

一种启发式知识约简算法

刘启和 闵帆 蔡洪斌 杨国伟

(电子科技大学计算机科学与工程学院 成都 610054)

摘要 属性约简是 Rough 集理论中的核心问题之一,找出所有的约简或最小约简是一个 NP 难题。本文证明了正区域和边界域的一些性质,指出在考虑正区域作为启发信息的同时,还应该考虑在不一致决策表中边界域对约简的影响,综合这两种信息,提出了不一致决策表约简的启发信息。并在此基础上,设计了一致决策表的启发式约简算法。实验证明,在多数情况下,该算法能够得到决策表的最小或次优约简。

关键词 Rough 集,属性约简,边界域,正区域

A Heuristic Algorithm of Knowledge Reduction

LIU Qi-He MIN Fan CAI Hong-Bin YANG Guo-Wei

(College of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

Abstract In rough sets theory, reduction of attributes is an important issue. It has been proved that computing all reductions or the minimal reduction of decision table is a NP-hard problem. Now, many algorithms for reduction of attributes are still heuristic algorithms. In this paper, new heuristic information is proposed. We consider boundary region (complement of positive region) can affect reducing attributes in inconsistent decision table, so we use not only positive region but also boundary region to calculate this heuristic information. Based on this new heuristic information. We develop heuristic algorithm for reduction of attributes in inconsistent decision table. In order to test efficiency of the algorithm, an example is analyzed and some experiments are made. The analysis and experimental results show that the algorithm is efficient and capable of finding the minimal or suboptimal reduction in most of cases.

Keywords Rough set, Reduction of attributes, Boundary region, Positive region

1 引言

由波兰学者 Pawlak 教授提出的 Rough 集理论是一种新的处理不完备、不精确和不确定信息系统的重要数学工具。其主要思想是在保持信息系统分类能力不变的前提下,通过知识约简来获得问题的分类或决策规则。它已经在机器学习、数据挖掘、模式识别与分类、故障检测等方面取得了广泛的应用。

属性约简(知识约简)是指保持信息系统的分类或决策能力不变的条件下,删除其中的冗余属性。属性约简是 Rough 集理论中的核心问题之一,找出所用的约简或最小约简(即属性个数最小)是一个 NP 难题^[3],故一般采用启发式信息找出最优或次优约简。这些算法的共同特点是利用属性的重要性作为启发式信息^[2],分别有基于正区域的^[2]、基于区分矩阵属性频率的^[5]和基于信息熵的约简算法^[3,4]。

本文在以正区域为启发式信息的约简算法基础上^[2],分析了不一致决策表在约简过程中边界域的变化情况,证明了边界域的一些性质。在此基础上,综合考虑正区域和边界域两者,提出了一种新的启发式信息,并以此设计了一个约简算法。对该算法进行了时间复杂度分析,结果说明该约简算法与基于正区域的约简算法^[2]的时间复杂度相同。实验说明,在不一致决策表中,该算法优于仅以正区域为启发信息的约简算法。在多数情况下,该算法能够搜索到最小或次优约简。

2 基本概念

关于 Rough 集的基本概念,如约简、核、集合的上下近似等,可参考有关文[1,6]。这里,为后面叙述方便,引入一些记

号并介绍一下正区域、不一致决策表、属性重要性等概念。

设信息表 S 为一个四元组 $S = \langle U, A, V, f \rangle$, 其中 U 是一个非空有限的对象集合,称为论域, A 是属性集合, $V = \bigcup_{a \in A} V_a$, V_a 表示属性 a 的值域, $f: U \times A \rightarrow V$ 是一个信息函数,它指定 U 中每一个对象 x 的属性值,即对任意的 $x \in U$, $a \in A$, 有 $f(x, a) \in V_a$ 。一般简记信息表 S 为 $\langle U, A \rangle$ 。如果属性集 A 被分为条件属性集合 C 和决策属性集合 D , 即

$$C \cup D = A, C \cap D = \emptyset \quad (1)$$

则称此信息表 S 为决策表,其中 D 一般只含有一个决策属性,决策表 S 简记为 $S = \langle U, C, D \rangle$ 。

在信息表 S 中,对任意非空属性子集 $B \subseteq A$, 定义一个不可区分关系 $IND(B)$:

$$IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, f(x, a) = f(y, a)\} \quad (2)$$

显然 $IND(B)$ 是一个等价关系,对象 x 在 $IND(B)$ 中的等价类 $[x]_B$ 定义为:

$$[x]_B = \{y \mid y \in U, (x, y) \in IND(B)\} \quad (3)$$

由此 B 在 U 上导出的划分记为 $U / IND(B)$ 。为简单起见,在不产生混淆情况下可以用 B 代替 $IND(B)$, 即 $U / IND(B)$ 可记为 U / B 。

设 $X \subseteq U$, 集合 X 的 B -下近似和 B -上近似分别定义如下:

$$B(X) = \underline{B}X = \bigcup \{[x]_B \mid [x]_B \subseteq X\} \quad (4)$$

$$\bar{B}(X) = \overline{B}X = \bigcup \{[x]_B \mid [x]_B \cap X \neq \emptyset\} \quad (5)$$

定义 1^[6] 设在决策表 $S = \langle U, C, D \rangle$ 中, $P \subseteq C$, 则决策属性 D 的 P -正区域 $POS_P(D)$ 定义为

$$POS_P(D) = \bigcup_{x \in U/D} P(x) \quad (6)$$

其中 $\underline{P}(X)$ 为 X 的 P -下近似集;记 $\gamma_p(D) = |\text{POS}_p(D)|/|U|$ 。

定义 2^[4] 设在决策表 $S = \langle U, C, D \rangle$ 中,若决策属性 D 的 C -正区域 $\text{POS}_C(D) = U$,则称决策表 S 为一致决策表,否则称 S 为不一致决策表。

定义 3^[2](属性重要性定义) 设在决策表 $S = \langle U, C, D \rangle$ 中, $P \subseteq C$,则对任意的条件属性 $a \in C - P$ 的属性重要性 $\text{SGF}_1(a, P, D)$ 定义为

$$\text{SGF}_1(a, P, D) = r_{PU(a)}(D) - r_P(D) \quad (7)$$

3 正区域和边界域的性质

以下假设在决策表 $S = \langle U, C, D \rangle$ 中,决策属性集 D 导出 U 上划分为 Y ,其中 $Y = \{Y_1, Y_2, \dots, Y_m\}$,若 $P \subseteq C$,对任意的 $Y_i \in Y$, $\underline{PY}_i$ 表示集合 Y_i 的 P 下近似集, $\text{POS}_P(D)$ 表示 D 的 P -正区域。

定理 1 在决策表 S 中, $P \subseteq C, x \in U$,用 $[x]_P$ 表示由属性集 P 所确定的等价类,如果

$$[x]_P \cap \text{POS}_P(D) \neq \emptyset, \text{则 } [x]_P \subseteq \text{POS}_P(D).$$

证明:因为

$$\text{POS}_P(D) = \bigcup_{i=1}^m \underline{PY}_i \quad (8)$$

又 $[x]_P \cap \text{POS}_P(D) \neq \emptyset$,所以存在 $i \in \{1, 2, \dots, m\}$,使得

$$[x]_P \cap \underline{PY}_i \neq \emptyset \quad (9)$$

则存在 y ,使得 $y \in [x]_P$ 和 $y \in \underline{PY}_i$ 。由 $y \in \underline{PY}_i$,可得 $[y]_P \subseteq \underline{PY}_i$,由 $y \in [x]_P$,可得 $[y]_P = [x]_P$,故

$$[x]_P \subseteq \text{POS}_P(D) \quad (10)$$

定理 2 在决策表 S 中, $P \subseteq C, U - \text{POS}_P(D)$ 是正区域 $\text{POS}_P(D)$ 的补集,则

$$\underline{P}(U - \text{POS}_P(D)) = U - \text{POS}_P(D), \underline{P}(\text{POS}_P(D)) = \text{POS}_P(D) \quad (11)$$

证明:由下近似的定义^[6],显然有

$$\underline{P}(U - \text{POS}_P(D)) \subseteq U - \text{POS}_P(D) \quad (12)$$

对任意的 $x \in U - \text{POS}_P(D)$,设 P 确定的 x 等价类为 $[x]_P$,必有

$$[x]_P \subseteq U - \text{POS}_P(D) \quad (13)$$

否则 $[x]_P \cap \text{POS}_P(D) \neq \emptyset$,由定理 1 有

$$[x]_P \subseteq \text{POS}_P(D) \quad (14)$$

这与 $x \in U - \text{POS}_P(D)$ 矛盾,所以有

$$[x]_P \subseteq U - \text{POS}_P(D) \quad (15)$$

因此, $[x]_P \subseteq \underline{P}(U - \text{POS}_P(D))$,即

$$x \in \underline{P}(U - \text{POS}_P(D)) \quad (16)$$

由 x 的任意性,有

$$U - \text{POS}_P(D) \subseteq \underline{P}(U - \text{POS}_P(D)), \text{所以} \quad (17)$$

同理可得:

$$\underline{P}(\text{POS}_P(D)) = \text{POS}_P(D) \quad (18)$$

定义 4 设在决策表 $S = \langle U, C, D \rangle$ 中, $U - \text{POS}_C(D)$ 是正区域 $\text{POS}_C(D)$ 的补集,称 $U - \text{POS}_C(D)$ 为决策属性集 D 相对于 C 的边界域。

显然,在一致决策表 S 中,边界域为空集。

由定理 1 和定理 2 可知,

$$\underline{P}(\text{POS}_P(D)) = \underline{P}(\text{POS}_P(D)) = \text{POS}_P(D) \quad (19)$$

所以 $\text{POS}_P(D)$ 是一个 P -可定义集。同理 $U - \text{POS}_P(D)$ 也是一个 P -可定义集。这样,正区域和边界域都是可定义集。根据约简的定义^[6],决策表约简后形成的规则不仅要能够覆盖

正区域($\text{POS}_C(D)$)中的对象,而且不能覆盖边界域($U - \text{POS}_C(D)$)中的对象。这类类似于机器学习中的概念学习,概念不仅要覆盖正例,而且不能覆盖反例。因此,在不一致决策表中,应该综合考虑正区域和边界域的信息来求相对约简。

根据正区域和等价关系的定义,易证定理 3 和定理 4。

定理 3 在决策表 S 中, $P \subseteq A \subseteq C$,则

$$\text{POS}_P(D) \subseteq \text{POS}_A(D) \subseteq \text{POS}_C(D) \quad (20)$$

定理 4 在决策表 S 中, $P \subseteq A \subseteq C$,令 $\mathcal{R}_P = \{[x]_P \mid x \in U \cup [x]_P \subseteq U - \text{POS}_C(D)\}$, $\mathcal{R}_A = \{[x]_A \mid x \in U \cup [x]_A \subseteq U - \text{POS}_C(D)\}$,则

$$\mathcal{R}_P \subseteq \mathcal{R}_A \subseteq U - \text{POS}_C(D) \quad (21)$$

在属性约简中,通常以核属性集为起点,采用 bottom-up 的方式逐步添加属性求相对约简。定理 3 说明了添加属性会让正区域增加;定理 4 说明了添加属性会让包含在边界域中 $U - \text{POS}_C(D)$ 的等价类增加。这样,可以以 $\text{POS}_P(D)$ 和 $\mathcal{R}_P$ 为启发信息来计算相对约简。

定义 5 设在决策表 $S = \langle U, C, D \rangle$ 中, $P \subseteq C, U - \text{POS}_C(D)$ 是正区域 $\text{POS}_C(D)$ 的补集,令

$$\mathcal{R}_P = U \setminus \{[x]_P \mid x \in U \cup [x]_P \subseteq U - \text{POS}_C(D)\} \quad (22)$$

记 $\theta_P(D) = |\mathcal{R}_P|/|U|$ 。

定义 6(新的属性重要性定义) 设在决策表 $S = \langle U, C, D \rangle$ 中, $P \subseteq C, a \in C - P$,则对任意的条件属性 $a \in C - P$ 的属性重要性 $\text{SGF}(a, P, D)$ 定义为:

$$\text{SGF}(a, P, D) = \lambda_1 (r_{PU(a)}(D) - r_P(D)) + \lambda_2 (\theta_{PU(a)}(D) - \theta_P(D)) \quad (23)$$

其中 $0 \leq \lambda_2 \leq \lambda_1 \leq 1, \lambda_1 + \lambda_2 = 1$ 。

λ_1 和 λ_2 是权重系数,用以调节正区域和边界域对 $\text{SGF}(a, P, D)$ 的影响。当 $\lambda_1 = 1, \lambda_2 = 0$ 时,有 $\text{SGF}(a, P, D) = \text{SGF}_1(a, P, D)$ 。当使用 $\text{SGF}(a, P, D)$ 为启发信息求相对约简时,系数 λ_1 和 λ_2 可以人为指定,也可以根据数据集的具体情况指定,或由机器学习方法来获得,如可以指定 λ_1 和 λ_2 的值如下:

$$\lambda_1 = \frac{|\text{POS}_C(D)|}{|U|}, \lambda_2 = \frac{|U - \text{POS}_C(D)|}{|U|}$$

在本文中,如果没有特别说明,我们指定 $\lambda_1 = 0.5, \lambda_2 = 0.5$,用以说明正区域和边界域有同等的重要性。

在一致决策表中,由于 $\theta_P(D) = 0$,有

$$\text{SGF}(a, P, D) = \lambda_1 \text{SGF}_1(a, P, D) \quad (24)$$

与 $\text{SGF}_1(a, P, D)$ 仅相差一个常数倍,此时 $\text{SGF}(a, P, D)$ 与 $\text{SGF}_1(a, P, D)$ 有相同的启发信息。

4 知识约简算法

在本文中,以属性个数来定义最优的程度,即属性个数最小(或较小)的约简为最优(或次优)约简。为了考虑边界域对约简的影响,看下面的一个例子。

表 1 一个决策表系统

U	a	b	c	e	f	d
1	1	0	1	0	1	0
2	1	1	0	0	1	1
3	0	0	0	0	1	1
4	1	1	0	1	0	1
5	1	0	1	0	1	1
6	1	0	1	0	0	1
7	0	0	1	1	0	0
8	0	1	1	1	1	0
9	1	0	0	1	1	0
10	0	1	1	1	1	1

例1 如表1所示的决策表系统,其中 $C = \{a, b, c, e, f\}$ 是条件属性集, $D = \{d\}$ 是决策属性集。其中

$U/D = \{\{1, 7, 8, 9\}, \{2, 3, 4, 5, 6, 10\}\}, U/C = \{\{1, 5\}, \{2\}, \{3\}, \{4\}, \{6\}, \{7\}, \{8, 10\}, \{9\}\},$

$POS_C(D) = \{2, 3, 4, 6, 7, 9\}, U - POS_C(D) = \{1, 5, 8, 10\}, CORE_D(C) = \{f\}$ 。最小约简是 $\{c, e, f\}$ 。

如果以核属性集为起点,以 SGF_1 为启发信息,则计算各属性的重要性:

$SGF_1(a, \{f\}, D) = 0.3, SGF_1(b, \{f\}, D) = 0.1, SGF_1(c, \{f\}, D) = 0.1, SGF_1(e, \{f\}, D) = 0.1$

因 $SGF_1(a, \{f\}, D)$ 最大,故增加属性 a ,可得 $\{a, f\}$ 。由于 $|POS_{\{a, f\}}(D)| \neq |POS_C(D)|$,因此又以 $\{a, f\}$ 为起点,分别计算属性 b, c, e 的 SGF_1 。如此反复,最终得到约简 $\{a, b, c, f\}$,而不是最小约简 $\{c, e, f\}$ 。

如果我们同样以核属性集为起点,但以 SGF 为启发信息,则计算各属性的重要性:

$SGF(a, \{f\}, D) = 0.15, SGF(b, \{f\}, D) = 0.05, SGF(c, \{f\}, D) = 0.25, SGF(e, \{f\}, D) = 0.05$

因 $SGF(c, \{f\}, D)$ 最大,故增加属性 c ,可得 $\{c, f\}$ 。同样,由于 $|POS_{\{c, f\}}(D)| \neq |POS_C(D)|$,故应反复上述过程,最终得到最小约简 $\{c, e, f\}$ 。

例1说明 $SGF(a, P, D)$ 可以作为决策表约简算法中属性重要性的度量。下面以 $SGF(a, P, D)$ 设计约简算法。

算法1

输入:决策表 $S = \langle U, C, D \rangle$,其中 U 为论域, C, D 分别为条件属性集和决策属性集。

输出:决策表 S 的一个相对约简 P 。

Setp1 计算条件属性集 C 相对于决策属性集 D 的核属性集 $CORE_D(C)$,并令 $B = C - CORE_D(C)$ 。

Step2 计算决策表 S 中条件属性集 C 相对于决策集 D 的正区域 $POS_C(D)$ 和边界域 $U - POS_C(D)$ 。

Setp3 令 $P = CORE_D(C)$,

Setp3.1 如果 $|P| \neq 0$,则计算 $POS_P(D)$,转Setp3.4;

Setp3.2 对每个属性 $a \in B$,计算属性 a 的重要性 $SGF(a, P, D)$;

Setp3.3 选择使 $SGF(a, P, D)$ 最大的属性 a ,把 a 从 B 中删除,并把 a 增加到 P 的尾部;

Setp3.4 如果 $|POS_P(D)| = |POS_C(D)|$,则转Setp4,否则转Setp3.2。

Setp4 从 P 的尾部开始,从后往前判别每个属性 a 是否可约:

如果 $a \in CORE_D(C)$,则从 a 开始往前的属性都是核中属性,不可约,算法终止;

否则,如果 $|POS_P(D)| = |POS_{P-\{a\}}(D)|$,则 a 是可约简的,把 a 从 P 中删除。

比较算法1与文[2]中的约简算法(简称为算法2),其不同之处在于算法2以 $SGF_1(a, P, D)$ 为启发信息,而算法1以 $SGF(a, P, D)$ 为启发信息。如果以文[2]中的方法计算等价类、正区域和核,那么计算 $SGF(a, P, D)$ 的时间复杂度为 $O((|D|+1)|U|\log(|U|))$,与计算 $SGF_1(a, P, D)$ 的时间复杂度相同。因此算法1和算法2具有相同的时间复杂度,即算法1的时间复杂度为 $O((|C|^2|U|\log(|U|)))$ 。其详细的时间复杂度分析可以参见文[2]。

5 实验结果

我们选用例1和UCI机器学习数据库中的部分决策表在PC机(eleron1.7G,256M RAM,WINXP)上进行实验,分别采用算法1和算法2进行知识约简,并且算法使用Java语言实现。实验结果如表2所示。

从表2可以看出,算法1与算法2的执行时间大致相同。在不一致决策表中,算法1约简后的条件属性数小于或等于算法2约简后的条件属性数,这说明算法1提前搜索到约简。在一致决策表中,算法1与算法2计算的结果相同。这是因为当决策表 $S = \langle U, C, D \rangle$ 是一致决策表时, $SGF(a, P, D) = \lambda_2 SGF_1(a, P, D)$,此时 $SGF(a, P, D)$ 与 $SGF_1(a, P, D)$ 有相同的启发信息,所以在一致决策表中算法1和算法2计算的结果相同。

表2 约简结果与执行时间比较

决策表	是否是一致决策表	实例数	约简前的条件属性数	算法1		算法2	
				约简后的条件属性数	执行时间(s)	约简后的条件属性数	执行时间(s)
例1	否	10	5	3	0.04	4	0.05
voting-records	是	435	16	9	0.27	9	0.26
tic-tac-toe	是	958	9	8	0.48	8	0.48
mushroom	是	8124	22	5	6.42	5	6.42
chess end-game	是	3196	36	29	4.67	29	4.67

因为UCI数据库中大量的是一致决策表,所以为了考察两种算法对不一致决策表约简的可能差异,我们编写了一个决策表随机生成器,用来生成不一致决策表,然后分别使用算法1和算法2进行约简,并比较约简结果中属性的个数。实验结果如表3所示。

从表3可以看出:在多数情况下(99.81%),算法1在约简后,其约简中属性数小于或等于算法2约简后结果中的属性数;在极少数情况下(0.19%),算法1在约简后,其约简中属性个数大于算法2约简后结果中的属性数。因此,算法1在寻找最优或次优约简上优于算法2。由于算法2能在多数情

况下获得最小约简^[1],因此算法1也能够多数情况下获得最小约简。

表3 约简结果比较

Number	Smaller or Equal	Larger
10000	9981	19

注:Number表示不一致决策表的个数;Smaller or Equal表示算法1得到的约简中属性数小于或等于算法2得到的约简中属性数的决策表个数;Larger表示算法1得到的约简中属性数大于算法2得到的约简中属性数的决

策表个数。

结论 本文分析了 Rough 集理论中正区域和边界域的一些性质,认为在不一致决策表的约简过程中,既应该考虑正区域对约简过程影响,又应该考虑边界域对约简过程的影响。这类类似于机器学习中的概念学习,概念不仅要覆盖正例,而且不能覆盖反例。在此基础上,给出了一种新的属性重要性定义,这种定义既含有正区域信息又含有边界域信息。

最后,以新的属性重要性为启发信息,给出了一个新的算法,该算法的时间复杂度与基于正区域的约简算法时间复杂度相同。通过例子和仿真实验说明,新算法在搜索最优或次

优约简上优于传统的算法。

参考文献

- Pawlak Z. Rough sets approach to multi-attribute decision analysis. *European Journal of Operational Research*, 1994, 72:443~459
- 刘少辉,盛秋骛,吴斌,史忠植,胡斐. Rough 集高效算法研究. *计算机学报*, 2003, 26(5):524~529
- 苗夺谦,胡桂荣. 知识约简的一种启发式算法. *计算机研究与发展*, 1999, 36(6):681~684
- 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简. *计算机学报*, 2002, 25(7):759~766
- Wang Jue, Wang Ju. Reduction Algorithms Based on Discernibility Matrix: The ordered Attributes Method. *Journal Computer Science & Technology*, 2001, 16(6):489~504
- 张文修,等. 粗糙集理论与方法. 北京:科学出版社, 2001
- Smirnova E S, So C M, Watt S M. An Architecture for Distributed Mathematical Web Services. In: MKM2004, LNCS 3119, 2004. 363~377
- Franke A, Hess S M, Jung C G, et al. Agent-Oriented Integration of Distributed Mathematical Services. *Journal of Universal Computer Science*, 1999, 5:156~187
- Kohlhase M. Towards a Knowledge-Centered Infrastructure for Web-Based Mathematics. Available at: <http://www.cs.cmu.edu/~kohlhase>, February 14, 2001
- Homann K, Calmet J. Combining Theorem Proving and Symbolic Mathematical Computing, Integrating Symbolic Mathematical Computation and Artificial Intelligence. In: Proc. of 2nd Intl. Conf. on Artificial Intelligence and Symbolic Mathematical Computing (AISMC-2), Karlsruhe, Springer, LNCS 958, 1995. 18~29
- Franke A, Kohlhasse M. System Description: MATHWEB, an Agent-Based Communication Layer for Distributed Automated Theorem Proving. FB Informatik, University des Saarlandes. In: Harald Ganzinger, ed. *Automated Deduction - CADE-16*, LNAI 1632, Springer Verlag, 1999. 217~221
- Noy N F, Musen M A. SMART: Automated Support for Ontology Merging and Alignment. In: Proc. of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management, Banff, Canada, July 1999
- Sure Y, Staab S, Angele J, et al. OntoEdit: Guiding Ontology Development by Methodology and Inferencing. In: Prestigious, Applications of Intelligence System (PAIS), ECAI 2002, Lyon, France, 2002
- 陆汝钊. 世纪之交的知识工程与知识科学. 北京:清华大学出版社, 2001
- Carlisle D, Ion P, Miner R, et al. Mathematical Markup Language (MathML) version 2.0. W3C Recommendation, World Wide Web Consortium, 2001. Available at: <http://www.w3.org/TR/MathML2>
- Extensible Markup Language (XML). Available at <http://www.w3.org/XML/>
- Bray T, Paoli J, Sperberg-McQueen C M. Extensible Markup Language (XML). W3C Recommendation TRXML, World Wide Web Consortium, December 1997. Available at: <http://www.w3.org/TR/PR-xml.html>
- Raggett D, Hors A L, Jacobs I. HTML 4.0 Specification. W3C Recommendation REC-html40, World Wide Web Consortium, April 1998. Available at: <http://www.w3.org/TR/PR-xml.html>
- Kohlhasse M. Creating OMDoc Representations from LATEX. Internet Draft. Available at: <http://www.mathweb.org/omdoc>, 2000
- Caprotti O, Cohen A M. Draft of the OpenMath Standard. The Open Math Society. Available at: <http://www.nag.co.uk/projects/OpenMath/.mstd/>, 1998
- Kohlhasse M. OMDoc: An Open Markup Format for Mathematical Documents (Version 1.1). Available at: <http://www.cs.cmu.edu/~kohlhasse>, 2001
- Kohlhasse M. OMDoc: An Open Markup Format for Mathematical Documents; [Seki Report SR-00-02]. Fachbereich Informatik, Universität des Saarlandes, 2000. Available at: <http://www.mathweb.org/omdoc>
- Kohlhasse M. OMDoc: An Infrastructure for Openmath Content Dictionary Information. Bulletin of the ACM Special Interest Group on Symbolic and Automated Mathematics (SIGSAM), 2000, 34(2):43~48
- Kohlhasse M, Franke A. MBase: Representing Knowledge and Context for the Integration of Mathematical Software System. Available at: <http://www.cs.cmu.edu/~kohlhasse>
- Asperti A, Padovani L, Coen C S, et al. Towards a Library of Formal Mathematics. Short presentation at TPHOLS2000. Available at: <http://helm.cs.unibo.it>
- 中国数学机械化研究中心. Available at: <http://www.mmrc.iss.ac.cn/mmrc-il.html>
- Cao Cungen. Technology Focus of 21st Century. *Computer World*, 1998, D1-D3 (in Chinese)
- Cao Cungen. Medical Knowledge Acquisition from the Electronic Encyclopedia of China. *Lecture Notes in Computer Science*, 2001, 2101:268~271
- Cao Cungen, et al. Progress in the Development of National Knowledge Infrastructure. *Journal of Computer Science & Technology*, 2002, 17(5): 1~16
- Si Jinxin, Cao Cungen, et al. An Environment for Multi-Domain Ontology Development and Knowledge Acquisition. EDCIS 2002, Beijing, LNCS2480, 2002. 104~116
- Zeng Qingtian, Cao Cungen. Ontology-base Mathematical Knowledge Multipurpose Representation, Acquisition, Analysis and Management; [Technical Report 2002]. Institute of Computing Technology, Chinese Academy of Sciences, 2002
- 曾庆田, 曹存根, 眭跃飞, 等. 基于本体的数学知识获取与知识继承机制研究. *微电子学与计算机*, 2003, 20(9):19~27
- 陈传璋, 朱学炎, 金福临, 欧阳光中. 数学分析(第二版). 高等教育出版社, 2003
- 数学百科全书. 中国大百科全书出版社, 1998

(上接第 123 页)