ORIGINAL ARTICLES

ACQUISITION OF *KANSEI* DECISION RULES OF COFFEE FLAVOR USING ROUGH SET METHOD

Tatsuo NISHINO^{*}, Mitsuo NAGAMACHI^{*} and Masatoshi SAKAWA^{**}

^{*}Department of Kansei Information, Hiroshima International University, 555-36 Kurose, Higashi-Hiroshima, 724-0695, Japan

^{**}Department of Artificial Complex Systems Engineering, Graduate School of Engineering, Hiroshima University, 1-3-2 Kagamiyama, Higashi-Hiroshima, 739-8511, Japan

Abstract: The acquisition of effective decision rules between design elements of products and human evaluations is significant, but difficult especially in the cases where we have to handle linearly inseparable and much ambiguous data. We have proposed the rough set method to handle these cases. In this paper, we also propose the utilization of gain chart that can visualize a relation between the effects of product attributes on decision and the number of evaluation events. Using the gain chart, we can set up a suitable parameter corresponding with gain property of each evaluation word. We show an application of our proposed rough set method to the extraction of decision rules between coffee flavor evaluations, i.e., taste and aroma, and the pattern of coffee manufacturing conditions. The results showed that our rough set method enabled more effectively to extract the combination rules of manufacturing conditions from much ambiguous data like sense and feeling. We also found out that it is easy to obtain more general rules by β -upper approximation as well as β -lower approximation.

Keywords: *Rough set method, Extraction of probabilistic decision rule, Ambiguity, Information gain, Gain chart*

1. INTRODUCTION

Rough sets theory was introduced as mathematical theory to handle uncertain or inconsistent data [1]. Since it is superior in dealing with linearly inseparable data, it has been used to extract the decision rules in many application areas, and its effectiveness has been shown.

On the other hand, one of the core technologies in *Kansei Engineering* (KE) is to identify the relational rules between design elements of products and human evaluation data such as sense and feeling [2]. So far, we have applied rough set methods to extract decision rules between human evaluation data and design elements from kansei experimental data set [3-5]. The extracted rules would enable product designer to design the products fitted to the sense of human. Recently, it has been shown that rough set approaches are very effective to extract human decision rules in KE [6].

However, the original rough sets approach is restricted to the case where there exist fully correct and certain classifications derived from the decision table. Unfortunately, we have many cases where there is no lower approximation of a classification. Consequently, combining rough sets approaches and probability concept, many research papers [7-10] have been published.

Human evaluation data such as sense or feeling is dependent on individual perception and cognition, so that

ambiguity appears in the variation of individual evaluation to product sets. If we consider the properties of human evaluation data such as ambiguity and linearly inseparability, we would notice that we have to treat the case where there is no lower approximation of a classification, and the case where the decision class occurs with different probabilities. However, there is few rough sets approach to kansei engineering to deal with the ambiguity and probabilistic properties of the human evaluation.

Consequently, we have developed the rough set method that can properly handle with ambiguity, and can extract the decision rules in the form of uncertain *if-then* rules [11-12]. Our method inspired by VPBRs-model [9-10] is based on a new information gain to an equivalent class suitable for handling totally ambiguous and probabilistic properties of human evaluation data. We also propose a two-stage method for simply extracting uncertain decision rules from probabilistic decision table, using decision functions of approximated classes. In this paper, we will also show the utility of the gain chart that supports to set up a suitable parameter of the model in practical applications.

Accordingly, the purpose of this paper is to apply our proposed rough set method to the extraction of decision rules from coffee flavor evaluation data, i.e., taste and aroma, in a coffee manufacturing company, and to ex-

amine the effectiveness of our proposed rough set method.

The paper is organized as follows: in Section 2 our rough set method is introduced with a simple example: we describe an application of our proposed rough set method to coffee flavor evaluation data in Section 3; Section 4 describes and discusses the extraction and evaluation of decision rules; final Section is conclusions and future work.

2. PROPOSED ROUGH SET METHOD

First of all, let us introduce our proposed rough set method. Our proposed rough set method is the following:

Step 1. Construction of data table

Step 2. Approximation of decision class

Step 3. Extraction of decision rules

Step 4. Evaluation of decision rules

In the following subsection, each step will be described with a simple example so that you can easily understand our rough set approach. Refer to the reference [11-12] about the theoretical side of our method.

2.1 Construction of data table

A simple example of human evaluation data to products is shown in Table 1, where a set of products, a set of design attributes (conditional attributes) of products and human evaluation (decision attribute) to products are denoted as $E = \{E_1, E_2, E_3, E_4\}$, $A = \{a_1, a_2, a_3\}$ and d , respectively.

Table 1: An example of evaluation data

Products	$U(\text{events})$	a_1	a_2	a_3	d
E_1	x_{11}	1	0	0	0
	x_{12}	1	0	0	0
	x_{13}	1	0	0	0
	x_{14}	1	0	0	0
	x_{15}	1	0	0	1
E_2	x_{21}	1	1	1	1
	x_{22}	1	1	1	0
	x_{23}	1	1	1	0
	x_{24}	1	1	1	1
	x_{25}	1	1	1	1
E_3	x_{31}	2	0	0	1
	x_{32}	2	0	0	1
	x_{33}	2	0	0	0
	x_{34}	2	0	0	0
	x_{35}	2	0	0	0
E_4	x_{41}	1	0	1	0
	x_{42}	1	0	1	0
	x_{43}	1	0	1	1
	x_{44}	1	0	1	0
	x_{45}	1	0	1	0

An evaluation event of j -th evaluator to i -th product is denoted as x_{ji} . There are four products and five human evaluators. Assuming that equivalence relation R is defined on U , E_i becomes an equivalent class with respect to the set A of conditional attributes because the same product has the same attribute values. Each attribute of A has a domain of its design attribute values, $V_{a_1} = \{1, 2\}$, $V_{a_2} = \{0, 1\}$ and $V_{a_3} = \{0, 1\}$. Human evaluation d has also a domain of its evaluation values, $V_d = \{0, 1\}$, which may be "good" and "no good". Any conditional attribute a_k is a mapping function $a_k(x) = v_k$ and has a set of its values, V_{a_k} . A decision attribute d is also a mapping function $d(x) = v_d$ and it has V_d . A set of decision classes is $D = \{D_0, D_1, D_2\}$ where $D_j = \{x | d(x) = j\}$, $j = 0, 1$.

Original rough set defines a lower and an upper approximation of concepts D_j in the following:

$$\begin{aligned} R_-(D_j) &= \{E_i | E_i \subseteq D_j\} \\ R_+(D_j) &= \{E_i | E_i \cap D_j \neq \emptyset\} \end{aligned} \quad (1)$$

In Table 1, we have:

$$R_-(D_0) = \emptyset, R_+(D_0) = \{E_1, E_2, E_3, E_4\} = U.$$

It should be noted that there is no lower approximation to any decision class, and then we cannot derive decision rules from Table 1 based on (1). Thus, we define an approximate lower approximation of decision class by introducing the conditional probability and the information gain of product E_i to decision class D_j .

Table 1 will be used to illustrate our approach with a numerical example.

2.2 Approximation of decision class based on information gain

The evaluation data includes at least two important probabilistic aspects. One is the probability of decisions dependent on the attributes of product E_i , and the other is the prior probability of decision class. Such probabilities are experientially acceptable in human evaluation data. These probabilities are well known as the conditional and prior probability respectively.

$$P(D_j | E_i) = \frac{|D_j \cap E_i|}{|E_i|}, \quad (2)$$

$$P(D_j) = \frac{|D_j|}{|U|}, \quad (3)$$

where $|\cdot|$ denotes cardinality.

According to the parameterized version of Bayesian Rough Set (BRS) model [9], let us consider the difference between probabilities $P(D_j)$ and $P(D_j|E_i)$ as a kind of information gain. We define the information gain of E_i with $P(D_j|E_i) \geq P(D_j)$ denoted as

$$G_{pos}(D_j, E_i) = 1 - \frac{P(D_j)}{P(D_j | E_i)}, \quad (4)$$

which means that the larger the conditional probability is, the larger the information gain is. The definition of information gain corresponds with our intuition that the larger increment of $P(D_j | E_i)$ being more than $P(D_j)$ should take larger information gain when $P(D_j)$ is low, while the larger increment of $P(D_j | E_i)$ should take smaller information gain when $P(D_j)$ is high.

Let the set of E_i with $P(D_j | E_i) \geq P(D_j)$ and the set of E_j with $P(D_j | E_i) \leq P(D_j)$ denote as POS region and NEG region of decision class D_j , respectively.

Assuming the symmetry between POS and NEG region, we have:

$$G_{neg}(\neg D_j, E_j) = 1 - \frac{P(\neg D_j)}{P(\neg D_j | E_j)}. \quad (5)$$

For example, from Table 1 we have Table 2.

Table 2: Conditional probabilities and gains

Product	$P(D_0 E)$	$G(D_0, E)$	$P(D_0)$
E_1	0.8	0.31	0.55
E_2	0.4	0.25	
E_3	0.6	0.08	
E_4	0.8	0.31	

Let us define the positive region by using the information gain with the parameter β as

$$POS^\beta(D_j) = \bigcup \{E_i | G_{pos}(D_j | E_i) \geq \beta\} \\ = \bigcup \{E_i | P(D_j | E_i) \geq \frac{P(D_j)}{1-\beta}\}, \quad (6)$$

which means the region that E_i would belong possibly to D_j with β .

Using duality of rough set $POS^\beta(D_j) = NEG^\beta(\neg D_j)$, we obtain automatically:

$$NEG^\beta(D_j) = \bigcup \{E_i | P(D_j | E_i) \leq \frac{P(D_j) - \beta}{1-\beta}\}, \quad (7)$$

which means the region that E_i would not belong possibly to D_j with β .

$$BND^\beta(D_j) = \bigcup \{E_i | P(D_j | E_i) \in (\frac{P(D_j) - \beta}{1-\beta}, \frac{P(D_j)}{1-\beta})\}, \quad (8)$$

which means the region that E_i would belong to neither of D_j or $\neg D_j$ with β .

In addition, we can define:

$$UPP^\beta(D_j) = POS^\beta(D_j) \cup BND^\beta(D_j), \quad (9)$$

which means the possible region being larger than $POS^\beta(D_j)$. We call $POS^\beta(D_j)$ and $NEG(D_j)^\beta$ as β -lower approximation, and $UPP^\beta(D_j)$ as β -upper approximation.

As the value of β increases, the positive and negative regions decrease, and boundary region increases. Furthermore, as the value of β increases, the information associated with D_j is strongly relevant to E_i .

For example, from Table 1 we have:

$$POS^{0.1}(D_j) = \bigcup \{E_i | P(D_0 | E_i) \geq 0.61\} = E_1 \cup E_4 \\ NEG^{0.1}(D_j) = \bigcup \{E_i | P(D_0 | E_i) \leq 0.5\} = E_2 \\ BND^{0.1}(D_j) = \bigcup \{E_i | P(D_0 | E_i) \in (0.5, 0.61)\} = E_3 \\ UPP^{0.1}(D_j) = \bigcup \{E_i | P(D_0 | E_i) \geq 0.5\} = E_1 \cup E_3 \cup E_4$$

2.3 Extraction method of decision rules from approximate regions

A decision matrix with respect to $POS^\beta(D_j)$ can be described as Table 3. The element in the matrix $M_{ij}^\beta(D_j)$ is defined:

$$M_{ij}^\beta(D_j) = \{\bigvee a_k = v_{ki} | a_k(E_i) \neq a_k(E_j), \forall a_k \in A\}, \quad (10)$$

where $\bigvee a_k = v_{ki}$ indicates a disjunction of the attribute value of E_i to discern E_i from E_j .

Table 3: A decision matrix from approximate regions

		$NEG^\beta(D_j)$	$BND^\beta(D_j)$
		$E_{n1} \dots E_j$	$E_{B1} \dots E_{bm}$
$POS^\beta(D_j)$	E_{p1}	$M_{ij}^\beta(D_j)$	
	$\vdots$		
	E_i		
	$\vdots$		
	E_{pn}		

We can derive minimal decision rules as conjunctions of the following decision function:

$$POS^{\beta-rule}(D_j) = \bigvee_{E_i \in POS^\beta(D_j)} \bigwedge_{E_j \notin POS^\beta(D_j)} M_{ij}^\beta(D_j). \quad (11)$$

For example, we have Table 4 from Table 1. Then, we have the following decision function:

$$POS^{0.1-rule}(D_0) = \{(a_2 = 0 \vee a_3 = 0) \wedge (a_1 = 1)\} \\ \vee \{(a_2 = 0) \wedge (a_2 = 1 \vee a_3 = 1)\} \\ = (a_2 = 0 \wedge a_1 = 1) \vee (a_3 = 0 \wedge a_1 = 1) \\ \vee (a_2 = 0 \wedge a_1 = 1) \vee (a_2 = 1 \vee a_3 = 1) \\ = (a_2 = 0 \wedge a_1 = 1) \vee (a_3 = 0 \wedge a_1 = 1) \vee (a_2 = 1 \vee a_3 = 1)$$

Lastly, we have the *if-then* decision rules:

Table 4: Decision matrix from $POS^{\beta}(D_j)$

		$NEG^{0.1}(D_0)$	$BND^{0.1}(D_0)$
		E_2	E_3
$POS^{0.1}(D_0)$	E_1	$a_2=0 \vee a_3=0$	$a_1=1$
	E_4	$a_2=0$	$a_2=1 \vee a_3=1$

r_1 : if $a_1 = 1$ and $a_2 = 0$ then $POS^{0.1}(D_1) \{E_1, E_2\}$

r_2 : if $a_1 = 1$ and $a_3 = 0$ then $POS^{0.1}(D_1) \{E_1\}$

r_3 : if $a_2 = 0$ and $a_3 = 1$ then $POS^{0.1}(D_1) \{E_2\}$

2.4. Evaluation factors of the extracted rules

With considering indexes in [8], we can define the following three evaluation factors of rule in the context of our applications by using the number of evaluation to products $|E_i|$ and the effects of products on decision $P(D_j|E_i)$. Let $Cond_k$ be a set of the equivalence classes E_i matched with the condition part of the extracted rule $rule_k$ represented as $cond_k \rightarrow D_j$.

The following *certainty factor* denoted as $cer(Cond_k; D_j)$ means the ratio of the number of events satisfied with if-then rule to the number of events satisfied with the condition part $cond_k$ of the rule.

$$cer(Cond_k : D_j) = \frac{\sum_{E_i \in Cond_k} |E_i| \cdot P(D_j | E_i)}{\sum_{E_i \in Cond_k} |E_i|}. \quad (12)$$

This certainty factor shows the degree to which $cond_k \rightarrow D_j$ holds. It should be noted that we derive if-then rules from $POS^{\beta}(D_j)$. In our applications, we can use this factor as confidence degree of decision to predict the human evaluation from any product design elements.

Inversely, when we have to estimate the attribute values of the product candidates from targeted human evaluation, the following *coverage factor* will be useful.

$$cov(Cond_k : D_j) = \frac{\sum_{E_i \in Cond_k} |E_i| \cdot P(D_j | E_i)}{|D_j|}, \quad (13)$$

which means the ratio of the number of events satisfied with constructed rule to the number of the events satisfied with D_j . This factor shows the degree to which $D_j \rightarrow cond_k$, i.e., the inverse of rule holds.

The following *strength factor* can be used to evaluate the set of decision rules.

$$\sigma(Cond_k : D_j) = \frac{\sum_{E_i \in Cond_k} |E_i| \cdot P(D_j | E_i)}{|U|}, \quad (14)$$

which means the ratio of the number of events satisfied with if-then rule to all the events.

In similar way, we can associate if-then rule from $NEG^{\beta}(D_j)$, $BND^{\beta}(D_j)$ or $UPP^{\beta}(D_j)$ with three factors mentioned above.

For example, the rule r_1 has the following values:

$$\begin{aligned} cer(E_1, E_2; D_0) &= \frac{|E_1| \cdot P(D_0 | E_1) + |E_2| \cdot P(D_0 | E_2)}{|E_1| + |E_2|} \\ &= \frac{5 \cdot 0.8 + 5 \cdot 0.8}{10} = 0.8 \end{aligned}$$

$$cov(E_1, E_2; D_0) = \frac{8}{|D_0|} = \frac{8}{13} = 0.62$$

$$\sigma(E_1, E_2; D_0) = \frac{8}{|U|} = \frac{8}{20} = 0.40$$

2.5 Setting up the parameter β

In order to set a suitable parameter β in applications, we propose to use a gain chart, which visualizes the relation between the gain value $G(D_j, E_i)$ of equivalent class and the accumulated percent α of events.

Fig.1: An example of gain chart

Fig. 1 illustrates the gain chart for Table 1. The gains were changed into descending order. We can define the precision of approximation by β as follows:

$$\begin{aligned} \gamma_{\beta} &= 1 - \alpha_{\beta} \\ &= 1 - \frac{\sum_{E_i \in POS^{\beta}(D_j)} |E_i| + \sum_{E_i \in NEG^{\beta}(D_j)} |E_i|}{|U|}, \end{aligned} \quad (15)$$

where the coverage of approximation denoted as α_{β} means the ratio of the sizes of positive and negative regions covered by the value of β to the total number of events. The larger the value of β is, the larger the value of γ_{β} is, but the smaller the value of α_{β} is. It should be noted that the relation between γ_{β} or β and α_{β} is trade-off.

For example, the arrow dotted line in the Fig.1 shows that by setting β to 0.25 or less (β_0), we can extract the positive and negative rules with the coverage of approximation beyond 75% (α_{β_0}). In order to derive the

least 45% ($\alpha_{\beta 1}$) conversely, you should set β to 0.31 or less (β_1), as shown by the arrow solid line.

It should be noted that as β increases, positive and negative regions shrink, while boundary region expands. Thus, as β increases, the average of certainties of the positive and negative rules becomes higher, but the average of coverages becomes lower. Inversely, as β decreases, boundary region expands, while positive and negative regions shrink. In this lower β case, we can derive positive and negative rules with lower certainty, but with higher coverage.

Generally, when there are small samples in applications, you might set up lower β so that you can extract the rules covering many events as much as possible. Using the gain chart for each sensory word, we can set up the suitable β according to the gain feature of each decision class.

3. APPLICATION TO COFFEE TASTE EVALUATION DATA

3.1 Coffee evaluation processes

In this application, we hypothesize the hierarchical processes of coffee evaluations from coffee manufacturing conditions to behavior process as shown Fig. 2. Three middle levels of the hierarchy show sensory, feeling and attitudinal processes of evaluators to products,

Fig 2: Hypothetical process of coffee evaluation

and it is assumed that these processes are measured by such *kansei* words as shown in Fig. 2. These words were selected by a coffee-manufacturing expert from the viewpoints of significant words to evaluate coffee flavor such as taste and aroma. In this paper, we will focus on the relations between sensory process and coffee manufacturing conditions. We assume that not only one of manufacturing conditions, but the pattern of manufacturing conditions affects the sensory process.

Accordingly, it is our purpose to find out if-then decision rules between sensory words such as aroma, fruity, bitter, sweet, sour, and harsh, and the patterns of manufacturing conditions such as the kinds of raw coffee beans and the roast levels.

The extracted if-then decision rules are easily understandable to human so that these rules would be useful for coffee manufacturing company to manufacture the coffee fitted to human sense of coffee by using these decision rules.

3.2 Experiment

We carried out an experiment to identify hidden relations between significant coffee manufacturing conditions and human sensory evaluations. The manufacturing conditions were the combinations of two conditional attributes of the kinds of raw coffee beans (a_1) and its roast level (a_2). $V_{a_1} = \{Colombia\ Excelsio, Brazil\ No2\ s\ 17/18, Mocha\ Lekempti\}$, which are the typical kinds of coffee raw beans and $V_{a_2} = \{Light, Medium, French\}$ whose roast levels were controlled by roast time of roast machine.

A coffee manufacturing expert made the 9-kinds of coffees by combining V_{a_1} and V_{a_2} , and 10-evaluators (4-male and 6-female) evaluated them on 5-points semantic differential (SD) scales about 12-*kansei* words such as ‘aroma’, ‘fruity’, ‘soft’, ‘light’, ‘like’ and ‘want to buy’ shown in Fig. 2.

In this paper, we attempt to derive the decision rules between these sensory words and the coffee manufacturing conditions from coffee evaluation data.

3.3 Evaluation data for analysis

The obtained data table for aroma is shown in Table 3 where E_i denotes one of the kinds of coffees, x_{ji} denotes an evaluation event of j -th evaluator to i -th product.

The kinds of coffee raw beans, *Colombia*, *Brazil*, and *Mocha*, are abbreviated as *C*, *B* and *M*, respectively, and roast levels, *light*, *medium* and *french* as *L*, *E*, and *F*. The last column shows the re-categorized classes of aroma. In

Table5: Data table of aroma

Products	$U(\text{events})$	Beans (a_1)	Roast (a_2)	Aroma
E_1	x_{11}	C	L	3
	x_{12}	C	L	1
	x_{13}	C	L	2
	x_{14}	C	L	2
	x_{15}	C	L	2
	x_{16}	C	L	2
	x_{17}	C	L	2
	x_{18}	C	L	1
	x_{19}	C	L	1
	$x_{1,10}$	C	L	1
E_2	x_{21}	C	E	1
	x_{22}	C	E	1
	x_{23}	C	E	1
	x_{24}	C	E	2
	x_{25}	C	E	1
	x_{26}	C	E	1
	x_{27}	C	E	2
	x_{28}	C	E	1
	x_{29}	C	E	2
	$x_{2,10}$	C	E	1
	:	:	:	:
	:	:	:	:
	:	:	:	:
	:	:	:	:
	:	:	:	:
E_9	x_{91}	B	F	1
	x_{92}	B	F	2
	x_{93}	B	F	2
	x_{94}	B	F	1
	x_{95}	B	F	1
	x_{96}	B	F	1
	x_{97}	B	F	2
	x_{98}	B	F	1
	x_{99}	B	F	1
	$x_{9,10}$	B	F	1

order to clear the partition of evaluations, 4or 5 point in SD scale was re-categorized as 1, 1 or 2 as 2, and 3 as 3. Next, the events with 3 point were excluded from data table. We obtained the data consisting of 78 evaluation events for analysis. Thus, we set two decision classes $D = \{\text{Good aroma } (D_1), \text{No good aroma } (D_2)\}$ for the rough set analysis. In similar way, we obtained the data tables for decision classes for another 5-sensory words, and set two decision classes for each sensory word.

3.4 Approximation of decision class

We obtained the gain chart of each word as shown in Fig. 3. In this application, we derive decision rules so

that the coverage of approximation α_β is more than 80% because the number of product samples is relatively smaller. From Fig. 3, we obtained the following value of β for each word: $\beta_{\text{aroma}} = 0.07$, $\beta_{\text{fruity}} = 0.14$, $\beta_{\text{bitter}} = 0.14$, $\beta_{\text{harsh}} = 0.07$, $\beta_{\text{sweet}} = 0.04$, $\beta_{\text{sour}} = 0.05$.

The prior probability of each word was the following:

$$\begin{aligned}
 P(\text{Aroma} = \text{good}) &= 0.72 & P(\text{Aroma} = \text{no good}) &= 0.28 \\
 P(\text{Fruity} = \text{yes}) &= 0.36 & P(\text{Fruity} = \text{no}) &= 0.64 \\
 P(\text{Bitter} = \text{yes}) &= 0.66 & P(\text{Bitter} = \text{no}) &= 0.34 \\
 P(\text{Harsh} = \text{yes}) &= 0.63 & P(\text{Harsh} = \text{no}) &= 0.37 \\
 P(\text{Sweet} = \text{yes}) &= 0.43 & P(\text{Sweet} = \text{no}) &= 0.67 \\
 P(\text{Sour} = \text{yes}) &= 0.66 & P(\text{Sour} = \text{no}) &= 0.34
 \end{aligned}$$

You can easily see that good aroma, bitter, harsh and sour tastes are higher prior probabilities than others.

Fig.3: Gain chart of sensory evaluations

The average gain of each word is as follows:

$$\begin{aligned}
 \bar{G}(\text{Aroma}) &= 0.267 & \bar{G}(\text{Fruity}) &= 0.184 \\
 \bar{G}(\text{Bitter}) &= 0.300 & \bar{G}(\text{Harsh}) &= 0.205 \\
 \bar{G}(\text{Sweet}) &= 0.172 & \bar{G}(\text{Sour}) &= 0.159
 \end{aligned}$$

We can interpret the average of gains as a kind of sensitivity measure of sensory word. In this case, the bitter and aroma tastes have higher sensitivity than the sour and sweet tastes.

Table 6 shows the approximated decision table of aroma where the rows indicate products, and the columns indicate the manufacturing conditions, the number of evaluation cases, the number of each decision class, the value of conditional probability to each decision class, the value of the gain, and the approximated region when $\beta_{\text{aroma}} = 0.07$. It should be noted that the number of cases excludes the events with 3 point on the SD scale to discriminate the decision classes clearly.

Table 6: Approximation of good aroma

Product	Beans (a_1)	Roast (a_2)	No. of case	No. of D_0	No. of D_1	$P(D_0 E)$	$P(D_1 E)$	$G(D_0, E)$	$\beta=0.07$
E_1	C	L	9	4	5	0.44	0.56	0.49	NEG
E_2	C	E	10	8	2	0.80	0.20	0.10	POS
E_3	C	F	10	9	1	0.90	0.10	0.20	POS
E_4	B	L	9	4	5	0.44	0.56	0.49	NEG
E_5	B	E	7	6	1	0.86	0.14	0.16	POS
E_6	B	F	9	9	0	1.00	0.00	0.28	POS
E_7	M	L	5	2	3	0.40	0.60	0.53	NEG
E_8	M	E	9	7	2	0.78	0.22	0.08	POS
E_9	M	F	10	7	3	0.70	0.30	0.06	NEG

Table 7: Approximation of fruity taste (yes)

Product	a_1	a_2	$P(D_0 E)$	$G(D_0, E)$	$\beta=0.14$
E_1	C	L	0.56	0.35	POS
E_2	C	E	0.43	0.15	POS
E_3	C	F	0.43	0.15	POS
E_4	B	L	0.29	0.11	BND
E_5	B	E	0.43	0.15	POS
E_6	B	F	0.25	0.15	NEG
E_7	M	L	0.44	0.18	POS
E_8	M	E	0.25	0.15	NEG
E_9	M	F	0.14	0.26	NEG

Table 10: Approximation of sweet taste (yes)

Product	a_1	a_2	$P(D_0 E)$	$G(D_0, E)$	$\beta=0.04$
E_1	C	L	0.44	0.03	BND
E_2	C	E	0.44	0.03	BND
E_3	C	F	0.38	0.09	NEG
E_4	B	L	0.44	0.03	BND
E_5	B	E	0.70	0.39	POS
E_6	B	F	0.56	0.23	POS
E_7	M	L	0.50	0.14	POS
E_8	M	E	0.13	0.35	NEG
E_9	M	F	0.22	0.27	NEG

Table 8: Approximation of bitter taste (yes)

Product	a_1	a_2	$P(D_0 E)$	$G(D_0, E)$	$\beta=0.14$
E_1	C	L	0.40	0.43	NEG
E_2	C	E	0.70	0.06	BND
E_3	C	F	0.88	0.25	POS
E_4	B	L	0.60	0.15	NEG
E_5	B	E	0.50	0.32	NEG
E_6	B	F	0.90	0.27	POS
E_7	M	L	0.11	0.62	NEG
E_8	M	E	1.00	0.34	POS
E_9	M	F	0.90	0.27	POS

Table 11 Approximation of sour taste (yes)

Product	a_1	a_2	$P(D_0 E)$	$G(D_0, E)$	$\beta=0.05$
E_1	C	L	0.88	0.25	POS
E_2	C	E	0.63	0.09	NEG
E_3	C	F	0.70	0.06	POS
E_4	B	L	0.78	0.15	POS
E_5	B	E	0.70	0.06	POS
E_6	B	F	0.67	0.01	POS
E_7	M	L	0.40	0.43	NEG
E_8	M	E	0.70	0.06	POS
E_9	M	F	0.50	0.32	NEG

Table 9: Approximation of harsh taste (yes)

Product	a_1	a_2	$P(D_0 E)$	$G(D_0, E)$	$\beta=0.05$
E_1	C	L	0.40	0.38	NEG
E_2	C	E	0.80	0.21	POS
E_3	C	F	0.78	0.19	POS
E_4	B	L	0.67	0.05	POS
E_5	B	E	0.60	0.08	NEG
E_6	B	F	0.60	0.08	NEG
E_7	M	L	0.25	0.51	NEG
E_8	M	E	0.89	0.29	POS
E_9	M	F	0.67	0.05	POS

Using (6), (7) and (8) in Section 2, the products with being $P(D_0|E_i) > P(D_1|E_i)$, $\{E_2, E_3, E_5, E_6, E_8\}$ are approximated to *POS* region, i.e., the decision class $\{Aroma = good\}$, while the products with being $P(D_1|E_i) > P(D_0|E_i)$, $\{E_1, E_4, E_7\}$ were approximated to *NEG* region, i.e., $\{Aroma = no\ good\}$. The product E_9 was approximated to *BND* region. If *BND* region is regarded as *POS* region, we have *UPP* region. In similar way, we obtained the approximation tables for other 5 sensory kansei words as shown from Table 7 to Table 11.

4. EXTRACTION AND EVALUATION OF DECISION RULES

4.1 Extraction of decision rules from β -lower approximation

We derived the decision rules about 6 sensory words. Using the following values of β , we extracted the decision rules with the coverage of approximation being $\alpha_\beta = 0.8$.

$\beta_{aroma} = 0.07$, $\beta_{fruity} = 0.14$, $\beta_{bitter} = 0.14$, $\beta_{harsh} = 0.07$, $\beta_{sweet} = 0.04$, $\beta_{sour} = 0.05$.

Next, we will show the extracted decision rule of each word. We have the following decision rules of aroma. The evaluation factors are shown in Table 12.

$ar\ r_1$: If Roast = medium, Then Aroma = good
 $ar\ r_2$: If Beans = Colombia, Roast = french, Then Aroma = good
 $ar\ r_3$: If Beans = Brazil, Roast = french, Then Aroma = good
 $ar\ r_4$: If Roast = light, Then Aroma = no good
 $ar\ r_5$: If Beans = Mocha, Roast = french, Then Aroma = good or no

Table 12: The decision rules by $\beta_{aroma} = 0.07$

	Certainty	Coverage	Strength
$ar\ r_1$	0.81	0.38	0.27
$ar\ r_2$	0.90	0.16	0.12
$ar\ r_3$	1.00	0.18	0.13
$ar\ r_4$	0.57	0.59	0.17
$ar\ r_5$	1.00	0.13	0.13

The $ar\ r_1$ has higher value of coverage indicating more general rule, and implicates that 81% of medium roast coffees has good aroma, and inversely 38% of good aroma coffees is medium roast. The $ar\ r_2$ and $ar\ r_3$ with lower value of coverage indicate more specific rules, and it implicates that the combination condition of *Brazil* and *french* as well as that of *Colombia* and *french* has good aroma. In similar way, the negative rule r_4 implicates that 57% of light roast coffees has no good aroma, and inversely 57 % of good aroma coffees is light. The sum

of strength values of these rules including those of boundary decision rules is 82%. This means that 82% of obtained data would be correctly classified by these rules.

We have the rules of fruity taste with Table 13:

$fr\ r_1$: If Beans = Colombia, Then Fruity = yes
 $fr\ r_2$: If Beans = Brazil, Roast = medium, Then Fruity = yes
 $fr\ r_3$: If Beans = Mocha, Roast = light, Then Fruity = yes
 $fr\ r_4$: If Beans = Brazil, Roast = french, Then Fruity = no
 $fr\ r_5$: If Beans = Mocha, Roast = medium, Then Fruity = no
 $fr\ r_6$: If Beans = Mocha, Roast = french, Then Fruity = no
 $fr\ r_7$: If Beans = Brazil, Roast = light, Then Fruity = yes or no

Table 13: The decision rules by $\beta_{fruity} = 0.14$

	Certainty	Coverage	Strength
$fr\ r_1$	0.48	0.44	0.16
$fr\ r_2$	0.43	0.12	0.04
$fr\ r_3$	0.44	0.16	0.06
$fr\ r_4$	0.75	0.14	0.09
$fr\ r_5$	0.75	0.14	0.09
$fr\ r_6$	0.86	0.14	0.09
$fr\ r_7$	1.00	0.10	0.10

The rule $fr\ r_1$ with higher coverage implicates that the bean *Colombia* would make coffee fruity. It should be noticed that its certainty 0.48 is higher than its prior probability 0.36. However, we can see many combination rules.

We have the rules of bitter with Table 14:

Table 14: The decision rules by $\beta_{bitter} = 0.14$

	Certainty	Coverage	Strength
$bi\ r_1$	0.89	0.45	0.29
$bi\ r_2$	1.00	0.14	0.09
$bi\ r_3$	0.62	0.62	0.21
$bi\ r_4$	0.50	0.17	0.06
$bi\ r_5$	1.00	0.12	0.12

$bi\ r_1$: If Roast = french, Then Bitter = yes
 $bi\ r_2$: If Beans = Mocha, Roast = medium, Then Bitter = yes
 $bi\ r_3$: If Roast = light, Then Bitter = no
 $bi\ r_4$: If Beans = Brazil, Roast = medium, Then Bitter = no
 $bi\ r_5$: If Beans = Colombia, Roast = medium, Then Bitter = yes or no

The rules $bi\ r_1$ and $bi\ r_3$ are contrary mutually. These rules indicate that the roast time affects bitter taste, strongly. In other words, as the longer the roast time is, the stronger the bitter taste is.

We have the following rules of harsh with Table 15.

$ha\ r_1$: If Beans = Colombia, Roast = medium Then Harsh = yes
 $ha\ r_2$: If Beans = Colombia, Roast = french Then Harsh = yes
 $ha\ r_3$: If Beans = Mocha, Roast = medium Then Harsh = yes
 $ha\ r_4$: If Beans = Colombia, Roast = light Then Harsh = no
 $ha\ r_5$: If Beans = Brazil, Roast = medium, Then Harsh = no
 $ha\ r_6$: If Beans = Brazil, Roast = french, Then Harsh = no
 $ha\ r_7$: If Beans = Mocha, Roast = light, Then Harsh = no
 $ha\ r_8$: If Beans = Brazil, Roast = light, Then Harsh = yes or no
 $ha\ r_9$: If Beans = Mocha, Roast = french, Then Harsh = yes or no

Table 15: The decision rules by $\beta_{\text{harsh}} = 0.07$.

	Certainty	Coverage	Strength
$ha\ r_1$	0.80	0.15	0.10
$ha\ r_2$	0.78	0.13	0.08
$ha\ r_3$	0.89	0.15	0.10
$ha\ r_4$	0.60	0.19	0.07
$ha\ r_5$	0.40	0.13	0.05
$ha\ r_6$	0.40	0.13	0.05
$ha\ r_7$	0.75	0.19	0.07
$ha\ r_8$	1.00	0.11	0.11
$ha\ r_9$	1.00	0.11	0.11

The rule $ha\ r_3$ with higher certainty implicates that the combination of *Mocha* and *medium* would affect harsh taste strongly. Inversely, the combination of *Colombia* and *light* would not affect harsh taste. We cannot see any general rule about harsh taste because one of any conditions is not included in one of any decision classes completely. It seems that there exist some linearly inseparable interactions between the kinds of coffee beans and the roast levels for harsh taste. This suggests that the sense of harsh is much linearly inseparable compared to the other sensory words mentioned above.

We have the following rules of sweet with Table 16.

$sw\ r_1$: If Beans = Brazil, Roast = medium Then Sweet = yes
 $sw\ r_2$: If Beans = Brazil, Roast = french Then Sweet = yes
 $sw\ r_3$: If Beans = Mocha, Roast = light Then Sweet = yes
 $sw\ r_4$: If Beans = Colombia, Roast = french Then Sweet = no
 $sw\ r_5$: If Beans = Mocha, Roast = medium Then Sweet = no
 $sw\ r_6$: If Beans = Mocha, Roast = french Then Sweet = no
 $sw\ r_7$: If Beans = Colombia, Roast = light Then Sweet = yes or no
 $sw\ r_8$: If Beans = Colombia, Roast = medium Then Sweet = yes or no
 $sw\ r_9$: If Beans = Brazil, Roast = light Then Sweet = yes or no

Table 16: The decision rules by $\beta_{\text{sweet}} = 0.04$.

	Certainty	Coverage	Strength
$sw\ r_1$	0.70	0.21	0.09
$sw\ r_2$	0.56	0.15	0.06
$sw\ r_3$	0.50	0.12	0.05
$sw\ r_4$	0.63	0.11	0.06
$sw\ r_5$	0.88	0.16	0.09
$sw\ r_6$	0.78	0.16	0.09
$sw\ r_7$	1.00	0.11	0.11
$sw\ r_8$	1.00	0.11	0.11
$sw\ r_9$	1.00	0.11	0.11

The rule $sw\ r_1$ indicates that the combination of *Brazil*

and *medium* would be significant for the sweet taste. These rules also suggest that sweet taste would seemingly be affected according to inseparable conditions like harsh taste.

We have the rules of sour with Table 17:

$so\ r_1$: If Beans = Colombia, Roast = light Then Sour = yes
 $so\ r_2$: If Beans = Colombia, Roast = french Then Sour = yes
 $so\ r_3$: If Beans = Brazil, Roast = light Then Sour = yes
 $so\ r_4$: If Beans = Brazil, Roast = medium Then Sour = yes
 $so\ r_5$: If Beans = Mocha, Roast = medium Then Sour = yes
 $so\ r_6$: If Beans = Colombia, Roast = medium Then Sour = no
 $so\ r_7$: If Beans = Mocha, Roast = light Then Sour = no
 $so\ r_8$: If Beans = Mocha, Roast = french Then Sour = no
 $so\ r_9$: If Beans = Brazil, Roast = french Then Sour = yes or no

Table 17: The decision rules by $\beta_{\text{sour}} = 0.05$

	Certainty	Coverage	Strength
$so\ r_1$	0.88	0.13	0.09
$so\ r_2$	0.70	0.13	0.09
$so\ r_3$	0.78	0.13	0.09
$so\ r_4$	0.70	0.13	0.09
$so\ r_5$	0.70	0.13	0.09
$so\ r_6$	0.38	0.11	0.04
$so\ r_7$	0.60	0.21	0.07
$so\ r_8$	0.50	0.18	0.06
$so\ r_9$	1.00	0.11	0.11

These rules indicate that the combination of the *Colombia* and *light* effects on strong sour taste, while the combination of *Mocha* and *light* effects on weak sour taste. We cannot see any general rule about sour taste.

4.2 Extraction of decision rules from β -upper approximation

Next, we derived the decision rules from β -upper approximation region $UPP^{\beta}(D_j)$. It considers that $BND^{\beta}(D_j)$ region has the possibility of $POS^{\beta}(D_j)$ or $NEG^{\beta}(D_j)$ region. Generally speaking, the rule extraction by β -upper approximation expands the approximate region of D_j , and it derives the rules with more general and higher coverage.

We show the positive rules with higher coverage than other rules shown above:

$ar'r_1$: If Roast = french Then Aroma = good
 with $cer = 0.86$, $cov = 0.45$, $st = 0.32$

$fr'r_1$: If Roast = light Then Fruity = yes
 with $cer = 0.44$, $cov = 0.44$, $st = 0.16$

$sw'r_1$: If Roast = light Then Sweet = yes
 with $cer = 0.46$, $cov = 0.35$, $st = 0.15$

$sw'r_2$: If Beans = Brazil Then Sweet = yes
 with $cer = 0.57$, $cov = 0.47$, $st = 0.20$

$so r_1 : \text{If Beans} = \text{Brazil Then Sour} = \text{yes}$
with $cer = 0.71$, $cov = 0.35$, $st = 0.23$

It should be noted that only the rule $ar' r_1$ has higher certainty and higher coverage than the other positive rules extracted by β -lower approximation. No general rule about harsh was derived by β -lower and upper approximations. This suggests that harsh taste may have the most linearly inseparable decision rules among 6- sensory words.

Lastly, these obtained decision rules were examined by interviewing with an expert of the coffee manufacturing company.

5. CONCLUSIONS

We introduced a rough set method, which enables to derive decision rules in the cases where there is no lower approximation and where decision classes suppose to occur with different prior probabilities. We also proposed a gain chart with which we can set up a suitable parameter in applications by visualizing the trade-off relation between the precision and coverage of approximation.

Using the proposed rough set method, we tried to derive the decision rules from coffee flavor evaluation data with much ambiguity. As a result, we obtained combination rules as well as general ones. It was shown that the decision rules of sensory words are mostly linearly inseparable, and the extracted decision rules was almost combination rules rather than simple general rules. However, we might need general rules as well as combination ones in applications. Therefore, It was shown that using β -upper approximation, we could extract simple general rules.

We will have some tasks in the future. As for the acquisition of decision rules, for example, in order to extract more effective decision rules of coffee preference, we have to extract the decision rules about feeling, attitude, and behavior process in the hypothetical hierarchical processes shown in Fig. 2. Secondly, as the processes within any level in the hierarchy can suppose to constitute some patterns, so we have to try to derive the decision rules between the patterns of one level and those of another one. Thirdly, we will have the tasks that we apply our rough set method to larger data set and test its effectiveness.

REFERENCES

1. Z. Pawlak: Rough Set Elements, In; L. Polkowski, A. Skowron, Rough Sets in Knowledge Discovery 1, Physica-Verlag, 10-30, (1998).
2. M. Nagamachi: Introduction to Kansei Engineering, Japan Standard Association, (1996).
3. T. Nishino, M. Nagamachi: Extraction of Design Rules for Basic Product Designing Using Rough Set Analysis, Proc. of 14th Triennial Congress of IEA, 3, 515-518, (2003).
4. T. Nishino, M. Nagamachi, S. Ishihara: Rough Set Analysis on Kansei Evaluation of Color; In; M.G. Helander, H.M. Khalid, T. M. Po (Eds.), Proc. of the International Conference on Affective Human Factor Design, Asean Academic press, 109-115, (2001).
5. T. Nishino, M. Nagamachi: Rough Set Analysis on Kansei Evaluation of Color and Kansei Structure, In; M. P. Dahlgaard, J. J. Dahlgaard (Eds.): Proc. of the 4th International QMOD Conference, 543-550, (2001).
6. N. Mori, H. Tanaka, K. Inoue (Eds.): Rough Sets and Kansei, Kaibundo (2004).
7. W. Ziarko; Variable precision rough set model, Journal of Computer and System Sciences, 46, 39-59, (1993).
8. Z. Pawlak: Decision rules, Bayes' rule and rough sets, RSFDGrC 1999, LNAI 1711, Springer Verlag, 1-9, (1999).
9. D. Slezak, W. Ziarko: Variable precision Bayesian rough set model, RSFDGrC 2003, LNAI 2639, Springer Verlag, 312-315, (2003).
10. D. Slezak, W. Ziarko: The investigation of the Bayesian rough set model, Int. J. of Approximate Reasoning, in press.
11. T. Nishino, M. Nagamachi, H. Tanaka: Variable Precision Bayesian Rough Set Model and its Application to Human Evaluation Data, LNAI, Springer, 294-303, (2005).
12. T. Nishino, M. Nagamachi, H. Tanaka: Extraction of Possible Kansei Decision Rules from Kansei Evaluation Data Using Variable Precision Bayesian Rough Set Model, Proc. of Symposium on Rough Sets and Kansei Engineering, 55-58, (2004).