第11期 2013年11月
电子学报 ACTA ELECTRONICA SINICA
Vol.41 No.11 Nov. 2013

基于逻辑合取的语言真值概念格合并算法

杨 丽[1],徐 扬[2]
(1.华北水利水电学院数学与信息科学学院,河南郑州 450046;2.西南交通大学数学学院,四川成都 610031)

摘 要: 针对人工智能领域中有关不确定性自然语言的研究,将格值逻辑引入概念格,提出了自然语言转换为语言真值的方法,并基于语言真值格蕴涵代数建立了语言真值概念格.通过对各语言真值形式背景及形式概念间关系的分析,演示了由多个语言真值概念格的合并而构造完整概念格的具体过程,进而给出了语言真值概念格基于逻辑合取的合并定理与算法.实验表明,与Bordat算法相比,当属性个数增多时,适当的合并算法能够使其时间复杂度呈指数型降低.

关键词: 概念格;语言真值概念格;不确定性自然语言;逻辑合取;合并算法
中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112(2013)11-2149-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2013.11.006

Union Algorithm of Linguistic Truth-Valued Concept Lattice Based on Logical Conjunction

YANG Li[1], XU Yang[2]
(1. *School of Mathematics and Information Science, North China University of Water Resources and Electric Power, Zhengzhou, Henan* 450046, *China*;
2. *College of Maths, Southwest Jiaotong University, Chengdu, Sichuan* 610031, *China*)

Abstract: For the uncertainty natural language processing research in the field of artificial intelligence, this paper, introducing the lattice-valued logic into the concept lattice, proposes the method of transforming the natural language into linguistic truth values and establishes the linguistic truth-valued concept lattice based on linguisitc truth-valued lattice implication algebra. In order to reduce the complexity of constructing the concept lattice, we demonstrate the specific process of constructing a complete linguistic truth-valued concept lattice through combining the multiple sub-concept lattices after analyzing the relationship among the linguistic truth-valued formal contexts and formal concepts, and then give the union theorem and algorithm based on logical conjunction. Experiments show that comparied with Bordat algorithm, the appropriate union algorithm can exponentially ruduce the time complexity when increasing the number of attributes.

Key words: concept lattice; linguistic truth-valued concept lattice; uncertainty natural language; logical conjunction; union algorithm

1 引言

概念格是由德国的Wille教授于1982年首先提出的[1],它作为一种知识表示和数据分析的形式化工具,以其清晰和易于理解的结构图形,实现了概念格的可视化,并在众多领域获得了成功的应用[2~5],这也使得概念格构造方面的研究成为其应用于大型复杂数据系统的前提.

概念格的分布式存储与并行处理正是以降低概念格构造的时空复杂度而发展起来的一种新的方法和手段,其思想就是通过形式背景的拆分,形成分布存储的多个子背景,然后构造相应的子概念格,再由子概念格的合并得到所需的概念格[4].其中,作为核心问题的概念格合并方法主要依赖于对形式背景的分解方式.目前,形式背景的分解有纵向和横向之分,对应地,概念格的合并也就有纵向合并和横向合并两种.文献[6,7]提出了纵向合并算法与横向合并算法,分别通过按内涵属性数目和外延对象数目的升序插入格节点的方式进行概念格的合并,避免了对更新和新生节点的重复比较,提高了算法效率;在此基础上,文献[8,9]对格节点作了

收稿日期:2012-12-14;修回日期:2013-04-02;责任编辑:马兰英
基金项目:国家自然科学基金(No.71101049,No.61175055);四川省科技厅应用基础研究项目资助(No.08JC0068);四川省科技支撑计划项目资助(No.2011FZ0051)

细分,分别提出了一种利用同义概念和同类概念更新节点的快速合并算法,进一步提高了概念格纵向合并以及横向合并的效率.

然而,上述概念格模型及相应的合并算法是无法用于处理以自然语言所刻画的不确定性信息.为了建立一种能够在人工智能领域中直接用于处理自然语言的数学模型,本文在语言真值格蕴涵代数的基础上建立了语言真值概念格[10,11],它不同于经典概念格和模糊概念格的关键之处在于:其取值域不再是$\{0,1\}$以及$[0,1]$,而是一个能够用于刻画不确定性信息的格蕴涵代数的完备格结构.该模型形成的理论依据是:用于刻画不确定性信息的格蕴涵代数[12]和用于形式化描述概念及概念间关系的经典概念格模型.对于本文所提出的适合于语言真值概念格的合并方法,其原理是通过各子背景中属性值间的关系来合并产生于各子背景的语言真值概念格,使最终基于逻辑合取关系合并而来的语言真值概念格同构于原形式背景上形成的语言真值概念格.

具体地,本文首先给出了一般自然语言转换为语言真值的具体步骤,并在语言真值格蕴涵代数上定义了伽罗瓦连接,进而建立了语言真值概念格模型;其次,在格值逻辑的合取关系下分别详细讨论了同一个语言真值概念格和不同语言真值概念格中不同语言真值形式概念间的关系;最后,给出了基于合并方法构造语言真值概念格的具体步骤以及合并算法,并通过实验与 Bordat 算法作了比较分析,进一步验证了该算法在降低时间复杂度上的有效性.

2 语言真值概念格

定义 1[1] 一个形式背景是一个三元组(U,A,F),这里 U 和 A 是集合,二元关系 $F\subseteq U\times A$,U 和 A 中的元素相应的被称为对象和属性.通常,我们用 uFa 表示$(u,a)\in F$,意思是"对象 u 具有属性 a".对于 $X\subseteq U$ 和 $Y\subseteq A$,定义伽罗瓦连接(f,g)如下:

$$f(X)=\{a\in A\mid uFa,\forall u\in X\},$$
$$g(Y)=\{u\in U\mid uFa,\forall a\in Y\}$$

因此,$f(X)$是 X 中所有对象所共有的属性之集,$g(Y)$是具有 Y 中所有属性的对象之集.形式背景(U,A,F)上的形式概念定义为元素对(X,Y),这里 $X\subseteq U$,$Y\subseteq A$,$f(X)=Y$ 和 $g(Y)=X$,形式概念(X,Y)的外延是 X,而它的内涵是 Y.

定义 2[1] 形式背景(U,A,F)上的所有形式概念的集用 $L(U,A,F)$表示,对于形式背景(U,A,F)上的形式概念(X_1,Y_1)和(X_2,Y_2)定义"子概念—父概念"为:$(X_1,Y_1)\leqslant(X_2,Y_2)\Leftrightarrow X_1\subseteq X_2(\Leftrightarrow Y_2\subseteq Y_1)$,即是说$(X_1,Y_1)$是$(X_2,Y_2)$的子概念,$(X_2,Y_2)$是$(X_1,Y_1)$的父概念.

定义 3[1] 如果对任意$(X_j,Y_j)\in L(U,A,F)$,上确界与下确界的定义如下:

$$\bigvee_{j\in J}(X_j,Y_j)=(g(\bigcap_{j\in J}Y_j),\bigcap_{j\in J}Y_j),$$
$$\bigwedge_{j\in J}(X_j,Y_j)=(\bigcap_{j\in J}X_j,f(\bigcap_{j\in J}X_j))$$

则称有序集$(L(U,A,F),\leqslant)$为(U,A,F)的概念格,记作 $L(U,A,F)$.

概念格是所有形式概念在子概念和父概念关系下的序集.因此,概念格可以图形化表示为其对应的 Hasse 图.这使得给定数据背景的概念结构变得清晰和易于理解,从而实现了概念格的可视化.

由于客观物理世界中以及人类对客观物理世界反映的过程中存在着大量的不确定性,并且人类的一些活动中也不得不经常处理这些不确定性信息.作为人类思维基本工具之一的自然语言,正是人们常常惯用的用以表示具有随机性、模糊性以及不可比较性等不确定性信息的方式,它同时也是人工智能领域中一个非常重要的研究内容.本文借助格蕴涵代数来表示不确定性信息,进一步用语言真值格蕴涵代数来刻画带有不确定性的自然语言信息.

定义 4[12] 设$(L,\wedge,\vee,')$是一个有泛界 O,I 的有余格.若映射$\rightarrow$:$L\times L\rightarrow L$ 满足:$\forall x,y,z\in L$,

(1)$x\rightarrow(y\rightarrow z)=y\rightarrow(x\rightarrow z)$.

(2)$x\rightarrow x=I$.

(3)$x\rightarrow y=y'\rightarrow x'$.

(4)若 $x\rightarrow y=y\rightarrow x=I$,则 $x=y$.

(5)$(x\rightarrow y)\rightarrow y=(y\rightarrow x)\rightarrow x$.

(6)$(x\vee y)\rightarrow z=(x\rightarrow z)\wedge(y\rightarrow z)$.

(7)$(x\wedge y)\rightarrow z=(x\rightarrow z)\vee(y\rightarrow z)$.

则称$(L,\wedge,\vee,',\rightarrow)$为一个格蕴涵代数.

用于表示属性值的自然语言有很多,属性的种类不同,所采用的自然语言就不同,比如,关于属性"程度"的属性值"比较好"、"有点好"、"非常差"等;关于属性"频度"的属性值"非常高"、"有点高"、"很低"等;关于属性"持续性"的属性值"非常长"、"有点短"、"很短"等.为了用一种统一的信息形式来刻画这些不同的自然语言,以下给出一种自然语言转换为语言真值的方法:

步骤 1 建立描绘性形容词所形成的集合 $S=$ {好,坏,大,小,长,短,高,低,多,少……} 以及元语言真值集合 $P=$ {真,假} $=\{p_j\mid j=1,2\}$.

步骤 2 在集合 S 与 P 之间建立映射 t,使得

$$\begin{cases}t(\text{好})=t(\text{大})=t(\text{长})=t(\text{高})=t(\text{多})=\cdots\cdots=\text{真},\\ t(\text{坏})=t(\text{小})=t(\text{短})=t(\text{低})=t(\text{少})=\cdots\cdots=\text{假}.\end{cases}$$

步骤 3　建立限定词所形成的集合 $R=\{$绝对,非常,很,相当,比较,差不多,有点……$\}=\{r_i \mid i=1,2,3\cdots\}$,并建立 R 与 P 间的笛卡尔积,称为语言真值集 $R\times P$,简记为 L,即 $L=R\times P=\{(r_i,p_j)\mid r_i\in R,p_j\in P\}$.

定义 5　设$(L,\wedge,\vee,',\rightarrow)$是一个格蕴涵代数,若集合 $L=\{l_1,l_2,\cdots,l_n\}$是语言真值集,其中 $l_i(i=1,2,\cdots,n)$为语言真值,则称$(L,\wedge,\vee,',\rightarrow)$为一个语言真值格蕴涵代数.

定义 6　称 $K=(U,A,L,F)$为语言真值形式背景,其中 U 为有限非空对象集,A 为有限非空属性集,$L=(L,\vee,\wedge,',\rightarrow)$是一个语言真值格蕴涵代数,$F$ 表示定义在 U 和 A 之间的格值模糊关系,即 $F:U\times A\rightarrow L$.

对于有限非空对象集 U 和语言真值格蕴涵代数$(L,\vee,\wedge,',\rightarrow)$,记 L^U 为定义在集合 U 上的所有格值模糊子集的集合,对任意的 $X_1,X_2\in L^U,u\in U,X_1\subseteq X_2\Leftrightarrow X_1(u)\leqslant X_2(u)$,则$(L^U,\subseteq)$是偏序集.

对于有限非空属性集 A 和语言真值格蕴涵代数$(L,\vee,\wedge,',\rightarrow)$,记 L^A 为定义在集合 A 上的所有格值模糊子集的集合,对任意的 $Y_1,Y_2\in L^A,a\in A,Y_1\subseteq Y_2\Leftrightarrow Y_1(a)\leqslant Y_2(a)$,则$(L^A,\subseteq)$是偏序集.

定理 1　设 $K=(U,A,L,F)$为语言真值形式背景,L 是一个语言真值格蕴涵代数,f,g 是定义在 L^U 与 L^A 之间的映射:

$$f:L^U\rightarrow L^A,$$
$$f(X)(a)=\bigwedge_{u\in U}(X(u)\rightarrow F(u,a));$$
$$h:L^A\rightarrow L^U,$$
$$g(Y)(u)=\bigwedge_{a\in A}(Y(a)\rightarrow F(u,a));$$

则对任意 $X\in L^U,Y\in L^A,(f,g)$是伽罗瓦链接.

证明　对任意,$X\in L^U,Y\in L^A$,

$$X\subseteq g(Y)\Leftrightarrow X(u)\leqslant\bigwedge_{a\in A}(Y(a)\rightarrow F(u,a)),\text{for any } u\in U,$$
$$\Leftrightarrow X(u)\leqslant Y(a)\rightarrow F(u,a),\text{for any } u\in U,a\in A$$
$$\Leftrightarrow Y(a)\leqslant X(u)\rightarrow F(u,a),\text{for any } u\in U,a\in A$$
$$\Leftrightarrow Y(a)\leqslant\bigwedge_{u\in U}(X(u)\rightarrow F(u,a)),\text{for any } a\in A$$
$$\Leftrightarrow Y\subseteq f(X),$$

因此,(f,g)是伽罗瓦链接.

定义 7　设 $K=(U,A,L,F)$为语言真值形式背景,若$\forall X\in L^U,Y\in L^A$,有 $f(X)=Y,h(Y)=X$,则定义语言真值形式概念为序对(X,Y);且对语言真值形式概念(X_1,Y_1)与(X_2,Y_2),定义

$$(X_1,Y_1)\leqslant(X_2,Y_2)\Leftrightarrow X_1\subseteq X_2(\text{or } Y_2\subseteq Y_1),$$

则称集合 $\mathscr{L}(K)=\{(X,Y)\mid f(X)=Y,g(Y)=X\}$为语言真值概念格.

3　语言真值概念格的合并

3.1　语言真值概念格的合并原理

语言真值概念格的合并方法与语言真值形式概念间的关系是密不可分的,以下,我们从逻辑合取的角度,具体讨论形式概念间存在的一些关系.

定义 8　设$(L^S,\subseteq)$是一个格,$\forall S_i,S_j\in L^S$,

(1)如果 $S_i\supseteq S_j$ 且不存在 $S_\alpha\in L^S$,使得 $S_i\supset S_\alpha\supset S_j$,则称 S_i 覆盖 S_j,记作 $S_i\succ S_j$.

(2)如果 $S_i\not\subseteq S_j$ 且 $S_i\not\supseteq S_j$,则称 S_i 与 S_j 不可比,记作 $S_i\parallel S_j$.

定义 9　设 $K=(U,A,L,F)$为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,$\forall C_i=(X_i,Y_i),C_j=(X_j,Y_j)\in\mathscr{L}(K)$.

(1)如果 $X_i=X_j$,或者 $Y_i=Y_j$,则称 C_i 与 C_j 相等,记作 $C_i=C_j$.

(2)如果 $X_i\supseteq X_j$,或者 $Y_i\subseteq Y_j$,则称 C_i 大于 C_j,或者称 C_j 小于 C_i,分别记作 $C_i\geqslant C_j,C_j\leqslant C_i$.

(3)如果 $X_i\supset X_j$,或者 $Y_i\subset Y_j$,则称 C_i 真大于 C_j,或者称 C_j 真小于 C_i,分别记作 $C_i>C_j,C_j<C_i$.

(4)如果 $X_i\parallel X_j$,或者 $Y_i\parallel Y_j$,则称 C_i 与 C_j 不可比,记作 $C_i\parallel C_j$.

定义 10　设 $K=(U,A,L,F)$为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,$\forall C_i=(X_i,Y_i),C_j=(X_j,Y_j)\in\mathscr{L}(K)$.

(1)如果 $C_i\geqslant C_j$ 且不存在 $C_\delta=(X_\delta,Y_\delta)\in\mathscr{L}(K)$,使得 $C_i>C_\delta>C_j$,则称 C_i 覆盖 C_j,记作 $C_i\succ C_j$.

(2)如果 $C_i<C_j$,或者 $C_i\parallel C_j$,则称 C_i 不大于 C_j,记作 $C_i\ngeq C_j$,或者 $C_j\nleq C_i$.

定义 11　设 $K_1=(U,A,L,F_1)$与 $K_2=(U,A,L,F_2)$为语言真值形式背景,如果 $U_1=U_2,A_1=A_2$ 且 $F_1\neq F_2$,则称 K_1 与 K_2 为同形语言真值形式背景,相应地,称 $\mathscr{L}(K_1)$与 $\mathscr{L}(K_2)$为同形语言真值概念格.

定义 12　设 $K_1=(U,A,L,F_1)$与 $K_2=(U,A,L,F_2)$为同形语言真值形式背景,$\mathscr{L}(K_1)$与 $\mathscr{L}(K_2)$是它们的语言真值概念格,$\forall C_1=(X_1,Y_1)\in\mathscr{L}(K_1),C_2=(X_2,Y_2)\in\mathscr{L}(K_2)$.

(1)如果 $X_1=X_2$,则称 C_1 与 C_2 同形相等,记作 $C_1\cong C_2$.

(2)如果 $X_1\supseteq X_2$,则称 C_1 同形大于 C_2,或称 C_2 同形小于 C_1,记作 $C_1\gtrsim C_2$,或 $C_2\lesssim C_1$.

(3) 如果 $X_1\supset X_2$,则称 C_1 同形真大于 C_2,或称 C_2 同形真小于 C_1,记作 $C_1\gnsim C_2$,或 $C_2\lnsim C_1$.

(4)如果 $X_1 > X_2$,则称 C_1 同形覆盖 C_2,记作 $C_1 \gtrsim C_2$.

(5)如果 $X_1 \parallel X_2$,则称 C_1 与 C_2 同形不可比,记作 $C_1 \Uparrow C_2$.

定理 2 设 $K=(U,A,L,F)$ 为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,$\forall C=(X,Y)\in\mathscr{L}(K)$,

(1)如果$\exists X_p \prec X$,且$(X_p,f(X_p))\in\mathscr{L}(K)$,则 $Y\prec f(X_p)$.

(2)如果$\exists X_p \prec X$,且$(X_p,f(X_p))\notin\mathscr{L}(K)$,则 $X_p\prec g(f(X_p))$.

证明 (1)$X_p\prec X\Rightarrow X_p\subset X\Rightarrow f(X)\subset f(X_p)$,即 $Y\subset f(X_p)$;假设$\exists Y_\beta\in L^A$,使得 $Y\subset Y_\beta\subset f(X_p)$,则 $g(f(X_p))\subset g(Y_\beta)\subset g(Y)$,即 $X_p\subset g(Y_\beta)\subset X$,这与 $X_p\prec X$ 矛盾,所以 $Y\prec f(X_p)$.

(2)根据伽罗瓦链接(f,g)的定义,$\forall u_k\in U$,

$$\begin{aligned}
&g(f(X_p))(u_k)\\
&=\bigwedge_{a\in A}(f(X_p)(a)\to F(u_k,a))\\
&=\bigwedge_{a\in A}(\bigwedge_{u_i\in U}(X_p(u_i)\to F(u_i,a))\to F(u_k,a))\\
&=\bigwedge_{a\in A}((X_p(u_k)\vee F(u_k,a))\vee(\bigwedge_{u_i\in U,i\neq k}(X_p(u_i)\\
&\to F(u_i,a))\to F(u_k,a)))\\
&>X_p(u_k),
\end{aligned}$$

即 $X_p\subset g(f(X_p))$.

假设$\exists X_\alpha\in L^U$,使得 $X_p\subset X_\alpha\subset g(f(X_p))$,若$(X_\alpha,f(X_\alpha))\in\mathscr{L}(K)$且$(g(f(X_p)),f(X_p))\in\mathscr{L}(K)$,则由 $X_\alpha\subset g(f(X_p))$,可知 $f(X_p)\subset f(X_\alpha)$,进而 $X_\alpha\subset X_p$,这与 $X_p\subset X_\alpha$ 矛盾;若$(X_\alpha,f(X_\alpha))\notin\mathscr{L}(K)$,则有$(g(f(X_\alpha)),f(X_\alpha))\in\mathscr{L}(K)$,即 $f(g(f(X_\alpha)))=f(X_\alpha)$,并且由$(X_p,f(X_p))\notin\mathscr{L}(K)$,可知$(g(f(X_p)),f(X_p))\in\mathscr{L}(K)$,即 $f(g(f(X_p)))=f(X_p)$,则又由 $X_\alpha\subset g(f(X_p))$,可知 $f(X_p)\subset f(X_\alpha)$,因此 $X_\alpha\subset X_p$,这仍与 $X_p\subset X_\alpha$ 矛盾.因此,不存在 $X_\alpha\in L^U$,使得 $X_p\subset X_\alpha\subset g(f(X_p))$,即 $X_p\prec g(f(X_p))$得证.

定理 3 设 $K=(U,A,L,F)$ 为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,$\forall C=(X,Y)\in\mathscr{L}(K)$.

(1) 如果$\exists$ $Y_p\prec Y$,使得$(g(Y_p),Y_p)\in\mathscr{L}(K)$,则 $X\prec g(Y_p)$.

(2) 如果$\exists$ $Y_p\prec Y$,使得$(g(Y_p),Y_p)\notin\mathscr{L}(K)$,则 $Y_p\prec f(g(Y_p))$.

证明 类似于定理 2 的证明过程.

定理 4 设 $K=(U,A,L,F)$ 为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,则$\forall(X_p,f(X_p))\notin\mathscr{L}(K)$,$\exists(X,Y)\in\mathscr{L}(K)$且 $X_p\prec X$,使得 $f(X_p)=Y$.

证明 由定理 2,$(X_p,f(X_p))\notin L(K)\Rightarrow(g(f(X_p)),f(X_p))\in\mathscr{L}(K)$且 $X_p\prec g(f(X_p))$,因此$\exists X=g(f(X_p))$,使得$(X,f(X_p))\in L(K)$,即 $f(X_p)=Y$.

定理 5 设 $K=(U,A,L,F)$ 为语言真值形式背景,$\mathscr{L}(K)$是语言真值概念格,则$\forall(g(Y_p),Y_p)\notin\mathscr{L}(K)$,$\exists(X,Y)\in\mathscr{L}(K)$且 $Y_p\prec Y$,使得 $g(Y_p)=X$.

证明 由定理 3,类似于定理 4 的证明过程.

定义 13 设 $K=(U,A,L,F)$,$K_1=(U,A,L,F_1)$,$K_2=(U,A,L,F_2)$为同形语言真值形式背景,如果$\forall u_i\in U$,$a_j\in A$,$F(u_i,a_j)=F_1(u_i,a_j)\wedge F_2(u_i,a_j)$,称 K 为 K_1 与 K_2 的逻辑合取形式背景,记作 $K=K_1\wedge K_2$.

定义 14 设 $K=(U,A,L,F)$,$K_1=(U,A,L,F_1)$,$K_2=(U,A,L,F_2)$为同形语言真值形式背景,$\mathscr{L}(K)$,$\mathscr{L}(K_1)$,$\mathscr{L}(K_2)$为同形语言真值概念格,且 $C=(X,Y)\in\mathscr{L}(K)$,$C_1=(X_1,Y_1)\in\mathscr{L}(K_1)$,$C_2=(X_2,Y_2)\in\mathscr{L}(K_2)$,如果 $X=X_1\cap X_2$,$Y=Y_1\cap Y_2$,则称 C 为 C_1 与 C_2 的逻辑合取形式概念.

对于两个语言真值概念格,通过合并的方法可将它们合二为一构造成为一个概念格,不妨设 $K_1=(U,A,L,F_1)$与 $K_2=(U,A,L,F_2)$是同形语言真值形式背景,其构造的具体过程如下:

步骤 1 对于 K_1,计算语言真值形式概念 $C_{1i}=(X_{1i},Y_{1i})$,$i=1,2,\cdots,m$,并构造语言真值概念格 $\mathscr{L}(K_1)$;对于 K_2,计算语言真值形式概念 $C_{2j}=(X_{2j},Y_{2j})$,$j=1,2,\cdots,n$,并构造语言真值概念格 $\mathscr{L}(K_2)$.

步骤 2 在语言真值概念格 $\mathscr{L}(K_1)$中添加语言真值形式概念 $C_{2j}=(X_{2j},Y_{2j})$,或者在语言真值概念格 $\mathscr{L}(K_2)$中添加语言真值形式概念 $C_{1i}=(X_{1i},Y_{1i})$,根据形式概念间的序关系"$C_{1i}\tilde{\leqslant}C_{2j}\Leftrightarrow X_{1i}\subseteq X_{2j}$",排列所有的语言真值形式概念 C_{1i}与 C_{2j},并建立它们的 Hasse 图,记作 $\mathscr{L}\{C_{1i},C_{2j}\}$.

步骤 3 在满足一定条件的语言真值形式概念 C_{1i}与C_{2j}之间执行逻辑合取运算,构造语言真值概念格 $\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$,并且其结构同构于 $\mathscr{L}(K_1\wedge K_2)$.

对于上述步骤 3 中,语言真值形式概念 C_{1i}与C_{2j}所要满足的条件,我们给出以下解释:

在 Hasse 图 $\mathscr{L}\{C_{1i},C_{2j}\}$中,任意两个语言真值形式概念 C_{1i}与C_{2j}之间至少存在一个链,其它不同于 C_{1i}与C_{2j}的语言真值形式概念存在于每条链上,这些语言真值形式概念的具体表示形式及数目将影响 C_{1i}与C_{2j}之间的逻辑合取运算,那么也就会直接影响到语言真值概念格 $\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$的结构.因此,我们有必要来讨论这些语言真值形式概念:

设 $C_{1p}=(X_{1p},Y_{1p})\in\mathscr{L}(K_1)$,$C_{2q}=(X_{2q},Y_{2q})\in\mathscr{L}(K_2)$,这里 $p\in\{1,2,\cdots,m\}$,$q\in\{1,2,\cdots,n\}$,对于 C_{1p}与C_{2q}之间链上的其它语言真值形式概念的个数 k,

记作 $C_{1p}\xrightarrow{[k]}C_{2q}$, $k\in\{0,1,\cdots,m+n-2\}$. 根据不同的 k,我们可以寻找到 C_{1p}与C_{2q}生成的新的语言真值形式概念:

(1)当 $k=0$ 时,$C_{1p}\tilde{\wedge}C_{2q}$是由$C_{1p}$与$C_{2q}$的逻辑合取运算而生成的语言真值形式概念,并且可以证明 $C_{1p}\tilde{\wedge}C_{2q}\in\mathscr{L}(K_1\wedge K_2)$.

(2) 当 $k\neq0$,分以下情况讨论:

① 若在该链上 $\exists C_{1p>}\in\mathscr{L}(K_1)$且 $C_{1p>}>C_{1p}$,则 $C_{1p}\tilde{\wedge}C_{2q}$是由 $C_{1p>}$与 C_{2q}生成的语言真值形式概念,并且可以证明 $C_{1p}\tilde{\wedge}C_{2q}\in\mathscr{L}(K_1\wedge K_2)$;类似的,若在该链上 $\exists C_{2q>}\in\mathscr{L}(K_2)$且 $C_{2q>}>C_{2q}$,同样可证明 $C_{1p}\tilde{\wedge}C_{2q}\in\mathscr{L}(K_1\wedge K_2)$;

② 若在该链上 $\exists C_{1p\not>}=(X_{1p\not>},Y_{1p\not>})\in\mathscr{L}(K_1)$且$C_{1p\not>}C_{1p}$,那么,当 $C_{1p\not>}\tilde{\geqslant}C_{2q}$时,由于 $X_{1p}\cap X_{2q}=X_{1p\not>}\cap X_{2q}$以及 $C_{1p\not>}\tilde{\wedge}C_{2q}=(X_{1p\not>}\cap X_{2q},Y_{1p\not>}\cap Y_{2q})\tilde{\in}\mathscr{L}(K_1\wedge K_2)$则 $C_{1p}\tilde{\wedge}C_{2q}\notin\mathscr{L}(K_1\wedge K_2)$;当 $C_{1p\not>}\tilde{\geqslant}C_{2q}$时,$C_{1p}\tilde{\wedge}C_{2q}\in\mathscr{L}(K_1\wedge K_2)$;对于该链上的 $C_{2q\not>}=(X_{2q\not>},Y_{2q\not>})\in\mathscr{L}(K_2)$,有类似的证明.

以下图 1 ~ 图 4 是由 $K_1=(U,A,L,F_1)$和 $K_2=(U,A,L,F_2)$分别生成的语言真值概念格 $\mathscr{L}(K_1)$和$\mathscr{L}(K_2)$的 Hasse 图,以及将 $\mathscr{L}(K_2)$中的语言真值形式概念添加到 $\mathscr{L}(K_1)$中的 $\mathscr{L}\{C_{1i},C_{2j}\}$,最后生成完整的语言真值概念格 $\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$的 Hasse 图:

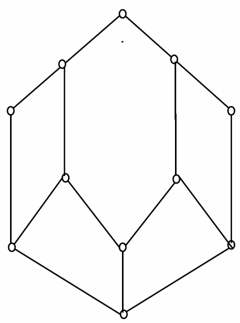

图1　$\mathscr{L}(K_1)$的Hasse图

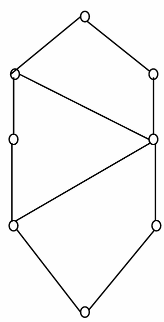

图2　$\mathscr{L}(K_2)$的Hasse图

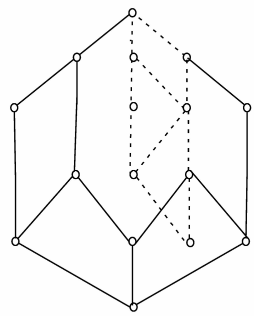

图3　$\mathscr{L}\{c_{1i},c_{2j}\}$的Hasse图

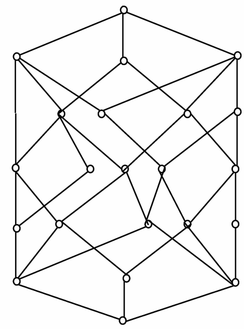

图4　$\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$的Hasse图

基于以上分析,对于任意的语言真值形式概念 C_{1i}与C_{2j},以及由此所形成的链结构,如果 $\exists C_{1i\not>}\in\mathscr{L}(K_1)$且同时满足条件 $C_{1i\not>}=C_{1_i}$与 $C_{1i\not>}\tilde{\geqslant}C_{2j}$,或者 $\exists C_{2j\not>}\in\mathscr{L}(K_2)$且同时满足条件 $C_{2j\not>}\not>C_{2j}$与 $C_{2j\not>}\tilde{\geqslant}C_{1i}$,则可以证得 $C_{1i}\tilde{\wedge}C_{2j}\notin\mathscr{L}(K_1\wedge K_2)$;如果 $C_{1i}\tilde{\wedge}C_{2j}\in\mathscr{L}(K_1\wedge K_2)$,我们可以得到 $C_{1i}\tilde{\wedge}C_{2j}\tilde{<}C_{1i}$.

下面的定义以及定理对上述的分析作了解释:

定义 15　设 $K=(U,A,L,F)$, $K_1=(U,A,L,F_1)$, $K_2=(U,A,L,F_2)$是同形语言真值形式背景,$\mathscr{L}(K)$, $\mathscr{L}(K_1)$, $\mathscr{L}(K_2)$是相应的同形语言真值概念格. 如果 $\forall C\in\mathscr{L}(K)$, $\exists C_1\in\mathscr{L}(K_1)$, $C_2\in\mathscr{L}(K_2)$,使得 $C=C_1\tilde{\wedge}C_2$,并且满足 $C_i\tilde{>}C$ 或者 $C_i\tilde{=}C$, $i=1,2$,称 $\mathscr{L}(K)$是 $\mathscr{L}(K_1)$与 $\mathscr{L}(K_2)$的逻辑合取语言真值概念格,记作$\mathscr{L}(K)=\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$.

定理 6　设 $K_1=(U,A,L,F_1)$与 $K_2=(U,A,L,F_2)$是同形语言真值形式背景,$\mathscr{L}(K_1)$与 $\mathscr{L}(K_2)$是相应的同形语言真值概念格,则 $\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)=\mathscr{L}(K_1\wedge K_2)$.

证明　由定义 13 可知,$K_1\wedge K_2=(U,A,L,F_1\wedge F_2)$,设$(f_1,g_1)$,$(f_2,g_2)$以及$(f,g)$分别为 K_1, K_2 以及 $K_1\wedge K_2$ 上的伽罗瓦连接.

根据定义 15, $\forall C=(X,Y)\in\mathscr{L}(K_1)\wedge\mathscr{L}(K_2)$, $\exists C_1=(X_1,Y_1)\in\mathscr{L}(K_1)$与 $C_2=(X_2,Y_2)\in\mathscr{L}(K_2)$,使得 $C=C_1\tilde{\wedge}C_2$,且满足 $C_i\tilde{>}C$ 或 $C_i\tilde{=}C$, $i=1,2$,也即,$X=X_1\cap X_2$ 与 $Y=Y_1\cap Y_2$,以及 $X_1>X$ 与 $Y_2>Y$,或者 $X_1=X$ 与 $Y_2=Y$,因此,由定理 4 知,在语言真值形式背景 K_1 上有 $f_1(X)=Y_1$,以及在 K_2 上有 $f_2(X)=Y_2$,并且对任意 $a\in A$,有

$$
\begin{aligned}
Y(a)&=(Y_1\cap Y_2)(a)\\
&=(f_1(X)\cap f_2(X))(a)\\
&=\bigwedge_{u_i\in A}(X(u_i)\rightarrow F_1(u_i,a))\wedge\bigwedge_{u_k\in A}(X(u_k)\rightarrow F_2(u_k,a))\\
&=\bigwedge_{u\in A}(X(u)\rightarrow F_1(u,a)\wedge F_2(u,a))\\
&=f(X)(a),
\end{aligned}
$$

所以,$C=(X,Y)\in\mathscr{L}(K_1\wedge K_2)$;

相反地,设$\forall C=(X,Y)\in\mathscr{L}(K_1\wedge K_2)$,则 $\exists C_1=(X_1,Y_1)\in\mathscr{L}(K_1)$以及 $C_2=(X_2,Y_2)\in\mathscr{L}(K_2)$,并且满足 $C_i\tilde{>}C$ 或 $C_i\tilde{=}C$, $i=1,2$.

一方面,若 $C_i\tilde{>}C$, $i=1,2$,则有 $X_1>X$, $X_2>X$ 以及 $Y_1>Y$, $Y_2>Y$. 因为 $C\notin\mathscr{L}(K_1)$且 $C\notin\mathscr{L}(K_2)$,由定理 4 知,在语言真值形式背景 K_1 上有 $f_1(X)=Y_1$,以及在

K_2 上有 $f_2(X)=Y_2$，并且在语言真值形式背景 $K_1\wedge K_2$ 上，对任意 $a\in A$，有

$$\begin{aligned}Y(a)&=f(X)(a)\\&=\bigwedge_{u\in U}(X(u)\rightarrow(F_1(u,a)\wedge F_2(u,a)))\\&=f_1(X)(a)\wedge f_2(X)(a)\\&=Y_1(a)\wedge Y_2(a)\\&=(Y_1\cap Y_2)(a),\end{aligned}$$

因此，$Y=Y_1\cap Y_2$；

其次，由定理 5 知，在语言真值形式背景 K_1 与 K_2 上，分别有 $g_1(Y)=X_1$ 与 $g_2(Y)=X_2$，并且在 $K_1\wedge K_2$ 上，对任意 $u\in U$，有

$$\begin{aligned}X(u)&=g(Y)(u)\\&=\bigwedge_{a\in A}(Y(a)\rightarrow(F_1(u,a)\wedge F_2(u,a)))\\&=g_1(Y)(u)\wedge g_2(Y)(u)\\&=X_1(u)\wedge X_2(u)\\&=(X_1\cap X_2)(u),\end{aligned}$$

因此，$X=X_1\cap X_2$.

另一方面，若 $C_1\tilde{=}C$ 且 $C_2\tilde{=}C$，则很显然有 $X=X_1=X_2$，即 $C=C_1\tilde{\wedge}C_2$，因此 $C=(X,Y)\in\mathcal{L}(K_1)\wedge\mathcal{L}(K_2)$；

综上，$\mathcal{L}(K_1)\wedge\mathcal{L}(K_2)=\mathcal{L}(K_1\wedge K_2)$.

3.2 语言真值概念格的合并算法

计算语言真值形式背景 K_i 上的语言真值形式概念 $C_{i_j}=(X_{i_j},Y_{i_j})$, $i=1,2,\cdots,m$; $j=1,2,\cdots,\|\mathcal{L}(K_i)\|$；其中，$\|\mathcal{L}(K_i)\|$ 是语言真值概念格中形式概念的个数 $\mathcal{L}(K_i)$.

Input：语言真值形式背景 $K_1,K_2,\cdots,K_m,(m\geqslant 2)$；

Output：语言真值形式背景 $K_1\wedge K_2\wedge\cdots\wedge K_m=K$ 上的语言真值形式概念；

Begin
  while $(K_1,K_2,\cdots,K_m\neq\Phi)$ do
  for $i\leftarrow 1$ to m do
    for $j\leftarrow 1$ to $\|\mathcal{L}(K_i)\|$ do
      $C_{i_j}:=C_{i_j}$
      for $p\leftarrow 1$ *to m* do
        if $p\neq i$ then
          for $q\leftarrow 1$ to $\|\mathcal{L}(K_p)\|$ do
            if $C_{p_q}\tilde{=}C_{i_j}$ then
              $C_{i_j}:=C_{i_j}\wedge C_{p_q}$
            endif
            if $C_{p_q}\tilde{>}C_{i_j}$ then $C_{i_j}:=C_{i_j}\wedge C_{p_q}$
            endif
          endfor
        else
        endfor
      else
    endfor
  endfor
 endfor
end

时间复杂度是检验一个算法好坏的主要因素之一，当语言真值形式背景随着属性个数的变化而发生变化时，构造相应的语言真值概念格所需的时间，我们可以通过简单的例子加以说明：假设 $K=(U,A,L_6,F)$ 为语言真值形式背景，其中属性的个数从 60 变到 80 个，下面的两组实验分别将形式背景 K 按照逻辑合取关系分解成两个和三个子背景，然后利用合并算法构造概念格，并分别与 Bordat 算法做比较，如图 5，图 6 所示：

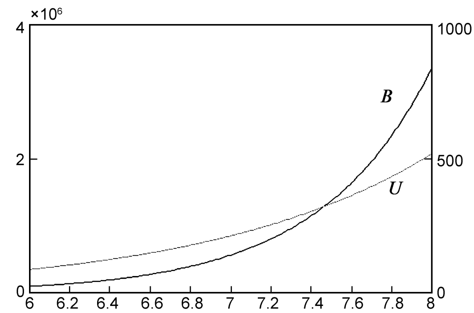

图5 时间复杂度比较I

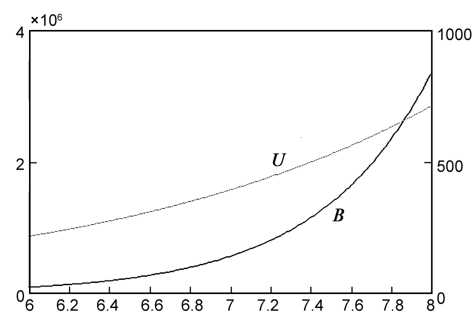

图6 时间复杂度比较II

在图 5 和图 6 中，曲线 B 和曲线 U 分别表示 Bordat 算法和合并算法（Union Algorithm）的时间复杂度，通过实验中所描绘的曲线，我们可以看到 Bordat 算法导致时间复杂度很大程度上的增加，并且增长率大于合并算法在时间复杂度上的增长率.另外，试验 I 与实验 II 中的绿色曲线分别代表由两个子概念格和三个子概念格通过合并算法合并形成完整概念格所用的时间，经比较发现，实验 I 的时间复杂度要大于实验 II 的，因此，选

择适当数目的子概念格作为合并的对象,并利用合并算法构建完整的语言真值概念格是可以提高建格速度的.

4 结束语

本文为人工智能领域中自然语言的处理提供了一种数学工具.从逻辑的角度出发,将自然语言转换为语言真值,并在语言真值格蕴涵代数的基础上建立了语言真值概念格;进而,为找到适合语言真值概念格构造及能够提高其构造速度的方法,提出了基于逻辑合取关系的合并原理及算法;最后,分析比较了该合并算法与 Borda 算法的时间复杂度,实验表明,该合并算法在很大程度上提高了语言真值概念格的构造速度.

参考文献

[1] R Wille. Reconstructing Lattice Theory: An Approach Based on Hierarchies of Concepts. In: Rival, I ed. Ordered Sets[M]. Dordrecht-Boston: Reidel, 1982, 445 – 470.

[2] B Ganter, R Wille. Conceptual Scaling. In: F. Roberts (ed.): Applications of Combinatorics and Graph Theory in the Biological and Social Sciences[M]. Springer-Verlag, NewYork, 1989, 139 – 167.

[3] 胡可云,陆玉昌,石纯一.基于概念格的分类和关联规则的集成挖掘方[J].软件学报,2000,11(11):1478 – 1484.
Hu Ke-yun, Lu Yu-chang, Shi Chun-yi. An integrated mining approach for classification and association rule based on concept lattice[J]. Journal of Software, 2000, 11(11): 1478 – 1484. (in Chinese)

[4] 魏玲,李强.面向属性概念格基于覆盖的压缩[J].电子科技大学学报,2012,141(2):299 – 304.
Wei Ling, Li Qiang. Covering-based reduction of property-oriented concept lattices[J]. Journal of University of Electronic Science and Technology of China, 2012, 141(2): 299 – 304. (in Chinese)

[5] 杨丽,王宇辉,徐扬.区间型决策概念格在质量——库存管理中的应用[J].计算机应用研究,2012,29(11):4228 – 4231.
Yang Li, Wang Yu-hui, Xu Yang. Application of interval decision concept lattice for quality-inventory cost[J]. Application Research of Computers, 2012, 29(11): 4228 – 4231. (in Chinese)

[6] 李云,刘宗田,等.多概念格的横向合并算法[J].电子学报,2004,32 (11):1849 – 1854.
Li Yun, Liu Zong-tian, et al. Horizontal union algorithm of multiple concept lattices[J]. Acta Electronica Sinica, 2004, 32(11): 1849 – 1854. (in Chinese)

[7] 张磊,沈夏炯,等.基于同类概念的概念格横向合并算法[J].计算机应用,2006,26(8):1900 – 1903.
Zhang Lei, Shen Xia-jiong, et al. Horizontal union algorithm of concept lattice based on congener concept[J]. Computer Applications, 2006, 26(8): 1900 – 1903. (in Chinese)

[8] 张磊,沈夏炯,等.基于同义概念的概念格纵向合并算法[J].计算机工程与应用,2007,43(2):95 – 98.
Zhang Lei, Shen Xia-jiong, et al. Vertical union algorithm of concept lattices based on synonymous concept[J]. Acta Eleltronica Sinica, 2007, 43(2): 95 – 98. (in Chinese)

[9] 智慧来,智东杰,刘宗田.概念格合并原理与算法[J].电子学报,2010,38(2):455 – 459.
Zhi Hui-lai, Zhi Dong-jie, Liu Zong-tian. Theory and algorithm of concept lattice union[J]. Acta Electronica sinica, 2010, 38 (2): 455 – 459. (in Chinese)

[10] Yang Li, Xu Yang, Decision making with uncertainty information based on lattice-valued fuzzy concept lattice[J]. Journal of Universal Computer Sciences, 2010, 16 (1): 159 – 177.

[11] Yang Li, Xu Yang, A decision method based on uncertainty reasoning of linguistic truth-valued concept lattice[J]. International Journal of General Systems, 2010, 39(3): 235 – 253.

[12] Xu Yang, Ruan Da, et al. Lattice-Valued Logic-an Alternative Approach to Treat Fuzziness and Incomparability [M]. Springer-Verlag, Berlin. 2003, 27 – 57.

作者简介

杨　丽　女,讲师,博士,主要研究方向为智能信息处理,概念格.
E-mail: yangli6672@sina.com

徐　扬　男,教授,博士生导师,主要研究方向为逻辑代数,代数逻辑,智能信息处理.
E-mail: xuyang@swjtu.edu.cn