J Electr Eng Technol.2016; 11(1): 247-256
http://dx.doi.org/10.5370/JEET.2016.11.1.247

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Context-Awareness Healthcare for Disease Reasoning Based on Fuzzy Logic

Byung-Kwan Lee*, Eun-Hee Jeong† and Sang-Sik Lee**

Abstract – This paper proposes Context-Awareness Healthcare for Disease Reasoning based on Fuzzy Logic. It consists of a Fuzzy-based Context-Awareness Module (FCAM) and a Fuzzy-based Disease Reasoning Module (FDRM). The FCAM computes a Correlation coefficient and Support between a Condition attribute and a Decision attribute and generates Fuzzy rules by using just the Condition attribute whose Correlation coefficient and Support are high. According to the result of accuracy experiment using a SIPINA mining tool, those generated by Fuzzy Rule based on Correlation coefficient and Support (FRCS) and Improved C4.5 are 0.84 and 0.81 each average. That is, compared to the Improved C4.5, the FRCS reduces the number of generated rules, and improves the accuracy of rules. In addition, the FDRM can not only reason a patient's disease accurately by using the generated Fuzzy Rules and the patient disease information but also prevent a patient's disease beforehand.

Keywords: Fuzzy-based context-awareness module(FCAM), Fuzzy-based disease reasoning module (FDRM), Fuzzy set, Fuzzy rules, Context-awareness information, Correlation coefficient, Support

1. Introduction

IT technologies such as Internet, Mobile, and Ubiquitous are now changing the paradigm of medical care from provider-oriented medicine to user-oriented medicine, from treatment medicine to preventive medicine, and from disease-oriented medicine to wellbeing-oriented medicine. Thanks to them, medical customers can be diagnosed at an early stage of disease with mobile devices that are connected to computers at any time or anywhere. For this, health care providers are doing a lot of researches on individually-customized medical service to provide their customers with the best medical service.

In particular, Data Mining Technique was used to analyze the medical information of a patient's systemic lupus erythematosus in the RX project of Stanford University in the 1980's [1]. Since then, Data Mining technique using machine learning has been applied to many fields of medical diagnosis such as thyrotoxic myopathy [2-4], rheumarthritis [5], cardiovascular diseases [6], Neuropsychosis [7], etc [8].

This paper proposes Context-Awareness Healthcare for Disease Reasoning based on Fuzzy Logic. It consists of the Fuzzy-based Context-Awareness Module (FCAM) and the Fuzzy-based Disease Reasoning Module (FDRM).

The FCAM fuzzifies disease information, manages a patient's health condition by collecting the context-awareness information such as a patient's physiological information, medical record, family history in real time, and generates the fuzzy rules on a disease. The FDRM reasons a disease with the Fuzzy rules and a patient's Context-Awareness information collected in the FCAM accurately and prevents a patient from an unexpected disease beforehand.

The remainder of this paper is organized as follows. Section 2 discusses the related works on a Fuzzy Logic System and C4.5. Section 3 shows the Context-Awareness Module and Disease Reasoning based on Fuzzy Logic. Section 4 shows the estimation result and finally in section 5, we describe our conclusions.

† Corresponding Author: Dept. of Regional Economics, Kangwon National University, Korea. (jeongeh@kangwon.ac.kr)
* Dept. of Computer Engineering, Catholic Kwandong University, Korea. (bklee@ cku.ac.kr)
** Dept. of Biomedical Engineering, Catholic Kwandong University, Korea. (lsskyj@cku.ac.kr)
Received: June 23, 2015; Accepted: November 3, 2015

2. Related works

2.1 Fuzzy logic system

"Fuzzy Logic" was introduced with the 1965 proposal of fuzzy set theory by Lotfi A. Zadeh [9]. The Fuzzy Logic has been applied to many fields, from control theory to artificial intelligence.

Specially, a Fuzzy Logic System can be defined as the nonlinear mapping of an input data set to a scalar output

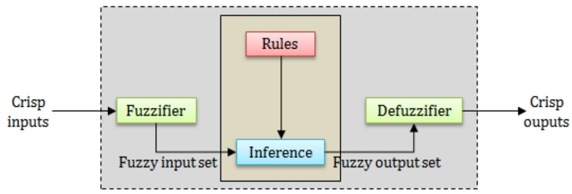


Fig. 1. A Fuzzy logic system

data [10, 11]. A Fuzzy Logic System consists of four main parts: fuzzifier, rules, inference engine, and defuzzifier. The general architecture of a Fuzzy Logic System is shown in Fig. 1.

The process of a Fuzzy Logic system is explained in Algorithm 1[9].

Algorithm 1. fuzzy logic algorithm

1. Define the linguistic variables and terms (initialization)
2. Construct the membership functions (initialization)
3. Construct the rule base (initialization)
4. Convert crisp input data to fuzzy values using the membership functions (fuzzification)
5. Evaluate the rules in the rule base (inference)
6. Combine the results of each rule (inference)
7. Convert the output data to non-fuzzy values (defuzzification)

2.1.1 Linguistic variables

Linguistic variables are the input or output variables of the Fuzzy Logic System whose values are words or sentences from natural language, instead of numerical values. For example in the Fuzzy Logic System the linguistic variables are expressed as follows according to a patient's temperature.

temperature(t)={hypothermia, normothermia,
mild fever, high fever}

2.1.2 Membership function

Membership functions are used in the fuzzification and defuzzification steps of the Fuzzy Logic System. Membership functions can have several different shapes such as Fig. 2 [9-11].

The most commonly used shapes are triangular, trapezoidal, Gaussian and bell shaped membership function.

This paper uses triangular, trapezoidal, and singleton membership function for a patient's disease reasoning.

2.1.3 Fuzzification

A Fuzzy Logic system uses linguistic variables instead

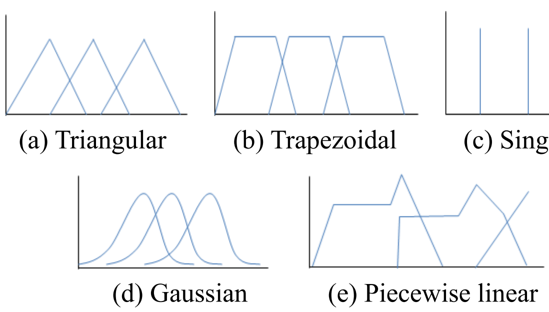


Fig. 2. Membership functions

of numerical variables. The process of converting a numerical variable (real number or crisp variable) to a linguistic variable (fuzzy number) is called fuzzification. The simplest form of membership function is triangular membership function.

2.1.4 Defuzzification

The reverse fuzzification is called defuzzification. The use of a Fuzzy Logic System inference engine produces the output in a linguistic form. According to real world requirements, the linguistic variables have to be transformed to crisp output. Weighted average method is the best well-known defuzzification method for Sugeno type fuzzy controller [12, 13].

This paper produces a patient's disease reasoning result in this weighted average method.

2.2 Context-awareness

Context-awareness is defined by Brown [14], Ryan [15], and Dey [16].

Brown et al. [14] defines context as location, identities of the people around a user, the time of day, season, temperature, etc. Ryan [15] defines context as a user's location, environment, identity and time. Dey [16] enumerates context as the user's emotional state, focus of attention, location and orientation, date and time, objects, and people in a user's environment [17].

This paper defines Context-Awareness as a patient's physiological information, personal information, medical record, and family history.

In addition, it provides a patient's disease accurately by combining Fuzzy Theory such as Fuzzy Set, Fuzzy Rule, etc with the Context-Awareness information.

2.3 C4.5 algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan [18, 19]. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier [20].

In general, the steps in C4.5 algorithm to build decision tree are [21, 22]:

- Choose attribute for root node
- Create branch for each value of that attribute
- Split case according to branches
- Repeat process for each branch until all cases in the branch have the same class.

Choosing which attribute to be a root is based on highest gain of each attribute. To count the gain, formula (1) is used to count the gain [21, 22].

$$\mathrm{Gain}(S, A) = \mathrm{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times \mathrm{Entropy}(S_i) \qquad (1)$$

Here, $\{S_1, \ldots, S_i, \ldots, S_n\}$ the partitions of S according to values of attribute A.

n is the number of attributes A.

$|S_i|$ is the number of cases in the partition S_i.

$|S|$ is the total number of cases in S.

The entropy is gotten by formula (2) [21, 22].

$$\mathrm{Entropy}(S) = \sum_{i=1}^{n} -P_i \times \log_2 p_i \qquad (2)$$

Here, S is case set, and n is the number of cases in the partitions S. P_i is the Proportion of S_i to S.

3. Context-Awareness Healthcare for Disease Reasoning based on Fuzzy Logic

The "Context-Awareness Healthcare for Disease Reasoning based on Fuzzy Logic" proposed in this paper consists of a Fuzzy-based Context-Awareness Module (FCAM) and a Fuzzy-based Disease Reasoning Module (FDRM). Its total structure and data flow are shown in Fig. 3.

3.1 Overview

The Fuzzy-based Context-Awareness Module (FCAM) fuzzifies all the disease information beforehand, generates Fuzzy Rules with the disease information and collects the Context-Awareness information such as physiological information, medical records, personal information, and family history to check a patient's current health condition. Then, it judges a patient's disease after comparing the fuzzified disease information with a patient's Context-Awareness information.

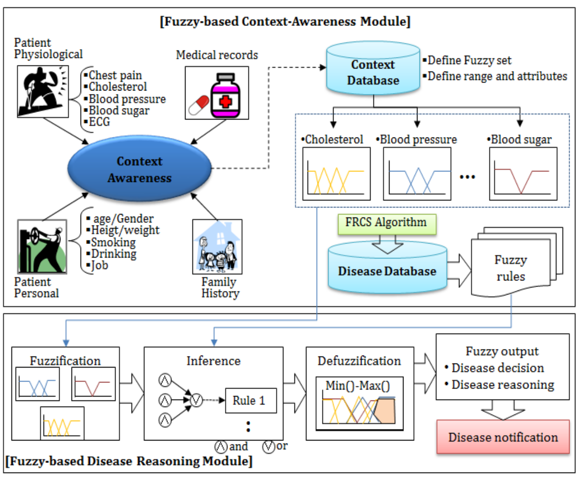


Fig. 3. The total structure and data flow

Awareness information.

In particular, when it generates Fuzzy Rules, it not only decreases the number of conditions within a rule by excluding the condition attributes whose Correlation coefficient and Support is the lowest about elements but also improves the accuracy of disease reasoning with the better accuracy of rules.

The Fuzzy-based Disease Reasoning Module (FDRM) reasons a patient's disease accurately by using the Fuzzy Rules and the Context-Awareness information collected in the FCAM and then prevents a patient from an unexpected disease by informing them of the result.

3.2 Fuzzy-based context-awareness module (FCAM) design

The FCAM collects a patient's Context-Awareness information and classifies it into the following 4 categories.

- A patient's physiological information such as chest pain, cholesterol level, blood sugar, blood pressure, ECG, etc
- A patient's medical records
- A patient's personal information such as age, weight, height, smoking, drinking, job, etc.
- A patient's family history

Table 1. Fuzzy set, range, and linguistic variables

Fuzzy set	Range	linguistic variables
Chest pain	1 2 3 4	typical angina atypical angina non angina asymptomatic
Cholesterol	<197 188-250 217-307 >281	low medium high very high
Blood pressure	<134 124-153 142-172 >154	low medium high very high
Blood sugar	<120 >=120	no yes
ECG (ST_depression)	<0.4 0.4-1.8 >1.8	normal abnormal hypertrophy
Thallium	3 6 7	Normal Fixed Defect Reversible Defect
Age	<38 33-45 40-58 >52	young mid old very old
Gender	1 0	male female
smoking year	<=30 >30	low high
drinking	0 1	no yes
Family history (diabetes, hypertension, ...)	<1 >=1	no yes
Medical records (diabetes, ...)	<1 >=1	no yes

Besides, the FCAM defines Fuzzy Set, data range and linguistic variables about each element with the collected Context-Awareness information in Table 1 [23, 24].

Fig. 4 shows the membership function and fuzzification of cholesterol [24].

Fig. 5 shows the procedure in which the FCAM is processed.

To begin with, the FCAM collects a patient's Context-Awareness information in real time, judges the patient's health condition by using Table1, and informs the patient of the result. At this time, if the patient's health condition is abnormal, the FCAM transfers the patient's family history and medical records to the FDRM so that the FDRM can reason the patient's disease.

For example, Bob has blood pressure 150 and both of his parents suffer from high cholesterol and diabetes. When the FCAM receives such cholesterol information about Bob, it judges his health condition as "abnormal" because the "linguistic variable in Table1 is "high". Therefore, because his health condition is abnormal, the FCAM transfers a "attention: message" to him and the information on the his medical records and family history to the FDRM.

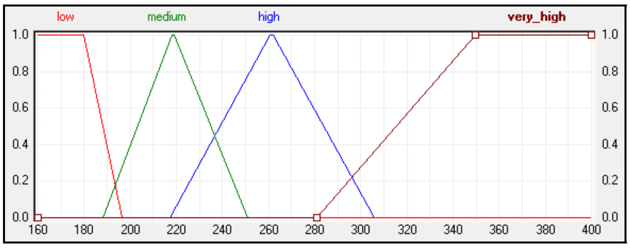
$$\mu_{low}(x) = \begin{cases} 1 & x < 151 \\ \dfrac{197 - x}{46} & 151 < x < 197 \end{cases}$$

$$\mu_{medium}(x) = \begin{cases} \dfrac{(x - 188)}{27} & 188 < x < 215 \\ 1 & x = 215 \\ \dfrac{250 - x}{35} & 215 < x < 250 \end{cases}$$

$$\mu_{high}(x) = \begin{cases} \dfrac{(x - 217)}{46} & 217 < x < 263 \\ 1 & x = 263 \\ \dfrac{307 - x}{44} & 263 < x < 307 \end{cases}$$

$$\mu_{veryhigh}(x) = \begin{cases} \dfrac{(x - 281)}{66} & 281 < x < 347 \\ 1 & x \geq 347 \end{cases}$$

(a) Membership function



(b) fuzzification

Fig. 4. Fuzzification and membership function

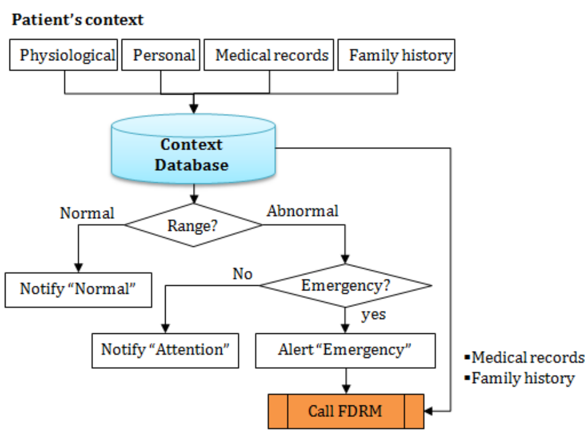


Fig. 5. The procedure of FCAM

3.2.1 Fuzzy rule generation

The FRCS (Fuzzy Rule based on Correlation coefficient and Support) algorithm proposed in this paper generates Fuzzy Rules by using the disease information stored in the database. At this time, the FRCS algorithm computes the Correlation coefficient and Support about a Decision Attribute and a Condition Attribute. Then, because the FRCS selects the Condition Attribute which has the highest Correlation coefficient and Support, it not only reduces the number of conditions within rules, but also improves the accuracy of rules.

Algorithm 2 shows the flow of the FRCS algorithm.

Algorithm 2. FRCS algorithm

---

1. Input disease information.
2. Calculate the Correlation Coefficient (CC) between a Condition attributes and a Decision attribute.
   CC=Correlation(a Condition.attribute, a Decision.attribute)
3. A Condition attribute with the highest correlation coefficient becomes a condition rule.
4. Calculate the support of a Decision attribute value.
   Support =

$$\begin{cases} \dfrac{n(y)}{n(y)+n(n)}, & \text{if a Decision attribute value is y} \\ \dfrac{n(n)}{n(y)+n(n)}, & \text{if a Decision attribute value is n} \end{cases}$$

5. A Condition attribute with the highest support becomes a condition rule.

6.if min(support)>0.35) then
 Go to 2.
 else
 Make fuzzy rules of attributes
 Save fuzzy rules in database
 end if

The following example shows that the FCAM generates Fuzzy Rules by using the FRCS algorithm and Heart Disease [24]. Table 2 shows the attributes of Heart Disease data and Fig. 6 shows its experimental data.

Table 2. The attributes of Heart Disease and statistical values

Field Name	Description	Statistical values: Average	Statistical values: Standard deviation
a1	Age	54	9.1
a2	Gender(1=male, 0=female)	-	-
a3	chest pain type	3.2	0.9
a4	blood pressure	131	17.8
a5	serum cholesterol in mg/dL	250	51.5
a6	fasting blood sugar (0=false, 1=true)	-	-
a7	ECG results	1	1
a8	maximum heart rate	150	23.1
a9	exercise induced angina (0=no, 1=yes)	-	-
a10	ST depression	1.05	1.14
a11	the slope of ST segment	1.59	0.61
a12	number of major vessels	0.67	0.94
a13	Thallium	4.7	1.93
a14	Class(n=absent, y=present)	-	-

Condition attributes (a1–a13) / decision attributes (a14)

a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14
70	1	4	130	322	0	2	109	0	2.4	2	3	3	y
67	0	3	115	564	0	2	160	0	1.6	2	0	7	n
57	1	2	124	261	0	0	141	0	0.3	1	0	7	y
64	1	4	128	263	0	0	105	1	0.2	2	1	7	n
74	0	2	120	269	0	2	121	1	0.2	1	1	3	n
65	1	4	120	177	0	0	140	0	0.4	1	0	7	n
56	1	3	130	256	1	2	142	1	0.6	2	1	6	y
59	1	4	110	239	0	2	142	1	1.2	2	1	7	y

Fig. 6. Experimental data (information table)

(1) The attribute selection of Fuzzy Rule

The FCAM selects Fuzzy rule's attributes by computing the Correlation coefficient and Support of each attribute with the FRCS algorithm.

[The 1st step] The FCAM classifies the value of a Decision attribute (a14) into no or yes and computes a Correlation coefficient and Support about each attribute.

At this time, the FCAM computes the Correlation Coefficient and Support by using the expression (3) and (4). Table 3 shows the result of the Correlation Coefficient and Support.

$$CC = \text{Correlation(CA, DA)} \tag{3}$$

$$S = \begin{cases} \dfrac{n(y)}{n(y)+n(n)}, & \text{if a Decision attribute value is y} \\ \dfrac{n(n)}{n(y)+n(n)}, & \text{if a Decision attribute value is n} \end{cases} \tag{4}$$

Here, CC is Correlation Coefficient, S Support, CA Condition Attribute, DA Decision Attribute.

For example, the number of data which is a1<60 among 270 experimental data is 186. The number of data whose

Table 3. The Conditional support of each attribute

Attribute	a1		a2		a3		a4	
Correlation	0.212		0.298		0.417		0.155	
Condition	<60		=0		<=3		<120	
Y	72	0.39	20	0.23	29	0.21	20	0.35
N	114	0.61	67	0.77	112	0.79	37	0.65
Condition	>60		=1		>3		>=120	
Y	48	0.57	100	0.55	91	0.71	100	0.47
N	36	0.43	83	0.45	38	0.29	113	0.53
Attribute	a5		a6		a7		a8	
Correlation	0.118		-0.016		0.182		-0.419	
Condition	<210		=0		<2		<140	
Y	19	0.33	103	0.45	47	0.35	55	0.71
N	39	0.67	127	0.55	86	0.65	22	0.29
Condition	>=210		=1		>=2		>=140	
Y	101	0.48	17	0.43	73	0.53	65	0.34
N	111	0.52	23	0.58	64	0.47	128	0.66
Attribute	a9		a10		a11		a12	
Correlation	0.419		0.418		0.338		0.455	
Condition	=0		<1.8		<=1		<1	
Y	54	0.30	67	0.33	32	0.25	40	0.25
N	127	0.70	136	0.67	98	0.75	120	0.75
Condition	=1		>=1.8		>1		>=1	
Y	66	0.74	53	0.79	88	0.63	80	0.73
N	23	0.26	14	0.21	52	0.37	30	0.27
Attribute	a13							
Correlation	0.525							
Condition	<=4							
Y	33	0.22						
N	119	0.78						
Condition	>4							
Y	87	0.74						
N	31	0.26						

Decision attribute (a14) is "YES" and "NO" among the 186 is 72 and 114 each. At this time the Support of "YES" by using expression (4) is 72/(72+114) = 0.39.

[The 2nd step] The FCAM selects as a Condition attribute of Fuzzy rules the attribute whose correlation coefficient and support is the highest. For example, because a13 meets this condition, it is selected as the condition of Fuzzy rules.

With these selected attributes, Correlation coefficient and Support are computed in the generation step of Fuzzy Rules repeatedly and only the attributes whose Correlation coefficient and Support are high are used as a Fuzzy Rule in the FCAM.

(2) Fuzzy Rule Generation

The FCAM computes the Correlation coefficient and Support about a Condition attribute with the selected attributes and decides the order of rule conditions about each Fuzzy Rules.

The FCAM generates the condition of Fuzzy rules with the attributes the FRCS algorithm is satisfied with.

[The 1st step] The FCAM selects as the condition of Fuzzy rules the attributes whose Correlation coefficient is the highest. That is, because a13 meets this condition in the Table 3, it is selected as the condition of Fuzzy rules.

Table 4. The result of correlation coefficient and support

Attribute	a1		a2		a3		a4	
Correlation	0.284		0.226		0.279		0.062	
Condition	<60		=0		<=3		<120	
y	15	0.14	9	0.12	10	0.10	10	0.27
n	91	0.86	65	0.88	91	0.90	27	0.73
Condition	>=60		=1		>3		>=120	
y	18	0.39	24	0.31	23	0.45	23	0.20
n	28	0.61	54	0.69	28	0.55	92	0.80
Attribute	a5		a6		a7		a8	
Correlation	0.136		-0.016		0.186		-0.388	
Condition	<210		=0		<2		<140	
y	5	0.15	29	0.22	11	0.14	14	0.47
n	28	0.85	103	0.78	65	0.86	16	0.53
Condition	>=210		=1		>=2		>=140	
y	28	0.24	4	0.20	22	0.29	19	0.16
n	91	0.76	16	0.80	54	0.71	103	0.84
Attribute	a9		a10		a11		a12	
Correlation	0.300		0.314		0.244		0.429	
Condition	=0		<1.8		<=1		<=1	
y	19	0.16	21	0.16	12	0.13	12	0.11
n	103	0.84	110	0.84	82	0.87	95	0.89
Condition	=1		>=1.8		>1		>1	
y	14	0.47	12	0.57	21	0.36	21	0.47
n	16	0.53	9	0.43	37	0.64	24	0.53
Attribute	a13							
Correlation	1							
Condition	<=4							
y	33	0.22						
n	119	0.78						
Condition	>4							
y	0	0						
n	0	0						

[The 2nd step] The FCAM arranges the experimental data on the basis of the attributes selected in the 1st step and computes the Support about each attribute with the expression (4).

For example, the FCAM extracts just the data whose attributes are a13<4. Consequently, the FCAM extracts 152 data and computes the Correlation Coefficient and Support with expression (3) and (4). The Table 4 shows the result of the 2nd step.

[The 3rd step] The FCAM selects as Fuzzy rules the attributes whose Support in the 2nd step is highest. That is, the FCAM selects as the condition of Fuzzy rules the a3 whose Support is 0.90.

Fig. 7 shows an example of the Fuzzy Rules generated by the FCAM. Its Fuzzy Rules in Fig. 7 can have several rule conditions and each rule condition can be called the 1st rule condition, the 2nd rule condition, etc according to each location.

[The 4th step] The FCAM arranges the experimental data on the basis of the a3 selected in the 3rd step and decides the conditions of Fuzzy rules by repeating the 1st step ~the 3rd step. At this time, the FCAM selects the conditions of Fuzzy whose Support is less than 0.35.

[The 5th step] After the FCAM selects the rule condition of each Fuzzy Rule in this way, it generates a Fuzzy Rule and stores it in the database. Because the Fuzzy Rule generated by the FRCS algorithm is selected by computing the Conditional Support among each attribute, their accuracy is greater than that generated by the Improved C4.5 algorithm.

The rules generated by the Improved C4.5 algorithm and the FRCS algorithm are summarized in Table 5.

Table 5. The rules of the improved C4.5 and the FRCS algorithm

Improved C4.5	if(a10>=1.7) then a14=y
	if(a10<1.7 and a13<4.5) then a14=n
	if(a10<1.7 and a13>=4.5 and a5<211.5) then a14=n
	if(a10<1.7 and a13>=4.5 and a5>=211.5 and a4>=122) then a14=y
	if(a10<1.7 and a13>=4.5 and a5<211.5 and a4<122 and a8<145.5) then a14=y
	if(a10<1.7 and a13>=4.5 and a5<211.5 and a4<122 and a8>=145.5 and a3<3.5) then a14=n
	if(a10<1.7 and a13>=4.5 and a5<211.5 and a4<122 and a8>=145.5 and a3>=3.5) then a14=y
FRCS	if (a13<=4 and a3<=3) then a14=n
	if (a13<=4 and a3>3 and a12<1) then a14=n
	if (a13<=4 and a3>3 and a12>=1) then a14=y
	if (a13>4 and a12>=1) then a14=y
	if (a13>4 and a12<1 and a9=0) then a14=n
	if (a13>4 and a12<1 and a9=1) then a14=y

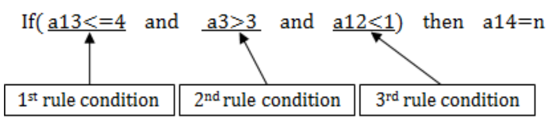


Fig. 7. The example of fuzzy rules

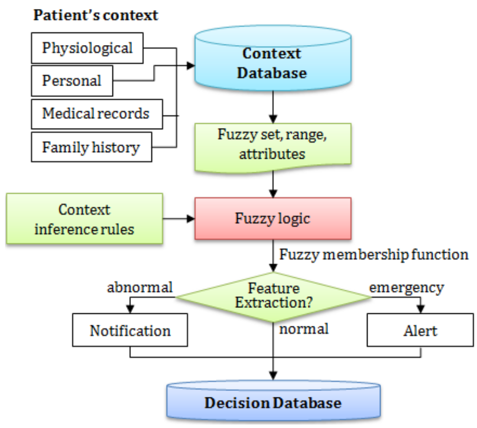


Fig. 8. The flowchart of the FDRM

3.3 Fuzzy-based disease reasoning module (FDRM) design

The Fuzzy-based Disease Reasoning Module (FDRM) reasons a patient's disease by using the Fuzzy Set, Context-Awareness and Fuzzy Rules. Fig. 8 shows the total flowchart of the FDRM.

[The 1st step] The FDRM applies the Fuzzy Set, range, linguistic variables and the Fuzzy Rules to Fuzzy Logic.

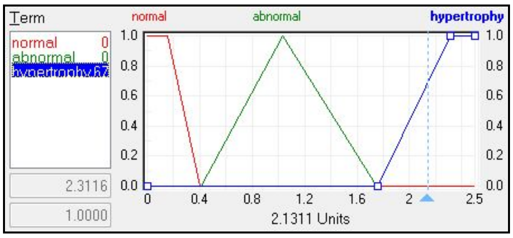
[The 2nd step] The Fuzzy Logic generates a membership

Table 6. Defuzzification

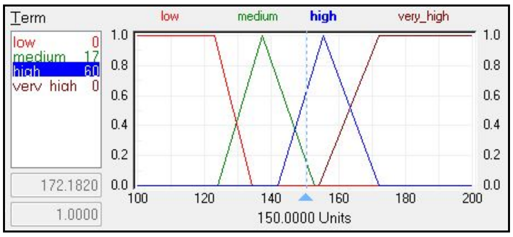
Operation	Formulae
Right most maximum	U=sup(u'), μ(u)=sup(μ(u))
Left most minimum	U=inf(u'), μ(u')= sup(μ(u))
Center of Area(Gravity)	$U = \dfrac{\int_{min}^{max} u \cdot \mu(u)du}{\int_{min}^{max} \mu(u)du}$

IF (ST_depression is hypertrophy) and (Blood pressure is high) then heart_disease is present)

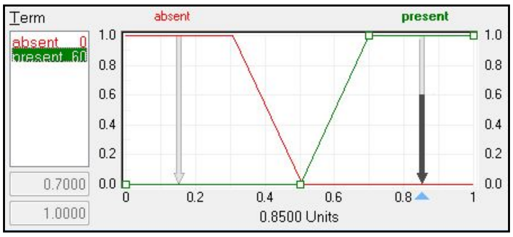
Fig. 9. Fuzzy Rule (ST depression and Blood pressure)



(a) ST depression=2.1311, max=67



(b)Blood pressure=150, max=60



(c) output, min=60, Gravity=0.8500

Fig. 10. The result after processing FDRM

function with Fuzzy set, range, and linguistic variables.

[The 3rd step] The FDRM assigns a patient's Context-Awareness information to the membership function and converts the result to Fuzzy value.

For example, in case the ST depression and blood pressure of Bob is 2.1311 and 150 each, he has the Fuzzy value of Fig. 10(a) and (b).

[The 4th step] The FDRM estimates the Fuzzy Rule of Fig. 9 and prints out the result value like Fig. 10(c).

[The 5th step] The FDRM computes a disease reasoning value about a patient's Context-Awareness information by using the defuzzification formulae [12] of Table 6.

[The 6th step] The FDRM notifies patients of the health information according to the computed result value. If the patient's health condition is abnormal, it prevents a patient from an unexpected disease beforehand by transferring the information to him rapidly.

4. Analysis

The FCAM and FDRM proposed in this paper are estimated based on the accuracy of Fuzzy Rules and Disease Reasoning by using the experimental data of Heart Disease [25].

4.1 The accuracy estimation for fuzzy rules

The Heart Disease consists of 14 attributes and the explanation about each attribute is shown in the Table 2. This experimental data is made up of total 270 data and the Fuzzy Rules are generated with the 270.

Table 7. The comparison of the FRCS and the Improved C4.5 algorithm

(a) The number of rule and rule attributes

Data Set	The number of rule Attributes		The number of rules	
	C4.5	FRCS	C4.5	FRCS
Training data	6	4	7	6

(b) The comparison of accuracy

Data Set	Tuples	Accuracy	
		C4.5	FRCS
Training data	270	0.82	0.86
Testing data	50	0.82	0.82
	100	0.79	0.83
	150	0.81	0.85
	200	0.80	0.85
	250	0.83	0.86
	Average	0.81	0.84

Table 8. Fuzzy reasoning rules

Rule 1	if(Thallium is low) and (Chest_pain is typical_angina) then (Heart_disease is absent)
Rule 2	if(Thallium is low) and (Chest_pain is atypical_angina) then (Heart_disease is absent)
Rule 3	if(Thallium is low) and (Chest_pain is non_angina) then (Heart_disease is absent)
Rule 4	if(Thallium is low) and (Chest_pain is asymptomatic) and (Vessel is low) then (Heart_disease is absent)
Rule 5	if(Thallium is low) and (Chest_pain is asymptomatic) and (Vessel is high) then (Heart_disease is present_2)
Rule 6	if(Thallium is high) and (vessel is low) and (angina is no) then (Heart_disease is absent)
Rule 7	if(Thallium is high) and (vessel is low) and (angina is yes) then (Heart_disease is present_2)
Rule 8	if(Thallium is high) and (vessel is low) then (Heart_disease is present_2)
Rule 9	if(family_history is true) then (Heart_disease is present_1)
Rule 10	if(medical_record is true) then (Heart_disease is present_1)
Rule 11	if(family_history is ture) and (medical_record is ture) then (Heart_disease is present_1)

Consequently, as shown in the Table 7, the number of the decision rules generated by the FRCS and the Improved C4.5 algorithm each is 6 and 7. This paper estimates the accuracy in comparison with the rules generated by the Improved C4.5 and the FRCS algorithm using SIPINA mining tool [26]. The Table 7 shows the estimated result. The number of the attributes used in the FRCS is smaller than that of the Improved C4.5, but the FRCS algorithm is more accurate. Therefore, it is concluded that the FRCS is more excellent than the Improved C4.5.

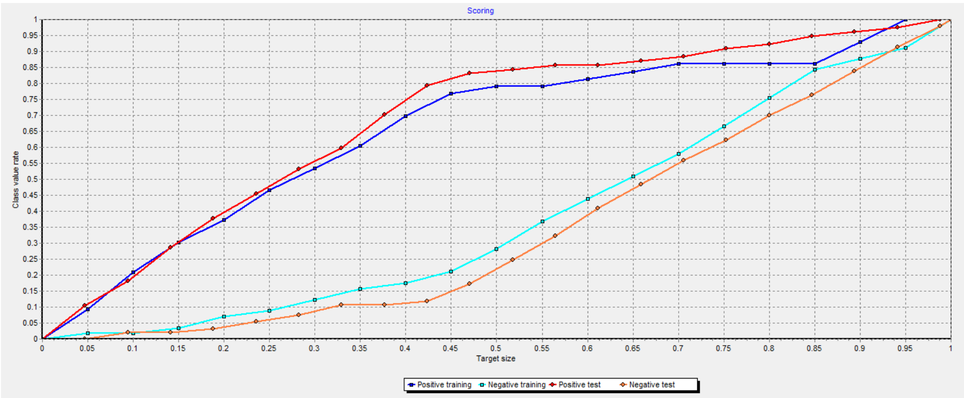
Fig. 11 shows the rule accuracy between Improved C4.5 and FRCS. The accuracy of rules is better if the gap in the graph is narrower. Therefore, Fig. 11 (b) shows that the rule accuracy of FRCS is the higher.

4.2 The accuracy estimation for disease reasoning

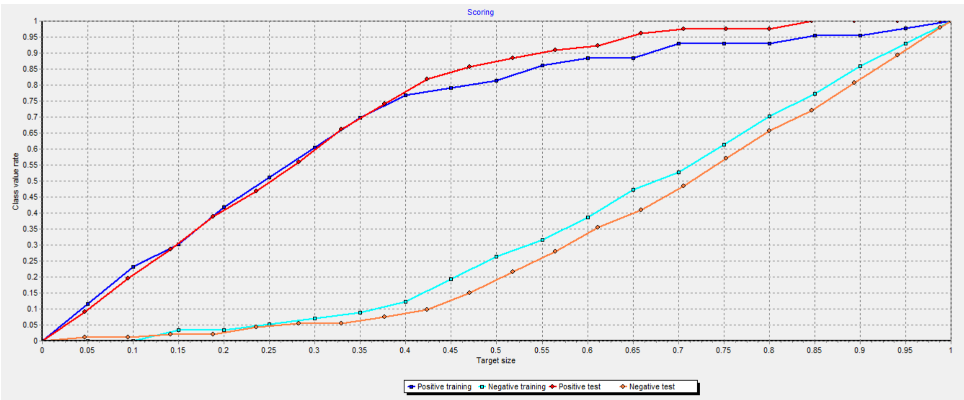
Table 8 shows that the rules generated by the FRCR

Table 9. Fuzzy set of heart disease

Output	Range
Heart_disease	<0.5 0.4-0.6 >0.5



(a) Improved C4.5



(b) FRCS

Fig. 11. The comparison of their accuracy

Table 10. Experimental data and reasoning result

no	Thallium	Chest pain	Vessels	Angina	Family history	Medical Record	Heart Disease	Inference results	Success /Fail
1	3	4	0	0	0.3	0	N	0.29	S (absent)
2	3	1	0	0	0	0	N	0.15	S (absent)
3	7	4	0	1	0	0.5	Y	0.68	S (present_2)
4	3	3	0	0	0	0.3	N	0.29	S (absent)
5	3	2	0	0	0.5	0	N	0.38	S (absent)
6	7	3	0	1	0	0	Y	0.83	S (present_2)
7	3	4	0	0	0	0	Y	0.15	F (absent)
8	7	1	0	1	0.3	0	N	0.80	F (present_2)
9	3	4	2	0	0	0	N	0.37	S (absent)
10	3	3	0	0	0.7	0.3	Y	0.47	S (present_1)

*a: absent, p: present

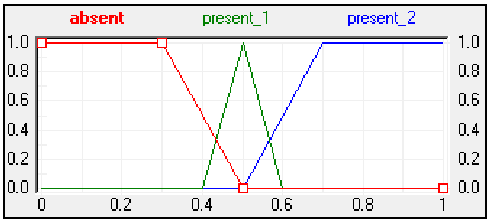


Fig. 12. Fuzzification of Heart_disease

Fig. 13. The accuracy estimation according to the number of experimental data

algorithm were converted to the linguistic variables of Fuzzy Set. The rules made out of family history and medical records were added to the Table 8.

Table 9 shows the Fuzzy Set on heart disease as a result of a disease reasoning rule, and Fig. 12 shows the fuzzification of heart disease.

The Accuracy of Disease Reasoning was estimated with the randomly extracted 10 values of Heart Disease data.

The result shows that the accuracy of Disease Reasoning in the FDRM is 80%. The accuracy estimation is done by applying the randomly selected values of experimental data [25] to the Fuzzy rules of Table 8. The accuracy is computed by using expression (5)

$$\text{Accuracy}(\%) = \frac{\text{the number of data}}{\text{the number of experimental data}} \times 100 \quad (5)$$

where, the number of data means the result which is true in comparison.

Therefore, when the number of experimental data is 200, the accuracy of Disease Reasoning in the FDRM is 86%.

As shown in Fig. 13, as the number of experimental data gets increased, the accuracy of Reasoning gets better.

5. Conclusion

This paper proposes Context-Awareness Healthcare for Disease Reasoning based on Fuzzy Logic. It consists of the Fuzzy-based Context-Awareness Module (FCAM) and the Fuzzy-based Disease Reasoning Module (FDRM).

The FCAM collects the Context-Awareness information such as physiological information, medical records, family history, etc and manages a patient's disease and the FDRM prevents a patient's disease by reasoning his disease with the information of the FCAM.

The FCAM and the FDRM have the following characteristics.

First, the FCAM manages patient's diseases by rapidly notifying a patient of his health condition after fuzzifying his collected Context-Awareness information.

Second, the FCAM computes the Correlation coefficient and Support between Condition attribute and Decision attribute. The FRCR algorithm generates Fuzzy Rules by using the attributes whose Correlation coefficient and Support are high and strengthens the relation between attributes, contrary to the Improved C4.5 algorithm. Therefore, the FRCR algorithm decreases the number of attributes used for Fuzzy Rules more than the Improved C4.5 algorithm and improves the accuracy of rules more.

Third, the FDRM reasons a patient's disease by using disease information, Fuzzy Set, Fuzzy Rule and predicts a patient's disease by the reasoning result

Acknowledgements

This work was supported by research fund of Catholic Kwandong University (CKURF-201407130001).

References

[1] Blum R. L., "Displaying clinical data from a time-oriented database," *Computers in Biology and Medicine*, vol.11, no.4, pp. 197-210, 1981.

[2] Hojker S. KI, Jauk A., Fidler V. and Porenta M., "Expert system's development in the management of thyroid diseases," *Paper presented at: European Congress for Nuclear Medicine*, Milano, Sep., 1988.

[3] Horn W., "AI in medicine on its way from knowledge-intensive to data-intensive systems," *Artificial Intelligence in Medicine*, vol.23, no.1, pp. 5-12, 2001.

[4] Quinlan R. CP, Horn KA and Lazarus L., "Inductive knowledge acquisition: a case study. In: JR Q.," *Applications of expert systems*. Boston: Addison-Wesley, pp. 137-156, 1987.

[5] Zupan B. and Dzeroski S., "Acquiring background knowledge for machine learning using function decomposition: a case study in rheumatology," *Artificial Intelligence Medicine*, vol.14, no.1-2, pp. 101-117, 1998.

[6] Cohen ME and Hudson DL, "Neural network models for biosignal analysis," *Conference Proceeding IEEE Engineering in Medicine Biology Society*, vol. 1, pp. 3537-3540, 2006.

[7] Chun FK, Karakiewicz PI, Briganti A., Walz J, Kattan MW, Huland H and Graefen M., "A critical appraisal of logistic regression-based nomograms,

artificial neural networks, classification and regression-tree models, lookup tables and risk-group stratification models for prostate cancer," *BJU International*, vol. 99, no. 4, pp. 794-800, 2007.
[8] Lee SM and Park RW, "Basic Concepts and Principles of Data Mining in Clinical Practice," *Journal of Korea Society Medical Informatics*, vol. 15, no. 2, pp. 175-189, 2009.
[9] Zadeh LA, "Fuzzy set," *Information and Control*, vol. 8, pp. 338-353, 1965.
[10] A Short Fuzzy Logic Tutorial, Available athttp:// cs. bilkent.edu.tr/~zeynep/files/short_fuzzy_logic_tutorial.pdf.
[11] Mendel J., "Fuzzy logic systems for engineering: a tutorial," *proceedings of the IEEE*, vol. 83, no. 3, pp. 345-377, 1995.
[12] Defuzzification formular Fuzzy control programming, Technical report, *International Electrotechnical Commision*, 1997.
[13] Nitin K. Dhote† and J. B. Helonde, "Improvement in Transformer Diagnosis by DGA using Fuzzy Logic", *J Electr Eng Technol*, vol. 9, no. 2, pp. 615-621, 2014.
[14] Brown PJ, Bovey JD and Chen X., "Context-Aware Applications: From the Laboratory to the Marketplace," *IEEE Personal Communications*, vol. 4, no. 5, pp. 58-64, 1997.
[15] Ryan N., Pascoe J. and Morse D., "Enhanced Reality Fieldwork: the Context-Aware Archaeological Assistant," Gaffney,V., van Leusen, M., Exxon, S. (eds.), *Computer Applications in Archaeology*, 1997.
[16] Dey AK, "Context-Aware Computing: The CyberDesk Project," *AAAI 1998 Spring Symposium on Intelligent Environments*, Technical Report SS-98-02. pp. 51-54, 1998.
[17] Anind KD and Gregory DA, "Towards a Better Understanding of Context and Context- Awareness," Available atftp://ftp.cc.gatech.edu/pub/gvu/tr/99-22.pdf/1999.
[18] Quinlan JR, "C4.5: Programs for Machine Learning," Morgan Kaufmann Pulishers Inc, San Francisco, 1993.
[19] Quinlan JR, "Learning decision tree classifiers," *ACM Computing Surveys* (*CSUR*), vol. 28, no. 1, pp. 71-72, 1996.
[20] http://en.wikipedia.org/wiki/C4.5_algorithm.
[21] Craw S, "Case Based Reasoning: Lecture 3: CBR Case-Base Indexing," Available at http://www.comp.rug.ac.uk/staff/smc/teaching/cm3016/Lecture-3-cbr-indexing.ppt.
[22] Kusrini and Sri Hartati, "Implementation of C4.5 Algorithm to evaluate the cancellation possibility of new student applications at stmik amikom yogyakarta," *Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung*, Indonesia June 17-19, pp. 623-626, 2007.
[23] Rehana Parvin and Abdolreza Abhari, "Fuzzy database for Heart Disease Diagnosis," available at July.15 2014.
[24] Ali Adeli and Mehdi Neshat, "A Fuzzy Expert System for Heart Disease Diagnosis," *Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 (IMECS 2010*), vol.1, Hong Kong, March 17-19, 2010.
[25] Heart disease, http://archive.ics.uci.edu/ml/machine/learning-databases/statlog/heart/, July. 15, 2014.
[26] SIPINA mining tool, http://eric.univ-lyon2.fr/~ricco/sipina.html , Nov. 10, 2014.

Byung-Kwan Lee He received his B.S. degree from Pusan National University in 1979, the M.S. degree in Computer Science from Chung-Ang University in 1986 and the ph.D. degree in Computer Science from Chung-Ang University in 1990 in Korea. He has been a professor of Computer Science at Catholic Kwandong University in Korea since 1988. He was a visiting professor at Saginaw Valley State university, Michigan, USA during 2000~2001. He is a permanent member of the KISS and KIPS. His current research interests are network security, Wireless Sensor Network Security, Internet of Things and Big Data.

Eun-Hee Jeong She received her B.S. degree from Kangneung National University in 1991, the M.S. degree in Computer Science from Kwandong University in 1998 and the Ph.D. degree in Computer Science from Kwandong University in 2003 in Korea. She has been a professor of department of Regional Economics at Kangwon National University in Korea since 2003, Sept. She is a regular member of the KSII. Her current research interests are Sensor Network, IT security, web programming, and e-commerce security.

Sang-Sik Lee He was born in Sancheong, Korea, in 1968. In 1994 he graduated in Bio-mechatronic engineering at the Sungkyunkwan University, Korea. He received his master degree in Engineering (Mechatronics system) in 1996 and his Ph.D in 2000 from Sungkyunkwan University, Korea. He is currently an assistant professor at the college of Engineering at Catholic Kwandong University in Gangneung, Korea. His main research interests are in the area of Biomedical engineering and u-Health systems.