

基于 Rough 集的模式语义数据分析

Modal Semantic Data Analysis Based on Rough Set Theory

商琳 陈世福

(南京大学计算机软件新技术国家重点实验室 南京210093)

Abstract This paper proposes the equivalence between upper approximation, lower approximation of rough sets, modal operators and relational operations. Based on the relational operations, the algorithms of modal operations and rough approximate operations are presented, and along with the studies of some instances of modal logic, the precise definitions with respect to peculiar relations are shown by means of rough sets. The approximate definitions are also described for the general accessibility relations.

Keywords Rough sets, Modal logic, Accessibility relation, Upper approximation, Lower approximation

1. 引言

知识的表示与管理一直是当今人工智能领域的研究热点课题,而逻辑表示是人工智能中使用最早的表示模式之一^[1],特别是模态逻辑的研究越来越得到人们的重视。模态逻辑中加入了“必然”和“可能”两种模态算子,从而对命题演算给出了恰当的非真值系统,使其从纯粹逻辑领域走向应用领域^[2]。虽然用模态表示的知识模式具有很好的优点,如证明过程的推理规则有效,具有很好的理解形式语义的能力,但是其最大的、难以克服的缺点是表示和处理的分离,如何用可操作的方法来处理逻辑表示的数据一直是研究者探讨的热点。目前关于 Rough 集和模态逻辑的工作主要是通过 Rough 集的近似计算和模态算子的相似性来对 Rough 理论进行研究^[3,4]。本文通过对 Rough 集理论的研究,证明了其近似算子、模态算子与关系运算的等价性,并从关系运算的角度给出了模态算子与近似集计算的步骤。特别地,我们通过对模态逻辑中特定实例的研究,得出其数据表示中知识的精确定义,并在一般情况下根据可达关系(accessibility relation),运用 Rough 集概念对知识进行近似描述。

2. Rough 集

Rough 集理论是由 Z. Pawlak 于1982年提出的一种刻画不完整性和不确定性的数学理论^[5]。Rough 集理论中所有的概念和运算都是通过代数学的等价关系和集合运算来定义的,我们称之为 Rough 集理论的代数表示。

2.1 Rough 集基本概念

在 Rough 集理论中,知识被认为是一种对对象进行分类的能力。假定给定我们感兴趣的对象的论域 U ,对于任何子集 $X \subseteq U$,称为 U 中的一个概念或范畴,空集也认为是一个概念,并且 U 中的任何概念族称为关于 U 的抽象知识,简称知识^[6]。

定义1 设 $U \neq \Phi$ 是论域, $C = \{X_1, X_2, \dots, X_n\}$, 使得 $X_i \subseteq U, X_i \neq \Phi, X_i \cap X_j = \Phi$, 对 $i \neq j, i, j = 1, 2, \dots, n$, 且 $\bigcup X_i = U$, 则称 C 为一个划分。 X_i 称为划分 C 的一个等价类。 U 上的一族划分,称为关于 U 的一个知识库。

U 上的一个划分与其上的一个等价关系是等价的。每一

个等价关系描述的是论域 U 上的某一个属性,即属性也可看作是一个等价关系。

定义2 设 $U \neq \Phi$ 是论域, R 是 U 上的等价关系(自反、对称和传递)或称不分明关系,则 U/R 表示 R 在 U 上导出的所有等价类; $[x]_R$ 表示包含元素 x 的 R 的等价类。

对于子集 $X, Y \subseteq U$,若根据关系 R, X 和 Y 由属性 R 不可分辨时,我们用 $[X]_R$ 表示或 $[Y]_R$ 表示,它代表子集 Y 和子集 X 都属于 R 中的一个范畴。

2.2 Rough 集的上、下近似及边界区

设 $U \neq \Phi$ 是论域, R 是 U 上的不分明关系(等价关系),它将 U 分解成若干互不相交的子集 $E_i, i = 1, \dots, n$, 称其为 R -基本集,对所有那些被包含在 U 上的子集 X 中的 R -基本集的并,定义为 X 的下近似集,简称 Rough 下近似,而对所有那些与 X 的交非空的 R -基本集的并,被称作 X 的上近似集,简称 Rough 上近似,形式化定义如下:

定义3 集合 X 关于 R 的下近似、集合 X 关于 R 的上近似和集合 X 的边界区分别是:

$$R_-(X) = \{x \in U \mid [x]_R \subseteq X\}$$

$$R^-(X) = \{x \in U \mid [x]_R \cap X \neq \Phi\}$$

$$Bn(X) = R^-(X) - R_-(X)$$

$R_-(X)$ 实际上是那些根据已有知识判断肯定属于 X 的对象所组成的最大集合, $R^-(X)$ 是那些可能属于 X 的对象所组成的最小集合。

$Bn(X)$ 为集合 X 的上、下近似之差。如 $Bn(X) = \Phi$, 则称 X 关于 R 是清晰的,或精确的;反之,如 $Bn(X) \neq \Phi$, 则称集合 X 是关于 R 的 Rough 集,即粗糙的。

3. 模态逻辑和模态算子

模态语言的表达式由以下集合的符号组成^[7]:

VARPROP:命题变量的集合

CONAC:可达关系(accessibility relation)

常量的集合

OPAC:可达关系运算的集合

OPPROP:命题演算的集合, $OPPROP =$

$$\{\neg, \cup, \cap, \rightarrow, \leftrightarrow, [], \langle \rangle\}$$

其中 $[]$ 和 $\langle \rangle$ 运算,我们称之为“必然”和“可能”模态算子。

关系表达式 EAC 是由满足以下条件的最小集合组成:

$\text{CONAC} \subseteq \text{EAC}$;

对任意的 $x, y \in \text{EAC}, U \in \text{OPAC}, xUy \in \text{EAC}$. 模态公式 FORMOD 是由满足以下条件的最小集合组成:

$\text{VARPROP} \subseteq \text{FORMOD}$;

$F, G \in \text{FORMOD} \Rightarrow F \cap G, F \cup G, F \rightarrow G, F \leftrightarrow G \in \text{FORMOD}$;

$F \in \text{FORMOD}, R \in \text{EAC} \Rightarrow$

$[R]F, \langle R \rangle F \in \text{FORMOD}$.

经典命题模态逻辑的标准模型为:

$M = \{ST, \{r_R\}_{R \in \text{CONAC}}, m\}$

其中 (1) ST 是状态的非空集合

(2) r_R 是集合 ST 上的二元关系, 称为可达关系

(3) m 是一个满足如下条件的映射:

$m(R) \subseteq r_R$, 对任意的 $R \in \text{CONAC}$;

$m(P) \subseteq \text{ST}$, 对任意的 $p \in \text{VARPROP}$;

$m(uR) = um(R)$, 对任意的一元关系运算符 $u \in \text{OPAC}$;

$m(RbS) = m(R) \circ m(S)$, 对任意的二元关系运算符 $b \in \text{OPAC}$;

特别地, 我们令 $m(r_R) = r_R$, 对任意的 $R \in \text{CONAC}$ 为等价关系; $m(P) = P$, 对任意的 $P \subseteq \text{ST}$.

根据必然 $[\]$ 和可能算子 $\langle \rangle$, 定义其作用在集合上的运算如下:

定义4

$[R]F = \{x \in \text{ST} \mid [x]_R \subseteq F\}$

$\langle R \rangle F = \{x \in \text{ST} \mid [x]_R \cap F \neq \emptyset\}$

其中 $[x]_R = \{y \in \text{ST} \mid xRy\}$

4. 模态算子与关系运算的等价性

4.1 集合与关系运算的定义

基于复合关系的运算早已由 Tarski 提出过, 后来 Hoare 和 He Jifeng 又增加了 Weakest Prespecification ($\backslash$) Weakest Postspecification ($/$) 运算^[8]. 在文[9]中, 将关系与关系之间的运算引申到集合与关系之间, 得到以下定义.

定义5 A, B 是两个集合, 二元关系 $f \subseteq A \times B, D \subseteq A, C \subseteq B$. 集合 D 和关系 f 的复合运算 $D \circ f$ 定义为: $D \circ f = \{x \mid \exists t \in D \text{ such that } (t, x) \in f\} \subseteq B$; 关系 f 和集合 D 的复合运算 $f \cdot D$ 定义为: $f \cdot D = \{x \mid \exists t \in C \text{ such that } (x, t) \in f\} \subseteq A$.

定义6 A, B 是两个集合, 二元关系 $f \subseteq A \times B$, 子集 $E \subseteq A$. 集合 E 和关系 f 的右除运算 E/f 定义为最大的子集 $C \subseteq B$, 使得 $f \cdot C \subseteq E$.

定义7 A, B 是两个集合, 二元关系 $f \subseteq A \times B$, 子集 $F \subseteq B$. 集合 F 和关系 f 的左除运算 $f \backslash F$ 定义为最大的子集 $D \subseteq A$, 使得 $D \circ f \subseteq F$.

4.2 模态算子与关系运算的等价性

在模态逻辑的证明系统中, 一般采用的方法是先将模态逻辑转换成非模态的语言, 然后再运用非模态定理证明方法. 在文[10]中, 提出将模态语言的公式转换成相应的关系运算公式, 定义转换函数 $t: \text{EAC} \cup \text{FORMOD} \rightarrow \text{EREL}$, 其中 $\text{EAC} \cup \text{FORMOD}$ 表示模态语言的关系表达式和模态公式的并集, EREL 为关系表达式集合, 于是有以下转换关系:

$t(R \cup S) = t(R) \cup t(S)$

$t(R \cap S) = t(R) \cap t(S)$

$t(R \circ S) = t(R) \cdot t(S)$

$t(\neg F) = \neg t(F)$

$t(F \rightarrow G) = t(\neg F \cup G)$

$t([R]F) = t(R) \backslash t(F)$

$t(\langle R \rangle F) = t(R) \cdot t(F)$

其中 $\cup, \cap, \cdot, \neg$ 等运算为关系之间的运算, 定义见文[8].

5. 模态语义数据分析中的近似概念研究

5.1 相关结论

由 Rough 集近似概念的定义(定义3)和模态算子的定义(定义4)可以得出:

结论1 给定任一子集 A 和等价关系 R , 关于 A 的上近似集 $R^+(A) = \langle R \rangle A$, 关于 A 的下近似集 $R_-(A) = [R]A$.

类似的结论在文[3, 11]中都提到过, 这些结论是通过集合表示, 建立近似运算和模态算子之间的关系. 在相应转换下, 由模态算子与关系运算的等价性(4.2节), 我们得出:

$[R]A = R \backslash A$

$\langle R \rangle A = R \cdot A$

根据结论1, 我们可以得出:

结论2

$R^+(A) = \langle R \rangle A = R \cdot A$

$R_-(A) = [R]A = R \backslash A$

本结论说明可以通过 Rough 集的近似概念来进行模态语义数据分析. 我们知道在命题模态逻辑中, 当可达关系 r_R 变化时, 形成了不同的模型实例^[8]. Flarinas del cerro 和 Or-lowska 提出的数据分析逻辑(DAL)^[12], 对应的可达关系是等价关系, 其模态算子的概念与 Rough 集近似运算是一致的.

下面我们研究当可达关系是一些特定的关系时的情况.

5.2 模态逻辑不同实例研究

相关定义如下^[13]:

定义8 G 是论域, 子集 $B \subseteq G, f \subseteq G \times G, B$ 称作 f 的向前封闭的集合当且仅当对任意的 $x, y \in G$ 如果 $x \in B$ 而且 $(x, y) \in f$, 则 $y \in B$.

定义9 $R \subseteq G \times G$ 是树关系, 当且仅当 R 是自反, 反对称和传递的, 并且满足条件: 如果 $(x, z) \in R$, 并且 $(y, z) \in R$, 则 $x = y$.

文[13]中, 给出了事件结构逻辑(ESL), 非确定信息逻辑(NIL), 以及时态信息逻辑(TIL)中相对于可达关系的子集不变性. 从 Rough 集理论出发, 这也就是对知识进行精确划分的情况.

结合5.1节, 我们可以得出以下结论

结论3 G 是论域, 子集 $X \subseteq G, R \subseteq G \times G$ 是树(看作关系). 如果 X 是 R 的向前封闭集合或向前封闭集合的并集, 则 $[R]X = R \backslash X = R_-(X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1} \cdot (A) = X$.

证明: 1) 根据定义4, $[R]X = \{x \in \text{ST} \mid [x]_R \subseteq X\}$, 设 $x \in [R]X$, 则 $[x]_R \subseteq X$, 所以 $x \in X$, 得出 $[R]X \subseteq X$;

2) 设 $x \in X$, 根据已知条件, X 是向前封闭集合的并集, 则由定义8可知, 如果有 $y \in G$, 使得 $(x, y) \in R$, 则 $y \in X$, 即 $[x]_R \subseteq X$, 所以 $x \in [R]X$, 得出 $X \subseteq [R]X$;

3) 由1和2可得 $[R]X = X$;

4) 其余证明见文[13].

结论4 G 是论域, 子集 $X \subseteq G, R \subseteq G \times G$ 是相容关系. 如果 X 是 R 的半等价关系或半等价关系的并集, 则 $[R]X = R \backslash$

$$X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$$

结论5 G 是论域, 子集 $X \subseteq G, R \subseteq G \times G$ 是格。如果 X 是 R 的向前封闭集合或向前封闭集合的并集, 则 $[R]X = R \setminus X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$

结论6 G 是论域, 子集 $X \subseteq G, R \subseteq G \times G$ 是线性序。如果 X 是 R 的向前封闭集合或向前封闭集合的并集, 则 $[R]X = R \setminus X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$

结论4、5、6的证明方法同结论3。

在事件结构逻辑(ESL)中, 可达关系 r_R 是树(看作关系), 根据性质结论3, 当集合 X 是 R 的向前封闭集合或向前封闭集合的并集, 则可以得到 $[R]X = R \setminus X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$ 。

在非确定信息逻辑中(NIL), 可达关系 r_R 包括相容关系和格, 根据结论4、5, 如果集合 X 是半等价关系或半等价关系的并集, 则可以得到 $[R]X = R \setminus X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$ 。

在时态信息逻辑中(TIL), 可达关系 r_R 是线性序, 根据结论6, 如果集合 X 是 R 的向前封闭集合或向前封闭集合的并集, 则可以得到 $[R]X = R \setminus X = R - (X) = X, \langle R^{-1} \rangle X = R^{-1} \cdot X = R^{-1-}(A) = X$

5.3 近似计算算法

根据结论2, 我们可以得出基于关系 R 的上近似 $R^-(A)$ 和下近似集 $R_-(A)$ 可以通过 $R \cdot A$ 和 $R \setminus A$ 运算得到。

算法1

基于关系 R 的上近似 $R^-(A)$;
设 $E = \{x_1, x_2, \dots, x_n\}, R \subseteq E \times E$,
 $A = \{y_1, y_2, \dots, y_m\} \subseteq E$

1. 设 $R^-(A)$ 结果初始为 B
2. $B = \emptyset$;
3. for $i = 1$ to n for $j = 1$ to m
if $(x_i, y_j) \in R$ then
 $B = B \cup \{x_i\}$;
4. return B

算法2

基于关系 R 的下近似 $R_-(A)$;
设 $E = \{x_1, x_2, \dots, x_n\}, R \subseteq E \times E$,
 $A = \{y_1, y_2, \dots, y_m\} \subseteq E$

1. 设 $R_-(A)$ 结果初始为 B
2. $B = \emptyset$;
3. for $i = 1$ to n
 $X_i = \emptyset$
for $j = 1$ to m
if $(x_i, y_j) \in R$ then
 $X_i = X_i \cup \{y_j\}$;
if $X_i \subseteq A$ then
 $B = B \cup \{x_i\}$;
4. return B

通过算法1和2我们可以对模态算子进行计算, 并运用相应的 Rough 集精度概念, 对不确定性进行描述。

在模态数据表示中, 定义由关系 R 分类的近似精度为:

$$d_R(X) = |[R]A| / |\langle R \rangle A|$$

其中 $|B|$, 表示集合 B 元素的个数。

根据 $d_R(X)$, 定义粗糙度 $P_R(X) = 1 - d_R(X)$ 。

同 Rough 集中对于知识的近似描述相同, X 的 R 粗糙度与其精确度相反, 精确度反映我们对于了解集合 X 的知识完全程度, 粗糙度表示的是集合 X 的知识的不完全程度。显然 $0 \leq d_R(X) \leq 1$, 于是可用 $d_R(X)$ 来定义 Rough 集^[6]。

结束语 Rough 集理论已被广泛地应用于知识分类和处理不完全信息, 其上近似和下近似计算与模态逻辑中的必然算子 $[\Box]$ 和可能算子 $\langle \Diamond \rangle$ 的语法性质相同, 本文结合关系运算, 证明其合成和右除运算与近似计算和模态算子的等价性, 并从关系运算的角度给出了模态算子与近似集计算的算法。通过对模态逻辑中特定实例的研究, 运用 Rough 集对其数据表示中特定的知识进行精确定义, 并在一般情况下根据可达关系, 对知识进行近似描述。

参考文献

- 1 陈世福, 陈兆乾, 等. 人工智能与知识工程. 南京: 南京大学出版社, 1997
- 2 [美]B. F 切莱士著, 郑文辉, 张宜生译. 模态逻辑导论. 广州: 中山大学出版社, 1989
- 3 Lin T Y, Liu Qing. First Order Rough Logic-revisited. In: Ning Zhong, Andrzej Skowron, Setsuo Ohsuga. eds. Lecture Notes in Artificial Intelligence 1711, Berlin: Springer-Verlag, 1999. 276~284
- 4 Liao Churn-Jung. An Overview of Rough Set Semantics for Modal and Quantifier Logics. International Journal of Uncertainty, Fuzziness and Knowledge-Based System, 2000, 8(1): 93~118
- 5 Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982, 11(5): 341~356
- 6 曾黄麟, 等. 粗糙集理论及其应用. 重庆: 重庆大学出版社, 1996
- 7 Orlowska E. A Logic of Indiscernibility Relations. Bulletin of the Polish Academy of Sciences, Mathematics, 1987, 33: 475~485
- 8 Hoare C A R, He Jifeng. The Weakest Prespecification I. Fundamenta Informaticae, 1986, 9(1): 51~84; The weakest prespecification II. Fundamenta Informaticae, 1986, 9: 217~252
- 9 刘莉, 李永礼, 商琳. The Generation on Operation of Set and Pansystems Relation. In: Advances in Sysytems Science and Applications; Special Issue, 1996. 138~141
- 10 Fari~nas del Cerro Luis. A Simple Deduction Method for Modal Logic. Information Processing Letters, 1982, 14(2): 49~51
- 11 Yao Y Y. Two views of the theory of rough sets in finite universes. International Journal of Approximate Reasoning, 1996, 15: 291~317
- 12 Fari~nas del Cerro Luis, Orlowska E. DAL-A logic for data analysis. Theoretical Computer Science, 1985, 36: 251~264
- 13 李永礼, 商琳, 刘莉. Fixed Pansystems Theorems of Sub-binary Relation. In: Advances in Sysytems Science and Applications; Special Issue, 1996. 145~149

(上接第65页)

- 3 Alexander D S, et al. A Secure Active Network Environment Architecture; Realization in Switch-Ware. IEEE Network, 1998
- 4 Arbaugh W A, et al. A secure and reliable bootstrap architecture. In: IEEE Security and Privacy Conference, May 1997. 65~71
- 5 Alexander D S. ALIEN: A Generalized Computing Model of Ac-

tive Networks. PhD thesis, University of Pennsylvania, 1998

- 6 Tennenhouse D L, et al. A Survey of Active Network Research. IEEE Communications Magazine, January 1997. 80~86
- 7 van der Merwe J E, Leslie I M. Switchlets and dynamic virtual ATM networks. In: Proc. of the Fifth IFIP/IEEE Intl. Symp. on Integrated Network Management, San Diego, CA., May 1997