《智能信息处理》课程作业

模糊概念格的构造方法研究

王娅菲

作业	分数[20]
得分	

2021 年 11 月 28 日

模糊概念格的构造方法研究

王娅菲
（大连海事大学　信息科学与技术学院　大连　116026）

摘　要　形式概念分析(Formal Concept Analysis，FCA)[1]是由 R.Wille 于 1982 年提出的一种从形式背景进行数据分析和规则提取的强有力工具，被广泛应用于信息检索、数据挖掘、软件工程等领域。概念格是形式概念分析理论中的核心数据结构，在信息检索、知识发现、推荐系统等方面得到了广泛的应用。在实际生活中，大多数概念都是模糊的或不确定的，即模糊形式背景。将概念格理论和模糊理论相结合，提出模糊概念格模型，可用于探索性地分析和处理模糊信息，而模糊概念格的构造是其得以广泛应用的前提。本文介绍了模糊概念格的基本理论知识，并提出了两种模糊概念格的构造方法。

关键词　形式概念分析;模糊集;模糊概念格;构造方法

Research on the Construction Method of Fuzzy Concept Lattice

Yafei Wang
(*Department of Information Science Technology, Dalian Maritime University, Dalian　116026*)

Abstract　Formal concept analysis (FCA) [1] is a powerful tool for data analysis and rule extraction from formal background proposed by R.wille in 1982. It is widely used in information retrieval, data mining, software engineering and other fields. Concept lattice is the core data structure in formal concept analysis theory, and it is put into wide use in information retrieval, knowledge discovery, recommendation system and so on. In real life, most concepts are vague or uncertain, that is, fuzzy formal background. Combining concept lattice theory and fuzzy theory, a fuzzy concept lattice model is proposed, which can be used to analyze and process fuzzy information, and the construction of fuzzy concept lattice is the prerequisite for its wide application. This article introduces the basic theoretical knowledge of fuzzy concept lattice, and proposes two construction methods of fuzzy concept lattice.

Keywords　formal concept analysis；fuzzy set；fuzzy concept lattice；construction method.

1　引　言

概念是人类认知的基本单元，代表了事物的基本特征。德国的数学家 Wille 教授于 1982 年提出了形式概念分析理论（简称 FCA）[1]，探讨了概念和概念格的性质。形式背景与形式概念是 FCA 中的两个基本概念。其中，形式背景是由对象集、属性集以及两者之间的二元关系构成的三元组。形式概念是各种概念的抽象，由外延（extent）与内涵（intent）组成，其中外延是概念中所有对象的集合，内涵表示这些对象的共同特征[2]。概念格的每个节点是一个形式概念，从形式背景中生成概念格的过程实质上是一种概念聚类的过程,通过 Hass 图可以清楚地反应出概念间的层次结构,从而实现数据的可视化。

人们从现实世界中抽取出来的往往都是精确的概念，然而，现实生活中存在着大量模糊的或不确定性的信息。例如，在评价一名学生成绩好坏时，我们会用“很好”，“一般”，“不太好”这样的语言进行表达，而这样的信息无法通过精确的形式概念来描述。为了适应客观需求，研究人员将概念格理论和模糊理论相结合，提出了模糊概念格模型。利用模糊概念格模型处理日常生活中大量的模糊语言问题，具有深远的研究意义。

2 模糊概念格

2.1 形式背景和概念格

定 义 1 设U是对象的集合，M是属性的集合，I是U与M间的关系，则称三元组$K=(U,M,I)$为一个形式背景。$(u,m)\in I$表示对象u具有属性m。形式背景可以用一个对象集合、属性集合以及他们之间的二元关系建立起来的表格来表示，它的每行表示某一对象，每列则表示某一属性。

定 义 2 设$K=(U,M,I)$为形式背景，对象集合$U=\{u_1,u_2,...,u_m\}$，属性集合$M=\{m_1,m_2,...,m_n\}$。

设$m\in M$,具有属性m的对象集合。定义

$$g(m)=\{u\in U\mid (u,m)\in I\}$$

设$u\in U$,具有属性m的对象集合。定义

$$f(u)=\{m\in M\mid (u,m)\in I\}$$

定 义 3 设$K=(U,M,I)$为形式背景，若

$A\subseteq U$，$B\subseteq M$，那么存在

$$f(A)=\{m\in M\mid \forall u\in A,(u,m)\in I\}$$ 和

$$g(B)=\{u\in U\mid \forall m\in B,(u,m)\in I\}$$

如果集合A，B满足$f(A)=B$，$g(B)=A$，则称二元组(A,B)为概念。A是概念(A,B)的外延，B是概念(A,B)的内涵。

定 义 4 设$K=(U,M,I)$为形式背景，存在两个概念(A_1,B_1)和(A_2,B_2)满足$(A_1,B_1)\le(A_2,B_2)\Leftrightarrow A_1\subseteq A_2(\Leftrightarrow B_1\supseteq B_2)$，则称$(A_1,B_1)$为$(A_2,B_2)$的子概念，$(A_2,B_2)$为$(A_1,B_1)$的超概念。由形式背景$(U,M,I)$中所有的概念根据它们之间的层次关系有序组成的集合，常称为(U,M,I)的概念格。从形式背景中生成概念格的过程实质上是一种概念聚类的过程,通过 Hasse 图可以清楚的反应出概念间的层次结构,从而实现数据的可视化。

概念格的形式背景通常是由表 1 所示的二维数表来表示,横向维表示属性, 纵向维表示对象。设形式背景为$K=(U,M,I)$，对象集$U=\{1,2,3,4\}$，属性集$M=\{a,b,c,d,e\}$。若对象具有某属性，就在对象行与属性列的交叉处用“1”标记，否则，用“0”标记。以形式概念$(\{2,4\},\{a,b,c\})$为例，$\{2,4\}$是外延，$\{a,b,c\}$是内涵，表示对象 2 和 4 共同具有的属性是a，b和c，而共同具有这三个属性的对象也恰为 2 和 4。

表 1 形式背景 $K=(U,M,I)$

	a	b	c	d	e
1	1	1	0	1	1
2	1	1	1	0	0
3	0	0	0	1	0
4	1	1	1	0	0

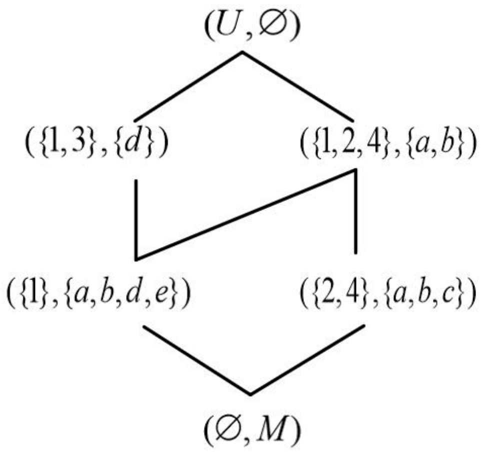

图 1 与表 1 对应的概念格的 Hasse 图

2.2 模糊集

精确的概念格描述的是外延与内涵之间确定性的关系，但是现实生活中有许多事物之间是模糊关系，是无法完全精确地表达出来的。例如某学生的成绩是很好、一般还是不太好，其成绩的等级无法用精确概念描述，这种模糊关系需要利用模糊概念来表示，才更符合客观情况。经典集合论中的元素要么属于某个集合，要么不属于某个集合，属于的情况用 1 表示，不属于的情况用 0 表示。Zadeh 与 1965 年提出了模糊集，将其作为经典集合论的一种扩展[3]。模糊集理论特征函数的取值范围由$\{0,1\}$推广到连续闭区间[0,1]，建立“隶属函数”的概念。

定 义 5 模糊集可以表示为(U,m)，其中论域U是一个集合，$m:U\to[0,1]$是一个隶属函数。对于每个$x\in U$，$m(x)$的值被称为x在(U,m)中的隶属度。对于有限集$U=\{x_1,x_2,...,x_n\}$而言，模糊集(U,m)通常用$\{m(x_1)/x_1,...,m(x_n)/x_n\}$来表示。对于$x\in U$，如果$m(x)=0$，那么$x$被视为不属于模糊集$(U,m)$；如果$m(x)=1$，那么$x$被视为完全属于模糊集$(U,m)$；如果$0<m(x)<1$，那么$x$被视为模糊集$(U,m)$中的模糊成员。集合$\{x\in U\mid m(x)>0\}$

被称为 (U,m) 的支集，集合 $\{x\in U\mid m(x)=1\}$ 被称为 (U,m) 的核。函数 m 被称为模糊集 (U,m) 的

隶属函数。

定 义 6 对于给定的论域 U，一个映射 $m_A:U\to[0,1]$ 可以确定 U 的一个模糊子集，其表示方法如下：

$$A=\begin{cases}\sum\limits_{xi\in U}\dfrac{m_A(x_i)}{x_i} & \text{论域}U\text{为有限集}\\ \int\limits_{x\in U}\dfrac{m_A(x)}{x} & \text{论域}U\text{为无限集}\end{cases}$$

其中，m_A 为 A 的隶属函数，$m_A(x)$ 表示 x 隶属于 A 的程度，称为 x 对 A 的隶属度。

2.3 模糊形式概念与模糊概念格

定 义 7 模糊形式背景可以用一个三元组 $K(U,D,I)$ 来表示,其中 I 是二元关系，描述了模糊形式概念的外延与内涵之间存在的模糊关系。

定 义 8 设 $F(U)$ 为模糊形式背景 $K(U,D,I)$ 的对象集的模糊幂集，即 $F(U)$ 为对象集 U 上的所有模糊集合组合的集合。设 $F(D)$ 为模糊形式背景 $K(U,D,I)$ 的属性集的模糊幂集，即 $F(D)$ 为属性集 D 上的所有模糊集合组成的集合。对于任意 $U'\in F(U)$ 和任意 $D'\in F(D)$，模糊属性映射可以表示为 $f:F(U)\to F(D), f(U')\supseteq D', D'\in F(D)$，模糊对象映射可以表示为 $g:P(D)\to P(U), g(D')\supseteq U', U'\in F(U)$。如果 $C(U',D')$ 能够同时满足 $U'=g(D')$ 和 $D'=f(U')$,则 $C(U',D')$ 被称为模糊形式概念。

一个模糊形式概念节点的外延中对象越少，对应的内涵中的属性就越多。比如，“肥胖的年轻人”与“肥胖的人”相比较，前者的对象更少，而属性更多。在模糊形式概念之间存在着一种偏序关系，通过这种偏序关系，可以产生一个完全格结构，称为模糊概念格。

与精确概念格类似，模糊概念格也可以通过 Hasse 图将模糊形式概念节点之间的偏序集关系直观地表示出来。如果在建格时的模糊算子是确定的，那么产生的 Hasse 图是唯一的。模糊概念格和精确概念格的主要区别在于概念节点的不同，模糊形式概念节点的内涵是模糊形式背景的属性集的模糊子集。例如，一个模糊形式背景的实例如表 2 所示，可以得到对应的如图 2 所示的模糊概念格 Hasse 图。

表 2 模糊形式背景 $K(U,D,I)$

	a	b	c	d
1	0.7	0.9	0.4	0.8
2	0.8	1.0	0.5	0.8
3	0.7	0.8	0.8	0.9

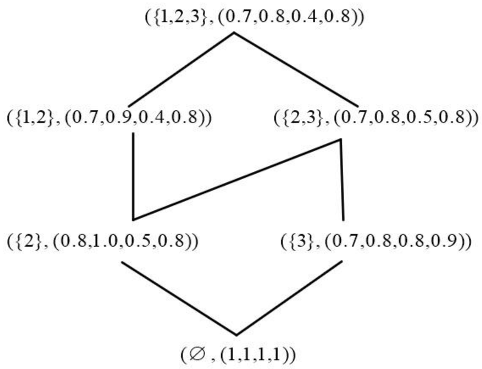

图 2 与表 2 对应的模糊概念格的 Hasse 图

3 模糊概念格的构造

概念格的构造算法是概念格研究的基础，也是当前研究的热点之一。概念格的结构通常具备完备性, 构造效率呈现指数级发展, 国内外很多学者专家 Bordat、Nourine、刘宗田、胡可云等从提高概念格构造效率的方向入手, 对概念格构造方法和规则算法进行研究,优化和改进了一些算法[4]。目前概念格构造的主流主要包括批处理算法、渐进式（增量式）算法和分布式算法。前两类是单机构造算法，随着数据规模的急剧增长，概念格的分布式构造算法也成为日益重要的研究内容。

概念格的渐进式构造算法的主要思想是：将对象逐个插入到概念格中，计算各个概念节点的内涵和待插入对象的属性集的交集，根据结果分别进行相应的处理。当形式背景被改变的时候，例如新增一个对象或者属性时，渐进式构造算法能够在已有的概念格的基础上进行构建，而不需要重复地从零开始构造概念格[5]。

很多科研人员对精确概念格的构造算法进行了比较深入的研究，而目前模糊概念格的构造算法方面的研究则比较少。与精确概念格类似，模糊概念格的构造算法也包括批处理和渐进式两种[6]。对精确概念格的渐进式构造算法

进行扩展和修改，可以设计出模糊概念格的渐进式构造算法。

3.1 基于对象的模糊概念格渐进式构造

在精确概念格的渐进式构造算法里面加入模糊集合的运算方法，就可以得到模糊概念格的渐进式构造算法。对于模糊形式背景 $K(U,D,I)$ 来说，模糊概念节点的内涵是 D 上的模糊集合，这些内涵之间存在偏序关系，结合模糊集合的运算方法，可以构造出模糊概念格。基于对象的渐进式构造模糊概念格的算法的步骤如下：

步骤一：初始化模糊概念格，里面只有一个底节点 $(\varnothing, f(\varnothing))$，即外延为空集的空概念，其内涵通常初始化为 $\sum_{j=1}^{|D|}\frac{1}{d_j}$，所以底节点可以记为 $(\varnothing,\sum_{j=1}^{|D|}\frac{1}{d_j})$。

步骤二：判断模糊形式背景中是否还有待处理对象，如果所有对象都已经插入了模糊概念格中，没有对象需要处理，则算法结束，否则就取出一个待插入对象 $(x, f(x))$。

步骤三：自顶向下按照外延势的降序来遍历模糊概念格，如果对于当前待插入对象 $(x, f(x))$ 来说，模糊概念格中的所有节点都已经被访问过或者被剪枝处理掉了，则执行步骤二，否则就从模糊概念格中按照外延势的降序取出一个待访问节点 $C(A,B)$。

步骤四：访问节点 $C(A,B)$，若 $B\cap f(x)=B$，则将 $C(A,B)$ 更新为 $(A\cup x,B)$，并为这个更新节点添加标注“*”，然后判断 $B=f(x)$ 是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤五：访问节点 $C(A,B)$，若 $B\cap f(x)\neq B$，并且存在 $C(A,B)$ 的某个最亲近父节点 $F^*(A_{F^*},B_{F^*})$ 满足 $B_{F^*}\supseteq B\cap f(x)$，则不必处理，然后判断 $f(x)\subset B$ 是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤六：访问节点 $C(A,B)$，若 $B\cap f(x)\neq B$，并且不存在 $C(A,B)$ 的某个最亲近父节点 $F^*(A_{F^*},B_{F^*})$ 满足 $B_{F^*}\supseteq B\cap f(x)$，则新增节点 $(A\cup x,B\cap f(x))$，并为这个新增节点添加标注“*”，设置新增节点 $(A\cup x,B\cap f(x))$ 为 $C(A,B)$ 的父节点，然后判断 $C(A,B)$ 是否有最亲近父节点 F^*。如果有，则最亲近父节点 F^* 的内涵为新增节点 $(A\cup x,B\cap f(x))$ 内涵的真子集，那么设置 F^* 为新增节点的父节点。接着判断 $f(x)\subset B$ 是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤七:运用剪枝技术，将 $C(A,B)$ 的所有子孙节点都剪掉，然后执行步骤三。

3.2 基于属性的模糊概念格渐进式构造

对于模糊形式背景 $K(U,D,I)$ 来说，基于属性的模糊概念节点 $C_i=(U_i,D_i)$ 的外延 U_i 是 U 上的模糊集合，所以这些外延之间存在偏序关系，结合模糊集合的运算方法，可以构造出基于属性的模糊概念格。基于属性的渐进式构造模糊概念格的算法的步骤如下：

步骤一：初始化模糊概念格，里面只有一个顶节点 $(g(\varnothing),\varnothing)$，其内涵为空集，外延通常初始化为 $\sum_{j=1}^{|U|}\frac{1}{x_j}$，所以顶节点可以记为 $(\sum_{j=1}^{|U|}\frac{1}{x_j},\varnothing)$。

步骤二：判断模糊形式背景中是否还有待处理属性，如果所有属性都已经插入了模糊概念格中，没有属性需要处理，则算法结束，否则就取出一个待插入属性 $(g(d),d)$。

步骤三：自底向上按照内涵势的降序来遍历模糊概念格，如果对于当前待插入属性 $(g(d),d)$ 来说，模糊概念格中的所有节点都已经被访问过或者被剪枝处理掉了，则执行步骤二，否则就从模糊概念格中按照内涵势的降序取出一个待访问节点 $C(A,B)$。

步骤四：访问节点 $C(A,B)$，若 $A\cap g(d)=A$，则将 $C(A,B)$ 更新为 $(A,B\cup d)$，并为这个更新节点添加标注“*”，然后判断 $A=g(d)$ 是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤五：访问节点 $C(A,B)$，若 $A\cap g(d)\neq A$，并且存在 $C(A,B)$ 的某个最亲近子节点 $S^*(A_{S^*},B_{S^*})$ 满足 $A_{S^*}\supseteq A\cap g(d)$，则不必处理，然后判断 $g(d)\subset A$ 是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤六：访问节点 $C(A,B)$，若 $A\cap g(d)\neq A$，并且不存在 $C(A,B)$ 的某个最亲近父节点 $S^*(A_{S^*},B_{S^*})$ 满足 $A_{S^*}\supseteq A\cap g(d)$，则新增节点 $(A\cap g(d),B\cup d)$，并为这个新增节点添加标注“*”，设置新增节点 $(A\cap g(d),B\cup d)$ 为 $C(A,B)$ 的子节点，然后判断 $C(A,B)$ 是否有最亲近子节点 S^*。如果有，则最亲近子节点 S^* 的外延为新增节点 $(A\cap g(d),B\cup d)$ 外延的真

子集，那么设置$S*$为新增节点的子节点。接着判断$g(d)\subset A$是否成立，如果成立，就执行步骤七，否则执行步骤三。

步骤七:运用剪枝技术，将$C(A,B)$的所有父节点都剪掉，然后执行步骤三。

4 结束语

概念格是一种优秀的数据分析工具，已经在数据挖掘、人工智能、知识表示、信息检索、模式识别、机器学习、专家系统等许多领域得到了广泛应用。实际应用中的数据信息大多具有模糊或不确定性的特征，要处理这样的问题，需要将概念格的应用加以推广。本文介绍了模糊概念格模型，将模糊理论与概念格理论结合起来，可用于探索性地分析和处理模糊信息，模糊概念格的研究具有重要意义。

参 考 文 献

[1] Ganter B, Stumme G, Wille R. Formal Concept Analysis: Theory and applications - J. UCS Special Issue. Journal of Universal Computer Science, 2004, 10(8): 926-926.

[2] 李金海、魏玲、张卓、翟岩慧、张涛、智慧来、米允龙. 概念格理论与方法及其研究展望[J]. 模式识别与人工智能, 2020, 33(7):24.

[3] Zadeh L. A. Fuzzy sets [J]. Information and control, 1965, 8(3): 338-353

[4] 降惠. 概念格理论研究进展与发展综述[J]. 办公自动化（办公设备与耗材）, 2019, 024(009):18-21,28.

[5] 邓硕. 多粒度标记粗糙集与概念格的规则比较[D].昆明理工大学,2019.

[6] 邹才凤. 模糊概念格构造与应用中的关键问题研究[D]. 华南理工大学.