ICIC Express Letters
Volume **11**, Number **2**, February **2017**
ICIC International ©2017 ISSN 1881-803X
pp. **325–331**

METHOD OF MINING ASSOCIATION RULES BASED ON INTERVAL CONCEPT LATTICE AND ITS APPLICATION IN MEDICAL DIAGNOSIS

Mingxia Li[1,2], Baoxiang Liu[1,2,*], Chunying Zhang[1,2,*] and Liya Wang[1,2]

[1]College of Science
North China University of Science and Technology
[2]Key Laboratory of Data Science and Application of Hebei Province
No. 46, Xinhua Road, Tangshan 063009, P. R. China
nicemingxia@sina.cn; wang_liya@126.com; *Corresponding authors: { liubx5888; Hblg_zcy }@126.com

Received July 2016; accepted October 2016

Abstract. *Traditional medical diagnosis method is based on the pathological analysis to judge the disease. With the era of big data and the development of Internet, the research methods of medical diagnosis based on data have been emerging. Especially for the disease which is hard to conduct pathological analysis and judgment, patients are generally diagnosed by processing and analyzing the characteristics of the disease. Therefore, mining disease characteristics and analyzing association rules between them are of great significance. Firstly, the association rules extraction algorithm based on interval concept lattice is introduced in the paper; by collecting lots of patients' characteristic data of pneumonia disease, features are extracted according to the decisional form of medical diagnosis, then interval concept lattice is built and further association rules are mined with high support and confidence; finally, medical diagnosis rule base is constructed. Compared with the method of mining association rules based on classical concept lattice, we find that this method has more advantages.*

Keywords: Medical diagnosis, Interval concept lattice, Feature extraction, Association rules mining

1. **Introduction.** Association rules mining, one of the most active research methods in data mining, was firstly proposed by Agrawal et al. in 1993, mainly for basket analysis problems [1]. Its purpose is to find trade association rules between items in the database. The association rules define interesting association or correlation between item-sets from transaction database, and two thresholds, the support and confidence, are the two important concepts to describe association rules. Support reflects the importance of association rules in database, and confidence measures the credible degree of association rules. The rules themselves are described by the relationship between intension sets in mining association rules, and concept node is the unity of the intension and extension. Parent-child relationships between nodes embody the generalized and specialized relations of concepts, and each node is really the most item-set. Therefore, concept lattice is suitable as the data structure of extracting rules. [2-6] provide many rules extraction algorithms based on classical concept lattice, which removes redundant candidate binary group based on built lattice structure and generates association rules. Using concept lattice to design algorithms of mining association rules has more advantages than traditional data mining tools both in mining classification rules and decision-making rules. Yao used rough set approximation operators to expand the definition of concept lattice and raised rough set approximation in formal concept [7]; Yang and Zhang put forward rough concept lattice [8], which uses the upper and lower approximation of rough set to describe extension in concept. The method based on rough concept lattice [9] can extract more uncertain information than classical concept lattice. However, these rules are of low confidence

and support. So it is necessary to raise a more efficient model of mining rules. Interval concept lattice [10] is a new kind of concept hierarchy structure based on rough concept lattice, and extension of concept in it is objects sets meeting the intension attributes within the interval $[\alpha, \beta]$ $(0 \le \alpha \le \beta \le 1)$, which can describe the practical problems of extracting rules meeting certain conditions. Association rule mining algorithm based on interval concept lattice has great practical significance for mining uncertain rules and making uncertain decisions.

Usually the doctor diagnoses the disease by many years of clinical experience, and their understanding about problems and comprehensive ability will directly affect the diagnosis system. Existing medical diagnostic system usually adopts inexact reasoning model, which requests that samples (symptoms) are independent or meeting certain constraint conditions [11], and it is hard to do. It is necessary to seek a method that can deal with fuzzy, inaccurate and incomplete data; at the same time it does not need to provide any prior knowledge of the data sets. Acquiring knowledge from vast medical diagnostic information and effectively extracting diagnosis rules of disease can assist clinical experts in diagnosis.

Applying the method of mining association rules based on interval concept lattice to medical aided diagnosis is a relatively new research subject. This paper has carried on the exploration of method of mining association rules in the application of pneumonia diagnosis, made good use of confirmed cases, and assisted clinical doctors in diagnosing. Firstly, the theories of classical concept lattice and association rules based on it are introduced. Secondly, association rules extraction algorithm based on interval concept lattice is proposed, and then the credibility of diagnosis rules derived from the lattice structure under interval parameters α and β is analyzed. Finally, a set of assistant diagnosis model is systematized.

2. Association Rules Based on Classical Concept Lattice.

2.1. Classical concept lattice.

Definition 2.1. [5] *(U, A, R) is a formal context. $U = \{x_1, x_2, \ldots, x_n\}$ is an object set. x_i $(i \le n)$ is an object. $A = \{a_1, a_2, \ldots, a_m\}$ is an attribute set and each a_j $(j \le m)$ is called an attribute. R is the binary relation between U and A, namely $R \subseteq U \times A$. If $(x, a) \in R$, call x has attribute a, namely xRa.*

Definition 2.2. [5] *Formal context (U, A, R), operators f and g are defined by:*
$\forall x \in U$, $f(x) = \{y | \forall y \in A, xRy\}$, namely f is a mapping of object x and its attribute;
$\forall y \in A$, $g(y) = \{x | \forall x \in U, xRy\}$, namely g is a mapping of attribute y and its object.

Definition 2.3. [5] *Formal context (U, A, R), because $X \subseteq U$ and $Y \subseteq A$, $f(X) = Y$ and $g(Y) = X$. Call binary group (X, Y) a formal concept, concept for short. Extension of concept is X and intension is Y.*

Definition 2.4. [5] *Using $L(U, A, R)$ to express all concepts in formal context (U, A, R): $(X_1, Y_1) \le (X_2, Y_2) \Leftrightarrow X_1 \subseteq X_2$ $(\Leftrightarrow Y_1 \supseteq Y_2)$, "$\le$" is the partial ordering relation. If all concepts in $L(U, A, R)$ meet the partial ordering relation, call $L(U, A, R)$ is the concept lattice on formal context (U, A, R). Each node C in lattice is a formal concept.*

Definition 2.5. *For node $C_1(X_1, Y_1)$, character subset Y_2 is called intension reduction of $C_1(X_1, Y_1)$, when (1) $g(Y_2) = g(Y_1) = X_1$ and (2) $\forall Y_3 \subset Y_2$, $g(Y_3) \supset g(Y_2) = X_1$.*

Among them, Condition (1) is called extension invariance with the intension reduction, namely intension of $C_1(X_1, Y_1)$ and its intension reduction have the same extension; Condition (2) is called minimum intension reduction, namely removing any attribute of intension reduction will lead to the increase of the extension.

2.2. **Association rules based on classical concept lattice.** Association rule $A \Rightarrow B$ is generated by the node binary group (C_1, C_2) $(C_1 \geq C_2)$ and $C_1 = (g(A), f(g(A)))$, $C_2 = (g(A \cup B), f(g(A \cup B)))$. From (C_1, C_2) we can directly calculate $supp(A \Rightarrow B) = |\text{Extension}(C_2)|/|U|$, $conf(A \Rightarrow B) = |\text{Extension}(C_2)|/|\text{Extension}(C_1)|$.

Definition 2.6. [5,6] *Set support θ and confidence φ, if node binary group (C_1, C_2) meets: (1)* $|\text{Extension}(C_2)| / |U| \geq \theta$*; (2)* $\text{Extension}(C_2) \subseteq \text{Extension}(C_1)$ *and* $|\text{Extension}(C_2)| / |\text{Extension}(C_1)| \geq \varphi$*, node binary group is called (θ, φ)-candidate binary group.*

Theorem 2.1. [5,6] *If $C \subset B$, $A \Rightarrow C$ can be derived from $A \Rightarrow B$. The number of generated association rules can be effectively reduced.*

3. Model of Mining Association Rules Based on Interval Concept Lattice.

3.1. Interval concept lattice.

Definition 3.1. [10] *The formal context (U, A, R) is given. $L(U, A, R)$ is a classic concept lattice based on it. (M, N, Y) is a rough concept lattice based on RL. Assume the interval $[\alpha, \beta]$, $0 \leq \alpha < \beta \leq 1$*

α-upper extension M^α: $M^\alpha = \{x | x \in M, |f(x) \cap Y|/|Y| \geq \alpha, 0 \leq \alpha \leq 1\}$

β-lower extension M^β: $M^\beta = \{x | x \in M, |f(x) \cap Y|/|Y| \geq \beta, 0 \leq \alpha \leq \beta \leq 1\}$

Among them, Y is the intension of the concept. $|Y|$ is the number of elements contained by set Y, namely cardinal number. M^α expresses the objects covered by at least $\alpha \times |Y|$ attributes from Y.

Theorem 3.1. *When $\beta = 1$, $|f(x) \cap Y|/|Y| = 1$; and $\alpha = 0$, $|f(x) \cap Y|/|Y| = 0$, interval concept lattice degenerates into classical concept lattice.*

Definition 3.2. [10] *Formal context (U, A, R), the ternary ordered pairs (M^α, M^β, Y) are called interval concept, among which, Y is the intension and concept description; M^α is the α-upper extension; M^β is the β-lower extension.*

Definition 3.3. [10] *Using $L_\alpha^\beta(U, A, R)$ to express all interval concepts within $[\alpha, \beta]$, record: $\left(M_1^\alpha, M_1^\beta, Y_1\right) \leq \left(M_2^\alpha, M_2^\beta, Y_2\right) \Leftrightarrow Y_1 \supseteq Y_2$, and "$\leq$" is the partial ordering relation. If all concepts in $L_\alpha^\beta(U, A, R)$ meet the partial ordering relation, call $L_\alpha^\beta(U, A, R)$ is the interval concept lattice on formal context (U, A, R).*

3.2. **Association rule mining algorithm based on interval concept lattice.** Due to the upper and lower extension of interval concept, α-upper association rules and β-lower association rules can be extracted respectively. The following shows extracting α-upper association rules, for example. The algorithm of extracting β-lower association rules is similar to it.

Algorithm: α-upper association rules mining algorithm based on interval concept lattice

Input: Concept lattice $L_\alpha^\beta(U, A, R)$, minimum support and confidence thresholds: θ and φ, interval parameter α.
Output: α-upper association rules
Process: Step 1, Breadth-first traverse interval concept lattice (built by [12]), and obtain α-upper frequent nodes set α-Fcset. For any concept node of interval concept C1 = (M^α, M^1, Y), if $|M^\alpha| \geq |U|\theta$, α-Fcset = α-Fcset $\cup$ {C1}, finally obtain upper frequent nodes set after traversing all the interval concepts.

Step 2, Generate all α-upper candidate binary groups. Assuming that set of nodes of candidate binary groups formed with any node C1 = (M_1^α, M_1^1, Y_1) in α-upper frequent nodes set α-Fcse is PAIRS(C1). For other node C2 = (M_2^α, M_2^1, Y_2) in α-Fcset, if C2 > C1 and $|M_1^\alpha|/|M_2^\alpha| \geq \varphi$, PAIRS(C1) = PAIRS(C1) $\cup$ {C2}. Repeat above steps until finding all PAIRS in α-Fcset.

Step 3, Eliminate superfluous candidate binary groups. Arrange the nodes of α-Fcset in the descending order of the intension cardinal numbers. If C1 > C2, PAIRS(C1) = PAIRS(C1) – PAIRS(C2). After eliminating superfluous of α-Fcset, obtain α-upper candidate binary groups, and then we can count α-upper frequent nodes set α-Fiset.

Step 4, Generate α-upper association rules: α-Rluesset.

Based on this algorithm, we can extract more refined uncertain association rules, and improve the reliability of the rules.

4. **Model Application.** Medical diagnosis system converts experts' diagnostic experience into rules. As long as users input some symptoms, it can quickly make judgments, which can reduce the errors of doctors' subjective judgments. Furthermore, the reasoning rules and conclusions of the system are designed in advance. Some patients' clinical manifestations may not be here, so it has some limitations. Using data mining technology to deal with a large number of historical data in the database can mine valuable diagnosis rules, where we can make a preliminary diagnosis being of objectivity according to the basic symptoms of patients. Due to large amount of data, the resulting diagnosis rules have well universal application.

4.1. **Association rules analysis of pneumonia symptoms.** According to patients' pneumonia cases and doctors' clinical diagnosis experience, attributes of pneumonia symptoms include headache, inappetence, emesis, dyspnea, fever, chest pain, cough, hemoptysis, muscle pain and shiver 10 characteristic attributes. Take 15 cases as the training samples and obtain decision formal context, including conditional attributes {a, b, c, d, e, f, g, h, i, j}, decisional attribute K and objects 1-15, as shown in Table 1.

Table 1. Decision formal context

object	a	b	c	d	e	f	g	h	i	j	K	object	a	b	c	d	e	f	g	h	i	j	K
1	0	0	0	1	1	0	0	1	0	1	1	9	0	0	1	1	0	1	1	1	1	0	1
2	1	0	0	0	1	0	0	0	1	1	1	10	0	1	1	0	1	0	0	1	0	1	1
3	0	1	1	1	0	0	0	1	0	1	1	11	0	0	0	1	0	0	0	0	1	0	0
4	0	0	1	1	0	1	0	1	0	0	1	12	0	0	1	1	0	1	1	1	1	0	1
5	1	1	0	1	0	0	1	0	0	0	0	13	0	0	0	0	1	1	0	1	0	1	1
6	0	0	1	1	1	1	0	0	0	1	1	14	0	1	1	1	1	0	0	1	0	1	1
7	1	1	0	0	0	1	0	0	0	0	0	15	1	0	1	1	1	0	0	0	0	1	1
8	0	0	0	1	1	1	0	1	0	1	1												

"1" expresses patients have this symptom or pneumonia; "0" expresses patients do not have this symptom or pneumonia. Interval parameters $\alpha = 0.5$, $\beta = 1$ in here. Build interval concept lattice structure [12] under interval parameters $[\alpha, \beta]$. Due to space limitations, here are some interval concepts shown by Table 2.

According to α-upper association rules mining algorithm based on interval concept lattice, set the minimum support is 0.5, and obtain the frequent nodes from Table 2, as shown in Table 3.

Set the minimum confidence is 0.6, and the α-upper candidate binary groups will be got. Finally, we can compute α-upper association rules between the attributes. In a similar way, β-lower association rules will be obtained. Above all, there are 50 association rules. Among them, the number of rules whose confidences are 1 reaches more than half. Extracted association rules are shown by Table 4.

Among these rules, we can easily see that confidences of rules ehj⇒k and dfh⇒k are 1, explained by: when someone starts to have symptoms like fever, shiver, hemoptysis or dyspnea, dyspnea, hemoptysis. The patient can be initially determined suspected pneumonia by system, and needs hospital review (blood test or CT), and compared with

TABLE 2. Some interval concepts

Concept	Upper extension	Lower extension	Intension
C1	$\{2, 5, 7, 15\}$	$\{2, 5, 7, 15\}$	a
C2	$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15\}$	$\{2, 15\}$	ak
C3	$\{2, 3, 5, 7, 10, 14, 15\}$		abk
C4	$\{2, 3, 4, 6, 9, 10, 12, 14, 15\}$	$\{15\}$	ack
C5	$\{1, 2, 6, 8, 10, 13, 14, 15\}$	$\{2, 15\}$	aek
C6	$\{2, 4, 6, 7, 8, 9, 12, 13, 15\}$		afk
C7	$\{2, 5, 9, 12, 15\}$		agk
C8	$\{1, 2, 3, 6, 8, 10, 13, 14, 15\}$	$\{2, 15\}$	ajk
C9	$\{2, 3, 4, 5, 6, 7, 9, 10, 12, 14, 15\}$		abck
C10	$\{1, 2, 3, 5, 6, 7, 8, 10, 13, 14, 15\}$		abek
C11	$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15\}$		abfk
C12	$\{2, 3, 5, 7, 9, 10, 12, 14, 15\}$		abgk

TABLE 3. α-upper frequent nodes

Concept	Upper extension	Intension
α-FC1	$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15\}$	ak
α-FC2	$\{2, 3, 4, 6, 9, 10, 12, 14, 15\}$	ack
α-FC3	$\{1, 2, 6, 8, 10, 13, 14, 15\}$	aek
α-FC4	$\{2, 4, 6, 7, 8, 9, 12, 13, 15\}$	afk
α-FC5	$\{1, 2, 3, 6, 8, 10, 13, 14, 15\}$	ajk
α-FC6	$\{2, 3, 4, 5, 6, 7, 9, 10, 12, 14, 15\}$	abck
α-FC7	$\{1, 2, 3, 5, 6, 7, 8, 10, 13, 14, 15\}$	abek
α-FC8	$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15\}$	abfk
α-FC9	$\{2, 3, 5, 7, 9, 10, 12, 14, 15\}$	abgk

TABLE 4. Association rules under $[0.5, 1]$ interval concept lattice

Rules	Conf	Rules	Conf	Rules	Conf	Rules	Conf	Rules	Conf
ahj⇒k	1	dfh⇒k	1	ehi⇒k	1	abdh⇒k	0.91	acdi⇒k	0.8
cde⇒k	1	dfj⇒k	1	ehj⇒k	1	abdj⇒k	0.9	acdj⇒k	0.92
cfj⇒k	1	dgh⇒k	1	eij⇒k	1	abef⇒k	0.89	acef⇒k	0.91
chj⇒k	1	dgj⇒k	1	fhj⇒k	1	abeh⇒k	0.8	aceh⇒k	1
cij⇒k	1	dhi⇒k	1	ghj⇒k	1	abej⇒k	0.82	acej⇒k	1
def⇒k	1	dhj⇒k	1	hij⇒k	1	abfh⇒k	0.9	acfh⇒k	0.91
deg⇒k	1	dij⇒k	1	abcd⇒k	0.9	abfj⇒k	0.9	c⇒k	1
deh⇒k	1	efh⇒k	1	abcf⇒k	0.9	acdf⇒k	0.8	d⇒k	0.9
dei⇒k	1	efj⇒k	1	abde⇒k	0.9	acdg⇒k	1	h⇒k	1
dej⇒k	1	egj⇒k	1	abdf⇒k	1	acdh⇒k	0.91	ej⇒k	1

experts' experiences, these conclusions have a certain credibility. According to the supports and confidences of different association rules, system can provide users with some reference of diagnosis. In general, association rules having higher support and confidence are more persuasive.

4.2. **Model comparison.** Compared with association rules extracted by classical concept lattice, the number of rules extracted by interval concept lattice is more and the precision of rules is more considerable. This method can not only flexibly deal with uncertain information in formal context, but also mine more potential association rules

Table 5. Association rules under $[0, 1]$ classical concept lattice

Rules	Conf	Rules	Conf
c⇒k	1	h⇒k	1
d⇒k	0.9	ej⇒k	1

in medical diagnosis. According to association rules extraction algorithm based on the classical concept lattice, some rules can be obtained as shown in Table 5.

The number of concept nodes in classical concept lattice is obviously less than $[0.5, 1]$ interval concept nodes'. The method of mining association rules based on classical concept lattice ignores a lot of useful and potential edge information. Due to the fact that the extensions of concepts are small, concept nodes meeting support are less, which will affect the number of rules. However, generally the confidences of association rules are higher shown by Table 5. According to that the confidence of h⇒k is 1, patients with a symptoms like hemoptysis are just determined suspected pneumonia, which is lack of persuasiveness and ignores some potential information.

5. **Conclusions.** Based on theory of interval concept lattice, a new association rules mining algorithm is proposed and applied on cases of medical diagnosis: pneumonia. By comparing with association rules mining algorithm based on classical concept lattice, we find that this new method of mining association rule not only improves the support and confidence of rules, but also increases the number of rules. When users use the pneumonia diagnosis system, they can obtain more accurate initial diagnostic rules as reference of hospital review. Obviously, this new method has more advantages. Changing interval parameters [13] to mine rules and enhancing medical diagnosis system performance could be considered in future studies.

Acknowledgements. This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61370168, 61472340), Hebei Province Natural Science Foundation (Grant No. F2016209344), the Graduate Student Innovation Fund of North China University of Science and Technology.

REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami, Database mining: A performance perspective, *IEEE Trans. Knowledge and Data Engineering*, no.5, pp.914-925, 1993.

[2] Z. H. Wang, K. Y. Hu et al., General and incremental algorithms of rule extraction based on concept lattice, *Chinese J. Computers*, vol.22, no.1, pp.66-70, 1999.

[3] K. Y. Hu, Y. C. Lu and C. Y. Shi, An integrated mining approach for classification and association rules based on concept lattice, *Journal of Software*, vol.11, no.11, pp.1478-1483, 2000.

[4] J. Y. Liang and J. H. Wang, A new lattice structure and method for extracting association rules based on concept lattice, *International Journal of Computer Science and Network Security*, vol.6, no.11, pp.107-114, 2006.

[5] Z. P. Xie and Z. T. Liu, Concept lattice and association rule discovery, *Journal of Computer Research & Development*, vol.37, no.12, pp.1416-1421, 2000.

[6] F. D. Xie and Z. F. Wang, Association rules mining based on concept lattice, *Computer Engineering and Applications*, vol.43, no.33, pp.170-179, 2007.

[7] Y. Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, *Information Sciences*, vol.111, pp.239-259, 1998.

[8] H. F. Yang and J. F. Zhang, Rough concept lattice and construction algorithm, *Computer Engineering and Application*, vol.43, no.24, pp.172-175, 2007.

[9] J. Z. Huang, Based on rough concept lattice attribute and rule acquisition, *Software*, vol.32, no.10, pp.16-23, 2011.

[10] B. X. Liu and C. Y. Zhang, A new concept lattice structure-interval concept lattice, *Computer Science*, vol.39, no.8, pp.273-277, 2012.

[11] Z. Y. Wang, *Research and Application of KDD in Traditional Chinese Medical Diagnosis*, Master Thesis, Huazhong University of Science and Technology, Wuhan, 2007.
[12] C. Zhang and L. Wang, An effective interval concept lattice construction algorithm, *ICIC Express Letters, Part B: Applications*, vol.5, no.6, pp.1573-1578, 2014.
[13] M. Li, C. Zhang, L. Wang and B. Liu, Parameters optimization and interval concept lattice update with change of parameters, *ICIC Express Letters*, vol.10, no.2, pp.339-345, 2016.