10[th] IMEKO TC7 International Symposium
June 30–July 2, 2004, Saint-Petersburg, Russia

COARSE-GRAINED INFORMATION AND ITS APPLICATION TO A FORMAL THEORY OF MEASUREMENT

Hiroshi Watanabe

Higashi-mukohjima 6—19—10, Sumida-ku, Tokyo, 131-0032, Japan
hwatanab@hkg.odn.ne.jp

Abstract – Measurement is an acquisition of the knowledge of the external world, and there is little doubt that our concept and scientific theories depend strongly on the relations such as correlation and probabilistic dependences we discover in a number of observed data. However, measurement is not determined only by the measured object but depends on various factors, one of which is the degree of the fineness of observation. The present paper is an effort to introduce this subject into a formal framework of measurement science. Taking advantage of entropic measure of interdependence, we examine the nature of information obtained by coarse nominal measurement, where some symbols are put into one group as the same symbol, regarding it as a kind of clustering task, and propose a possible formal approach to group formation procedure as coarse observation of objects. Information-theoretical interdependence is often measured by entropy function. However, in grouping of objects, our result suggests that we should take rather the average entropy as the measure of the degree of interdependence.

Keywords: measurement science, nominal scale, entropic measure of interdependence

1. INTRODUCTION

Measurement is an acquisition of the knowledge of the external world, and there is little doubt that our concept and scientific theories depend strongly on the relationship such as correlation and probabilistic dependences we discover in a great number of observed data. The accuracy of measurement, however, is not determined by only the measured object but depends on various factors, such as instruments of measurement, which may be adopted intentionally or made available accidentally, and on how those instruments are connected and operated. In this sense the structure of measured data depends on the accuracy of measurement, and we have to admit that our theoretical picture of the external world is influenced in a very essential fashon by the additional factors which cannot be attributed to the external world. Thus, from a practical point of view, it is important to choose a suitable degree of the accuracy of observation for the intention of measurement.

Suppose that we observe a stochastic chain, or a sequence of numbers being produced, digit after digit, according to some stochastic rule, assuming that those numbers represent physical states of a system under consideration. Then it can be expected without calculation that grouping of numbers, or states, will cause the information content of the chain to decrease. However, if the original stochastic chain has some redundancy in information-theoretical sense then the loss of information due to grouping would be small, or could be made vanish.

There are innumerable examples where such a grouping of variables plays a decisive role in our understanding of surroundings. In the context of communication theory `noise' has been understood as such that a certain number of emitted symbols are received as an identical symbol at the receiving end. Then the amount of information that can be sent through a communication channel with the `noise' of this kind may be recapitulate from the point view of grouping of variables.

We may call such a grouping of variables *coarse observation* of objects. The present paper of a rather preliminary nature is an effort to introduce this subject of coarse observation into a formal framework of measurement science. In the context of measurement theory, forming a group of similar objects is a function of the nominal scale, while from the point of view of pattern recognition, putting some entities together in accordance with their similarity is nothing but a kind of clustering task. Recently, within a formal framework we have examined the basic nature of nominal measurement, and discussed it as a basis of pattern recognition and clustering [5][6]. Putting forward this formal approach, in the present paper, we examine the properties of information obtained by coarse nominal measurement where some symbols are put into one group as the same symbol, and discuss its implications to some basic problems of measurement.

At the end of 1970s, in relation to his `Fuzzy Sets' Zadeh introduced a notion of information granule [10] to reduce heavy computational burden by representing information in the form of aggregates defined by a certain numerical level of similarity. Aiming at enlarging the scope of granule computing, Pedrycz newly defined the notion of information granule, and discussed the fundamentals of the subject [3]. The following consideration will also include some insights to this respect [7].

2. INFORMATION OBTAINED BY MEASUREMENT

Within the formal framework of measurement theory four types of scales are discerned; nominal, ordinal, interval, and ratio scales. In nominal scaling, numbers are used to represent identity or difference with respect to some attribute, and a measurement in nominal scale is, in effect, a matching or identifying of numbers. In current understanding of measurement theory, therefore, the nominal scale is supposed to be the least informative scale of measurement [1][2]. Recently, however, in an effort to establish a new framework of measurement science, the function of the nominal scale came to be reconsidered, and be received a new treatment as basis of testing, diagnosis, identification, and pattern recognition [4][5]. In this respect, some resarcher argues that measurement in the nominal scale yields, in a sense, more information than measurement in the ratio scale [4]. How should we undertand these two views of the nominal scale?

To answer this problem, it is helpful to consider from where information comes. In a framework of formal measurement theory information is supposed to be provided by *scales*. That is, in our scaling practice, we first evaluate the measured object with respect to some attribute, and assign a number, or a symbol, to it according to some rule. Then we obtain information by referring the symbol to the scale of measurement. In this context, the measurement scale works as so-called *frame of reference*. Intuitively speaking, the more structured such a frame is, the more information we would obtain. Among four scales of measurement, the nominal scale has the most simple structure. It deals with only the equivalence relation, and in this sense nominal measurement is uninformative.

On the other hand, when one argues that nominal measurement is informative, emphasizing its function, one seems to bear in mind some acutual application such as testing and diagnosis, and under this implicit assumption, information via nominal measurement means the consequenses of nominal matching in those applications. Let us take a case of medical diagnosis, for example. In medical diagnosis our chief interest is to know the disease from which a patient suffers, and to decide which treatment we sholud take. In what we are interested is whether the given symptom implies, say, disease A or disease B, and not the `distance' between disease A and disease B, nor `order relation' between them (Rigorously speaking, we do not need even the `name' of the disease!). Once the disease is specified by nominal matching of the result of medical tests, then we choose a suitable prescription for treatment based on our medical knowledge. That is, in this case information obtained by nominal matching is provided not by the formal scale structure but by the system of medical knowledge itself. When one says that nominal measurement is informative he or she is often talking about information derived from some *additional, relevant system of knowledge* via nominal matching, taking such a system as a frame of reference.

This observation suggests that, first, in nominal level of measurement a description of the matching problem under consideration plays an important role. We have to choose, for instance, a suitable set of predicates, and a suitable level of the fineness of description of objects so that we can derive as much information as possible from the additional source of information. Sometimes such a choice may be not determined only by formal, objective factors, but may depend on some subjective, semantic factors such as our preference of predicates and athetic taste, which beyond our formal framework of measurement.

Besides, secondly, in so-called weak measurement, or measurement-like procedure based on some weak scale structure, the subjective factor mentioned above seems to play also a crucial role. In those types of measurement, we have to derive information also from some additional source of information, since the mathematical structure of those scales are limited.

It should be remarked that the subjective factor we have discussed here determines only the scope of measurement, reflecting its intention. The importance of those factors is not necessarily incompatible with the objectivity of measurement, and the importance of calibration.

3. A FORMAL THEORY OF CLUSTERING

3.1. Clustering as coarse observation of objects

In the following, we restrict ourselves to the case where the objects are represented by a suitable set of predicates as binary vectors of *m*-components as follows: Suppose that we are given a set $X = \{x_1, x_2, \ldots, x_n\}$of objects, and a set $Y = \{y_1, y_2, \ldots, y_m\}$ of predicates, or attributes, which are relevant and meaningful to describe the members of *X*, and that each predicate is applied to each object either affirmatively or negatively. That is, we assume for all pairs of x_i and y_j the proposition `the object x_i satisfies, or negates, the predicate y_j' is meaningful whether it is true or false. This object-predicate relationship can be identified with a table *T*, which is called *object-predicate table*, with n rows and m columns, whose (i,j)-component is 1 or 0 according to whether an object x_i affirms or negates a predicate y_j. Such a table may be the result of a set of *m* observations, or the result of an appropriate method of dimensionality reduction. The triplet $<X, Y, T>$ thus defined is sometimes called *object-predicate system*.

Such a formal framework allows us to examine basic properties of classification procedure as a possible reinterpretation of nominal scaling [5][6]. If we regard class formation a sort of coarse observation of objects then we can discuss some information-theoretical aspect of coarse observation within this framework by formulating coarse observation as clustering.

3.2. Entropic measure of interdependence

Suppose that we are given an object-predicate system <*X, Y, T*>. Then we can define formal entropy function for a group of objects as follows: Let us take a subset of *X*, say X_I , and form a subtable of *T*, say T_I, by picking up rows of *T* corresponding to the members of X_I . In such a subtable T_I various kinds of column types would appear, some of which would appear repetitively. Then, denoting the relative frequency of appearing of the *i*-th column type p_i, we can define formal entropy of a group X_I by

$$S(X_I) \equiv -\sum_i p_i \log p_i \tag{1}$$

Let us take another subset X_J from *X*. Then formal entropy of X_J and of $X_I \cup X_J$ are also defined in a similar way, and the *interdependence J* between two subgroups X_I and X_J is given by

$$J(X_I, X_J) \equiv S(X_I) + S(X_J) - S(X_I \cup X_J). \tag{2}$$

This quantity *J* is nothing but a counterpart of redundancy in the communication theory [6], and is supposed to be a good measure of the degree of interdependence between two groups X_I and X_J . If we asked, therefore, to divide the entire set *X* into two groups then it seems natural to divide it so that the value of this *J*-function is minimized. In the context of pattern recognition this prescription has been known as a minimum entropy approach to clustering [8][9]. This method is capable of dealing with complicated examples which usual methods of clustering cannot cope with. A large number of practical algorithms to carry out clustering tasks are based on the notion of \`distance' between two objects, which reflects only \`one-to-one' relationship. The entropic measure of interdependence *J*, defined by (2), can take \`more-than-two-elements-correlation' into account.

This minimum entropy approach, sometimes called *Interdependence analysis* [8][9], has a number of theoretical merits. It enable us to examine complicated inter-group relationship reflected in the observed data. However, as we shall see later in Example, sometimes this method does not function well. In order to understand from where this difficulty stems, let us consider a partitioning of *X* into a subset that has a single member, say $\{x_i\}$, and the remaining members $X \setminus \{x_i\}$, and the interdependence $J(\{x_i\}, X \setminus \{x_i\})$ between them. If a cardinality of *X* is not so small and if patterns of the observed states has a wide variety then formal entropy of the entire group $S(X)$ often coincides with formal entropy $S(X \setminus \{x_i\})$. That is, in such a case, we can remove some members from *X*, preserving the value of formal entropy function. Then, in (4), two terms in the right-hand side are cancelled, and the interdependence $J(\{x_i\}, X \setminus \{x_i\})$ is reduced to $S(\{x_i\})$. A small subset sometimes gives a very small entropy value, and in such a case the method of Interdependence analysis tells us to remove such a small subset from the entire set regardless of its relationship to other members.

3.3. Clustering as average entropy minimization

One possible way to avoid this difficulty is to take the cardinality of subset into account, and to modify the entropic measure of the interdependence as follows:

$$\begin{aligned} K(X_1, X_2) \equiv\ & n_1 S(X_1) + n_2 S(X_2) \\ & - (n_1 + n_2) S(X_1 \cup X_2) \end{aligned} \tag{3}$$

where n_1 and n_2 are cardinalities of X_1 and X_2 , respectively. Thus a new approach to clustering is: when we divide the entire set we should choose a partitioning which minimizes the value of *K*-function defined by (3) [6][7].

How should we understand the meaning of our *K*-value and its minimization? Dividing (3) by the cardinality $n_1 + n_2$, we have

$$k(X_1, X_2) \equiv \frac{n_1 S(X_1) + n_2 S(X_2)}{n_1 + n_2} - S(X_1 \cup X_2) \tag{4}$$

That is, the *K*-value per member is the difference of the average entropy of the divided subsets from the entropy of the entire set. Then an immediate, intuitive interpretation is to regard the *K*-value as a measure of the entropy reduction by partitioning, and its minimization as a minimization of the degree of \`post-partitioning' disorder, or uncertainty. When we consider a loss of entropy as a gain of information, this can be paraphrased also as the maximization of information gain by partitioning.

Another, more convincing explication is as follows: By using the Bayesian formula we can rewrite (3) as

$$\begin{aligned} K(X_1, X_2) &= n_1 \{S(X_1) - S(X_1 \cup X_2)\} \\ &\quad + n_2 \{S(X_2) - S(X_1 \cup X_2)\} \\ &= -\{n_1 S(X_2 | X_1) + n_2 S(X_1 | X_2)\} \end{aligned} \tag{5}$$

or, in a similar way, (4) as

$$k(X_1, X_2) = -\frac{n_1 S(X_2 | X_1) + n_2 S(X_1 | X_2)}{n_1 + n_2} \tag{6}$$

where the conditional entropy $S(X_i \mid X_j)$ can be understood as a measure of disorder, or uncertainty, of X_i under the condition that the knowledge of X_j is given ($i, j = 1, 2$). In other words, $S(X_i \mid X_j)$ can be regarded as the amount of information on X_i provided by X_j . If two groups are bound by close relationship then we have much information on the state of one group by observing the state of the other. If not so then the knowledge of the state of one group does not give any clues regarding the state of the other.

Thus the quantity $-S(X_i \mid X_j)$ can be regarded as a measure of the degree of interdependence between X_i and X_j. Our *K*-function is, as (5) shows, a weighted sum of these terms, and measures the degree of interdependence between two groups by the amount of information which the one provides for the other. This is the gist of our approach. The proposed algorithm directs us to divide a group of objects into some subgroups so that the amount of information obtained through the inter-subgroup relationship is minimized. As for the details of algorithm, see [6][7].

We first apply this method to the entire set X, and then apply the same method to the obtained subsets. Repeating this procedure until finally each resulting subset consists of a single element, we obtain a suitable partitioning of X, which can be described as a complete taxonimic tree, as we shall see later in Example. Then, tracing back such a tree from each member to the entire set, we have a procedure of class formation, or coarse observation, of the given set of objects. As a common feature of algorithm of this sort, we cannot decide in which stage we should stop our procedure without introducing some additional criteria. However, the proposed algorithm allows us to calucu-late change of the amount of information at each branchig point of the taxonimic tree as follows.

Let X be a set of n objects, $X = \{x_1, x_2, \ldots, x_n\}$. We denote X newly $X^{(0)}$, and first divide it into m_1 subgroups,

$$X^{(0)} = X_1^{(1)} \cup X_2^{(1)} \cup \cdots \cup X_{m_1}^{(1)}$$

where $X_i^{(1)} \cap X_j^{(1)} = \phi$, for $i \neq j$. Then we divide each subgroup $X_i^{(1)}$ into some sub-subgroups in a similar way. We continue this procedure until finally each resulting group consists only of one member $\{ x_i \}$, and obtain a complete taxonomic tree whose trunk is $X^{(0)}$, and whose peripheral branches are $\{ x_i \}$'s. Let us suppose that the k-th branching point of the tree divides a subgroup $X^{(k-1)}$ into m_k sub-subgroups, $X_1^{(k)}$, $X_2^{(k)}$, … , $X_{mk}^{(k)}$. Then we can define the interdependence at this branching point by

$$\begin{aligned} K&\left(X^{(k-1)};\, X_1^{(k)}, X_2^{(k)}, \cdots, X_{m_k}^{(k)}\right) \\ &= \sum_{i=1}^{m_k} n_i^{(k)} S\left(X_i^{(k)}\right) - n^{(k-1)} S\left(X^{(k-1)}\right) \\ &= -\sum_{i=1}^{m_k} n_i^{(k)} S\left(X^{(k-1)} \mid X_i^{(k)}\right) \end{aligned} \quad (7)$$

The total sum of such interdependences taken at all branching points in the complete taxonomic tree is independent of the choice of the tree. That is, we have

$$\sum_{\substack{all\ branching \\ points}} K\,(\text{branching point}) = \sum_{i=1}^{n} S(x_i) - n\,S\left(X^{(0)}\right)$$

The right-hand side can be further transformed into

$$-\sum_{i=1}^{n} S\left(X^{(0)} \mid x_i\right)$$

(8)

This formula also clarifies the gist of our approach. The total interdependence measured by K is the sum of information on the state of the whole system provided by each constituent component.

Table I. Interferon treatment of chronic hepatitis C [6].

	a_1	a_2	a_3	a_4	a_5	a_6
x_1	1	0	0	0	0	0
x_2	1	0	1	0	0	0
x_3	1	1	1	0	0	0
x_4	1	0	0	0	1	1
x_5	1	0	1	0	1	1
x_6	1	0	1	1	1	1
y_1	1	1	1	0	0	1
y_2	1	0	1	1	1	0
y_3	1	1	1	0	1	1
y_4	1	0	0	1	1	1
y_5	1	1	1	1	1	1

4. EXAMPLE

The following example, taken from the actual medical data, shows that the proposed method works fairly reasonable [5][6]. Table I shows the efficacy of interferon treatment of chronic hepatitis C. Chronic hepatitis C has been defined arbitrarily by persistence of elevated serum alanine aminotransferase (ALT) levels for more than six months in a person infected with hepatitis C virus (HCV). In some cases interferon (INF) is effective. The patients treated by INF are classified into two groups, responders and non-responders, according to whether serum ALT levels are normalized or not. The responder group can be divided further into two subgroups, complete responders and incomplete responders, according to whether lower serum HCV RNA levels are sustained or not. In non-responder group viremia persists, though in some patients clearance of HCV RNA may occur transiently. The eradication of HCV implies normalization of serum ALT levels, while the latter does not mean the former in general.

Table I shows eleven types of patients treated by INF. x_i (i=1,…, 4) corresponds to the responders, y_j (j=1,…, 5) to the non-responders, while x_5 and x_6 appear in both groups. a_k is a predicate `is a patient in whose serum HCV was detectable at the k-th blood test' (0: negative, 1: positive). The first test was made before INF treatment, the second just after completion of the treatment, others every six month for two years.

Now let us consider a clustering of these eleven patient types without going into medical details. For simplicity, we abbreviate $\{x_i\}$ as X_i, $\{ x_i , x_j \}$ as X_{ij}, $\{ x_i , x_j , y_k \}$ as $X_{ij}Y_k$, and so on, and denote the average entropy of the right-hand side of (4), $\langle S \rangle$. The original method of interdependence analysis (IDA), minimization of the J-function, directs us first to remove $\{y_5\}$ from the entire set $X \cup Y$, and then to divide the remaining ten members into four groups, three of which have one member, and one has seven as shown in Fig. 1. Since all entries of the row of y_5 are 1, the first step may be rather trivial, while the second step is seemingly unbalanced. It seems rather difficult

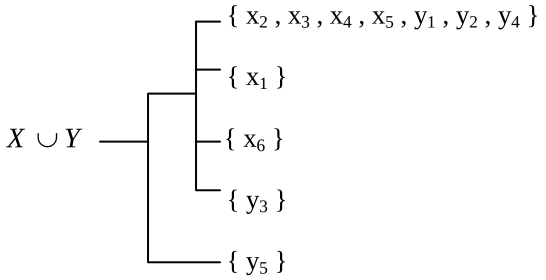

Fig. 1. Clustering of the eleven patient types by the original method of inter-dependence analysis

to give a natural, convincing interpretation to such a clustering.

Let us consider a grouping of these eleven patient types by the proposed method. We first apply our method to the entire set *XY*, and obtain the dichotomy $\{ XY_{345}, Y_{12}\}$, which give the minimum of $\langle S \rangle$ =1.4767. Then we apply the same procedure to nine patient types XY_{345}, and have two ways of dichotomies, $\{ X_{123}, X_{456}Y_{345}\}$ and $\{X_{45}Y_3, X_{1236}Y_{45}\}$, which give the minimum of $\langle S \rangle$ = 1.2425. These two dichotomies are of the same type: a dichotomy into a granule of three elements and a granule of six elements. Hence we adopt these two simultaneously, and obtain a trichotomy as shown in Fig. 2

Observing the fifth and the sixth columns of Table I, we can regard two patient types y_1 and y_2 as exceptional cases. The remaining nine types are classified into three groups $Z_1 =\{x_1, x_2, x_3\}$, $Z_2=\{x_4, x_5, y_3\}$ and $Z_3=\{y_4, x_6, y_5\}$, and Table II shows that these three groups are characterized by the right hand side of the table. Since our predicate a_i corresponds to a result of the *i*-th blood test, these three groups can be understood as different types of the progress of, or recovery from, the hepatitis. In the members of Z_1 lower serum HCV levels are sustained, and hence Z_1 can be understood as a complete responder group, while other two, Z_2 and Z_3, have no such counterparts. Thus the proposed method allows us a natural, convincing interpretation [5][6].

It should be remarked that at the second stage of our procedure two dichotomies $\{X_{14}Y_4, X_{2356}Y_{35}\}$ and $\{X_{256}, X_{134}Y_{345}\}$ take the second minimum $\langle S \rangle$ = 1.330, and give another trichotomy as shown in Table III, where three groups $W_1=\{x_1, x_4, y_4\}$, $W_2=\{x_2, x_5, x_6\}$ and $W_3=\{x_3, y_3, y_5\}$ correspond to different types of response to INF treatment at the earlier periods after a completion of the treatment. Should we discard this alternative just because the corresponding entropy is slightly larger than the minimum?

Thus once again we come back to the importance of subjective, semantic factor. The minimum entropy approach proposed above provides us with a unique solution in principle. However, whether the prescribed group formation is really suitable or not depends on the intention of measurement after all, which is beyond a grasp of our formal consideration [5][6][7].

Table II. Clustering of nine patient types

	a_1	a_2	a_3	a_4	a_5	a_6
x_1	1	0	0	0	0	0
x_2	1	0	1	0	0	0
x_3	1	1	1	0	0	0
x_4	1	0	0	0	1	1
x_5	1	0	1	0	1	1
y_3	1	1	1	0	1	1
y_4	1	0	0	1	1	1
x_6	1	0	1	1	1	1
y_5	1	1	1	1	1	1

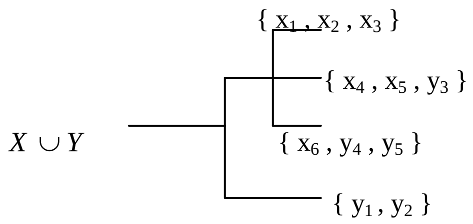

Fig. 2. Clustering of the eleven patient types by the proposed method

5. DISCUSSION

Our formal approach to coarse observation has some interesting implications. First of all, considering the convenience of information processing, it seems natural to form groups so that each group consists of more or less equal numbers of members. However, as we observed in our example, if we pay attention to the intrinsic nature of the empirical data then different sizes of groups may give a suitable coarse description.

Besides, secondly, our formal consideration also clarifies the value-oriented nature of group formation procedure. As we have discussed elsewhere [5] within a formal framework of object-predicate system same classification can be formed via different sets of predicates, and in so far as we take various predicates with the same importance we cannot choose a suitable set of predicates for the classification uniquely. As for forming of groups of objects, our example shows that from a purely formal point of view, we cannot choose a suitable grouping uniquely. To decide which is suitable, or which is the best, we have to bring in some

Table III. Another possible clustering of nine patient types

	a_1	a_2	a_3	a_4	a_5	a_6
x_1	1	0	0	0	0	0
x_4	1	0	0	0	1	1
y_4	1	0	0	1	1	1
x_2	1	0	1	0	0	0
x_5	1	0	1	0	1	1
x_6	1	0	1	1	1	1
x_3	1	1	1	0	0	0
y_3	1	1	1	0	1	1
y_5	1	1	1	1	1	1

additional factor such as a priority among predicates or the intention of grouping.

Third, in our formal theory of clustering we have introduced a modified concept of indiscernibility [5], called *weak indiscernibility*, between two groups of objects. When we apply this concept to measure the distance between two binary sequences we have to take one binary sequence, for instance, 1101001…., and the sequence obtained by inverting 0 and 1 in that sequence, 0010110…., to be essentially identical. This may seem somewhat queer, considering the usual information measure such as Hamming distance, but, a principle of minimum average entropy justifies this seemingly counter-intuitive consequence [6].

6. CONCLUSION

Formal measurement theory has been formulated and elaborated over the notion of quantity. However, the growth of information processing technology enlarged the scope of our measurement practice, and nowadays we are faced with measurement non-physical, non-quantitative and ill-defined. The nominal scale may be given a prominent place as a basis of a wide class of non-quantitative measurement. Here we have discussed some information-theoretical aspect of `coarse' nominal matching in connection with a grouping of objects, proposed a possible formal approach to form groups of objects, and pointed out the importance of subjective factor in `coarse' nominal matching. As a consequence of value-oriented nature of evaluation, such a factor seems to appear also, sometimes more clearly, in other weak measurement.

REFERENCES

[1] L.Finkelstein, "Measurement: fundamental principles", in *Concise encyclopedia od measurement and instrumentation*, L.Finkelstein and K.Grattan (eds.), Oxford, Pergamon, 1994.

[2] L.Mari, "Beyond the representational viewpoint: a new formalization of measurement", *Measurement*, 27, 2, pp.71--84, 2000.

[3] W. Pedrycz, "Granular computation: an introduction", *Proceedings of IFSA/NAFIPS 2001*, pp.1708—1713.

[4] I.van Biesen, "Advances of measurement science", *IMEKO-TC7 Workshop*, pp.13--19, Kyoto, Japan, 1999.

[5] H.Watanabe, "Modified concept of indiscernibility and a basis of classification", *IMEKO-TC7 Symposium*, pp.56--61, Cracow, Poland, 2002.

[6] H.Watanabe, "Clustering as average entropy minimization and its application to structure analysis of complex systems", *IEEE SMC 2001*, pp.2408--2414, Tucson, Arizona, USA, 2001.

[7] H.Watanabe, "Elicitation of information granules as clustering", Proceedings of 5th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, pp.103—108, Japan, 2002.

[8] S.Watanabe, *Knowing and Guessing*, Wiley, 1969.

[9] S.Watanabe, *Pattern recognition: human and mechanical*, Wiley, 1985.

[10] L. A. Zadeh, "Fuzzy sets and information granuality", in M. M. Gupta, R.K. Ragade, R.R. Yager eds. *Advances in Fuzzy Set Theory and Applications*, pp.3--18, North Holland, 1979.

Author: Hiroshi Watanabe. Higashi-mukohjima 6--19--10, Sumida-ku, Tokyo, 131-0032, Japan.