然后下级决策者在这一前提下选取自己的目标函数最优的决策 y ，这一过程也会影响到上层决策问题的约束和目标函数，因为在上层决策问题中出现了 y ，进一步，上层决策者可以再调整它的决策变量 x ，直到它的目标函数达到最优。



基于决策规则集的双层决策算法

TGBD

Step 1: 求解上层决策问题;

Step 1.1: 去掉目标决策规则集中不满足约束规则集的对象;

Step 1.2: 在上层决策目标规则集中选出决策最优的几条规则, 并根据相应排序准则进行排序;

Step 1.3: 设 **OS** 为满足排序后第一条规则的对象集, 并对该对象集中的对象排序;

Step 1.4: 上层决策的解是 **OS** 中的第一个对象;

基于决策规则集的双层决策算法

TGBD

Step 2: 求解下层问题

Step 2.1: 去掉下层目标决策规则集中不满足上层决策的规则;

Step 2.2: 去掉下层目标决策规则集中不满足约束的规则;

Step 2.3: 选择下层目标决策规则集中决策最优的几条规则, 并进行排序;

Step 2.4: 设 **OS'** 为满足排序后第一条规则的对象集, 并对该对象集中的对象排序;

Step 2.5: 下层决策的解是 **OS'** 中的第一个对象;

基于决策规则集的双层决策算法

TGBD

Step 3: 如果上下层问题的解相同，则结束；
否则，选择上层求得的排序后的规则集的
下一个规则，继续**Step 2**直至结束。

基于相容粒集的双层决策问题

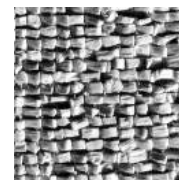
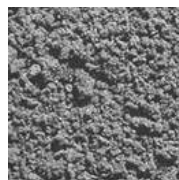
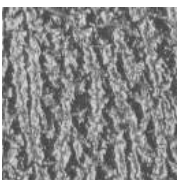
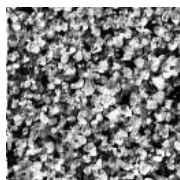
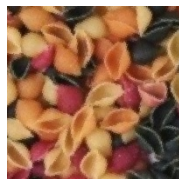
- 即用相容粒度空间求取决策规则集
- 进一步，从广义上来说，基于相容粒集的双层决策可以做为双层决策问题的一个泛化，即只要相容粒集表示条件和决策之间的关系，就可以构造一类双层决策问题。

报告提纲

- ✓ 引言
- ✓ 相容粒度空间模型
- ✓ 基于相容粒度空间模型的信息分类
- ✓ 基于相容粒集的双层决策模型
- ✓ 相容粒度空间的其它应用
- 结束语

基于相容粒度空间模型的 图像纹理识别

- 纹理是描述图像时常用的一个概念(类似于颜色,也常取决于感知)。纹理常被看作是图像的某种局部性质,或是对局部区域中像素之间关系的一种度量。



16:54:33

史忠植: 相容粒度空间模型

67

图像纹理识别示例

对象集系统

原始对象 $\mathbf{o}_{0,xy} = (x, y, R, G, B)$ 是图像中的一个像素，其中 x, y 是 x 坐标值和 y 坐标值， R, G, B 是这个像素的 RGB 颜色值。子集对象 \mathbf{O}_1 是一个图像，且

$$\mathbf{O}_1 = \begin{bmatrix} \mathbf{O}_{0_{11}} & \mathbf{O}_{0_{12}} & \cdots & \mathbf{O}_{0_{1m}} \\ \mathbf{O}_{0_{21}} & \mathbf{O}_{0_{22}} & \cdots & \mathbf{O}_{0_{2m}} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{O}_{0_{n1}} & \mathbf{O}_{0_{n2}} & \cdots & \mathbf{O}_{0_{nm}} \end{bmatrix},$$

\mathbf{O}_2 可以看作是一个视频中的一组关键帧。

相容关系系统

$cp(\alpha, \beta | DIS, D) = (dis_1(\alpha, \beta | \omega) \leq d_1) \wedge (dis_2(\alpha, \beta | \omega) \leq d_2)$ 和

$dis_1(\alpha, \beta | \omega') = |\alpha_1 - \beta_1|, dis_2(\alpha, \beta | \omega) = |\alpha_2 - \beta_2|,$

其中 $\alpha = (\alpha_1, \dots, \alpha_n), \beta = (\beta_1, \dots, \beta_n)$ 。

嵌套相容覆盖系统

一个二层相容粒的内涵和外延可以有下面的公式得到：

$$EG_2^1(\eta_2^1 | tr) = \{x / (x, \eta_2^1) \in tr_{(cp, \omega, DIS, D)} \wedge (x \in O_1)\}, \text{ and}$$

$$IG_2^1 = \eta_2^1,$$

其中 $\eta_2^1 \in Grid_2^1$ 和 $Grid_2^1 \subseteq O_1$ 。例如，我们可以设

$$Grid_2^1 = \{O_{0,xy} \mid x=3i+2, y=3j+2, \text{ 且 } i \text{ 和 } j \text{ 为 } 0 \text{ 或任意正整数}\},$$

则 $Grid_2^1 = \{O_{0_{22}}, O_{0_{25}}, O_{0_{28}}, O_{0_{52}}, \dots\}$ 。通过上面的步骤，原始对象被粒化为大小为 $d_1 \times d_2$ 的矩形粒。当然，我们也可以通过修改距离函数来改变粒形状。

嵌套相容覆盖系统

设 $IGS_{i+1}^k = \bigcup_p IG_{i_p}^k$ ，则第 $i+1$ 层相容粒的内涵和外延可以有

下面的公式得到：

$$EG_{i+1}^1(\eta_{i+1}^1 | tr_{i+1}^1) = \bigcup \{ EG / (IG, \eta_{i+1}^1) \in tr_{(cp, \omega, DIS, D)} \wedge (IG \in IGS_{i+1}^1) \},$$

$$EG_{i+1}^1 = \eta_{i+1}^1,$$

其中 $\eta_i^1 \in Grid_i^1$ 和 $Grid_i^1 \subseteq IGS_i^1$ 。

基于相容粒度空间模型的 图像纹理识别

- Step 1: 建立每个种子图象的相容粒度空间;
- Step 2: 选择其中一层（设为 i 层）上的相容粒做为模板;
- Step 3: 建立每个待识别图象的相容粒度空间;
- Step 4: 选择第 $i+1$ 层上的相容粒做为识别粒单元;

基于相容粒度空间模型的 图像纹理识别

Step 5:

对于每个 $i+1$ 层的相容粒 G_{i+1}^1 ，我们可以分配一个向量：

$$MG_{i+1}^1 = (mg_{(i+1)p_1}^1, \dots, mg_{(i+1)p_m}^1)$$

给这个相容粒，其中 n 是模板的数目且

$$mg_{(i+1)p}^1 = \text{Card}(\{G_i^1 \mid G_i^1 \text{ 是 } G_{i+1}^1 \text{ 的子粒且 } \text{sim}(M_{i_p}^1, G_i^1) \leq d^p\})$$

其中 $\text{Card}()$ 是计算集合维度的函数。

Step 6:

用类似的方法，假设 SG_{i+1}^1 是种子图像 S 中一个 $i+1$ 层的粒，然后我们可以分配一个向量

$$MSG_{i+1}^1 = (msg_{(i+1)1}^1, \dots, msg_{(i+1)m}^1)$$

给这个粒，其中

$$16:54 \quad msg_{(i+1)p}^1 = \text{Card}(\{SG_i^1 \mid SG_i^1 \text{ 是 } SG_{i+1}^1 \text{ 的子粒且 } \text{sim}(M_{i_p}^1, SG_i^1) \leq d^p\})$$

基于相容粒度空间模型的 图像纹理识别

Step 7:

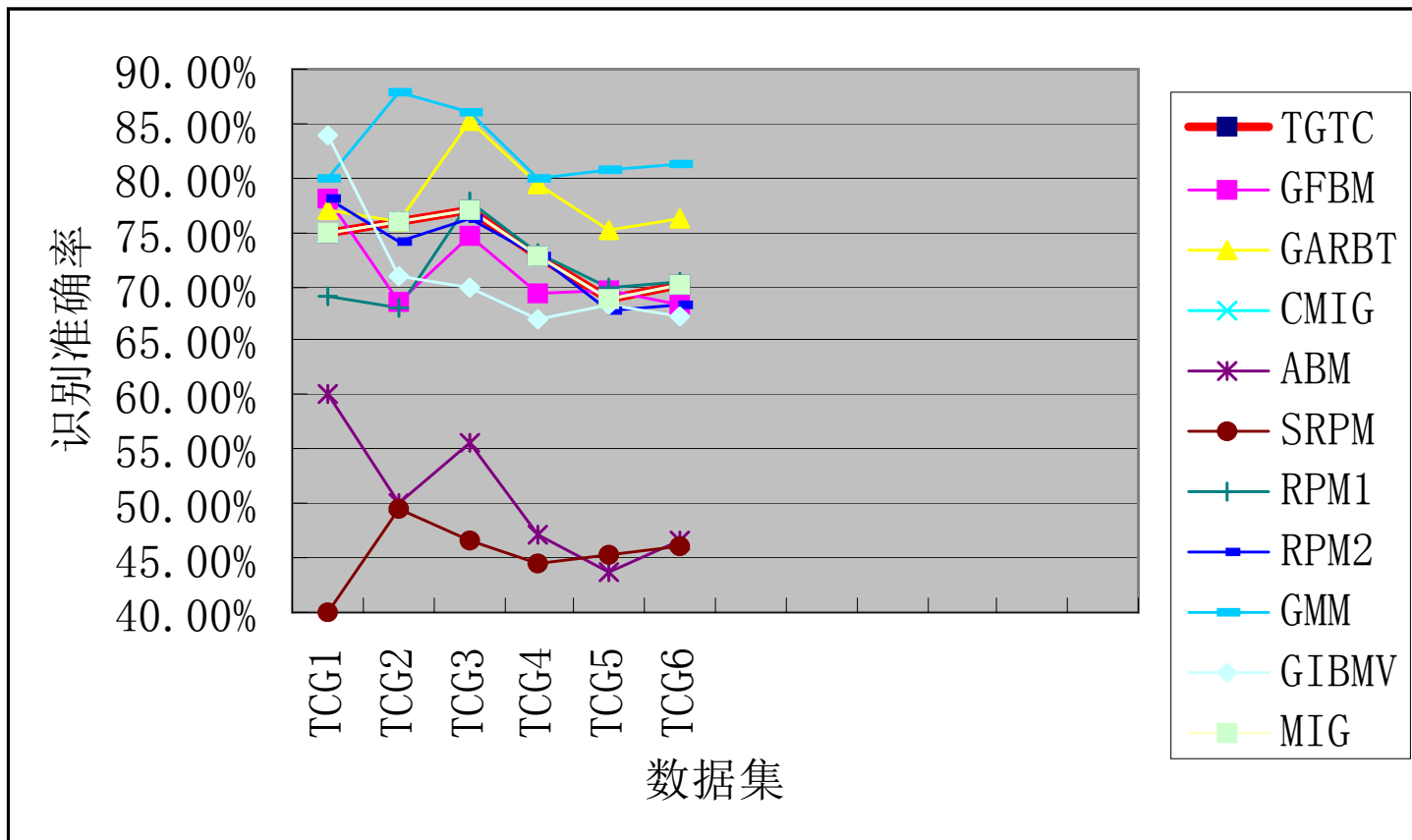
待识别图像 O_1 和一个种子图像 S_j 的相似性可以基于 i 层的粒来计算, 如下

$$SIM(O_1, S_j) = \sum_{p=1}^l \sum_{q=1}^m ((mg_{pq}^1)^2 - (msg_{pq}^1)^2)^2,$$

其中 l 为 O_1 上 i 层粒的数目。

如果 $SIM(O_1, S_j) \leq d$, 则图像 O_1 被分配到 S_j 所代表的类中。

试验结果



基于相容粒度空间模型的 图像纹理识别

- 结论

我们的算法在大部分数据上的识别率好于大多数算法，但由于粒的固定大小以及粒度空间的固定层次，我们的算法的识别率并不是最好的。然而，这个应用表明使用相容粒度空间来解决图像问题是一个有效的方法。

相容粒度空间在数据挖掘中的应用

- 概念描述
- 关联分析
- 分类和预测
- 聚类

相容粒度空间在数据挖掘中的应用

- 目前还没有非常具体的这方面的研究，我们的工作通过更加具体的理论框架展现了这种定义的合理性和有效性。
- 我们认为挖掘的任务不仅是寻找最合适的知识粒和结构知识，还需要通过反馈和优化的思想寻找在粒度空间中合适的跳跃路径，实现在不同的粒之间，粒集之间以及粒度层次之间的交互，从而通过这种交互来完成数据挖掘任务。

基于相容粒度空间的问题求解系统

根据具体问题建立相容粒度空间模型的一般性过程为：

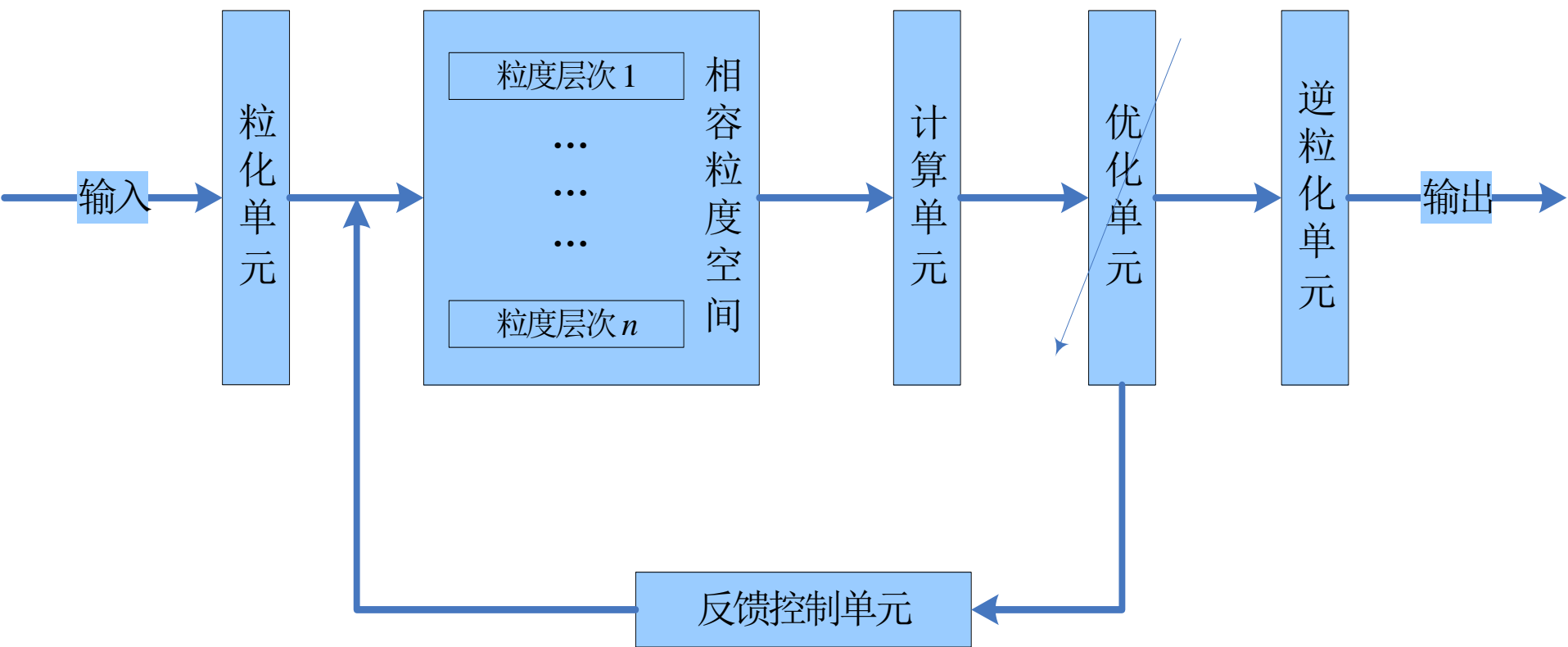
1. 分析问题，构建对象系统；
2. 分析问题，建立对象之间的关系，进而构建相容关系系统；
3. 根据具体问题背景，定义相容粒的内涵和外延的表示；
4. 基于对象系统、相容关系系统以及相容粒的表示来构建嵌套覆盖系统；
5. 在嵌套覆盖系统中选择合适的粒度层次停止构造；
6. 结束

问题求解的一般性过程

用相容粒度空间进行问题求解的一般性过程为：

1. 选择合适的粒度层次或者粒集进行问题求解；
2. 对求解的结果进行评价；
3. 如果结论不满足优化准则或者该问题没有妥善解决，则返回1；
4. 结束。

基于相容粒度空间的问题求解系统



结束语

取得的成果

- 相容粒度空间模型
- 基于相容粒度空间模型的信息分类方法
- 基于相容粒集的双层决策模型及其建模和决策算法
- 基于相容粒度空间的图像纹理识别方法
- 基于相容粒度空间模型的建模和问题求解的一般性方法

发表的文章

- Zhongzhi Shi. Tolerance Relation Based Granular Computing Model. Invited Speaker, IEEE ICGrC 2005
- Zheng, Z., Hu, H., Shi, Z.Z., Granulation Based Image Texture Recognition, Lecture Notes in Computer Science, Vol.3066, pp.659~664, 2004; (SCI检索)
- Zheng, Z., Hu, H., Shi, Z.Z., Rough set based image texture recognition algorithm, Lecture Notes in Computer Science, Vol.3213, p.772~780, 2004; (SCI检索)
- Zheng, Z., Hu, H., Shi, Z.Z., Tolerance Relation Based Information Granular Space, Lecture Notes in Computer Science, Vol.3641, pp.682~691, 2005; (SCI检索)
- Shi, Z.Z., Tolerance Granular Space and Its Applications, Invited speaker, IEEE International Conference on Granular Computing, 2005;
- Zheng, Z., Hu, H., Shi, Z.Z., Tolerance Granular Space and Its Applications, IEEE International Conference on Granular Computing, pp.367~372, 2005; (EI检索)
- Zheng, Z., Zhang, G., He, Q., Lu, J., Shi, Z.Z., Rule Sets Based Bilevel Decision Model, the Twenty-Ninth Australasian Computer Science Conference, 113-120, 2006
- 史忠植, 郑征. 相容粒度空间模型及其应用. 粒计算: 过去、现在与展望, 科学出版社, 42~82

谢谢！



中科院计算所
INSTITUTE OF COMPUTING
TECHNOLOGY



Intelligence Science

<http://www.intsci.ac.cn/>