International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:1, No:11, 2007

Modeling the Symptom-Disease Relationship by Using Rough Set Theory and Formal Concept Analysis

Mert Bal, Hayri Sever, and Oya Kalıpsız

***Abstract*—**Medical Decision Support Systems (MDSSs) are sophisticated, intelligent systems that can provide inference due to lack of information and uncertainty. In such systems, to model the uncertainty various soft computing methods such as Bayesian networks, rough sets, artificial neural networks, fuzzy logic, inductive logic programming and genetic algorithms and hybrid methods that formed from the combination of the few mentioned methods are used. In this study, symptom-disease relationships are presented by a framework which is modeled with a formal concept analysis and theory, as diseases, objects and attributes of symptoms. After a concept lattice is formed, Bayes theorem can be used to determine the relationships between attributes and objects. A discernibility relation that forms the base of the rough sets can be applied to attribute data sets in order to reduce attributes and decrease the complexity of computation.

***Keywords*—**Formal Concept Analysis, Rough Set Theory, Granular Computing, Medical Decision Support System.

I. Introduction

MDDs is successfully applied to evaluate diagnostic information and test results which are stored in hospital management systems and to determine diseases and find the most suitable treatment for patients [1] by screening millions of patient records and to disclose confidential symptoms [2]. When uncertainty is the matter, the use of these systems which facilitate decision making are enormously growing. Within the framework of the study, symptom-disease relationships are modeled by the structure of the concept lattice. However, before modeling rough sets can be used to reduce the complexity in computation. The rough sets theorem was first recommended by Pawlak at the beginning of 1980s and is established on the hypothesis that information can be obtained from every object in the universe [3]-[4].

In the rough sets theory objects that are characterized with the same information and their current data are identical, what is meant, they cannot be discernibility. Based upon this indiscernibility relationship that is produced as the fore said method form the mathematical base of the rough set theory.

II. Rough Sets

At following subsections the basic concepts of the rough sets theory shall be scrutinized.

A. Information Systems

A data set is represented as a table; each line represents a condition, an event, a disease or simply, an object. Each column represents a measurability characteristic of each object (e.g. a variable, an observation). This table is called "*An Information System*". More formally, an information system is represented by the $\mathrm{A} = (U, \mathsf{A})$ binary formula. U is the set of non-empty finite of objects that are named as the universe and A, is a set of non-empty finite attributes. Here, for $\forall a \in \mathsf{A}$ $a : U \to V_a$. The V_a set is called the valued Set of a. Another type of information systems is called the *"Decision Systems"*. A decision system is a specific type of $\mathrm{A} = (U, \mathsf{A} \cup \{d\})$ of any other information system where, $d \notin \mathsf{A}$ is decision attributes. Other attributes are called as $a \in \mathsf{A} - \{d\}$ *conditional attributes*. Decision attributes can receive many values, but in general they will earn values as true or false [5]-[6].

B. Indiscernibility

Decision systems which constitute a very unique state of information systems is capable to cover all information related a model (event state). In decision systems, identical or indiscernibility objects can be represented more than once or their attributes could be more than requisites. Therefore, in such condition the table that represents the decision system could be larger than necessities. In this section we shall describe the relation related with indiscernibility.

If a binary relation $R \subseteq X \times X$, if reflected (e.g. if an object is related with itself, then it is xRx) symmetrical (if xRy then yRx) and transitive (if xRy and yRz, then it is xRz) then this will be a equivalence relation. Equivalence

Manuscript received October 12, 2007.

Mert Bal, Yildiz Tehnical University, Mathematical Engineering Department, Davutpasa Campus, Esenler, İstanbul, 34210 TR (corresponding author to provide phone: +90212-449 17 63, e-mail: mertbal@ yildiz.edu.tr).

Hayri Sever, Cankaya University, Computer Engineering Department, Balgat, Ankara, 06530 TR (phone: +90312-284 45 00-4022 sever@cankaya.edu.tr)

Oya Kalıpsız, Yildiz Technical University, Computer Engineering Department, Yildiz Campus, Yildiz, Istanbul, 34349 TR (phone: +90212-259 70 70-2251, kalipsiz@yildiz.edu.tr)

classes of a $x \in X$ element shall contain the entire objects of $y \in X$. In other words, it is then xRy. Then let $\mathrm{A} = (U, \mathrm{A})$ be an information system, there will be an $IND_{\mathrm{A}}(B)$ equivalence relation with any $B \subseteq \mathrm{A}$.

$$IND_{\mathrm{A}}(B) = \{(x, y) \in U^2 \mid \forall a \in Ba(x) = a(y)\} \quad (1)$$

$IND_{\mathrm{A}}(B)$ is described as a $B-$Indiscernibility relation.

If $(x, y) \in IND_{\mathrm{A}}(B)$ then x and y time objects may not be discernibility from each other by the attributes of B. $B-$Indiscernibility relation of the equivalence classes is shown as $[x]_B$ [5]-[7].

The $IND_{\mathrm{A}}(B)$ indiscernibility relation will separate a U universe sets to equivalence classes of $\{X_1, X_2, ..., X_r\}$ which was given as a binary equivalence relation. The family of the whole $\{X_1, X_2, ..., X_r\}$ equivalence classes that are called as an $IND_{\mathrm{A}}(B)$ in the U sets shall form a classification of the U sets and is then expressed as B^*. The family of B^* equivalence classes are called *"classification"* and is expressed as $U / IND_{\mathrm{A}}(B)$.

However, objects that belong to the same X_i equivalence classes indiscernibility. Otherwise, objects shall be discernibility according to the lower sets of B attributes. An A information that belongs to X_i $(1,2,..........,r)$ equivalence classes of $IND_{\mathrm{A}}(B)$ is called as elementary sets. $[x]_B$ demonstrate a B elementary set which includes an x element and is described with the (1) equality as below:

$$[x]_B = \{y \in U \mid xIND_{\mathrm{A}}y\} \quad (2)$$

A $(U, IND_{\mathrm{A}}(B))$ ordered pair called an *"approximation space"*. The finite combination of a elementary sets in an approximation space is ratified as "described set at an approximation space". The A elementary sets of an $\mathrm{A} = (U, \mathrm{A})$ information system is called *"atoms"* of the A information system [8].

C. Discernibility Matrix

A study of the discernibility of objects was carried out by Skowron and Rauszer. In the mentioned study [9], to describe the entire concepts in a given information system, a discernibility matrix and a discernibility function was presented to form effective algorithms related with the formation of lower sets of adequate number of minimum attributes.

Let's say A is an information system with n item of objects. The discernibility matrix M_{A} of an A information system will be a symmetric $n \times n$ matrix that are the elements of c_{pq} as shown below. Each c_{pq} element of this matrix forms from the attribute sets that differentiate x_p and x_q objects.

$$c_{pq} = \{a \in \mathrm{A} \mid a(x_p) = a(x_q)\} \quad (3)$$

$$(p, q = 1,2,............,n)$$

Conceptually, the M_{A} discernibility matrix is a $|U| \times |U|$ matrix. To form the discernibility matrix, we must consider the pairs of different objects. As $c_{pq} = c_{qp}$ and $c_{pp} = \varnothing$, for all x_p and x_q objects, it shall be not necessary to calculate the half of the elements during the formation of the M_{A} discernibility matrix.

D. Approximations of Set

The main idea that lies under the rough sets theory is to form the approaches of sets by using the $IND_{\mathrm{A}}(B)$ binary relation. If X cannot be defined precisely by using the attributes of A, then it must be expressed as lower and upper approximations. Let us assume that $\mathrm{A} = (U, \mathrm{A})$ is an information system and a $B \subseteq \mathrm{A}$ and $X \subseteq U$. X can be approached by using information which is included only in B by forming B-lower and B-upper approximations which are demonstrated respectively, as $\underline{B}X$ and $\overline{B}X$. Here, lower and upper approximation can be defined as below:

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

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

Objects in $\underline{B}X$, are classified as the exact X elements on the base of the information in B. Objects in $\overline{B}X$ can be classified as the presumptive elements of only X on the base of the information in B.

$$BN_B(X) = \overline{B}X - \underline{B}X \quad (6)$$

Equation (6) is called the "B-boundary region of X" and thus now is formed by unclassified objects of the exact X on the base of information B.

III. Formal Concept Analysis

Formal Concept Analysis (FCA) is a theoretical method for the mathematical analysis of scientific data and was found by Wille in the middle of 80s during the development of a framework to carry out the lattice theory applications. FCA models the real world as objects and attributes. FCA will define concepts in their given content and study the inter-concept relationship regarding the structure of the lattice that corresponds to the content. As a mathematical notion concept is rooted from formal logic. This common definition can be made by two routes, extent and intent. The intent provides the attributes of context while extent covers the objects that are included in the concept. Below a basic definition of a formal concept analysis shall be carried out and the concepts shall be emphasized.

A. *Definitions*

Some terms of Formal Concept Analysis will be defined here.

Definition 1: Formal Context is described as the (G, M, I)triplet between the G, M sets and the I sub set of $\mathrm{I} \subseteq G \times M$ and a binary relation between G and M. The elements of G are called objects while the elements of M are called attributes that the mentioned objects own or it can be simply considered as the characteristics of objects. For g object and the m characteristic, $(g, m) \in \mathrm{I}$ or $g \,\mathrm{I} m$ shall indicate the following: g object owns the m attribute [10]-[11].

Definition 2: Let us assume that (G, M, I) is the formal context. Then if $A \subseteq G$ and $B \subseteq M$ is as, then the α and β operators can be defined as the following:

$$\beta : 2^G \rightarrow 2^M,\ \beta(A) = \{\ m \in M \mid g \,\mathrm{I} m \ \ \forall g \in A\ \} \quad (7)$$

$$\alpha : 2^M \rightarrow 2^G,\ \alpha(B) = \{\ g \in G \mid g \,\mathrm{I} m \ \ \forall m \in B\ \} \quad (8)$$

$\beta(A)$, will take us to the attributes sets that are common in the entire objects in the A set. Similarly $\alpha(B)$ function will take us to the attribute elements of G that owns the entire attributes of B. In other words, $\beta(A)$, shall give the maximum object set that is hared by the entire objects in A while $\alpha(B)$ still give the maximum object set that owns the entire attributes in B. The (β, α) binary shall form a Galois connection between 2^G and 2^M. When the (β, α) binary is given as a Galois binary, then the Lemma below shall become valid.

Lemma: Let us assume that $(G, M, \mathrm{I})$is a context. Under such state, the following propositions are valid: [11]

1. For $\forall A_1, A_2 \subseteq G$, $A_1 \subseteq A_2 \Rightarrow \beta(A_1) \supseteq \beta(A_2)$ (9)

For $\forall B_1, B_2 \subseteq M$, $B_1 \subseteq B_2 \Rightarrow \alpha(B_1) \supseteq \alpha(B_2)$ (10)

2.For all $A \subseteq G$, $A \subseteq \alpha(\beta(A))$

and $\beta(A) = \beta(\alpha(\beta(A)))$ (11)

For all $B \subseteq M$, $B \subseteq \beta(\alpha(B))$

and $\alpha(\beta) = \alpha(\beta(\alpha(B)))$ (12)

Definition 3: When $A \subseteq G$ and $B \subseteq M$ are formed, an atomic concept in the (G, M, I) is a (A, B)binary. Thus, $\beta(A) = B$ and $\alpha(B) = A$ will form. For the (A, B) concept, A is called *extent* and B is called *intent.*

Definition 4: If the $B \subseteq M$ set is a $\beta(\alpha(B)) = B$, in other words, if the B set is only the context of the $(\alpha(B), B)$ concept, then it is called the *feasible internal* (or set of attributes feasible to B). Similarly, if $A \subseteq G$ set is $\alpha(\beta(A)) = A$ or in other words, if $(A, \beta(A))$ is a concept and if the A set is only the extent of this context, then it is called a feasible set for A (or feasible set of objects).

Let us assume that the whole set of concepts of the (G, M, I) context is $\mathcal{C}(G, M, \mathrm{I})$. The following ordered relation can be defined on $\mathcal{C}(G, M, \mathrm{I})$.

Let us assume that (A_1, B_1) and (A_2, B_2) are two concepts in $\mathcal{C}(G, M, \mathrm{I})$.

Then: $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$ (or $\cong B_1 \supseteq B_2$)

In the expression above, (A_1, B_1) is called the sub-concept of (A_2, B_2).

Similarly, (A_2, B_2) is called the upper-concept of (A_1, B_1).

The $\mathcal{L}(G, M, \mathrm{I})$ $\leq$ partial ordered relation and $\mathcal{C}(G, M, \mathrm{I})$ which is the set of entire concepts of the (G, M, I) will indicate the partial ordering set formed by $\mathcal{C}(G, M, \mathrm{I})$.

$\mathcal{L}(G, M, \mathrm{I})$ can be demonstrated as $\mathcal{C}(G, M, \mathrm{I})$, $\leq$. The basic theory of Wille on concept lattices state that $\mathcal{L}(G, M, \mathrm{I})$ is a whole lattice. This lattice is called the concept lattice of (G, M, I) context. The upper and lower concept relation between concepts may lead to a concept hierarchy, therefore the ordered sets theory, may form a study medium compatible to formally solve the hierarchy in concepts. In an application where the data is interpreted as a content, FCA can be applied in data combining, analysis of attribute dependencies, classification of objects, understanding or explanation of data [10].

Theorem: Let us assume that (G, M, I) is a context. (G, M, I) on $\mathcal{C}(G, M, \mathrm{I})$ shall form a complete lattice of a definable set of concepts. Therefore, ($\mathcal{C}(G, M, \mathrm{I})$, $\leq$) is a complete lattice. Suprema and Infima are described as below: [12]

$$\bigvee_{j \in J} (A_j, B_j) = \left(\alpha \left(\bigcap_{j \in J} B_j \right), \bigcap_{j \in J} B_j \right) \quad (13)$$

$$\bigwedge_{j \in J} (A_j, B_j) = \left(\bigcap_{j \in J} A_j, \beta \left(\bigcap_{j \in J} A_j \right) \right) \quad (14)$$

IV. Modeling Symptom-Disease by Concept Lattice

In this section we shall explain the modeling of the lattice structure within a given context regarding Symptom-Disease relationship. In the table below, letters show attributes (symptoms) and numbers show objects (diseases). The lattice structure of a context is shown as in Figure 1. Here, if we take a look to the table below, we can assume the set of objects $G = \{1,2,3,4,5,6,7,8\}$ and the symptoms as set of attributes $M = \{a,b,c,d,e,f,g,h,i\}$ which corresponds to the mentioned objects.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
1	x	x					x		
2	x	x					x	x	
3	x	x	x				x	x	
4	x		x				x	x	x
5	x	x		x		x			
6	x	x	x	x		x			
7	x		x	x	x				
8	x		x	x		x			

Fig. 1 Context for symptom-disease relationship [13]

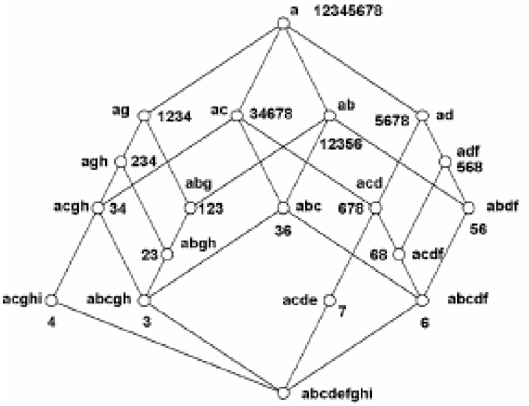

Fig. 2 Lattice structure that belong to symptom-disease context [13]

Firstly, let us define a $\Im(D,S)$ relationship between diseases and symptoms. As an example, if we carefully consider the context in Table I, we can demonstrate the relationship between disease and symptoms as:

$$\Im(D,S) = \{\langle 1,a\rangle,\langle 1,b\rangle,\langle 1,g\rangle,...............,\langle 8,f\rangle\} \quad (15)$$

Here let us define a new relationship as

$$\Im_1(S') = \{\langle a,b,g\rangle,\langle a,b,g,h\rangle,..........,\langle a,c,d,f\rangle\} \quad (16)$$

And let us define the sets here as Symptom-Disease relationship. The Symptom-Disease problem records that consist from 1's and 0's can be easily transformed into a Boolean relationship. Let us assume that the number of symptoms n, is the record number m in $\Im_1(S')$. Let us take a look to the transformation process below:

$$f : X \in \Im(S) \rightarrow \langle D_1, D_2,....., D_n\rangle \quad (17)$$

such that,

$$D_i = \{i \in X \Rightarrow 1, i \notin X \Rightarrow 0\} \quad (18)$$

The defined f function above, owns n items of attributes and if the value of each attribute is one of the element the X of i symptom, then if not 1, then it shall be transformed to a records of 0. As an example, if we apply the f function to the $\langle a,b,d,f\rangle$ record that is included in the $\Im_1(S')$ relationship, then a f ($\langle a,b,d,f\rangle$) = $\langle 1,1,0,1,0,1,0,0,0\rangle$ shall be obtained. Consequently, the f function will transform the $\Im_1(S')$ relationship into $\Im_2(D_1, D_2,....., D_n)$ relationship. The $\Im_2$ relationship shall consist from n items of attributes and m items of records.

V. Formation of Concept Lattice

In a given context, let us assume that the concept lattice is modeled together with a C=<V, E> undirected graph. Here, V represents the nodes and E represents the edges. Each node of the graph consists from a set of attributes. Let us assume that there is an $a : V \rightarrow \aleph(S)$ function that transforms the attributes in the $v \in V$ node. Let us also assume that a J index set defined on the V nodes. Every edge located on the graph will demonstrate the relationship between two nodes. If $(v_i, v_j) \in E$, then there will be a R matching present between the attributes of the v_i and v_j nodes. The R matched relationship located on the C graph shall display a reflected, transitive, cross symmetrical row.

To form the concept lattice of the C = <V, E> undirected graph, firstly we must find the V, E sets. The algorithm to form a concept lattice readily is based on an $a(v_i) \subset a(v_j)$ or an $a(v_j) \subset a(v_i)$ idea on an edge between the v_i and v_j nodes. To obtain a set of V nodes by using a whole lattice of the C = <V, E> graph, a node must be created for the each record in the $\Im_1(S')$ relationship and nodes are added till a whole lattice is formed. To determine the nodes located on the graph, the algorithm is as shown below.

```
Algorithm: V (C) presence of nodes on the graph.
Inputs: Records included in the $\Im_1(S')$relationship.
Outputs: V (C), J
Find Node ( $\Im_1(S')$, V (C), J) Routine;
Exterior Structures;
Records related with $\Im_1(S')$ relationship, ENTRANCE;
Set of nodes that form the V (C), C graph, EXIT;
```

```
Index that can change till the numbers of elements between 1 and V (C) on the borders of J, V (C), EXIT;
INTERIOR STRUCTURES;
COMMANDS;
    Add the records in the $\Im_1(S')$ onto the V (C) set;
    Add the sets of Ø, {1, 2, ………, n} onto the V (C) set;
    J:= |V(C)| +2;
    FIRST REITERATE, IF NO ANY CHANGE IN THE
    V (C) SET, THEN EXIT;
        IF $(i \neq j)$ AND if $(a(v_i) \cap a(v_j) \notin V(C))$, then [
            $V(C) := V(C) \cup a(v_i) \cap a(v_j)$
            J: = J + 1;
        ]
FINDING NODE IS OVER
```

The most excellent state of the FINDING NODE Algorithm complexity is O (m^2). However, the most worse working time algorithm complexity is O (2^n). To find the E (C) set of edges between nodes in the graph there must be node pairs which are directly related with each other on V (C). If there isn't a $v_k \in V(C)$ node that provides the $a(v_i) \subset a(v_k) \subset a(v_j)$ condition between $a(v_i) \subset a(v_j)$ and nodes, then there will be an edge between v_i and v_j. This is called the *"Inclusion Property"*. If v_i , v_j posses an inclusion property , then it is recognized that the v_i node will completely cover the v_j node. Entire edges that are located on E (C) are node pairs that provide the V (C) inclusion property. A necessary algorithm is shown below to determine the edges on the graph. The complexity of the FINDING EDGE Algorithm is O (h^3) [14]-[15].

```
Algorithm: Finding E (C) presence of nodes on the graph.
Inputs: V (C)
Outputs: E (C)
Find Node (V (C)) Routine;
Exterior Structures;
Sets of nodes that form the V (C) , C graph, ENTRANCE;
INTERIOR STRUCTURES;
The number of elements of the h, V (C) set
COMMANDS;
    h:= |V(C)|;
    E(C):= $\varnothing$;
    REPEAT BY COUNTING (i: = 1, I < = h, I: = i + 1)
        REPEAT BY COUNTING (j: = 1, j< = h, j: = i + 1)
            IF $(i \neq j)$ AND IF $a(v_i) \subset a(v_j)$, THEN [
                Adding an edge = TRUE;
                REPEAT BY COUNTING (k: = 1, adding an edge AND k< = h, k+1)
                    IF $(i \neq j)$ AND IF $(k \neq j)$ AND IF $a(v_i) \subset a(v_k) \subset a(v_j)$ THEN
                        adding an edge = FALSE
                IF add an edge THEN
                    E (C): = E (C) $\cup \{(v_i, v_j)\}$
            ]
ADDING AN EDGE IS OVER
```

VI. Conclusion

In this study, symptom-disease relationships, disease objects and symptoms were modeled by the aid of a concept lattice using a formal concept solution method as attributes. Rough sets were applied to attribute data to reduce the complexity of computation the lattice structure. A theoretical framework for symptom-disease relationship is presented in this paper.

References

[1] C. Zhang, S. Zhang, *Association Rule Mining.* Berlin: Springer-Verlag, 2002, ch. 1.

[2] S. Pal, P. Mitra, *Pattern Recognition Algorithms for Data Mining*, Boca Raton: Chapman Hall, Boca Raton, 2004, ch.1.

[3] H. S. Nguyen, D. Slezak, "Approximate reducts and association rules correspondence and complexity results", *Lecture Notes in Computer Science*, vol. 1711, pp. 137-145, November 1999.

[4] Z. Pawlak, "Rough sets and intelligent data analysis", *Information Sciences*, vol. 147, pp. 1-12, November 2002.

[5] J. Komorowski, L. Polkowski, A. Skowron. (1998). Rough sets: a tutorial [Online]. Available: http://citeseer.ist.psu.edu/komorowski98rough.html

[6] S. Hui. (2002), Rough set classification of gene expression data [Online]. Available: http://www.cs.uwaterloo.ca/~s2hui/RoughSetProject.pdf

[7] J. Komorowski, Z. Pawlak, L. Polkowski. (1999). A rough set perspective on data and knowledge [Online]. Available: http://citeseer.ist.psu.edu/336073.html

[8] H. S. Binay, "Rough set approaches in investment decisions", Ph.D. dissertation, Dept. Business Adm., Ankara Univ., Ankara, 2002.

[9] Intelligent Decision Support: Handbook of Advance and Applications of the Rough Set Theory", Kluwer Academic Publishers, 1992, pp, 311-362.

[10] J. M. Saquer, "Formal concept analysis and applications", Ph.D. dissertation, Dept. Comp. Sci. and Eng. Univ. of Nebraska, Lincoln, 2000.

[11] J. M. Saquer, J. S. Deogun, "Concept approximations based on rough sets and similarity measures", *Int. Journal App. Math. Computer Science*, vol. 11, pp. 655-674, 2001.

[12] J. S. Deogun, J. M. Saquer, "Monotone concepts for formal concept analysis", *Discrete Applied Mathematics*, vol. 144, pp. 70-78.

[13] W. Ganter, R. Wille, *Formal Concept Analysis: Mathematical Foundation.* Berlin: Springer-Verlag, 2002, ch.1.

[14] J. S. Deogun, V. V. Raghavan, H. Sever, "Association mining and formal concept analysis", in *Proc. 6th Int. Workshop on Rough Set, Data Mining and Granular Computing"*, North Carolina, 1998, pp. 335-338.

[15] B. Oğuz, H. Sever, M. Tolun, "Eşleştirme Sorgularının Modellenmesi", in *The 9th Turkish Symposium on Artificial Intelligence and Neural Networks"*, Izmir, 2000, pp. 381-390.