Journal of Korean Institute of Intelligent Systems, Vol. 23, No. 5, October 2013, pp. 400-405
http://dx.doi.org/10.5391/JKIIS.2013.23.5.400 빅데이터 특집 논문

Knowledge Extractions, Visualizations, and Inference from the big Data in Healthcare and Medical

Jin Sung Kim

School of Business Administration, Jeonju University

Abstract

The purpose of this study is to develop a composite platform for knowledge extractions, visualizations, and inference. Generally, the big data sets were frequently used in the healthcare and medical area. To help the knowledge managers/users working in the field, this study is focused on knowledge management (KM) based onData Mining (DM), Knowledge Distribution Map (KDM), Decision Tree (DT), RDBMS, and SQL-inference. The proposed mechanism is composed of five key processes. Firstly, in *Knowledge Parsing*, it extractslogical rules from a big data set by using DM technology. Then it transforms the rules into RDB tables. Secondly, through *Knowledge Maintenance*, it refinesand manages the knowledge to be ready for the computing of knowledge distributions. Thirdly, in *Knowledge Distribution* process, we can see the knowledge distributions by using theDT mechanism.Fourthly, in *Knowledge Hierarchy*, the platform shows the hierarchy of the knowledge. Finally, in*Inference*, it deduce the conclusions by using the given facts and data.This approach presents the advantages of diversity in knowledge representations and inference to improve the quality of computer-based medical diagnosis.

Key words: Data mining, Decision Tree, Knowledge distribution, Knowledge management, RDB.

1. Introduction

In the area of KM, a lot of researches explored its concept, nature, tools, methodologies, frameworks, architectures, and the real world implementations [1-10]. Recently, however, the focus of the related studies was shifted into the practical approaches finding better ways to manage the organizational knowledge [11-12].

From this view, the medical system could be considered as an emerging market where the human expert knowledge is used as an important organizational resource [6][8]. Basically, the knowledge in the field of healthcare and medical is not strictly factual, but is mainly based on the experience and judgment of the human experts [13]. Therefore, the importance of the human experts' knowledge is growing rapidly. However, the supporting systems and tools were not perceived as an essential part of the clinical information process [14].

As an attempt to support the experts working in the area, this study is focused on the methodologiesto construct the composite knowledge management platform. Especially, it is concentrated on the development of the knowledge extractions, visualizations, and inferences. In the process of working, Data Mining (DM), Decision Tree (DT), RDBMS, and SQL-based knowledge inference were used.

The platform consists of the five main processes updated from Kim's [3] research. Firstly, in *Knowledge Parsing*, it extracts domain rules as the forms of the logical rules from a data setand transform/divide the rules into the RDB tables. Secondly, through *Knowledge Maintenance*, it refines the knowledge to showthe knowledge distributions. Thirdly, in *Knowledge Distribution*, it canshow the knowledge distributions by using thedistribution chart.Fourthly, *Knowledge Hierarchy*, the platform shows the hierarchies of the knowledge in the form of the DT. Finally, in *Inference*, it supports the users to start the logical inference and follow up the process until they can find the final goal or conclusion.

접수일자: 2013년 8월 19일
심사(수정)일자: 2013년 9월 10일
게재확정일자 : 2013년 9월 28일
†Corresponding author

Thisapproach has some advantages. First, it canextract logical decision rules and reduce the number of the rules by using the rule extraction and RDBMS functionalities.The advantages of the logical expressions (rules) are the clear interpretation and concise knowledge base (KB). Second, it can expand the KB easily by using the data maintenance functions of RDBMS. The advantages of the RDBMS-based KM are huge-sized knowledge manipulation and combinations.Third, it can help the knowledge managers and users to confirm the distributions of the knowledge they are using.It would be another advantage of this approach. Fourth, it supports the knowledge managers to confirm the knowledge hierarchies on the platform. It has advantages over text/frame-based models in expressing certain forms of knowledge framework modeling and in providing a visual reasoning [15]. Fifth, the logical rules can be used in the knowledge retrieval, predictions or classifications.It has the advantage of providing accurate information results with a certain level of probability of causation.

Using the platform, the knowledge managers can manage the various knowledge those were extracted from the clinical/medical data set, social networks, former prediction cases, historical data, classification problems, and other decision support problems.

To validate the performance of theplatform, a prototype system tool was implemented on the Windows 7 environment except the DM. The other mechanisms including the Knowledge Distribution Map (KDM), Decision Tree (DT), and Inference mechanisms were newly implemented in this study.In the validation, the dermatology data set was referred from the Machine Learning Data Repository at UC Irvine [16]. The medical diagnosis data was very difficult to predict, and they all share the clinical features of erythema and scaling with very little differences [17-18].

This paper is organized as follows. Section 2 shows the related works. In Section 3, the proposed methodologies would be described in details. The implementation of the platform and its applications with a big data set are presented in section 4. The summary of the study, conclusions, and further research topics are discussed in section 5.

2. Research background

2.1 Medical knowledge acquisition

In order to effectively build a medical knowledge base, knowledge mangers are required to help the physicians to extract and

formalize their tacit knowledge using the formal structures [14]. However it's not the easy work since the knowledge is strongly depended on their experiences and tacit knowledge.

In medical knowledge management, MYCIN [19] was well known as an early knowledge-based system to identify the bacteria causing severe infections. Then the major trend in the area was the development of tools to assist the physicians in solving the specific problems.

More recently, some of medical knowledge management systems focused on the general medical knowledge derived from the terminology and standardization [14]. For example, *LinKBase* [20] is an expansive medical knowledge base that enables cross-mappings from one classification system to another. It is based on the medical natural language processing.

In ontology development, one of the most cited generic-domain knowledge management systems which help the knowledge managers is Protégé [21].

In another view, the knowledge that describes the temporal evolution of patient's diseases could be formalized by the Temporal Behavioral Model (TBM) [22]. In this model, the causal and temporal relations between diseases and their abnormal findings are represented. Especially, the Causal and Temporal Knowledge Acquisition Tool (CATEKAT2) represents medical knowledge by the use of the TBM [14]. The significant advances came from the previous researches in this field could support the clinical decisions effectively.

2.2 Knowledge representation models

There are various knowledge representation languages that include the graph-based approach (e.g. conceptual graphs (CG) and semantic networks) and the frame-based approach (e.g. frame and description logics (DLs) [15].

First, the frame-based knowledge representations can represent the knowledge using an object-like structure with attached properties. The semantics of the frames are not entirely formalized, whereas the fully defined set-theoretic semantics of DLs can support the specialized deductive services.

Second, in the graph-based approaches, the knowledge can be represented as labeled direct graphs. Where the nodes denote conceptual entities and arcs indicate the relationships among them.

Especially, the graph-based approach has advantages over frame-based models in expressing certain forms of modeling and in visual reasoning that facilitates an intuitive understanding [15].

In this study, therefore, I had tried to help the physicians/knowledge managers/users using these methodologies in acquiring previous knowledge and visual reasoning.

3. Methodology

To construct the platform, the five key processes were designed and implemented. Fig. 1 shows the whole structure of the proposed platform and the processes.

• *Knowledge Parsing*:

In *Knowledge Parsing*, the knowledge managers would be required to ready for the Social Knowledge/Former Cases / Historical Data set / Clinical Data Set / and other data set used for the decision support.

Using the data set we can extract the initial knowledge in the forms of the logical rules. The knowledge used here is OAV (Object-Attribute-Value) typedlogical rules just like an association rule. The text knowledge base (Text-KB) contains the rules. Then the Text-KB would be transformed into the RDB tables. For this purpose, we developed a prototype knowledge transformation tool on the Windows 7 environment by using the Visual Studio 2010.

• *Knowledge Maintenance:*

The main functionalities of the *Knowledge Maintenance* are to *fetch*, *maintain*, and *restore*the knowledge elicited from the previous process. In maintaining, *Redundancy check*, *Sort*, *Keyword Search*, *Relation check*, and other knowledge checking functions would be executed by the knowledge managers.

• *Knowledge Distribution:*

To shows the *Knowledge Distributions* of the knowledge the Knowledge Distribute Map (KDM) was used. The construction algorithm for the KDM was newly implemented in this study. Before the constructions of the KDM, the knowledge distribution proportions would be computed by using the basic numerical computations. According to the portions of knowledge, the KDM would be shaped.

• *Knowledge Hierarchy:*

In the process, the platform shows the entire hierarchies of the knowledge in the form of the DT. Through the process, the physicians may try a visual reasoning that facilitates an intuitive understanding.

• *Inference:*

In the *Inference*, the knowledge managers or the physicians can start the logical inference and follow up the process until they can reach a final goal. Then they can match the result with their experiences and tacit knowledge.

4. Implementation and application

4.1 Dermatology data set

For the validation of the functionalities of the platform, the Dermatology data set was used. The data set contains 34 attributes (properties), 33 of which are linear valued (0~3) and one of them is nominal. The 12 attributes (V1~v11, V34) are clinical attributes, and 22 attributes are histopathological attributes.The final goal or decision variable (class) is *Disease* and it was classified into sixclassessuch as*Psoriasis*, *Seboreic dermatitis*, *Lichen planus*, *Pityriasis rosea*, *Cronic dermatitis*, and *Pityriasis rubra pilaris* [16].

Table 1 and Table 2 show the attributes and classes of the data set. Then the Table 3 shows the table format of the data set.

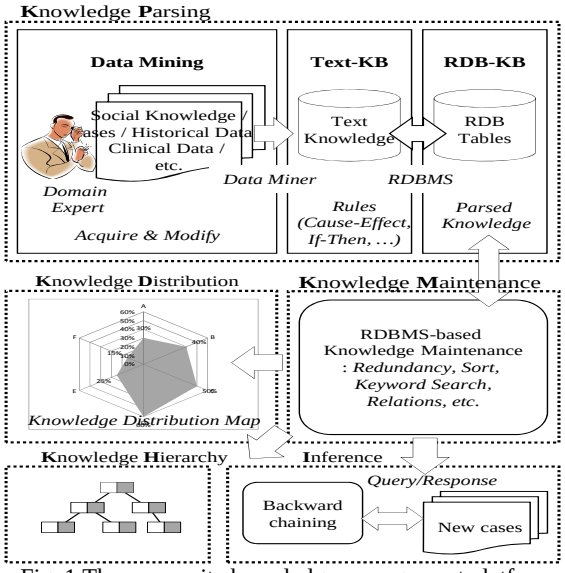


Fig. 1.The composite knowledge management platform

Table 1. The dermatopathy data set

(a) Clinical attributes

Clinical attributes(takes values 0,1,2,3)	
At. #	Descriptions
v1	Erythema
v2	Scaling
v3	Definite borders
v4	Itching
v5	Koebner phenomenon
v6	Polygonal papules
v7	Follicular papules
v8	Oral mucosal involvement
v9	Knee and elbow involvement
v10	Scalp involvement
v11	Family history (0, 1)
v34	Age (*linear*)

(b) Histopathological attributes

Histopathological attributes(values 0,1,2,3)	
v12	Melanin incontinence
v13	Eosinophils in the infiltrate
v14	PNL infiltrate
v15	Fibrosis of the papillary dermis
v16	Exocytosis
v17	Acanthosis
v18	Hyperkeratosis
v19	Parakeratosis
v20	Clubbing of the rete ridges
v21	Elongation of the rete ridges
v22	Thinning of the suprapapillary epidermis
v23	Spongiform pustule
v24	Munro microabcess
v25	Focal hypergranulosis
v26	Disappearance of the granular layer
v27	Vacuolisation and damage of basal layer
v28	Spongiosis
v29	Saw-tooth appearance of retes
v30	Follicular horn plug
v31	Perifollicular parakeratosis
v32	Inflammatory monoluclear inflitrate
v33	Band-like infiltrate

Table 2. The number of instances in each class

Value	Class	The number of instances
1	Psoriasis	112
2	Seboreic dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Cronic dermatitis	52
6	Pityriasis rubra pilaris	20
-	Total	366

Using the data set shown in the Table 3, we could extract the association rules as shown in the Table 4. Some of the rules extracted from the previous process were modified to transform/divide into the RDB tables.

Table 3. The table format of the data

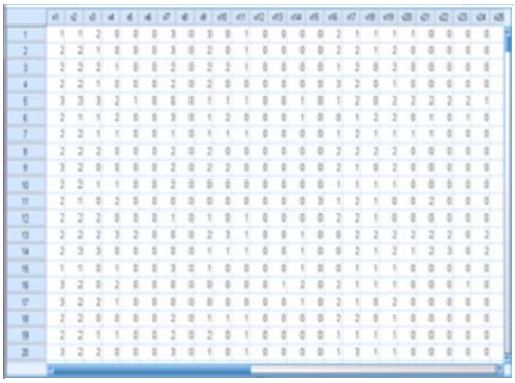


Table 4. Association rules

```
ACTIONS FIND Class;

ASK V1: "Erythema?";
CHOICES V1: 0, 1, 2, 3;

ASK V2: "Scaling?";
CHOICES V2: 0, 1, 2, 3;

ASK V3: "Definite borders?";
CHOICES V3: 0, 1, 2, 3;
…

ASK V33: "Band-like infiltrate?";
CHOICES V33: 0, 1, 2, 3;

RULE 1
IF  V33 = 0
THEN Class = 1;
…

RULE 14
IF  V32 = 0
AND  V33 = 2
THEN Class = 2;

RULE 15
IF  V32 = 0
AND  V33 = 3
THEN Class = 2;
…

RULE 20
IF  V32 = 3
AND  V33 = 2
THEN Class = 3;

RULE 21
IF  V32 = 3
AND  V33 = 3
THEN Class = 3;
…

RULE 28
IF  V14 = 0
```

AND V15 = 0
AND V20 = 0
AND V31 = 0
AND V33 = 0
THEN Class = 4;

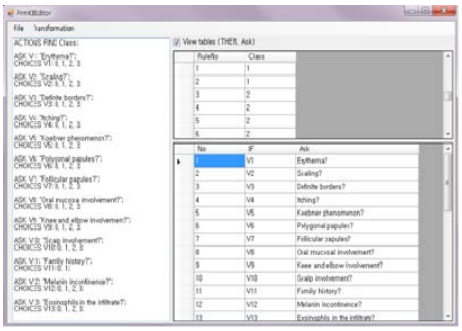
RULE 29
IF V14 = 0
AND V15 = 0
AND V20 = 0
AND V31 = 0
AND V33 = 1
THEN Class = 4;
…

RULE 34
IF V15 = 3
AND V33 = 0
THEN Class = 5;
…

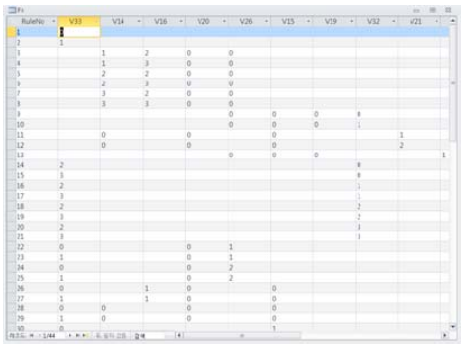
RULE 43
IF V18 = 2
AND V31 = 2
THEN Class = 6;

4.2 Simulation

To transform the knowledge, I have developed a prototype tool K-Expert. The tool was developed by using the Visual Studio 2010 on Windows 7. The text knowledge was transformed into the RDB table. The Fig. 2 shows the text rules and RDB tables in the window of the K-Expert.



(a) The text rules and RDB table



(b) RDB table of attributes
Fig. 2. The execution window of K-Expert

With the tables above, knowledge distributions would be computed. Table 5 shows the knowledge distributions of Properties (Attributes) and Classes.

Table 5. Knowledge distributions

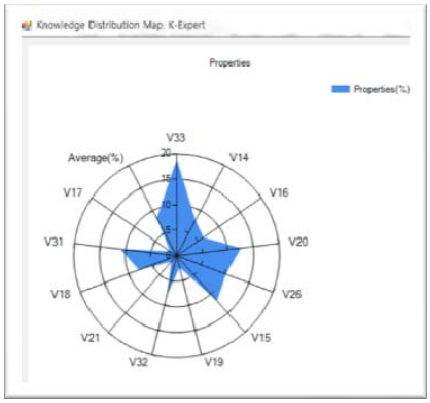
(a) Knowledge distributions of Properties

Attribute	V33	V14	V16	V20	V26	V15
Distribution	54.5%	22.7%	18.2%	36.4%	29.5%	34.1%
Attribute	V19	V32	V21	V18	V31	V17
Distribution	6.8%	22.7%	4.5%	22.7%	31.8%	4.5%

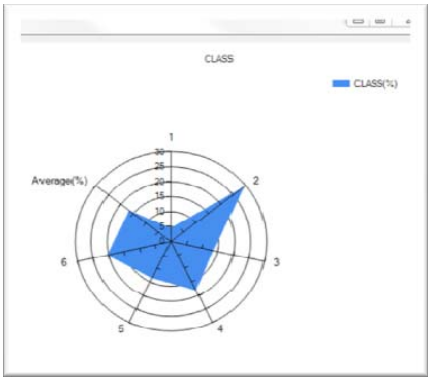
(b) Knowledge distribution of Classes

Class	1	2	3	4	5	6
Distribution	4.5%	29.5%	13.6%	18.2%	13.6%	20.5%

With the distributionshown above, the KDM would be shaped just like Fig. 3.



(a) KDM of attributes (properties)



(b) KDM of classes
Fig. 3. Knowledge Distribution Map (KDM)

At the next process, you can confirm the visualized knowledge hierarchy. The Fig. 4 shows the hierarchy as the form of the DT.

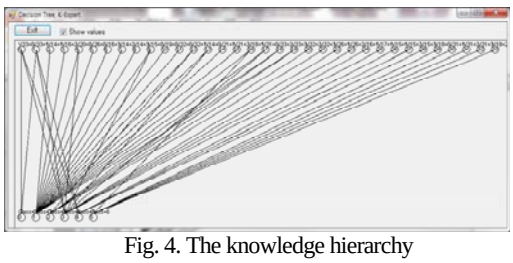


Fig. 4. The knowledge hierarchy

After the confirmation of the knowledge hierarchy, the knowledge manager or the users can try a logical reasoning by using at the *Inference* process. The Fig. 5 shows the process and result of inference.

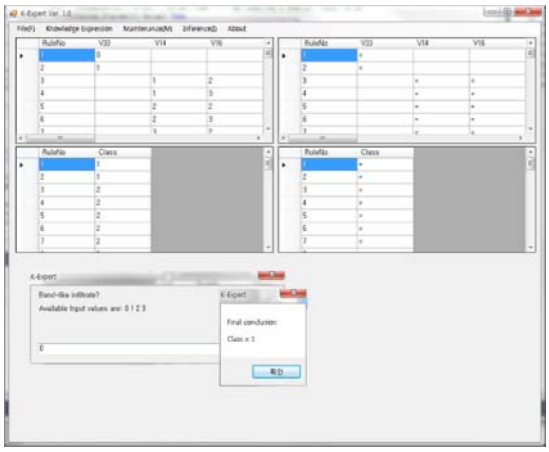


Fig. 5. The result of Inference

5. Conclusion

In this study, I haveproposed and implemented a composite knowledge management platform for the efficient knowledge manipulations, visualizations and inference.The platform consists of five main processes such as *Knowledge Parsing*, *Knowledge Maintenance*, *Knowledge Distribution*, *Knowledge Hierarchy*, and *Inference*.

The algorithms used in the processesweredeveloped based on the traditionalDM, RDBMS framework.Which was mainly aimed at expandthe adaptability, expandability, and usability of the KMS.Especially, the study is focused on the confirmations of the knowledge distributions, hierarchy, and inference. Therefore, it would be expected that the proposed platform have significant contributions on the advances of the researches related tothe development of KMS.

In the further works, theknowledge acquisitions and visualization methodologiesshould be improved by combining with other artificial intelligence mechanisms. Then the knowledgeinference mechanisms should be refined more elaborately.

References

[1] J.S. Kim, "RDB-based Automatic Knowledge Acquisition and Forward Inference Mechanism for Self-Evolving Expert Systems,"*Journal of Fuzzy Logic and Intelligent Systems*, vol. 13, no. 6, pp. 743-748, 2003a.

[2] J.S. Kim, "Development of Expert Systems using Automatic Knowledge Acquisition and Composite Knowledge Expression Mechanism,"*Proceedings of the 4th International Symposium on Advanced Intelligent Systems (ISIS) 2003*, Jeju (Korea), pp. 447-450, 2003b.

[3] J.S. Kim, "A construction of knowledge distribution map based on data mining and RDBMS methodologies,"*Proceedings of the KIIS Spring Conference*, Daegu University (Daegu, Korea), Daegu (Korea), pp. 189-190, 2013.

[4] S. Liao, "Knowledge management technologies and applications – literature review from 1995 to 2002,"*Expert Systems with Applications*, vol. 25, no. 2, pp. 155-164, 2003.

[5] K.M. Lee and K.M. Lee, "Candidate marker identification from gene expression data with attribute value discretization and negation,"*Journal of Korean Institute of Intelligent Systems*, vol. 21, no. 5, pp. 575-580, 2011.

[6] H.S. Kim, H. Cho, and I.K. Lee, "Design and development of an EHR platform based on medical informatics standards,"*Journal of Korean Institute of Intelligent Systems*, vol. 21, no. 4, pp. 407-535, 2011.

[7] C. Son, A. Shin, I. Lee, H. Park, H. Park, &Y. Kim, "Fuzzy discretization with spatial distribution of data and its application to feature selection,"*Journal of Korean Institute of Intelligent Systems*, vol. 20, no. 2, pp. 165-172, 2010.

[8] Y.I. Cho, "ITS: Intelligent tissue mineral analysis medical information system,"*Journal of Korean Institute of Intelligent Systems*, vol. 15, no. 2, pp. 257-263, 2005.

[9] C. Bukhari and Y. Kim, "Incorporation of fuzzy theory with heavyweight ontology and its application on vague information retrieval for decision making,"*Journal of Korean Institute of Intelligent Systems*, vol. 11, no. 3, pp. 171-177, 2011.

[10] H. Bae, Y. Kim, S. Kim, &G. J. Vachtsevanos, "Datamining roadmap to extract inference rules and design data models from process data of industrial applications,"*Journal of Korean Institute of Intelligent Systems*, vol. 5, no. 3, pp. 200-205, 2005.

[11] D. Nevo and and Y.E. Chan,"A Delphi study of knowledge management systems: Scope and requirements,"*Information & Management*, vol. 44, no. 6, pp. 583-597, 2007a.

[12] D. Nevo and Y.E. Chan, "A temporal approach to expectations and desires from knowledge managements systems,"*Decision Support Systems*, vol. 44, no. 1, pp.298-312, 2007b.

[13] A. Armoni, "Knowledge acquisition for medical diagnosis systems,"*Knowledge-Based Systems*, vol. 8, no. 4, pp. 223-226, 1995.

[14] J.M. Juarez, T. Riestra, M. Campos, A. Morales, J. Palma, &R. Marin, "Medical knowledge management for specific hospital department,"*Expert Systems with Applications*, vol. 36, no. 10, pp. 12214-12224, 2009.

[15] B. Kamsu-Foguem, G. Diallo, &C. Foguem, "Conceptual graph-based knowledge representation for supporting reasoning in African traditional medicine,"*Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1348-1365, 2013.

[16] UC Irvine, *Machine Learning Data Repository*, Center for Machine Learning and Intelligent Systems, University of California, Irvine, http://archive.ics.uci.edu/ml/datasets.html, 2013.

[17] L.M.Brasil, F.M.Azevedo, &J.M. Barreto, "Hybrid expert system for decision supporting in the medical area: complexity and cognitive computing,"*International Journal of Medical Informatics*, vo. 63, no. 1-2, pp. 19-30, 2001.

[18] A.Rafea, H.Hassen, & M. Hazman, "Automatic knowledge acquisition tool for irrigation and fertilization expert systems,"*Expert Systems with Applications*, vol. 24, no. 1, pp. 49-57, 2003.

[19] E.H. Shortliffe, *Computer-based medical consultations, MYCIN*, Elsevier, 1976.

[20] M. Casella Dos Santos, F. Montyne, &C. Dhaen, "Medical natural language processing enhancing drung ordering and coding," in R. Istepanian, S., Laxminarayan, and S. Pattichia (Eds.), *M-Health: Emerging mobile health systems*, McGraw-Hill, 2007.

[21] J.H. Gennari, M.A. Musen, R.W. Fergerson, W.E. Grosso, M. Crubézy, H. Eriksson, N.F. Noy, &S.W. Tu, "The evolution of

Protégé: An environment for knowledge-based systems development,"*International Journal of Human-Computer Studies*, vol. 1, no.58, pp. 89-123, 2003.
[22] J. Palma, J.M. Juárez, M. Campos, &R. Martín, "A fuzzy theory approach for temporal model-based diagnosis,"*Artificial Intelligence in Medicine*, vol. 38, pp. 197-218, 2006.

저 자 소 개

김진성 (Jin Sung Kim)
2002년: 성균관대학교 경영학과
경영학 박사
2002년~현재: 전주대학교 경영학부 교수
2007년~2009년: Southeast Missouri State University
교환교수

관심분야 : 퍼지이론, 인공신경망, 자연어처리, 의사결정지원, 전문가시스템, 시멘틱 웹
Phone : +82-63-220-2932
E-mail : kimjs@jj.ac.kr