

模糊概念格在智能疾病诊断中的应用

石津妮

(大连海事大学 信息科学技术学院, 辽宁省大连市 中国 116026)

摘 要 本文提出了一种新的基于模糊概念格的智能疾病诊断方法。可以提取每种疾病的症状和相应程度(例如, 频率、严重程度和持续时间)以形成模糊概念格。需要构建待诊断症状及其程度的模糊概念格, 以匹配可能疾病的模糊概念格。可以计算上述两种类型的概念格之间的相似性, 并用于帮助进行有效的诊断。自然, 相似性最大的疾病是发现非智力因素诊断。在此框架下, 将探索更有效的模糊概念格构造方法和更新算法, 这被认为是非常复杂的。

关键词 模糊概念格; 模糊形式背景; 智能疾病诊断; 相似性

中图法分类号 TP311

文献标识码 A

Using Fuzzy Concept Lattice for Intelligent Disease Diagnosis

SHI Jin Ni

(School of Information Science and Technology, Dalian maritime university, Liaoning Dalian 116026 China)

Abstract This paper proposes a novel intelligent disease diagnosis method based on fuzzy concept lattice. Symptoms and the corresponding extents (e.g., frequency, severity, and duration) of each disease can be extracted to form a fuzzy concept lattice. The fuzzy concept lattice of the symptoms and their extents to be diagnosed needs to be constructed to match the fuzzy concept lattice of possible diseases. The similarity between the above two types of concept lattices can be calculated and used to aid for effective diagnosis. Naturally, the disease with the largest similarity is the finding of intelligent diagnosis. In the future, more efficient fuzzy concept lattice construction method and update algorithm will be explored, which are presumed to be very complicated.

Key words Fuzzy concept lattice; fuzzy formal context; intelligent disease diagnosis; similarity

1 介绍

医疗保健部门目前面临许多挑战, 如慢性病、疾病预防、昂贵的医疗费用等。为了应对各种挑战, 医疗机构必须以医疗数据为资源, 分析数据, 获取有价值的信息, 从而提高治疗效果, 降低成本。数据挖掘和数据分析可以帮助医生做出准确的临床诊断、有效的疾病爆发预测、有用的个性化治疗等。在物联网时代, 数据呈指数级增长。医学数据已经成为不可或缺的重要资源。传统的手工数据分析方法已经不能满足要求。高效的基于计算机的智能数据分析方法已经变得不可或缺。

医疗保健部门的数据来源包括医院信息系统(HIS)、各种设备收集的用户身体体征数据、社交网络中隐含的健康信息等。医疗保健领域的数据处理技术包括数据融合、分类、聚类、关联规则挖掘、机器学习、模拟、可视化等^{[1][2]}。其中, 关联规则挖掘的目的是发现变量之间的关系; 聚类分析的目的是根据相似性数据将数据分成几个小数据集; 可视化的目标是通过图像、动画和其他直观的方式呈现数据。

“概念”被定义为由外延和内涵组成的基本单位。形式概念分析(FCA)是由德国著名科学家威 Wille 在 1982 年提出的^[3]。Wille 提出了用二元关系构造相应概念格的基本思想, 为形式概念

分析 (FCA) 的发展奠定了理论基础。

FCA 强调人的认知, 提供了一种有效的数据分析和知识表示方法。概念格作为 FCA 的关键数据结构, 是一种外部认知手段。概念格由一些形式概念节点组成, 涉及外延和内涵部分。概念的内涵是外延中所有对象的共同属性的集合, 概念的外延是可以由内涵确定的最大的对象集合^[4]。概念格是对客观世界的高度简化的描述, 揭示概念之间的关系, 具有良好的数学性质。概念格可以被图形化地表示为标记线图, 称为 Hasse 图, 这有助于数据可视化。

2 相关工作

相关工作包括形式背景、概念格、Hasse 图、模糊概念格、智能疾病诊断等。

2.1 形式背景

形式上下文 $K(U, A, I)$ 由对象集 U 、属性集 A 和它们的二元关系 I 组成, 即 $I \subseteq U \times A$ 。如果 $(x, a) \in I$, 则 xIa 表示对象 x 拥有属性 a , 形式上下文也可以表示为二维数据表, 其中行表示对象, 列表示属性。当给定

$$X^* = \{a \mid a \in A, \forall x \in X, xIa\}, X \subseteq U$$

$$B^* = \{x \mid x \in U, \forall a \in B, xIa\}, B \subseteq A$$

如果 $\exists X^* = B$ 和 $B^* = X$, 那么 (X, B) 被视为一个形式概念^[5]。 X 代表概念 (X, B) 的外延, 表示这个概念中所有对象的集合。 B 代表概念 (X, B) 的内涵, 它表示这个概念中所有对象的共同属性的集合。例如, $C((2, 3, 4), \{a, c\})$ 表明概念 C 涉及三个对象 2, 3 和 4。这三个对象的公共属性集是 $\{a, c\}$ 。

2.2 概念格

概念格理论是知识表示和发现的有效工具。主要用于认知计算、机器学习、模式识别、专家系统、决策分析、web 搜索等领域。概念格本质上描述了对象和属性之间的关联。

概念格的节点之间可以建立一种偏序关系。如果 $C_1(X_1, B_1), C_2(X_2, B_2)$ 满足 $X_1 \subseteq X_2$ 或 $B_2 \subseteq B_1$, 那么 $C_1(X_1, B_1)$ 称为子概念, $C_2(X_2, B_2)$ 称为父概念, 可以表示为 $C_1(X_1, B_1) \leq C_2(X_2, B_2)$ 。关系 \leq 称为概念的偏序。对于形式语境中的所有概念, 偏序集产生概念格 $L(U, A, I)$, 对于特定的形式语境 $K(U, A, I)$ 是唯一的。

2.3 HASSE 图

Hasse 图是概念格的可视化表示, 有效地展示了概念格中所有概念的偏序关系。Hasse 图中的每个节点代表一个概念, 节点之间的连接表示泛化-专门化关系。Hasse 图中节点的位置根据它们的顺序从下到上排列。对于部分有序集 (S, \leq) , 如果 y 包含 x (即 $x < y$ 并且 $\neg \exists z(x < z < y)$), 则添加从 x 到 y 的线。

2.4 模糊概念格

人的思维中有很多模糊的概念, 比如年轻、大的、温暖等。在现实世界中, 有很多模糊现象。模糊理论和 FCA 结合起来可以处理不确定信息。1994 年, A. Burusco 和 R. Fuentes 提出了模糊概念格理论^[6], 它是概念格和扎德模糊集理论的结合。

2.5 智能疾病诊断

智能医疗诊断是专家系统试图实现的目标。研究人员已经做了一些有意义的工作。Miller 等人^[7]开发了著名的 Internist-I 内科医疗计算机辅助诊断系统。Barnett 等人^[8]在哈佛医学院, 开发了解释软件, 包括大约 2200 种疾病和 5000 种症状。Umbaugh 等人^[9]开发了一个皮肤癌辅助诊断系统。Provan 和 Clarke^[10]构建了一个辅助诊断慢性腹痛的决策支持系统。Keith 和 Greene^[11]开发了智能 CTG 分析系统, 它是数字算法和人工神经网络方法的结合。Birndorf 等人^[12]开发了一个贫血诊断专家系统。Wells 等人^[13]开发了一个基于知识的乳腺癌治疗计划系统。近年来, 一些研究者在疾病智能诊断领域开展了广泛的研究。专家系统和概念格领域也取得了很好的成果。概念格在医学领域的应用值得进一步研究。

3 基于疾病的模糊概念格

根据临床医学诊断的一般过程, 模糊概念格可以用于医学信息检索和疾病诊断。

3.1 疾病的模糊形式背景

每种疾病的症状和对应程度可以提取出来形成模糊形式上下文 $K(U, A, I)$, 其中 U 表示疾病的一组症状, A 表示症状对应程度的一组, I 为隶属函数, $U \times A \rightarrow [0, 1]$ 。症状的程度主要包括发生的频率、患者遭受的严重程度以及持续的时间。

在模糊形式背景 $K(U, A, I)$ 中, 对于每个 $a \in A$, 给出了 $\phi_a (0 \leq \phi_a \leq 1)$ 。假设 $\theta K = (U, A, I_\theta)$, I_θ 被定义为

$$I_{\theta}(u, a) = \begin{cases} I(u, a), & I(u, a) \geq \theta\varphi_{ai} \\ 0, & I(u, a) < \theta\varphi_{ai} \end{cases}$$

模糊概念 $C_i = (U_i, A_i)$ ($i = 1, \dots, n$, n 是模糊概念的个数) $U_i \subseteq U, A_i \subseteq A$, 满足:

$$\forall U_i \subseteq U, g(U_i) = \{a \mid \forall u \in U_i, I(u, a) \geq \theta\varphi_{ai}\}$$
$$\forall A_i \subseteq A, g(A_i) = \{u \mid \forall a \in A_i, I(u, a) \geq \theta\varphi_{ai}\}$$

表 1 显示了普通感冒的模糊形式背景。假设 φ_{ai} 是每一列的平均值, 并且 $\theta = 0.8$ 。 $\theta \varphi_{ai}$ 在每列的值保持原始值, 否则取值为零。然后将对应属性值大于 $\theta \varphi_{ai}$ 的对象合并, 得到感冒的模糊形式概念, 如表 2 所示。

表 1 普通感冒的模糊形式背景

Symptom	Frequency			Severity		Duration	
	a	b	c	d	e	f	g
1	0.1	0.1	0.8	0.2	0.8	0.2	0.8
2	0.1	0.7	0.2	0.1	0.9	0.2	0.8
3	0.8	0.1	0.1	0.8	0.2	1.0	0.0
4	0.7	0.2	0.1	0.7	0.3	0.8	0.2
φ_{ai}	0.43	0.28	0.3	0.45	0.55	0.55	0.45
$\theta\varphi_{ai}$	0.34	0.22	0.24	0.36	0.44	0.44	0.36

表 2 普通感冒的模糊形式概念

Fuzzy Concept	Fuzzy Concept	Fuzzy Concept	Fuzzy Concept
$(\{1\}, \{c, e, g\})$	$(\{2\}, \{b, e, g\})$	$(\{3, 4\}, \{a, d, f\})$	$(\{1, 2\}, \{e, g\})$

3.2 基于疾病的模糊概念格

疾病的症状可以视为对象, 症状的程度可以视为属性。对象-属性对(即症状-程度)可以定义为概念。然后得到模糊概念格。图 1 表示出了对应于表 1 的模糊概念格。

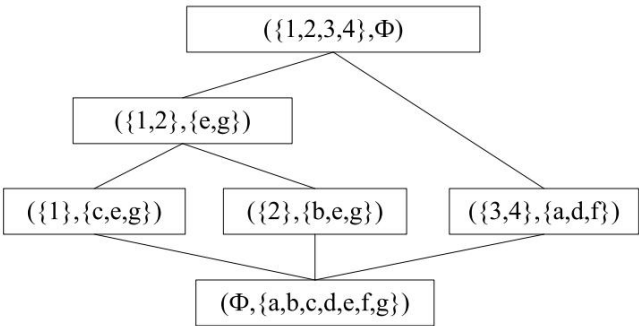


图 1 感冒的模糊概念格

4 基于模糊概念格的智能疾病诊断

需要构建待诊断症状及其程度的模糊概念格, 以匹配可能疾病的模糊概念格。那么就可以推测出病人最有可能被诊断的疾病。

4.1 两个概念节点之间的相似性

对于一个或多个形式上下文中的两个概念 $C_1 = (U_1, A_1)$ 和 $C_2 = (U_2, A_2)$, 设 r 是 U_1 和 U_2 的基数的最大值, m 是 A_1 和 A_2 的基数的最大值, 即

$r = \max(|U_1|, |U_2|); m = \max(|A_1|, |A_2|)$

那么两个概念 $C_1 = (U_1, A_1)$ 和 $C_2 = (U_2, A_2)$ 之间的相似度可以计算如下:

$$Sim(C_1, C_2) = \begin{cases} \frac{|U_1 \cap U_2|}{r} \times p + \frac{|U_1 \cap U_2|}{r} \times \frac{|A_1 \cap A_2|}{m} \\ \times (1 - 2p) + \frac{|A_1 \cap A_2|}{m} \times p, & U_1 \cap U_2 \neq \emptyset \\ 0, & U_1 \cap U_2 = \emptyset \end{cases}$$

其中 p 为加权因子 ($0 \leq p \leq 0.5$)。

4.2 概念节点和概念格的相似性

可以计算一个概念格 L_1 的概念节点 C_i 和另一个概念格 L_2 的每一层中的概念节点之间的相似性。然后选取层内相似度最大的概念节点作为候选节点, 依次计算候选节点的子概念节点。通过逐层递归, 可以得到一条或多条路径。概念节点 C_i 和概念格 L_2 之间的相似性定义如下^[14]:

$$Sim(C_i, L_2) = \max \left(\frac{\sum_{C_k \in R_j} sim(C_i, C_k)}{n_j} \right)$$

其中 R_j 是由候选概念组成的路径 ($j = 1, \dots, s$, s 是候选路径数), n_j 是路径 R_j 中的节点数, C_k 是路径 R_j 中的节点 ($k = 1, \dots, t$, t 是路径 R_j 中的节点数)。节点 C_i 和路径 R_j 之间的相似性是 C_i 和路径 R_j 中所有候选节点的相似性的平均值。

4.3 两个概念格之间的相似性

两个概念格的相似性被视为概念格 L_1 和概念格 L_2 中所有节点的相似性的平均值^[40]。

$$Sim(L_1, L_2) = \frac{\sum_{C_i \in L_1} Sim(C_i, L_2)}{n_1}$$

其中 n_1 是 L_1 中概念节点的数量。假设患者的症状如表 3 所示。可以得到患者症状的模糊概念格, 如图 2 所示。

表 3 病人症状的模糊形式背景

Symptom	Frequency			Severity		Duration	
	a	b	c	d	e	f	g
1	0.6	0.2	0.2	0.7	0.3	0.7	0.3
2	0.2	0.3	0.5	0.4	0.6	0.4	0.6
4	0.2	0.5	0.3	0.3	0.7	0.8	0.2
φ_{ai}	0.33	0.33	0.33	0.47	0.53	0.63	0.37
$\theta\varphi_{ai}$	0.26	0.26	0.26	0.38	0.42	0.50	0.30

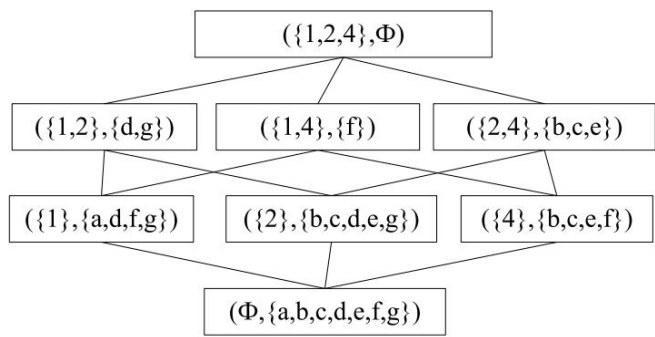


图 2 病人症状的模糊概念格

假设加权因子 p 设置为 $1/4$, 那么患者症状和感冒的模糊概念格的相似度计算结果是

$$\text{Sim1} = (0.229 + 0.513 + 0.25 + 0.459 + 0.292 + 0.375) / 6 = 0.353.$$

4.4 基于模糊概念格的智能疾病诊断

给出了一个基于模糊概念格的智能疾病诊断实例。可能的疾病包括普通感冒、病毒性咽炎和肺炎。病毒性咽炎的模糊形式背景见表 4。肺炎的模糊形式背景见表 5。病毒性咽炎的模糊概念格如图 3 所示, 肺炎的模糊概念格如图 4 所示。

患者症状的模糊概念格需要匹配可能疾病的模糊概念格, 并计算相应的相似度。按照相同的计算步骤, 计算患者症状与病毒性咽炎的模糊概念格的相似度

$$\text{Sim2} = (0.323 + 0.588 + 0.172 + 0.35 + 0.368 + 0.309) / 6 = 0.352.$$
 然后计算患者症状和肺炎的模糊概念格的相似度

$$\text{Sim3} = (0.623 + 0 + 0.459 + 0.257 + 0.282 + 0.219) / 6 = 0.307.$$
 相似性最大的疾病是发现非智力因素诊断。因此病人最有可能患普通感冒。

表 4 病毒性咽炎的模糊形式背景

Symptom	Frequency			Severity		Duration	
	a	b	c	d	e	f	g
1	0.8	0.1	0.1	0.8	0.2	0.8	0.2
2	0.3	0.5	0.2	0.7	0.3	0.7	0.3
3	0.0	0.1	0.9	0.1	0.9	0.1	0.9
4	0.8	0.1	0.1	0.9	0.1	0.8	0.2
5	0.9	0.1	0.0	0.9	0.1	0.9	0.1
φ_{ai}	0.56	0.18	0.26	0.68	0.32	0.66	0.34
$\theta\varphi_{ai}$	0.45	0.14	0.21	0.54	0.26	0.53	0.27

表 5 肺炎的模糊形式背景

Symptom	Frequency			Severity		Duration	
	a	b	c	d	e	f	g
1	0.9	0.1	0.0	0.9	0.1	0.8	0.2
4	0.9	0.1	0.0	0.9	0.1	0.9	0.1
5	0.8	0.1	0.1	0.8	0.2	0.9	0.1
φ_{ai}	0.87	0.1	0.03	0.87	0.13	0.87	0.13
$\theta\varphi_{ai}$	0.70	0.08	0.02	0.70	0.104	0.70	0.104

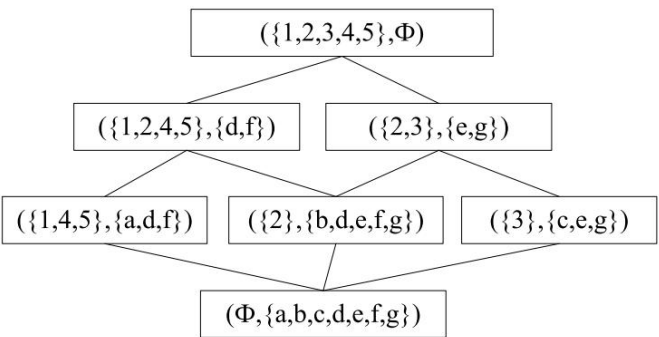


图 3 病毒性咽炎的模糊概念格

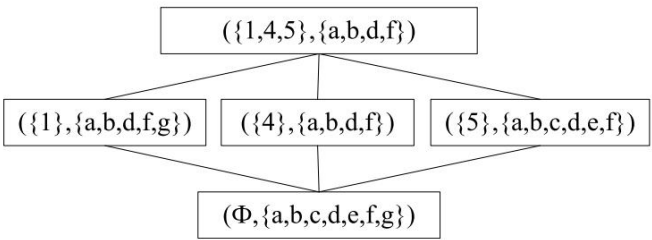


图 4 肺炎的模糊概念格

5 总结

本文提出了一种基于模糊概念格的智能疾病诊断方法。根据临床医学诊断的一般过程,模糊概念格可以用于医学信息检索和疾病诊断。可以提取每种疾病的症状及其程度,以形成模糊的形式背景。疾病的症状可以作为对象,症状的程度可以作为属性。需要构建待诊断症状及其程度的模糊概念格,以匹配可能疾病的模糊概念格。那么就可以推测出被诊断患者最有可能的疾病。可以计算患者症状和可能疾病的模糊概念格的相似性。相似度最大的疾病是智能诊断的发现。

在将来,我们将在计算模糊概念格相似度的基础上,考虑每种疾病症状的权重值,进一步探索更加高效的模糊概念格构造方法和更新算法,并进一步研究基于概念格的智能疾病诊断系统。

参考文献

- [1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.
- [2] F. Chen, P. Deng, J. Wan, D. Zhang, A. V. Vasilakos, and X. Rong, "Data mining for the Internet of things: Literature review and challenges," *Int. J. Distrib. Sensor Netw.*, vol. 11, Jan. 2015, Art. no. 12, doi: 10.1155/2015/431047.
- [3] R. Wille, "Restructuring lattice theory: An approach based on hierarchies of concept, ordered sets," *Rival, I.* (ed.), pp. 445–470, 1982.
- [4] Q. Yang and M. Zhao, "Progress in concept lattice research," *Comput. Eng. Design*, vol. 29, no. 20, pp. 5293–5296, 2008.
- [5] R. Bräijggemann and G. P. Patil, "Formal concept analysis," in *Ranking Prioritization for Multi-indicator System*. New York, NY, USA: Springer, 2011, pp. 117–133.
- [6] A. B. Juandeaburre and R. Fuentes-González, "The study of the L-fuzzy concept lattice," *Mathware Soft Comput.*, vol. 1, no. 3, pp. 209–218, 1994.
- [7] R. A. Miller, P. H. Jr., and J. D. Myers, *Internist-I, an Experimental Computer-Based Diagnostic Consultant for General Internal Medicine*. New York, NY, USA: Springer, 1985.
- [8] G. O. Barnett et al., "DXPLAIN—demonstration and discussion of a diagnostic clinical decision support system," in *Proc. Symp. Comput. Appl.*, 1991, p. 878.
- [9] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Applying artificial intelligence to the identification of variegated coloring in skin tumors," *IEEE Eng. Med. Biol. Mag.*, vol. 10, no. 4, pp. 57–62, Dec. 1991.
- [10] G. M. Provan and J. R. Clarke, "Dynamic network construction and updating techniques for the diagnosis of acute abdominal pain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 3, pp. 299–307, Mar. 1993.
- [11] R. D. F. Keith and K. R. Greene, "4 Development, evaluation and validation of an intelligent system for the management of labour," *Baillière's Clinical Obstetrics Gynaecol.*, vol. 8, no. 3, pp. 583–605, 1994.
- [12] N. I. Birndorf, J. O. Pentecost, J. R. Coakley, and K. A. Spackman, "An expert system to diagnose anemia and report results directly on hematology forms," *Comput. Biomed. Res.*, vol. 29, no. 1, pp. 16–26, 1996.
- [13] D. M. Wells, D. Walrath, and P. S. Craighead, "Improvement in tangential breast planning efficiency using a knowledge-based expert system," *Med. Dosimetry*, vol. 25, no. 3, pp. 133–138, 2000.
- [14] X. Li, *WordNet Based the Semantic Matching of Concept Lattice on the Search Engine*. Chengdu, China: Xihua Univ. 2008.
- [15] X. Yang and Z. Xu, "Hesitant Fuzzy Concept Lattice and its Application," in *IEEE Access*, vol. 8, pp. 59774–59786, 2020, doi: 10.1109/ACCESS.2020.2982892.