퍼지 및 지능시스템학회 논문지 2003, Vol. 13, No. 2, pp. 222-230

Fuzzy Relational Method를 이용한 CLINAID의 Knowledge Source 신뢰성 조사

Investigation of the Reliability of Knowledge Source in CLINAID using Fuzzy Relational Method

노찬숙
Chan-Sook Noe

우송대학교 컴퓨터전자정보공학부

요 약

의료 시스템이 개발되면 시스템이 사용하는 knowledge source의 신뢰도가 시스템의 수행능력에 큰 영향을 미치게 되므로, knowledge source의 신뢰도를 검증해야한다. 본 논문은 의료 시스템 CLINAID의 knowledge source의 신뢰성 조사에 대한 연구의 방법과 결과를 발표하였다. 그 방법으로는 CLINAID에 사용된 Cardiovascular body system 데이터에 fuzzy relational method를 적용하여 구조적 분석을 통해 만들어진 인공의 syndrome을 knowledge base에 저장되어있는 의료 전문가의 syndrome과 비교하였다. 7 가지 fuzzy implication operator를 사용하여 거의 비슷한 결과들을 산출해 냈으며, 그 결과들이 전문가가 제공한 syndrome과 거의 일치하였다.

ABSTRACT

Once the medical knowledge-based system has been developed, it is essential to investigate the knowledge sources of the system because knowledge sources can affect the performance of the system in great deal. This paper presents the method and the results of the reliability test done on the medical knowledge-based system CLINAID. A knowledge source tested is Cardiovascular body system data used in CLINAID. The reliability test will be done by investigating structural relationships revealed by fuzzy relational method between the components of the knowledge sources of individual body systems using syndromes as its main component. These partitions are going to be compared with the syndromes elicited from the medical experts. This paper also reports the outcome of the computations using 7 implication operators performed on Cardiovascular body system data.

Key Words : CLINAID, fuzzy relational products, structural analysis, fuzzy implication operators, medical knowledge-based systems

1. Introduction

A prototype of Diagnostic Unit of CLINAID has been successfully developed and is in the process of fine-tuning the system. Generally, once the medical knowledge-based system has been developed, it is essential to investigate the knowledge sources of the system. In the case of CLINAID, however, it is much more difficult and different from the other medical systems because it deals with the whole human body systems instead of a single body system [1][2]. Thus, CLINAID must deal with the complexity problem caused by the multiplicity of contexts and the huge amount of the medical knowledge it has to contain.

To reduce the complexity of inference, CLINAID uses the concepts of a hierarchy of diagnostic levels and syndromes in the Diagnostic Unit. Each level of hierarchy contains a complete structure, which may be called *granule* [3]. The hierarchy subdivides the whole activity of diagnostic process into several level-based activities. That is, it creates a number of different granules in the diagnostic process and so decreases the size of the problem to be resolved at each granule. This diagnostic hierarchy uses syndromes as the main instrument.

A syndrome is a set of signs and symptoms which distinguishes a set of diseases. Syndromes are essential when the system deals with a multiplicity of contexts of different medical specialties. Using the syndromes instead of the whole set of signs & symptoms in the diagnostic process eliminates the facts irrelevant to a particular context that may cause ambiguities in inference. Using syndromes also reduces the complexity of inference and increases the reliability of inference.

접수일자 : 2002년 11월 1일
완료일자 : 2003년 2월 7일

Therefore, it would be fittingly logical that reliability test on CLINAID be done by investigating structural relationships between the components of the knowledge sources of individual body systems using syndromes as its main component. Furthermore, it is desirable to compare information contained in syndromes that were given explicitly by experts, with intrinsic groupings of signs & symptoms characterizing individual diseases in a body system that are extracted computationally. These comparisons could give the breakthrough of finding the artificial syndromes that refer to the intrinsic clusters of signs and symptoms made computationally by the system.

Comparison of the structures of the knowledge sources will be done by computing the fuzzy relational products between a relation and a transpose(or, inverse) of that relation. For example, if SD is a relation of signs & symptoms to diseases, then we compute ($SD \lhd SD^T$), where SD^T is the transpose of a relation SD, to analyze the structure of the relation between signs & symptoms and diseases. This processing should reveal hidden groupings by discovering equivalences and preorders of signs & symptoms in a body system. The intrinsic groupings are represented by the partitions or the equivalence classes in the hierarchical structure of the Hasse Diagram extracted from the preorder of signs and symptoms over diseases. What is concerned about here is not the classification of diseases, but the groupings of signs & symptoms extracted mechanically, in order to examine whether these groupings are comparable to the real-life medical expertise given by the expert sources such as doctors, clinicians, or medical encyclopedias.

In this paper, these partitions are going to be compared with the syndromes elicited from the medical experts in order to investigate the reliability of the knowledge structures in the Diagnostic Unit. Cardiovascular system data provided by Dr. Anderson will be used to accomplish this task. This data has 181 signs & symptoms and 20 diseases. Thus, the composed relation using the relational products will produce the matrix the size of which is 181 x 181. Names of signs & symptoms with associated syndromes and names of diseases in the actual data are listed in [4].

Actual data have 60 signs & symptoms (1 through 60) that are related to syndrome 2, Acute Cardiac Failure, 22 signs & symptoms (61 through 82) related to syndrome 3, Right Sided Heart Failure, and 18 signs & symptoms (83 through 100) related to syndrome 4, Left Sided Heart Failure. Signs & symptoms of 101 through 181 are related specifically to 1 or more of 20 diseases in the data, but are not commonly related to any of syndromes in the data.

This paper reports the outcome of the computations using 7 implication operators, Standard Star (S^*), Gaines 43 ($G43$), Modified Gaines 43 ($G43'$), Łukasiewicz ($Ł$), Kleene-Dienes-Łukasiewicz ($KDŁ$), Kleene-Dienes (KD) and Early Zadeh (EZ), performed on this data. The definitions of the fuzzy implication operators most widely used including the above 7 operators are listed in the section 2.3. Comparisons will be made on how closely the groupings computed by the relational products match the syndromes given by the medical experts.

2. Theoretical Background

2.1 General Overview of CLINAID

CLINAID is a medical fuzzy knowledge-based system that is designed to capture the complete semantics of medical activity in a hospital environment. It is intended to assist in not only the diagnosis of a patients illness, but also other types of hospital activities such as consultation, the prescription of medication, and update of patients records, etc. Each type of activity is conducted in a different environment. In addition to this, unlike other medical knowledge-based systems, CLINAID is designed to work with all body systems rather than the narrow, restricted domain of a single body system [5]. Hence, it is a system that operates in a multi-environmental situation and makes decisions within a multiplicity of contexts of different medical specialties.

At present, CLINAID is designed to handle 11 body systems representing the multiple branches of medicine, including Cardiovascular, Respiratory, Central Nervous System, Endocrine, and Gastrointestinal System [6]. Thus, its knowledge base contains a large amount of medical expertise and must be able to handle the following problems:

Incompleteness of medical data:

It is rarely the case that the patients' signs & symptoms observed by the clinician possess the exact characteristics of the disease, as they are not always present or observable.

Locality of inference:

Not all signs & symptoms of the patient are relevant in a given context. Sometimes, it is possible that irrelevant signs & symptoms may lead to a wrong conclusion.

Complexity of inference:

Because of the huge amount of the medical knowledge and the multiplicity of contexts, the complexity of inference can grow so rapidly that it becomes unmanageable.

In order to be used for decision making across several knowledge domains, CLINAID must handle all of the above problems. Hence its knowledge base must have appropriately structured medical knowledge. These desired characteristics are provided by the fuzzy relational method [7], which is explained in the next section.

The conceptual structures as well as the basic architecture of CLINAID have been described in [7][8]. To give complete cover for all hospital activities, the basic architecture of CLINAID consists of four main cooperating units, each unit comprising a complex autonomous subsystem. The four main units of CLINAID are:

a) *Diagnostic Unit*, which infers a working diagnosis based on the information given by a user, or alternatively sends an indication that the information provided is not sufficient to produce a working diagnosis.
b) *Patient Clinical Record Unit*, which maintains and updates patients medical records. It also allows the information to be retrieved as required in order to ensure that all the relevant patient information can be provided at critical moments.
c) *Treatment Recommendation Unit*, which advises appropriate treatment and medication for the patient based on the working diagnosis provided by the Diagnostic Unit. These recommendations take account of both the personal history of the patient and any possible adverse effects from the available treatments.
d) *Planning and Co-ordination Unit*, which controls communication and interaction with the other units. It also ensures that the information necessary to the current activity of each unit is always provided.

2.2 Fuzzy Relational Method

The fuzzy relation theory was originally introduced in 1971 by Zadeh as an extension of fuzzy sets [9]. Since then a lot of productive research has been accomplished in this field. Especially Bandler and Kohout have developed the fuzzy relational method consisting of a mathematical theory, methodology, epistemology, and supporting computational algorithms for representing and processing knowledge.

The fuzzy relational method provides a means of analyzing real-world scientific data as an effective mathematical tool for structural analysis. It helps to identify meaningful structures implicit in real-world data. This method can also be used in Knowledge Engineering. A fuzzy knowledge-based system, such as CLINAID, which utilizes the fuzzy relational method to capture, represent, and infer knowledge, can be operated and manipulated in a unified computational framework.

Practically, the fuzzy relational method was established using fuzzy relational products, and related computational procedures such as fast fuzzy relational algorithms [10]. A brief explanation of the fuzzy relational method is given below.

Given two relations, the relation R, which is an element of the lattice of relations from set A to set B, i.e., $R \in \mathcal{R}(A \to B)$ and the relation $S \in \mathcal{R}(B \to C)$, a product relation $(R * S)$ is a relation from A to C, determined by R and S. There are several types of product used to yield product relations, each having distinctive mathematical properties and applicability. Of those four definitions will be presented by the scheme of $(R * S) \in \mathcal{R}(A \to C)$, where $* \in \{ \circ, \triangleleft, \triangleright, \square \}$.

Given two relations, $R \in \mathcal{R}(A \to B)$ and $S \in \mathcal{R}(B \to C)$,

1. The circle product is defined as:

$$a(R \circ S)c \Leftrightarrow (aR \cap Sc) \neq 0.$$

In this product relation, an element a is related to an element c by the composed relation $(R \circ S)$ if and only if the intersection of the afterset of a and the foreset of c is non-empty. That is, there exists at least one element common to the afterset of a and the foreset of c.

2. The triangle subproduct is defined as:

$$a(R \triangleleft S)c \Leftrightarrow aR \subseteq Sc.$$

In this product relation, an element a is related to an element c by the composed relation $(R \triangleleft S)$ if and only if the afterset of a is a subset of the foreset of c.

3. The triangle superproduct is defined as:

$$a(R \triangleright S)c \Leftrightarrow aR \supseteq Sc.$$

In this product relation, an element a is related to an element c by the composed relation $(R \triangleright S)$ if and only if the afterset of a is a superset of the foreset of c.

4. The square product is defined as:

$$a(R \square S)c \Leftrightarrow aR \equiv Sc.$$

In this product relation, an element a is related to an element c by the composed relation $(R \square S)$ if and only if the afterset of a is exactly equal to the foreset of c. Mathematically, the square product is the intersection of two triangle products, that is $(R \triangleleft S) \sqcap (R \triangleright S)$.

As the matrix notation is more convenient algorithmically because of its explicit handling of logic values, the above definitions of relational products may also be defined in terms of the logical connectives of the matrices of relations R and S.

1. $(R \circ S)_{ik} \equiv \vee_j (R_{ij} \wedge S_{jk})$
2. $(R \triangleleft S)_{ik} \equiv \wedge_j (R_{ij} \to S_{jk})$
3. $(R \triangleright S)_{ik} \equiv \wedge_j (R_{ij} \leftarrow S_{jk})$
4. $(R \square S)_{ik} \equiv \wedge_j (R_{ij} \equiv S_{jk})$

where R_{ij} and S_{jk} represent the fuzzy degrees to which the respective statements a_iRb_j and b_jSc_k are true.

The applied meaning of these product relations depends on the domain of the problems. Each product relation can be given a specific knowledge domain and semantic interpretation according to its application. For example, if R is the relation between a set of patients and a set of signs and symptoms represented as in (Patients → Signs & Symptoms) and S is the relation between a set of signs and symptoms and a set of diseases represented as in (Signs & Symptoms → Diseases), then the above product relations would have the following meanings [4]:

1. $a(R \circ S)c \Leftrightarrow$ patient a has *at least* one sign or symptom of disease c.
2. $a(R \triangleleft S)c \Leftrightarrow$ the signs and symptoms of patient a are *among* the signs and symptoms which characterize disease c.
3. $a(R \triangleright S)c \Leftrightarrow$ the signs and symptoms of patient a *include* the signs and symptoms which characterize disease c.
4. $a(R \square S)c \Leftrightarrow$ the signs and symptoms of patient a are *exactly* those of disease c.

2.3 Fuzzy Implication Operators

The definitions described in the previous section work for both classical two-valued logic(Boolean logic) and fuzzy logic. For Boolean logic computations, the logical connectives used in definitions($\wedge$, $\vee$, $\rightarrow$) are Boolean logic operators; such as, *and* for $\wedge$, *or* for $\vee$, and a material implication operator for $\rightarrow$.

However, to be of any real use, above product relations must be fuzzified so that the meanings of the product relations can have varying degrees of truth instead of absolutely true or false values. Thus, for the fuzzy logical computations discussed in this work, the following logical connectives are used: *min* for $\wedge$, *max* for $\vee$, and fuzzy implication operators for $\rightarrow$.

Various fuzzy implication operators exist which use different fuzzy logics. The standard condition, commonly accepted in literature which requires the compliance of each fuzzy implication operator, is that it must agree with Boolean logic values at the corners of its own implication table. That is, fuzzy variables must take the crisp value of 0 or 1 at the corners. In this paper, I will consider 10 fuzzy implication operators listed in Table 1. For detailed explanations, see [11].

Since a large number of many-valued logic implication operators exist, the choice of the implication operator becomes a critical step of the fuzzy relational method. Empirical research shows that the choice is all dependent on the nature of the data and knowledge domain of the particular application under investigation.

This determines the selection of the appropriate implication operator(s) in a specific class of application. In other words, one implication operator could be the best choice in one application, yet unusable in another. Or there may exist a number of implications that produce equivalent results. Thus, comparative studies are required for each application dealt with.

There exist two versions of computation for triangle and square products: *harsh and mean.* The harsh criterion takes the minimum value over all the elements, whereas the mean criterion takes the arithmetic mean. For instance, the triangle subproduct can be computed either using the harsh criterion as $(R \triangleleft S)_{ik} = \wedge_j (R_{ij} \rightarrow S_{jk})$, or the mean criterion as $(R \triangleleft S)_{ik} = \frac{1}{N} \sum_{j=1}^{N} (R_{ij} \rightarrow S_{jk})$, where N is the number of elements which are involved in the computation.

Table 1. Fuzzy Implication Operators

1.	$S\#$	Standard Sharp $a \rightarrow_1 b = \begin{cases} 1 & \textit{iff } a \neq 1 \textit{ or } b=1 \\ 0 & \textit{otherwise} \end{cases}$
2.	S	Standard Strict $a \rightarrow_2 b = \begin{cases} 1 & a \leq b \\ 0 & \textit{otherwise} \end{cases}$
3.	S^*	Standard Star $a \rightarrow_3 b = \begin{cases} 1 & a \leq b \\ b & \textit{otherwise} \end{cases}$
4.	$G43$	Gaines 43 $a \rightarrow_4 b = \min\left(1, \frac{b}{a}\right)$
4′.	$G43'$	Modified Gaines 43 $a \rightarrow_{4'} b = \min\left(1, \frac{b}{a}, \frac{1-a}{1-b}\right)$
5.	$Ł$	Łukasiewicz $a \rightarrow_5 b = \min(1, 1-a+b)$
5.5.	$KDŁ$	Kleene-Dienes-Łukasiewicz $a \rightarrow_{5.5} b = \min(1, 1-a+ab)$
6.	KD	Kleene-Dienes $a \rightarrow_6 b = (1-a) \vee b$
7.	EZ	Early Zadeh $a \rightarrow_7 b = (a \wedge b) \vee (1-a)$ $= (a \rightarrow_6 b) \wedge ka$, where $ka = (1-a) \vee a$
8.	W	Willmott $a \rightarrow_8 b = ((1-a) \vee b) \wedge (a \vee (1-b) \vee (b \wedge (1-a)))$ $= (a \rightarrow_7 b) \wedge kb$ $= (a \rightarrow_6 b) \wedge ka \wedge kb$

Thus the computation of fuzzy relational products depends on the choice of the fuzzy implication operator and the version of the two computational criteria used in computing it. As soon as these two are selected, the computation of each element in the matrix of the composed relation becomes a well-determined task.

3. Experimental Results

Comparisons will be made on three syndromes, *Acute Cardiac Failure, Right Sided Heart Failure, and Left Sided Heart Failure*, at each of the four α-cuts computed by TRISYS for each implication operator. After making individual comparisons, the overall performance of each implication operator will be examined by comparing three syndromes in the data. The results of the experiment and the inclusion of each class generated for three syndromes are summarized in the Tables 2 to 4 and Figures 1 to 3.

3.1 Syndrome 2 (Acute Cardiac Failure)

The half-lower α-cut of S^*, $G43$, $G43'$, and $Ł$ implication operators produced a class c1 of 40 signs and symptoms of the syndrome 2, which has 60 signs and symptoms in the data. They all produced the class of the same 40 signs and symptoms of the syndrome 2. The mean α-cut of *G43* and *Ł* implication operators produced a class c1 of 38 signs and symptoms of the syndrome 2. The half-upper α-cut of G43 produced a class c1 of 31 signs and symptoms of the syndrome 2. The mean α-cut of *EZ* and *KDŁ* implication operators produced a class c1 of 21 and 19 signs and symptoms of the syndrome 2, respectively. At the same α-cut, *KD* implication operator produced a class c2 of 14 signs and symptoms. Implication operators *KDŁ, KD,* and *EZ* produced exactly the same class (c1) of 159 signs and symptoms at the half-lower α-cut. Of these 159 signs and symptoms, 40 are related to the syndrome 2 and these 40 signs and symptoms are exactly the same as those produced by the half-lower α-cut of S^*, *G43*, *G43'*, and *Ł* implication operators. The classes generated at the various α-cut levels of different implication operators are completely contained in the syndrome 2 of the actual data. In other words, the same signs and symptoms are clustered together to form different partitions of the syndrome 2. Furthermore, the classes that have smaller numbers of signs and symptoms are contained in the classes of larger numbers of signs and symptoms. The summary and the inclusion of each class generated for the Acute Cardiac Failure syndrome are shown in Table 2 and Figure 1.

The convention we are using here through Table 3 and 4 are as follows. The first column shows the number of signs and symptoms of each syndrome, in this case for the Syndrome 2. Remaining columns show the names of implication operators with different levels of α-cuts. Four α-cut levels are represented as *H* for height α-cut, *HU* for half-upper α-cut, *M* for mean α-cut, and *HL* for half-lower α-cut. For example, S^*_{HL} represents Standard Star(S^*) implication operator using half-lower α-cut.

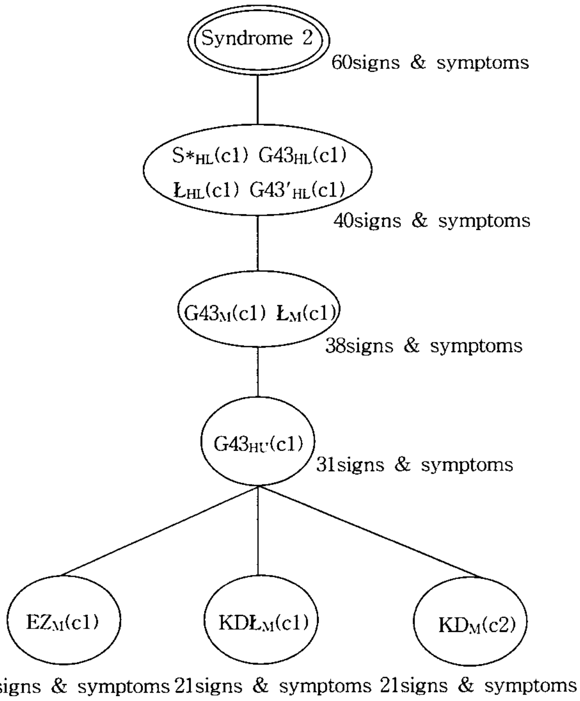


Figure 1. Inclusion of Operators for Syndrome 2

Table 2. Comparison with Syndrome 2

Signs & Symptoms	S^*_{HL} (c1)	$G43_{HL}$ (c1)	$G43'_{HL}$ (c1)	$Ł_{HL}$ (c1)	$G43_M$ (c1)	$Ł_M$ (c1)	$G43_{HU}$ (c1)	EZ_M (c1)	$KDŁ_M$ (c1)	KD_M (c2)
1	x	x	x	x	x	x	x	x	x	
2	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	x	x	x	x	x
5	x	x	x	x	x	x	x			
6	x	x	x	x	x	x	x			
7	x	x	x	x	x	x		x		x
9	x	x	x	x						
10	x	x	x	x	x	x	x			
12	x	x	x	x	x	x	x	x	x	x
13	x	x	x	x	x	x			x	x
14	x	x	x	x	x	x	x			
15	x	x	x	x	x	x		x		x
16	x	x	x	x	x	x	x			
17	x	x	x	x	x	x	x	x	x	x
18	x	x	x	x	x	x	x	x	x	x
19	x	x	x	x	x	x	x	x	x	x
20	x	x	x	x	x	x	x			
22	x	x	x	x	x	x	x	x	x	x
23	x	x	x	x	x	x	x	x	x	x
28	x	x	x	x	x	x	x			
29	x	x	x	x	x	x	x	x	x	x
31	x	x	x	x	x	x	x	x	x	x
33	x	x	x	x	x	x	x			
34	x	x	x	x	x	x		x		
36	x	x	x	x						
40	x	x	x	x	x	x	x	x	x	
41	x	x	x	x	x	x	x	x	x	
42	x	x	x	x	x	x	x	x	x	
43	x	x	x	x	x	x				
44	x	x	x	x	x	x	x			
45	x	x	x	x	x	x				
46	x	x	x	x	x	x	x			
48	x	x	x	x	x	x	x			
49	x	x	x	x	x	x	x			
50	x	x	x	x	x	x				
51	x	x	x	x	x	x	x	x	x	
53	x	x	x	x	x	x	x	x	x	
54	x	x	x	x	x	x	x	x	x	
57	x	x	x	x	x	x	x			
58	x	x	x	x	x	x	x	x	x	x

Also the name of the class inserted in the parenthesis represents the number of the smallest element in that partition. For instance, class c1 of $KDŁ_M$(c1) shows the partition that consists of 19 signs and symptoms, 1, 2, 3, 12, 13, 17, 18, 19, 22, 23, 29, 31, 40, 41, 42, 51, 53, 54, and 58. So, we represent this class as c1 because 1 is the smallest signs and symptoms number in that partition.

3.2 Syndrome 3 (Right Sided Heart Failure)

The mean α-cut of *G43* and *Ł* implication operators and the half-lower α-cut of S^* and *G43'* implication operators generated the same class (c62) which has 20 signs and symptoms of the syndrome 3. This result is almost identical to the syndrome 3 of our data, which has 22 signs and symptoms. At the half-upper α-cut, the *G43* implication operator generated a class c62 of 19 signs and symptoms of the syndrome 3. The mean α-cut of the *KDŁ* implication operator produced a class c64 which has 9 signs and symptoms of the syndrome 3. At the same α-cut, *KD* and *EZ* implication operators generated a class c62 of 7 signs and symptoms of the syndrome 3. As mentioned for the syndrome 2, *KDŁ*, *KD*, and *EZ* implication operators produced exactly the same class (c1) which has 159 signs and symptoms at the half-lower α-cut, and 20 of 159 signs and symptoms are exactly the same as those produced by *G43*, *Ł*, S^*, and *G43′* implication operators. As was the case for the syndrome 2, the classes generated at the various α-cut levels of different implication operators are completely contained in the syndrome 3 of our actual data, and the classes of smaller numbers of signs and symptoms are contained in the classes of larger numbers of signs and symptoms. The summary and the inclusion of each class generated for the Right Sided Heart Failure syndrome are shown in Table 3 and Figure 2.

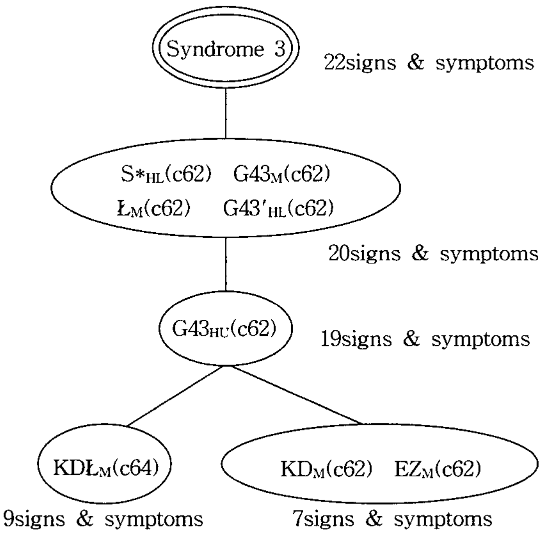


Figure 2. Inclusion of Operators for Syndrome 3

Table 3. Comparison with Syndrome 3

Signs & Symptoms	S^*_{HL} (c62)	$G43_M$ (c62)	$G43'_{HL}$ (c62)	$Ł_M$ (c62)	$G43_{HU}$ (c62)	$KDŁ_M$ (c64)	KD_M (c62)	EZ_M (c62)
62	x	x	x	x	x		x	x
63	x	x	x	x	x			
64	x	x	x	x	x	x		
65	x	x	x	x	x		x	x
66	x	x	x	x	x		x	x
67	x	x	x	x	x	x		
68	x	x	x	x	x	x		
70	x	x	x	x	x	x		
71	x	x	x	x	x	x		
72	x	x	x	x	x	x		
73	x	x	x	x	x		x	x
74	x	x	x	x	x		x	x
75	x	x	x	x	x	x		
76	x	x	x	x	x	x		
77	x	x	x	x	x			
78	x	x	x	x	x	x		
79	x	x	x	x	x		x	x
80	x	x	x	x				
81	x	x	x	x	x			
82	x	x	x	x	x		x	x

3.3 Syndrome 4 (Left Sided Heart Failure)

The mean α-cut of the *KDŁ* implication operator produced a class c83 of 15 signs and symptoms related to the syndrome 4 of our data, which has 18 signs and symptoms related to it. At the same α-cut, the *G43* implication operator generated a class c83 of 13 signs and symptoms of the syndrome 4. The mean α-cut of S^*, *EZ*, and *G43′* implication operators generated the same class (c83) which has 12 signs and symptoms of the syndrome 4. The half-upper α-cut of *G43* and *Ł* implication operators produced a class c84 containing 10 signs and symptoms of the syndrome 4. At the same α -cut, the *KD* implication operator yielded a class c84 of 6 signs and symptoms of the syndrome 4. Just as for the above two syndromes, the same signs and symptoms are clustered together to form several different partitions of the syndrome 4, and the largest class c83 of *KDŁ* is also completely contained in the syndrome 4 of the actual data. The summary and the inclusion of each class generated for the Left Sided Heart Failure syndrome are shown in Table 4 and Figure 3.

Table 4. Comparison with Syndrome 4

Signs & Symptoms	$KDŁ_M$ (c83)	$G43_M$ (c83)	S^*_M (c83)	$G43'_M$ (c83)	EZ_M (c83)	$G43_{HU}$ (c83)	$Ł_{HU}$ (c83)	KD_{HU} (c84)
83	x	x	x	x	x	x	x	
84	x	x	x		x	x	x	x
86	x	x	x	x	x	x	x	x
87	x	x	x	x	x			
88	x	x	x	x	x	x	x	x
89	x	x	x	x	x	x	x	
90	x							
91	x	x	x	x	x	x	x	x
92	x							
93	x	x	x	x	x	x	x	x
95	x	x	x	x	x	x	x	x
96	x	x	x	x	x	x	x	
98	x	x	x	x	x			
99	x	x	x	x	x	x	x	
100	x	x		x				

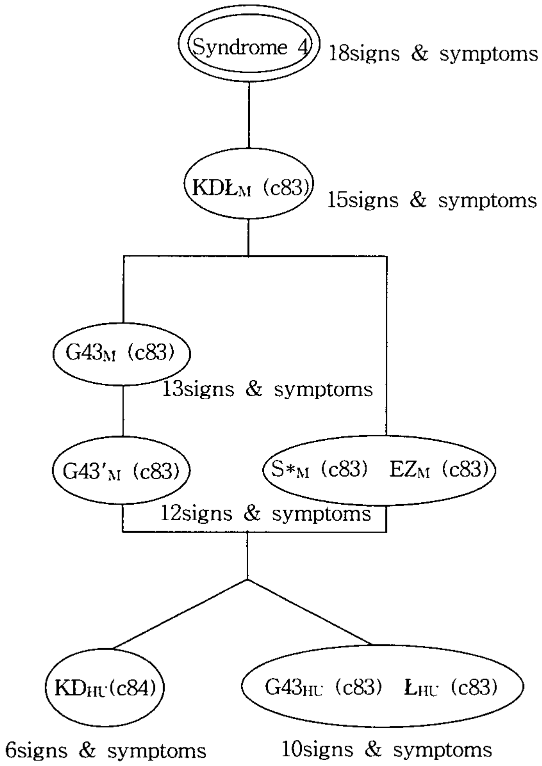


Figure 3. Inclusion of Operators for Syndrome 4

3.4 Analysis of the comparison outcome

In order to make a meaningful comparison, the following criteria are used by which the results are ranked:

1. How close the partition is to the actual syndrome?
   The larger the number of elements in the partition are, the better the performance is.
2. What level of α-cut the partition is clustered?
   The higher the α-cut is, the better the performance is for as the value of α decreases, each α-cut within the sequence of α-cuts contains its predecessors with the gradual addition of more elements.
3. Are elements in the partition contained in the actual syndrome of the data?
   The more elements in the partition are contained, the better the performance is.

The following results have been found from the above comparison based on these criteria:

1. For all 3 syndromes, we have found three classes, one for each syndrome, which are very close to each syndrome in the actual data given by the medical expert.
2. Signs & symptoms of these classes are completely contained in the respective syndromes of the actual data.
3. The same signs & symptoms of each syndrome are clustered together to form several different partitions of their respective syndromes in different hierarchical structures.
4. The *G43* implication operator performed the best among 7 implication operators on all 3 syndromes, the *Ł* implication operator being the close second. Both implication operators consistently yielded partitions that are very close to the actual syndromes.
5. The results of S^* and *G43′* implication operators are very similar, and the performance of both implication operators is closely comparable to that of *G43* and *Ł* implication operators.
6. Both *KD* and *EZ* implication operators produced very similar hierarchical structures, so was their performance on all 3 syndromes. They produced partitions that are very closely matched to each other. The *EZ* implication operator, however, performed somewhat better than the *KD* implication operator on most cases in terms of comparison criteria.
7. The hierarchical structures generated by the *KDŁ* implication operator are very different from those of *KD* and *EZ* implication operators.
   However, the performance of the *KDŁ* implication operator is similar to that of *KD* and *EZ* implication operators on all cases, except the mean α-cut computation of the syndrome 4. In that computation, *KDŁ* produced the best result of all the computations for the syndrome 4.

To show the hierarchical structures generated by each implication operator, Hasse Diagram of seven implication operators at half-lower α-cut are shown in Figures 4 through 10 as the representatives of each implication operator. A bold-faced node represents the equivalence class and the node number indicates the number of signs and symptoms that class represents.

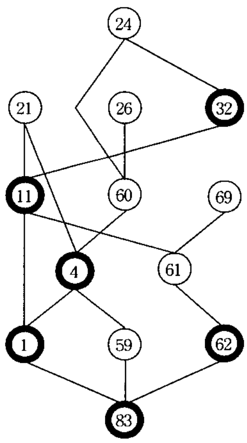


Figure 4 : Hasse Diagram of S^*_{HL}

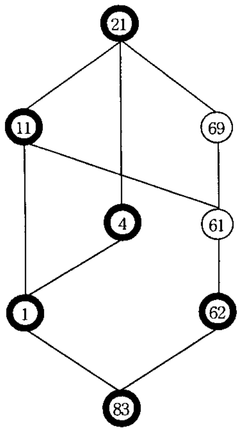


Figure 5 : Hasse Diagram of $G43_{HL}$

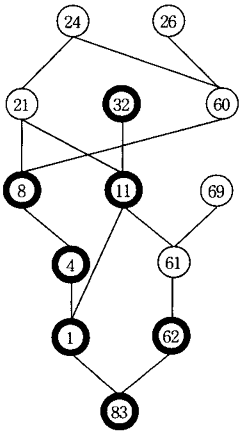


Figure 6 : Hasse Diagram of $G43_{HL}$

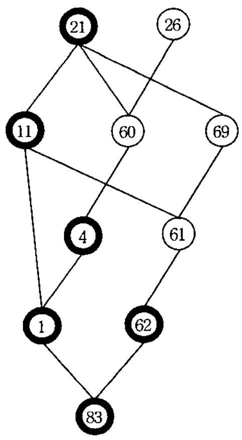


Figure 7 : Hasse Diagram of Ł_{HL}

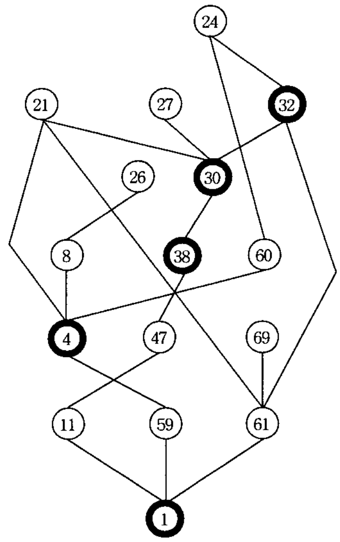


Figure 8 : Hasse Diagram of $KD\text{Ł}_{HL}$

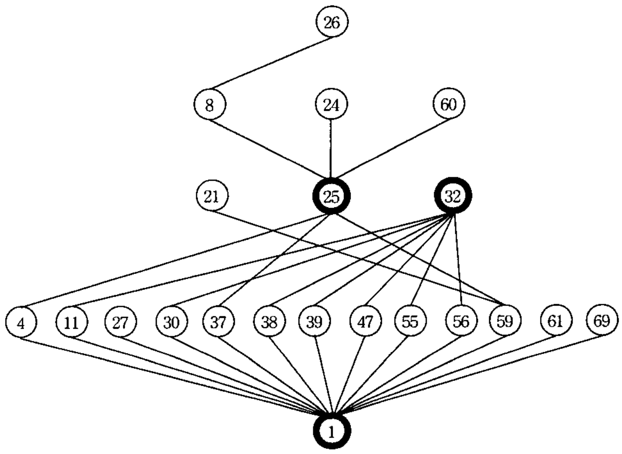


Figure 9 : Hasse Diagram of KD_{HL}

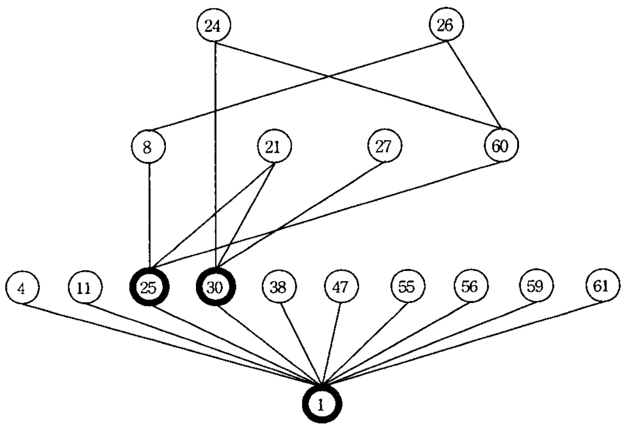


Figure 10 : Hasse Diagram of EZ_{HL}

4. Conclusions

In conclusion, I have been able to find the consistent patterns of partitions through all 7 implication operators. In particular, the largest partitions for each syndrome are the refinements of their respective sets of signs & symptoms forming artificial syndromes in the actual data. If the data are robust, the results from the computations of fuzzy relational products should be consistent among various implication operators. In other words, the more agreements in the outcome of computations using different implication operators, the more reliable the data are.

Thus, in the study of investigating structural relationships of the knowledge sources in the Diagnostic Unit, reliable means that there is no conflict between the knowledge structure elicited from the experts, and the structure relating syndromes to diseases acquired computationally. Therefore, we can conclude that the knowledge source relating signs & symptoms with diseases used for the Cardiovascular body system of the CLINAID system is reliable. Both these structures are compatible and can be used jointly in the inference system of the Diagnostic Unit of CLINAID. Also, this method of generating artificial syndromes is a contribution to computational aspects of engineering of knowledge-based systems connected with the so called data mining. In this case, it is discovering implicit syndrome granule in non-syndrome relational data.

References

[1] J. Anderson, "Theory, models and representation in medicine," *Journal of Fuzzy Logic and Intelligence Systems*, vol.3, no.1, pp. 5-14, 1993.

[2] G. Marshall, *Advanced students' guide to expert systems,* Oxford, U.K.: Heinemann Professional Publishing Ltd., 1990.

[3] L.A. Zadeh, "Fuzzy Sets and Information Granularity", In M.M. Gupta, R. Ragade and R.R. Yager, editors, *Advances in Fuzzy Set Theory and Applications*, North-Holland, pp. 3-18, 1979.

[4] C.S. Noe, *Using BK-products of fuzzy relations for management of granular structures in relational architectures for knowledge-based systems,* Ph.D Dissertation, Florida State University, Department of Computer Science, 1997.

[5] J. Anderson, W. Bandler, L.J. Kohout, and C. Trayner, "The Design of a Fuzzy Medical Expert System", In M.M. Gupta, A. Kandel, and W. Bandler, editors, *Approximate Reasoning in Expert Systems*, North-Holland, pp. 689-703, 1985.

[6] L.J. Kohout and W. Bandler, "Use of Fuzzy Relations in Knowledge Representation, Acquisition, and Processing," In L.A. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, John Wiley and Sons, pp. 415-435, 1992.

[7] L.J. Kohout, I. Stabile, W. Bandler, and J. Anderson, "CLINAID: Medical Knowledge-based System Based on Fuzzy Relational Structures," In M. Cohen and D. Hudson, editors, *Comparative Approaches in Medical Reasoning*, World Scientific, pp. 1-25, 1995.

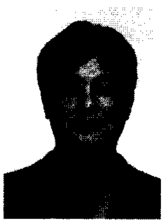
[8] L.J. Kohout, J. Anderson, W. Bandler, S. Gao, and C. Trayner, "An Overview of CLINAID," In L.J. Kohout, J. Anderson, and W. Bandler, editors, *Knowledge-Based Systems for Multiple Environments*, Ashgate Publishing, pp. 173-201, 1992.

[9] L.A. Zadeh, "Similarity relations and fuzzy orderings," *Information Sciences*, vol.1 no.3, pp. 177-200, 1971.

[10] W. Bandler and L.J. Kohout, "Semantics of implication operators and fuzzy relational products," *Internat. Journal of Man-Machine Studies*, vol. 12, 89-116, 1980.

[11] W. Bandler and L.J. Kohout, "Fuzzy Implication Operators," In M.G. Singh, editor, *Systems and Control Encyclopedia*, Pergamon Press, pp. 1806-1810, 1987.

저 자 소 개



노찬숙(Chan-Sook Noe)

1987년: 북 일리노이 주립대학 전산학과 (학사)
1991년: 플로리다 주립대학 전산학과(석사)
1997년: 플로리다 주립대학 전산학과(박사)
1998년~ 현재: 우송대학교 컴퓨터전자정보공학부 조교수

관심분야: 퍼지 이론, 뉴로-퍼지 및 소프트 컴퓨팅, 지식기반 시스템, 지능형 에이전트

Phone : 042-630-9715
E-mail : csnoe@woosong.ac.kr