*ICIC Express Letters* ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **1**, January **2016** pp. 167–175

# FEATURE ANALYSIS ON SEVERAL KNOWLEDGE CHANGE RATE-BASED ATTRIBUTES IMPORTANCE MEASURE METHODS

Fachao Li[1], Qihui Hu[2] and Bokai (William) Zhao[3]

[1]School of Economics and Management
[2]College of Sciences
Hebei University of Science and Technology
No. 26, Yuxiang Street, Shijiazhuang 050018, P. R. China
lifachao@tsinghua.org.cn; 15832170220@163.com

[3]Monte Vista Christian School
2 School Way, Watsonville, California 95076, United States

Received June 2015; accepted August 2015

Abstract. *For the problem of attributes importance measure, in this paper, we take the knowledge hiding in decision information system as a carrier, and the knowledge change caused by attributes set change as a basis. Firstly, we discuss associated features among positive region, lower and upper approximations of decision classes and knowledge in system, and then give several knowledge change rate-based attributes importance measure methods (BKCR-AIM for short) satisfying structural features of fuzzy measure. Secondly, we discuss their structural features and constructed strategies, and further analyze their features combining with a specific case. Finally, theoretical analysis and example calculation show that it is a feasible way which is using knowledge change in system to consider attributes importance, and comprehensive knowledge change rate-based attributes importance measure (BCKCR-AIM for short) has a good structural feature and strong interpretability. Moreover, BCKCR-AIM also has a wide application in information fusion, fuzzy decision, comprehensive evaluation and so on.*
**Keywords:** Decision information system, Lower and upper approximations, Importance measure, Decision classes

1. **Introduction.** Information fusion, as a strategy and method obtaining knowledge, is a problem faced by resource allocation, system optimization, transportation and other areas. Many scholars have studied on the strategy and method of information fusion by combining with different backgrounds and theoretical research, and many good research results having a good structural feature are obtained. The weighted average model based on a quantification platform is the most representative and widespread among the results. However, it is worth noting that the model requires mutual independence among various fusing indicators. And the kind independence is often difficult to meet in real problem and the importance (or weight) of indicators is also difficult to determine. Therefore, many scholars further studied on the construction of index system and the determination of weight system, such as the index decomposition and synthesis method based on analytic hierarchy process (AHP), the index selection method based on cluster analysis, the determining method of weight system based on relevance theory. However, these methods are essentially unable to solve the correlation among indicators. For this shortcoming, some scholars proposed many information fusion models which use fuzzy measure [1] to describe the importance of index system and take fuzzy integral as a comprehensive operator. Further the rationality and feasibility of models were analyzed and successfully applied to fault diagnosis, pattern recognition and other fields [2,3]. Because information fusion methods based on fuzzy measure and integral theoretically have solved the associated problems among indicators, they are widely recognized by academic field. While,

constructing fuzzy measure method based on relevant field experts are questioned, so the determination of fuzzy measure on index system is also a bottleneck for this kind problem.

With the development of information science and technology, the collection, storage and dissemination capability of information undergone tremendous change, and various industries have accumulated diverse data which can be considered as accumulation of past experience or observation results of some regulars (or phenomena). Therefore, it is widespread research content to mine hidden knowledge in data in the field of academic and application. Many scholars have had many beneficial discussions and obtained a lot of important theories and application results through combining with different theories. For example, most of the existing fuzzy decision tree algorithm did not systematically consider the impact of the non-linear characteristics of the membership degree of fuzzy set, and they were unable to integrate uncertain processing preference in the selection of extended attributes, therefore, the literature [13] gave an utility description system of membership state, future presented the generalized Hartley metric concept and the generalized fuzzy partition entropy concept which was applied to the selection process of extended attributes for fuzzy decision tree, and finally proposed a generalized fuzzy partition entropy-based fuzzy ID3 algorithm (abbreviated as GFID3). For the information security issue in e-government, [5] analyzed k-means algorithm, decision tree algorithm and artificial neural network algorithm, and took the company scale as the breakthrough point; finally it constructed the Associated Analysis Model based on multiple algorithm fusion, which was successfully applied to enterprise information security assessment and decision analysis.

The essence of data mining is to discover hidden rules in data based on attributes value. Different attributes play different roles in knowledge discovery process. And the more considered attributes they meet, the more discovered knowledge is. Therefore, there is a close connection between hidden knowledge in data system and attributes structure in system. And it is a feasible way to measure the comprehensive importance of attributes set through the change of knowledge carrier. The measurement not only reflects dependence of past experience, but also weakens effect of subjective preference to a certain extent. In fact, many fields all involve attributes importance problems, such as multiple attributes decision and attributes reduction, so some existing studies have partially related to this aspect. For example, **1) to the attributes reduction based on data.** According to knowledge benchmarks in data system that is positive domain and the breakthrough that is knowledge changes caused by attributes change, literature [6] gave an attribute importance measure method based on the covering information system. For the feature subset selection issue in the areas of pattern recognition, machine learning and data mining, [7] showed a simple and efficient feature subset selection technique based on a proposed fuzzy-rough model based on rough set theory and fuzzy-rough set theory, and then it constructed a forward hybrid attributes reduction algorithm (named FAR-VPFRS). **2) to the forecast problem based on data.** For the flood forecasting issue, literature [8] predicted the labor market by implementing Naive Bayesian Classifier, Decision Tree and Decision Rules techniques, and then determined which was the best predictor requirement for the labor market through comparing three techniques. **3) to the fault diagnosis based on data.** The literature [9] constructed a distance standardized method based on attributes importance by discussing the core attributes selection, which was successfully applied to fault diagnosis problems in the nuclear power plant, here, fault types were decision attributes values, relevant influential factors were condition attributes and the dependent relationship between fault types and relevant influential factors was as a foundation. Although the existing attributes importance researches have successfully applied in many special issues, it is still worth noting that they have some great limitations, such as, lacking systematic theoretical system, or lacking adequate operability. Therefore, how to combine with relevant data mining technology and accumulated data information to establish an attributes importance measure mode with structural feature of fuzzy measure, is an important research topic with important theory and application value.

According to the analysis, for the attributes importance measure problems, in this paper, regarding the knowledge hiding in decision information system as a carrier, we mainly have the following work. In Section 2, we give some definitions such as fuzzy measure, positive region and lower and upper approximations, and simply discuss their structural features. In Section 3, we give several knowledge change rate-based attributes importance measure methods (BKCR-AIM for short) satisfying structural features of fuzzy measure, and at the same time, discuss their associated features and constructed strategies. In Section 4, we further analyze their structural features and associated features through some theorems and their corollaries. In Section 5, by combining with a specific case, the above theories are verified. In Section 6, it is easy to get the conclusion of this paper. Moreover, the theoretical analysis and example calculation also show that it is a feasible way that is using knowledge change in system to consider attributes importance. And comprehensive knowledge change rate-based attributes importance measure (BCKCR-AIM for short) has a good structural feature and strong interpretability. Meanwhile, it also has a wide application in information fusion, fuzzy decision, comprehensive evaluation and so on.

## 2. **Preliminaries.**

### 2.1. **Fuzzy measure.** 
Fuzzy measure was proposed by Sugeno [1] in 1974, and its essence is to use monotonicity and continuity in place of the additivity in classic measure. As a result, this change does not weaken basic features of measure, and it also lays a foundation of building measure method conforming to real needs.

**Definition 2.1.** [10] *Let $X$ be a nonempty universe, and $\mathscr{B}$ be a nonempty class of subsets of $X$, $\mu : \mathscr{B} \to [0, \infty)$. If $\mu$ satisfies: 1) when $\phi \in \mathscr{B}$, $\mu(\phi) = 0$; 2) (monotonicity) for arbitrary $E, F \in \mathscr{B}$, when $E \subset F$, $\mu(E) \le \mu(F)$; 3) (continuity from above) for $\{E_n\}_{n=1}^{\infty} \subset \mathscr{B}$, when $E_1 \subset E_2 \subset \cdots E_n \subset \cdots$, and $\bigcup_{n=1}^{\infty} E_n \in \mathscr{B}$, $\lim_{n\to\infty} \mu(E_n) = \mu(\bigcup_{n=1}^{\infty} E_n)$; 4) (continuity from below) for $\{E_n\}_{n=1}^{\infty} \subset \mathscr{B}$, when $E_1 \supset E_2 \supset \cdots E_n \supset \cdots$, $\bigcap_{n=1}^{\infty} E_n \in \mathscr{B}$, and there exists a natural number $n_0$ such that $\mu(E_{n_0}) < \infty$, $\lim_{n\to\infty} \mu(E_n) = \mu(\bigcap_{n=1}^{\infty} E_n)$, then $\mu$ is called a fuzzy measure on $(X, \mathscr{B})$, and $(X, \mathscr{B}, \mu)$ is called a fuzzy measure space. Especially, when $\mu(X) = 1$, $\mu$ is called a normalized fuzzy measure.*

It is easy to see that when $X$ is a finite universe, the above and below continuity are certainly satisfied. Therefore, constructing fuzzy measure based on finite universe only considers monotonicity. However, when $X$ is an infinite universe, the above and below continuity are essential elements for the tightness of constructed measure mode. Many scholars have discussed structural features and constructed problems of fuzzy measure, and formed a relatively complete theoretical system; the specific content can be seen in literature [10].

### 2.2. **Pawlak rough set model.** 
Rough set was proposed by Pawlak in 1984, and its basic idea is a theoretical tool that uses incomplete information to seek complete conclusion and uses some division (or knowledge) of universe to consider the concept description problems on universe. Since rough set can simply describe essential features of data mining, scholars generalize Pawlak's rough set by combining different theories with application backgrounds. Then, it not only has formed a relatively complete theoretical system, but also has many successful applications in many fields, and the details can be found in [11-13].

For convenience, we assume in the following: 1) $(U, A, d, F_A)$ represents a decision information system. Here, $U = \{x_1, x_2, \cdots, x_n\}$ is a finite nonempty set, $A = \{a_1, a_2, \cdots, a_s\}$

denotes condition attributes set, and $d$ denotes decision attribute; $V(a_i)$ denotes the range of $a_i$, $V(d)$ denotes the range of $d$, and $F_A = \{f_a, f_d | a \in A\}$ denotes information function (here, $f_a$ is a mapping from $U$ to $V(a)$, and $f_d$ is a mapping from $U$ to $V(d)$); 2) For the equivalence relation $R$ over $U$ (namely, $R \subset U \times U$ and satisfied: i) $(x,x) \in R$ always holds for arbitrary $x \in U$; ii) if $(x,y) \in R$, $(y,x) \in R$; iii) if $(x,y), (y,z) \in R$, $(x,z) \in R$), $[x]_R = \{y | y \in U \textit{ and } (x,y) \in R\}$ denotes $R$-equivalence class of $x$, $U/R = \{[x]_R | x \in R\}$; 3) For $(U, A, d, F_A)$ and $B \subset A$, $R_B = \{(x,y) | (x,y) \in U \times U \textit{ and } f_a(x) = f_a(y) \textit{ for arbitrary } a \in B\}$ denotes an equivalence relation over $U$ induced by attributes set $B$, $[x]_B$ denotes $R_B$-equivalence class of $x$ (especially, when $B$ is a single set $\{a\}$, $[x]_B$ and $[x]_a$ denote $R_B$ and $R_a$, respectively), $R_d = \{(x,y) | (x,y) \in U \times U \textit{ and } f_d(x) = f_d(y)\}$ denotes an equivalence relation over $U$ induced by decision attribute $d$, $U/B$ denotes $U/R_B$, and $U/d$ denotes $U/R_d$; 4) when $[x]_A \subset [x]_d$ always holds for arbitrary $x \in U$, we call $(U, A, d, F_A)$ is coordinated.

**Definition 2.2.** *Given a decision information system $(U, A, d, F_A)$, $B \subset A$ and $X \subset U$, then, lower and upper approximations set of $X$ with respect to $B$ are respectively defined as*

$$\underline{B}(X) = \{x | x \in U \textit{ and } [x]_B \subset X\},\ \overline{B}(X) = \{x | x \in U \textit{ and } [x]_B \bigcap X \neq \phi\}. \tag{1}$$

*Positive region of $B$ with respect of $d$ is defined as*

$$\underline{B}(U/d) = \bigcup \{\underline{B}(X) | X \in U/d\}. \tag{2}$$

In the decision information system $(U, A, d, F_A)$, lower and upper approximations set of decision classes and positive region are an important basis for mining knowledge from different angles, and they change with $B$. Therefore, to some extent, this change reflects comprehensive importance of each attribute in $B$. Because $R_E \supset R_F$ for arbitrary $E \subset F \subset A$ always holds, we can draw the following conclusions.

**Theorem 2.1.** *Given a decision information system $(U, A, d, F_A)$, $E \subset F \subset A$ and $X \subset U$, then: 1) $\underline{E}(X) \subset \underline{F}(X)$, $\overline{E}(X) \supset \overline{F}(X)$; 2) $\underline{E}(U/d) \subset \underline{F}(U/d)$; 3) when $(U, A, d, F_A)$ is a coordinated decision information system, $\underline{A}([x]_d) = \overline{A}([x]_d)$ always holds for arbitrary $x \in U$.*

3. **Several Knowledge Change Rate-based Attributes Importance Measure Methods.** Lower and upper approximations and positive region are all a measurement basis about characterizing knowledge in decision information system $(U, B, d, F_B)$ (here, $\underline{B}(X)$ means the whole object that surely accords with $X$, and $\overline{B}(X)$ means the whole object that probably accords with $X$). Moreover, lower and upper approximations of decision classes and positive region change with attributes set $B$. Therefore, it is a feasible way to construct an importance measure of attributes set based on change rules of lower and upper approximations of decision classes and positive region. In this section, we focus on the constructed strategy about importance measure of attributes set based on knowledge change rate. For convenience, we assume in the following: $(U, A, d, F_A)$ is a decision information system, $U/d = \{D_1, D_2, \cdots, D_s\}$, $\mathscr{P}(A)$ is the power set of $A$ (namely, the set consists of all subsets of $A$), and $|C|$ is the element number of set $C$ (called the cardinality of $C$).

In $(U, A, d, F_A)$, for $B \in \mathscr{P}(A)$, the hidden knowledge is described from different angles. For example, the literature [6] proposed an importance measure based on knowledge change rate of $B$ by taking the surely knowledge as a scribed method (here, $\underline{B}(U/d) =$

$\bigcup\{\underline{B}(X)|X \in U/d\}$), which is called as positive region change rate-based attributes importance measure, BPRCR-AIM for short, namely

$$\mu_0(B) = \begin{cases} \frac{|\underline{B}(U/d)|}{|\underline{A}(U/d)|}, & |\underline{A}(U/d)| \neq 0, \\ 1, & |\underline{A}(U/d)| = 0 \text{ and } B \neq \phi, \\ 0, & B = \phi. \end{cases} \tag{3}$$

We can see that $\mu_0(B)$ only loosely considers surely knowledge change in system, and it is the effect of each decision class in system without considering. Therefore, if $\underline{B}(D_k)$ is a scribed method of $k$-class surely knowledge, $\frac{|D_k|}{|U|}$ is the importance of $k$-decision class in total sample, and we also can have an importance measure called as surely knowledge change rate-based attributes importance measure, BSKCR-AIM for short, that is

$$\mu_1(B, D_k) = \begin{cases} \frac{|\underline{B}(D_k)|}{|\underline{A}(D_k)|}, & |\underline{A}(D_k)| \neq 0, \\ 1, & |\underline{A}(D_k)| = 0 \text{ and } B \neq \phi, \\ 0, & B = \phi. \end{cases} \tag{4}$$

$$\mu_1(B) = \sum_{k=1}^{s} \frac{|D_k|}{|U|} \cdot \mu_1(B, D_k). \tag{5}$$

However, both $\mu_0(B)$ and $\mu_1(B)$ only consider the surely knowledge in system, and they do not consider the effect of relevant knowledge in system. Here, we take $k$-class relevant knowledge ($\overline{B}(D_k)$) as a scribed method, and $\frac{|D_k|}{|U|}$ is still the importance of $k$-decision class in total sample. Then the following mode (7) is also an importance measure called relevant knowledge change rate-based attributes importance measure, BRKCR-AIM for short, namely

$$\mu_2(B, D_k) = \begin{cases} \frac{|\overline{A}(D_k)|}{|\overline{B}(D_k)|}, & B \neq \phi, \\ 0, & B = \phi. \end{cases} \tag{6}$$

$$\mu_2(B) = \sum_{k=1}^{s} \frac{|D_k|}{|U|} \cdot \mu_2(B, D_k). \tag{7}$$

According to the above modes, it is easy to see that the effect between surely knowledge and relevant knowledge during decision process is different. Here, $w_1$ and $w_2$ denote the importance of the 'surely' and 'relevant' knowledge respectively (namely, $w_1, w_2 \in [0, 1]$ and $w_1 + w_2 = 1$), the same $\frac{|D_k|}{|U|}$ is the importance of $k$-decision class in total sample, and we have

$$\mu(B, D_k) = w_1\mu_1(B, D_k) + w_2\mu_2(B, D_k), \tag{8}$$

$$\mu(B) = \sum_{k=1}^{s} \frac{|D_k|}{|U|} \cdot \mu(B, D_k). \tag{9}$$

Mode (9) is also an importance measure based on knowledge change rate of $B$ (called comprehensive knowledge change rate-based attributes importance measure, BCKCR-AIM for short).

It is easy to see, 1) $\mu(B)$ is a comprehensive measure mode of attributes importance, and simultaneously considers feature and knowledge change of each decision class. Moreover, it is a popularization of $\mu_1(B)$ and $\mu_2(B)$ (namely, when $w_1 = 1$ and $w_2 = 0$, $\mu(B) = \mu_1(B)$, when $w_1 = 0$ and $w_2 = 1$, $\mu(B) = \mu_2(B)$); 2) when $X \subset D_k$ and $k = 1, 2, \cdots, s$, $\mu_0(B) = \mu_1(B)$. For convenience, we assume in the following that BPRCR-AIM, BSKCR-AIM, BRKCR-AIM and BCKCR-AIM are collectively referred to as knowledge change rate-based attributes importance measures, BKCR-AIM for short.

4. **Feature Analysis of BKCR-AIM.** This section mainly analyzes structural features and value rules of BKCR-AIM.

Combining with the discussion in Section 3 and Theorem 2.1, we can know that the following conclusions always hold for arbitrary $E, F \in \mathscr{P}(A)$ and $i \in \{0, 1, 2\}$: 1) when $E \subset F \subset A$, $\mu_i(E) \leq \mu_i(F)$, $\mu(E) \leq \mu(F)$; 2) $0 \leq \mu_i(E) \leq 1$, $0 \leq \mu(E) \leq 1$; 3) $\mu_i(\phi) = 0$, $\mu(\phi) = 0$, $\mu_i(A) = 1$, $\mu(A) = 1$. Therefore, we can draw the following theorems combined with Definitions 2.1 and 2.2.

**Theorem 4.1.** *Given a decision information system $(U, A, d, F_A)$, $U/d = \{D_1, D_2, \cdots, D_s\}$, $\mathscr{P}(A)$ is the power set of $A$. Then, both $\mu_i(B)$ and $\mu(B)$ are the normalized fuzzy measure on $(A, \mathscr{P}(A))$ for arbitrary $i \in \{0, 1, 2\}$.*

**Theorem 4.2.** *Given a decision information system $(U, A, d, F_A)$, $U/d = \{D_1, D_2, \cdots, D_s\}$, $B \in \mathscr{P}(A)$, then, 1) when $w_1 = 1$, the necessary and sufficient conditions of $\mu(B) = \mu_1(B) = 1$ are that $\underline{B}(D_k) = \underline{A}(D_k)$ holds for arbitrary $k \in \{1, 2, \cdots, s\}$; 2) when $w_2 = 1$, the necessary and sufficient conditions of $\mu(B) = \mu_2(B) = 1$ are that $\overline{B}(D_k) = \overline{A}(D_k)$ holds for arbitrary $k \in \{1, 2, \cdots, s\}$; 3) when $w_1, w_2 \in (0, 1)$, the necessary and sufficient conditions of $\mu(B) = 1$ are that $\underline{B}(D_k) = \underline{A}(D_k)$ and $\overline{B}(D_k) = \overline{A}(D_k)$ hold for arbitrary $k \in \{1, 2, \cdots, s\}$; 4) the necessary and sufficient conditions of $\mu_0(B) = 1$ are that $\underline{B}(U/d) = \underline{A}(U/d)$ holds.*

**Proof:** We only give the proof of (1) in the following, (2)$\sim$(4) can be similarly proved. By Formula (9), the necessary and sufficient conditions of $\mu(B) = 1$ are that $\mu(B, D_k) = 1$ holds for arbitrary $k \in \{1, 2, \cdots, s\}$, the necessary and sufficient conditions of $\mu(B, D_k) = 1$ are $\underline{B}(D_k) = \underline{A}(D_k)$ (actually, when $w_1 = 1$, we can follow $\mu_1(A, D_k) = \mu(A, D_k) = 1$ and $\underline{B}(D_k) \subset \underline{A}(D_k)$ to know that the necessary and sufficient conditions of $\mu(B, D_k) = 1$ are $\underline{B}(D_k) = \underline{A}(D_k) \neq \phi$ *or* $\underline{B}(D_k) = \underline{A}(D_k) = \phi$), and then (1) sets up.

**Corollary 4.1.** *Given a decision information system $(U, A, d, F_A)$, $U/d = \{D_1, D_2, \cdots, D_s\}$, $B \in \mathscr{P}(A)$, we always have $\mu(B) = \mu_i(B) = 1$ for arbitrary $i \in \{0, 1, 2\}$ if and only if $(U, B, d, F_B)$ is coordinated.*

**Corollary 4.2.** *Given a decision information system $(U, A, d, F_A)$, $U/d = \{D_1, D_2, \cdots, D_s\}$, $B \subset C \subset A$, then:*

*1) The necessary and sufficient conditions of $\mu(B) = \mu(C)$ are $\underline{B}(D_k) = \underline{C}(D_k)$ and $\overline{B}(D_k) = \overline{C}(D_k)$ for arbitrary $k \in \{1, 2, \cdots, s\}$ when $w_1, w_2 \in (0, 1)$;*

*2) The necessary and sufficient conditions of $\mu_1(B) = \mu_1(C)$ are $\underline{B}(D_k) = \underline{C}(D_k)$ for arbitrary $k \in \{1, 2, \cdots, s\}$;*

*3) The necessary and sufficient conditions of $\mu_2(B) = \mu_2(C)$ are $\overline{B}(D_k) = \overline{C}(D_k)$ for arbitrary $k \in \{1, 2, \cdots, s\}$;*

*4) The necessary and sufficient conditions of $\mu_0(B) = \mu_0(C)$ are $\underline{B}(U/d) = \underline{C}(U/d)$.*

5. **A Comparative Analysis of BKCR-AIM.** In this section, we will further expound and analyze the determining process and basic features of BKCR-AIM by combining with a specific disease diagnosis example.

**Case description**. In order to improve the efficiency of medical workers and reduce the misdiagnosis rate in disease diagnosis process, a medical institution decides to construct a comprehensive diagnosis system by taking the existed clinical cases as a basis. And its theme is the associated degree between disease and relevant characterization. Moreover, it provides an accordance for doctors and disease diagnosis and treatment of patients. The characterizations for different patients of same diseases are not sure exactly the same, and the partial characterizations for different diseases may be the same. Therefore, the nature of constructing comprehensive diagnosis system is to determine the associated measure problem between relevant disease and relevant characterizations. Namely, we can regard existed cases as universe $U$, associated diseases as

Table 1. Case system of patients

| $patients$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $d$ | $patients$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $d$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 0 | 11 | 1 | 1 | 1 | 1 | 0 |
| 2 | 2 | 1 | 1 | 0 | 1 | 12 | 1 | 2 | 0 | 1 | 0 |
| 3 | 1 | 2 | 1 | 1 | 1 | 13 | 2 | 0 | 1 | 1 | 1 |
| 4 | 2 | 2 | 0 | 0 | 1 | 14 | 0 | 0 | 1 | 1 | 0 |
| 5 | 2 | 1 | 1 | 0 | 0 | 15 | 1 | 1 | 1 | 0 | 0 |
| 6 | 0 | 2 | 1 | 1 | 0 | 16 | 0 | 1 | 0 | 1 | 0 |
| 7 | 0 | 1 | 1 | 1 | 0 | 17 | 2 | 1 | 1 | 0 | 1 |
| 8 | 2 | 1 | 0 | 0 | 1 | 18 | 1 | 2 | 1 | 0 | 1 |
| 9 | 1 | 0 | 1 | 1 | 0 | 19 | 2 | 2 | 1 | 1 | 2 |
| 10 | 2 | 0 | 1 | 0 | 1 | 20 | 2 | 1 | 0 | 1 | 0 |

Table 2. The calculation results of $\mu(B)$ and $\mu_i(B)$

| $B$ | $\mu(B)$ | | | $\mu_0(B)$ | $\mu_1(B)$ | $\mu_2(B)$ |
|---|---|---|---|---|---|---|
| | $w_1 = 0.2, w_2 = 0.8$ | $w_1 = 0.5, w_2 = 0.5$ | $w_1 = 0.8, w_2 = 0.2$ | | | |
| $\{a_1\}$ | 0.5144 | 0.4040 | 0.2936 | 0.2353 | 0.2200 | 0.5881 |
| $\{a_2\}$ | 0.4367 | 0.2729 | 0.1092 | 0 | 0 | 0.5458 |
| $\{a_3\}$ | 0.4329 | 0.2705 | 0.1082 | 0 | 0 | 0.5411 |
| $\{a_4\}$ | 0.4330 | 0.2708 | 0.1083 | 0 | 0 | 0.5417 |
| $\{a_1, a_2\}$ | 0.7322 | 0.6726 | 0.6130 | 0.5882 | 0.5733 | 0.7719 |
| $\{a_1, a_3\}$ | 0.5644 | 0.4765 | 0.3886 | 0.3529 | 0.3300 | 0.6230 |
| $\{a_1, a_4\}$ | 0.5233 | 0.4096 | 0.2958 | 0.2353 | 0.2200 | 0.5992 |
| $\{a_2, a_3\}$ | 0.4400 | 0.2750 | 0.1100 | 0 | 0 | 0.5500 |
| $\{a_2, a_4\}$ | 0.6335 | 0.5740 | 0.5146 | 0.4706 | 0.4750 | 0.6731 |
| $\{a_3, a_4\}$ | 0.5734 | 0.4909 | 0.4084 | 0.3529 | 0.3533 | 0.6285 |
| $\{a_1, a_2, a_3\}$ | 0.9122 | 0.8995 | 0.8868 | 0.8824 | 0.8783 | 0.9207 |
| $\{a_1, a_2, a_4\}$ | 0.8717 | 0.8492 | 0.8267 | 0.8235 | 0.8117 | 0.8867 |
| $\{a_1, a_3, a_4\}$ | 0.7027 | 0.6336 | 0.5644 | 0.5294 | 0.5183 | 0.7488 |
| $\{a_2, a_3, a_4\}$ | 0.6959 | 0.6587 | 0.6215 | 0.5882 | 0.5967 | 0.7207 |
| $\{a_1, a_2, a_3, a_4\}$ | 1 | 1 | 1 | 1 | 1 | 1 |

values of decision attributes and associated characterizations as condition attributes set $A$. Moreover, various special symptoms are regarded as values of condition attributes set. Then, the constructing problem of comprehensive diagnosis system is to construct the measure problem on $\mathscr{P}(A)$ with some special structural features based on $(U, A, d, F_A)$. The following section combining with a special disease system (see Table 1) is to illustrate BKCR-AIM. Here: 1) example space $U = \{1, 2, \cdots, 20\}$ concludes 20 patients; 2) condition attributes set $A = \{fever\ characterization$ $(a_1)$, $cough\ characterization$ $(a_2)$, $headache\ characterization$ $(a_3)$, $runny\ nose$ $(a_4)\}$, the range of $a_1$ is $\{$ $no\ fever$ (0), $low\ fever$ (1), $high\ fever$ (2)$\}$, the range of $a_2$ is $\{dry\ cough$ (0), $slight\ cough$ (1), $severe\ cough$ (2)$\}$, the range of $a_3$ is $\{occasional\ headache$ (0), $splitting\ headache$ (1)$\}$, and the range of $a_4$ is $\{no\ runny\ nose$ (0), $runny\ nose$ (1)$\}$; 3) decision attribute $d$ means actual disease, and its range is $\{catarrh\ (0), pneumonia\ (1), flu\ (2)\}$.

Combining with the above discussions, we can know that when values of decision attribute $d$ are 0, 1 and 2 respectively, their corresponding decision classes are $D_1 = \{1, 5, 6, 7, 9, 11, 12, 14, 15, 16, 20\}$, $D_2 = \{2, 3, 4, 8, 10, 13, 17, 18\}$, $D_3 = \{19\}$, and their lower and upper approximations of attributes set $A$ is $\underline{A}(D_1) = \{1, 6, 7, 9, 11, 12, 14, 15, 16, 20\}$, $\underline{A}(D_2) = \{3, 4, 8, 10, 13, 18\}$, $\underline{A}(D_3) = \{19\}$, $\overline{A}(D_1) = \{1, 2, 5, 6, 7, 9, 11, 12, 14, 15, 16, 17, 20\}$, $\overline{A}(D_2) = \{2, 3, 4, 5, 8, 10, 13, 17, 18\}$ and $\overline{A}(D_3) = \{19\}$. The values of $\mu(B)$ and $\mu_i(B)$ will be shown in Table 2 (here, $i \in \{0, 1, 2\}$).

It is easy to see from Table 2 that: 1) When $E \subset F$, $\mu_i(E) \leq \mu_i(F)$ and $\mu(E) \leq \mu(F)$ always hold for arbitrary $i \in \{0, 1, 2\}$, and it is consistent with the analysis in Section 4. It indicates that during the disease diagnosis process, the more aspects are considered, the more helpful for you to make the correct diagnosis; 2) there is an interaction among attributes, namely, the portion importance cannot completely represent the whole importance (for example, when $w_1 = 0.8, w_2 = 0.2$, $\mu(\{a_4\}) < \mu(\{a_2\}) < \mu(\{a_1\})$ and $\mu_i(\{a_4\}) < \mu_i(\{a_2\}) < \mu_i(\{a_1\})$, but $\mu(\{a_1, a_4\}) < \mu(\{a_2, a_4\})$ and $\mu_i(\{a_1, a_4\}) < \mu_i(\{a_2, a_4\})$; 3) though $\mu_0(B) \leq \mu(B) \leq \mu_2(B)$ always holds, their different amplitude is difference (for instance, when $w_1 = 0.8, w_2 = 0.2$, $\mu(\{a_1\}) - \mu_0(\{a_1\}) = 0.0583$, $\mu(\{a_1, a_2\}) - \mu_0(\{a_1, a_2\}) = 0.0248$, $\mu(\{a_1, a_2, a_3\}) - \mu_0(\{a_1, a_2, a_3\}) = 0.0044$, and when $w_1 = 0.2, w_2 = 0.8$, $\mu_2(\{a_1\}) - \mu(\{a_1\}) = 0.0737$, $\mu_2(\{a_1, a_2\}) - \mu(\{a_1, a_2\}) = 0.0397$, $\mu_2(\{a_1, a_2, a_3\}) - \mu(\{a_1, a_2, a_3\}) = 0.0085$); 4) for arbitrary $i \in \{0, 1\}$, $\mu_i(\{a_2\}) = \mu_i(\{a_3\}) = \mu_i(\{a_4\}) = 0$ indicates that $\mu(B)$ formally eliminates the irrationality of $\mu_i(B)$. Moreover, $\mu_i(B)$ and $\mu(B)$ have essential difference; 5) $\mu(B)$ changes with $w_1$ and $w_2$ (for example, when $w_1 = 0.8$ and $w_2 = 0.2$, $\mu(\{a_3\}) = 0.1082$, $\mu(\{a_1, a_3\}) = 0.3886$, and when $w_1 = 0.2$ and $w_2 = 0.8$, $\mu(\{a_3\}) = 0.4329$, $\mu(\{a_1, a_3\}) = 0.5644$ ). This indicates that BCKCR-AIM not only has a good structural feature, but also can take decision consciousness into decision process simply; 6) though the amount of data in this example is less and the calculation process is general, the decision problem has wide application in many areas having the features of above case, such as resource management, production process optimization, forecast of uncertain or incomplete environment. It indicates that the discussion has a broad applied prospect.

6. **Conclusions.** In this paper, we take the knowledge hiding in decision information system as a carrier, the knowledge change caused by attributes set change as a basis, and decision classes as a basic measure unit of knowledge. Firstly, we discuss associated features among positive region, lower and upper approximations of decision classes and knowledge in system, and further give several knowledge change rate-based attributes importance measure methods (BKCR-AIM for short) satisfying structural features of fuzzy measure. Secondly, we analyze their structural features and constructed strategies, and then compare features of various BKCR-AIM combining with a specific case. Finally, theoretical analysis and example calculations show that it is a feasible way by using knowledge change in system to consider attributes importance, and BCKCR-AIM has a good structural feature and strong interpretability. Moreover, it has a wide application in information fusion, fuzzy decision, comprehensive evaluation and so on.

Because many decision problems must face the importance of attributes, and various industries have accumulated a large of data information. Constructing attributes importance measure mode is urgently solved based on a large of data, which have a good structural feature and strong interpretability. Therefore, we discussed this problem in the paper, and further improved and modified it in many aspects. Moreover, we may recently carry out the following three aspects studies: 1) We can combine with inclusion degree and fuzzy set theories, and modify the basis knowledge factor in BCKCR-AIM to construct an attribute importance measure mode which has a good structural feature and strong robustness; 2) for a big data set, we will construct BCKCR-AIM system based on mathematical statistics theory; 3) combine medical diagnostic problem with data encrypted system to develop a decision system of BCKCR-AIM which is regarded as a support system.

**Acknowledgment.** This work is supported by the National Natural Science Foundation of China (71371064) and the Natural Science Foundation of Hebei Province (F2015208099, F2015208100).

## REFERENCES


[1] M. Sugeno, Theory of fuzzy integral and its applications, *Tokyo Institute of Technology: Tokyo*, 1974.

[2] G. Yang, X. P. Wu et al., Multi-sensor information fusion fault diagnosis method based on rough set theory, *Systems Engineering and Electronics*, vol.31, no.8, pp.2013-2019, 2009.

[3] H. Ge and L. F. Tian, Application of information fusion in pattern recognition, *Application Research of Computers*, vol.26, no.1, pp.19-24, 2009.

[4] C. X. Jin, F. C. Li et al., A generalized fuzzy ID3 algorithm using generalized information entropy, *Knowledge-Based Systems*, vol.64, pp.13-21, 2014.

[5] T. W. Yuan and P. Chen, Data mining application in E-government information security, *Procedia Engineering*, vol.29, pp.235-240, 2012.

[6] F. C. Li, Z. Zhang et al., A study of comprehensive evaluation method based on neighborhood covering information system, *ICIC Express Letters, Part B: Applications*, vol.4, no.3, pp.595-602, 2013.

[7] Q. H. Hu, Z. X. Xie et al., Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, *Pattern Recognition*, vol.40, pp.3509-3521, 2007.

[8] Y. A. Alsultanny, Labor market forecasting by using data mining, *Procedia Computer Science*, vol.18, pp.1700-1709, 2013.

[9] Y. K. Liu, C. L. Xie et al., Research data processing and attribute reduction algorithm based data mining of nuclear power plant failure, *Nuclear Power Engineering*, vol.31, pp.24-27, 2010.

[10] Z. Y. Wang and G. J. Klir, *Generalized Measure Theory*, Springer-Verlag, New York, 2008.

[11] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, 1991.

[12] W. X. Zhang and G. F. Qiu, *Uncertain Decision Making Based on Rough Sets*, Tsinghua University Press, 2005.

[13] D. Wei, *Rough Set Theory and Its Applications in Data Mining*, Northeastern University Press, 2009.