XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

THEORY OF COARSE-GRAINED INFORMATION

Hiroshi Watanabe

Higashi-mukohjima 6—19—10, Sumida-ku, Tokyo, 131-0032, Japan

Abstract – Measurement should be objective in principle. However, the accuracy of measurement often depends on various factors, including instruments of measurement, adopted intentionally or made available accidentally, and from a practical point of view, it is important to choose a suitable degree of the accuracy of measurement for the intention of measurement. The present paper discusses some information-theoretical aspects of the relationship between the degree of the accuracy of nominal description of objects and the interdependence among groups of those objects within a formal framework of clustering, where a new entropic measure of interdependence is applied to estimate the change of the interdependence when the nominal description becomes more `coarse' so that several objects appear to be identical, and are grouped into the same class. The degree of interdependence is often measured by (sum of entropies of partial systems) - (entropy of the whole). Close examination shows, however, that instead of simple formal entropy we should take rather (the number of the members of a group)×(entropy of the group). The result also implies that in integration of information channels the average entropy plays an important role. A brief discussion is also made on elicitation of so-called information granules.

Keywords: clustering, elicitation of information granule, interdependence analysis, measurement science, minimum average entropy.

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. However, the accuracy of measurement is not determined by 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. While measurement should be objective, if the structure of measured data depends on the accuracy of measurement then we shall have to admit that our theoretical picture of the external world is influenced in a very essential way by the additional factors which cannot be attributed to the 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 chain has 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 states plays a decisive role in our understanding of our surroundings. However, this subject of general interest has not been much discussed, excepting a few theoretical papers. As for mathematical investigation of the properties of stochastic chain, the papers of Blackwell [1] and Harris [2] in the middle 1950s are among the earliest studies of this subject. In 1960s, from the information-theoretical point of view, Satosi Watanabe compared a stochastic chain and a coarsely redefined chain, bearing in mind `microscopic' description and `macroscopic' description in physics, and discussed the effect of coarse observation of the states on the interdependence in stationary stochastic chain [8]. In the context of communication technology `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 our coarse observation. 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, indiscernibility or the degree of cohesion. Aiming at enlarging the scope of granule computing, Pedrycz newly defined the notion of information granule, and discussed the fundamentals of the subject [3][7].

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. On several occasions so far we have discussed the basic nature of classification procedure, regarding it as a form of nominal, non-quantitative measurement [4], and proposed a new approach to clustering [6] based on a modified concept of indiscernibility [5]. In the following consideration we extend our approach to integration of stochastic chains. After a brief review of our formal theory of clustering, we observe the basic properties of our entropic measure of interdependence, and then discuss the possibility of the extension of our approach to integration of stochastic chains.

The notion of information granule also can be recapitulated within our formal framework, and we make some comments and remarks on elicitation of information granules [7].

2. A FORMAL THEORY OF CLUSTERING

2.1. Entropic measure of interdependence

Suppose that we are given a set $X = \{x_1, x_2, \ldots, x_n\}$of objects, which we are going to classify. Each object is supposed to take any one of N states, 1, 2, …, N, and we observe the object to check in which state it is repeatedly, say m times, with a fixed time interval. We further assume that there exists a unique probability $p(x_1, x_2, \ldots, x_n)$ that the group of those n objects will take a set of n states. Then we have

$$\sum_{x_n=1}^{N} p(x_1, x_2 \cdots, x_n) = p(x_1, x_2 \cdots, x_{n-1}), \qquad \sum_{x_1=1}^{N} p(x_1, x_2 \cdots, x_n) = p(x_2, x_3 \cdots, x_n) \tag{1}$$

and

$$\sum_{x_1=1}^{N}\sum_{x_2=1}^{N}\cdots\sum_{x_n=1}^{N} p(x_1, x_2 \cdots, x_n) = 1 \tag{2}$$

Indeed, since our observation data are finite, we can define such a probability $p(x_1, x_2, \ldots, x_n)$ by the relative frequency of the appearance of a set of n states in the whole data.

Now let us divide the entire group X into two subgroups, X_I and X_J. Then we can define a probability $p(X_I)$, and in a similar way, according to this partitioning, we have $p(X_J)$, $p(X_I, X_J)$, and hence we can define formal entropy functions for these groups of objects, respectively. For instance:

$$S(X_I) \equiv -\sum_i p_i \log p_i \tag{3}$$

Suppose that we are asked to divide the entire group X into two subgroups according to the result of our observation. Then how should we divide it? The *interdependence* J between two subgroups X_I and X_J can be defined by

$$J(X_I, X_J) \equiv S(X_I) + S(X_J) - S(X_I \cup X_J) \tag{4}$$

as a counterpart of redundancy in the communication theory [8]. Then it seems natural to divide the entire set 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], and indeed this method is capable of dealing with complicated example 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 (4) can take \`more-than-two-elements-correlation’ into account.

This minimum entropy approach, called *Interdependence analysis* [8][9], has a number of theoretical merits, and often 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.

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 (4) as follows:

$$K(X_1, X_2) \equiv n_1 S(X_1) + n_2 S(X_2) - (n_1 + n_2) S(X_1 \cup X_2) \tag{5}$$

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 (5) [5][6].

2.2. Some basic properties of K-function

Before we go to the next section, and observe an example, it may be appropriate to examine here some basic properties of the K-function, and to clarify the gist of our approach.

The formal entropy function is infra-additive:

$$S(X_1) + S(X_2) \geq S(X_1 \cup X_2) \tag{6}$$

for two subsets such that $X_1 \cap X_2 = \phi$, and J-function is non-negative. This allows us to regard the interdependence J as a \`branching cost' in a taxonomic tree, such as we shall see in Figure 1, and justifies the minimization of J-value in clustering. Formal entropy function multiplied by the cardinality, however, is supra-additive:

$$n_1 S(X_1) + n_2 S(X_2) \leq (n_1 + n_2) S(X_1 \cup X_2) \tag{7}$$

and our K-function is non-positive. Then how should we understand the meaning of K-value and its minimization?

Dividing (4) by the cardinality $n_1 + n_2$, we have

$$k(X_1, X_2) \equiv \frac{n_1 S(X_1) + n_2 S(X_2)}{n_1 + n_2} - S(X_1 \cup X_2) \tag{8}$$

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 (4) as

$$K(X_1,\ X_2) = n_1\{S(X_1) - S(X_1 \cup X_2)\} + n_2\{S(X_2) - S(X_1 \cup X_2)\} = -\{n_1 S(X_2 | X_1) + n_2 S(X_1 | X_2)\} \quad (9)$$

or, in a similar way, (8) as

$$k(X_1,\ X_2) = -\frac{n_1 S(X_2 | X_1) + n_2 S(X_1 | X_2)}{n_1 + n_2} \quad (10)$$

where the conditional entropy $S(X_i | X_j)$ 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, $S(X_i | X_j)$ 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 $-S(X_i | X_j)$ can be regarded as a measure of the degree of interdependence between X_i and X_j. Our *K*-function is, as (9) 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. This view can be paraphrased also by regarding $-K$ as a measure of the degree of mutual independence.

As for *J*-function, we can rewrite also (4) in terms of conditional entropy as

$$J(X_1, X_2) = S(X_1) + S(X_2) - S(X_1 \cup X_2) = S(X_1) - S(X_1 | X_2) = S(X_2) - S(X_2 | X_1)$$

which shows that to minimize the value of *J* is to minimize the difference of $S(X_i | X_j)$ from $S(X_i)$ (i, j = 1, 2). This does not necessarily mean, however, to minimize the weighted sum of $-S(X_2 | X_1)$ and $-S(X_1 | X_2)$ in general.

It should be remarked that we can define the interdependence *J* for a partitioning of *X* into more than two subgroups. 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, $X_1^{(k)}, X_2^{(k)}, \ldots, X_{mk}^{(k)}$. Then we can define the interdependence *J* at this branching point by

$$J\left(X^{(k-1)};\ X_1^{(k)}, X_2^{(k)}, \cdots, X_{m_k}^{(k)}\right) = \sum_{i=1}^{m_k} S\left(X_i^{(k)}\right) - S\left(X^{(k-1)}\right) \quad (11)$$

Satosi Watanabe showed that the total sum of such interdependences taken at all branching points in a complete taxonomic tree is independent of the choice of the tree, and is equal to $J(X^{(0)}; \{x_1\}, \{x_2\}, \ldots, \{x_n\})$ [8]. That is,

$$\sum_{\substack{all\ branching\\ points}} J\,(\text{branching point}) = \sum_{i=1}^{n} S(x_i) - S\left(X^{(0)}\right) \quad (12)$$

In a similar way, we can define our interdependence K by

$$K\left(X^{(k-1)};\ X_1^{(k)}, X_2^{(k)}, \cdots, X_{m_k}^{(k)}\right) = \sum_{i=1}^{m_k} n_i^{(k)} S\left(X_i^{(k)}\right) - n^{(k-1)} S\left(X^{(k-1)}\right) = -\sum_{i=1}^{m_k} n_i^{(k)} S\left(X^{(k-1)} | X_i^{(k)}\right) \quad (13)$$

and obtain, in accordance with the above mentioned formula,

$$\sum_{\substack{all\ branching\\ points}} K\,(\text{branching point}) = \sum_{i=1}^{n} S(x_i) - n\,S\left(X^{(0)}\right) \quad (14)$$

and then, the right-hand side can be further transformed into

$$-\sum_{i=1}^{n} S\left(X^{(0)} | x_i\right) \quad (15)$$

The comparison between the expressions (12) and (14), or (15), also clarifies the gist of our approach. The total interdependence measured by *J* is nothing but the difference between the amount of total information on the state of the whole system and the sum of information on the state of each constituent component, while the total interdependence measured by *K* is the sum of information on the state of the whole system provided by each constituent component.

3. EXAMPLE

The proposed approach seems to be fairly reasonable. Let us give an example. The following example is a slight modification of the example which we discussed elsewhere, and then dubbed `a problem of seven red herrings' because the example was first introduced to illustrate the difficulty of the original method of interdependence analysis.

Suppose that we are given a set $X = \{x_1, x_2, \ldots, x_7\}$ of objects. For simplicity, we assume that each of these seven objects takes one of two states, say 0 and 1, and that we observe the states of these objects eight times with a fixed time interval. The result of observation is supposed to be given by Table 1, where the *j* th column a_j corresponds to the result of the *j* th observation. Now let us consider clustering of these seven members based on the data as shown in this table.

Applying the original method of interdependence analysis and our new method, we obtain a polychotomic tree of Figure 1*a* and that of Figure 1*b*, respectively. At the first stage of clustering, the minimum of *J* is given by a dichotomy $\{X_7, X_{123456}\}$, abbreviating $X_i = \{x_i\}$, $X_{ij} = \{x_i, x_j\}$, and so on, and the method of interdependence analysis directs us first to remove $\{x_7\}$ from the entire set. Since x_7 has only two 1's in its row, while others have four, this prescription may seem rather reasonable. In the second stage, however, this method cannot find any intergroup relationship among the remaining six members.

Table 1. A modification of a problem of seven red herrings

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
x_1	1	1	1	1	0	0	0	0
x_2	1	1	0	0	0	0	1	1
x_3	0	0	1	1	0	0	1	1
x_4	0	1	0	1	1	0	1	0
x_5	1	0	1	0	1	0	1	0
x_6	1	0	0	1	1	0	0	1
x_7	1	0	0	0	1	0	0	0

The proposed method first divides the entire set into two groups, { x_1, x_2, x_4 } and { x_3, x_5, x_6, x_7 }. It should be remembered that our approach is based on formal entropy function defined on a group of objects, and that we can replace a member of the group, say x_i, with $\overline{x_i}$ which is obtained by inverting 0 and 1 in a row corresponding to x_i , remaining the value of entropy function to be the same. Then, inverting 0 and 1 in rows of x_3, and of x_4, and rearranging the order of the members, we obtain Table 2. In the lower half of the table, an arrangement of 0 and 1 in the left-hand side is repeated in the right-hand side, while in the upper half of the table the inverted pattern of 0 and 1 is repeated. That is, our result corresponds to a partitioning of the table into a periodic and an inverted-periodic subtables.

In passing, as for the numerical values, we have

$$J(X_7 , X_{123456}) = .5623 < 1.3863 = J(X_{124} , X_{3567})$$

$$K(X_7 , X_{123456}) = -1.5171 > -2.7695 = K(X_{124} , X_{3567})$$

In the second stage our method divides { x_1, x_2, x_4 } into three subsets, each of which has only one member, while {x_3, x_5, x_6, x_7 } into two, discerning x_7 from other members. This example shows that our entropic measure of interdependence is indeed capable of finding out very subtle inter-group relationship among groups of objects, which the original method of interdependence analysis may not uncover. Other examples are discussed in [5][6][7].

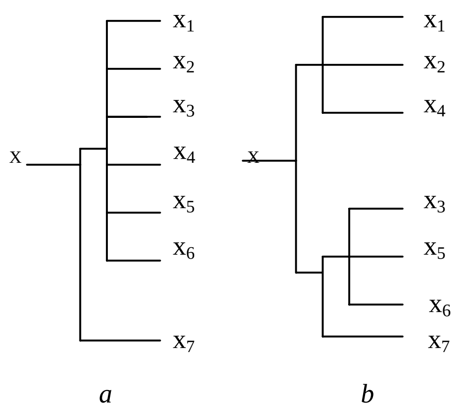

Figure 1. Two ways of clustering of the seven members illustrated in Table 1.

Table 2. Rearranged table according to the result of clustering

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
x_1	1	1	1	1	0	0	0	0
x_2	1	1	0	0	0	0	1	1
$\bar{x}_4$	1	0	1	0	0	1	0	1
$\bar{x}_3$	1	1	0	0	1	1	0	0
x_5	1	0	1	0	1	0	1	0
x_6	1	0	0	1	1	0	0	1
x_7	1	0	0	0	1	0	0	0

4. DISCUSSION

4.1 Interdependence analysis in stochastic chains

So far we have discussed the interdependence analysis in a finite sequence of observed data. Under a suitable condition we can extend our approach to more general case as follows: Suppose that we are given an infinite sequence of numbers, which represent `states' of the system under consideration. Any phenomenon, as presented to our cognizance, can be regarded as such a sequence of `states' or `situations' arranged on a time axis t. For simplicity, we further assume that each position of this infinite sequence takes any one of N states, 1, 2, …, N, as before. If there exists a unique probability $p(x_1, x_2, \ldots, x_n)$ that a sequence of n consecutive positions will take a set of states x_1, x_2, …, x_n, and if this probability is independent of the location of the segment in the infinite sequence, that is, if this stochastic chain is stationary, then we can apply our method to this sequence, and analyse the interdependence existing among various groups of positions of these n consecutive positions.

As a counterpart of redundancy in communication theory, the J-function, defined by (4), has been accepted as a measure of the degree of interdependence. However, our exapmle observed above, with other examples discussed in [5][6][7], strongly suggests us that we should take rather the K-function, defined by (5), as a measure of the degree of interdependence. It can be expected without calculation that in so far as we evaluate the degree of interdependence among *objects* the original method based on J-function and our modified method based on K-function give more or less similar results. Indeed, for instance, as we see in (12) and (14), the difference is in the term of the total entropy, which is constant when we fix the entire set of objects, or segments. However, when we evaluate the degree of interdependence among *groups of objects*, that is, groups of different size or segments of various lengths, the proposed method seems to provide us with a natural, convincing result.

Of course there are many ways of thinking about how we look at `coarse-grained' observation of objects. In our formal framework sketched above we have discussed the interdependence among groups of objects, regarding clustering as a form of `coarse-grained' observation. In his information-theoretical approach to pattern recognition and clustering, in 1960s, Satosi Watanabe discussed another form of `coarse-grained' observation as follows: Let us go back to the stochastic sequence as we discussed above, and consider grouping of N states ,1, 2, …, N, into N' classes, 1, 2, …, N', where $N' \leq N$, so that no class is empty and each state belongs to one and only one of those classes. If we introduce `macroscopic' variables $\xi_i$, $i$ = 1, 2, …, $k$, to label the class to which the state of the $i$ th position belongs then the `macroscopic' state of this chain is described in terms of these ξ_i 's as ξ_1, ξ_2, ..., ξ_k. With such a formalization Watanabe evaluated the loss of information of the original `microscopic' chain due to macroscopization in the presence of redundancy, and discussed the change of the range of correlation by grouping of the states [8]. Our approach is a possible modification of the method of interdependence analysis, and will shed a new light upon this line of interest.

4.2 Elicitation of information granules

Another form of `coarse-grained' observation can be found in the context of so-called soft computing approach, as an extension of fuzzy set theory. At the end of 1970s, in order to reduce heavy computational burden, Zadeh introduced a heuristic notion of information granule [10], and proposed representing information in the form of aggregates defined by a certain numerical level of similarity, indiscernibility or cohesion [3][7]. Our formal approach to clustering has some interesting implications also as to the formation of granules.

First of all, considering the convenience of computation, it seems natural to form information granules in a step of processing so that each granule consists of more or less equal numbers of objects. However, as we observed in our example, if we pay attention to the intrinsic nature of the empirical data under consideration, and take its structure into account, then different sizes of information granules may give a suitable granulation.

Besides, secondly, according to the current understanding of soft computing approach, granulation procedure seems to be regarded as a conventional preprocessing of numerical data to cope with a huge data structure. However, our example shows that a suitable granulation enable us to examine a subtle and complicated relationship among groups of objects, decomposing the entire system under consideration into some elementary constituent components. Thus granulation procedure helps us to understand complicated structure of complex systems [6].

Third, in our formal theory of clustering we have introduced a modified concept of indiscernibility [5], called weak indiscernibility, between two groups of objects. If we apply this concept to measure the distance between two binary sequences then 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, in particular, considering the usual information measure such as Hamming distance. However, a principle of minimum average entropy justifies this seemingly counter-intuitive consequence, and our modified concept of indiscernibility between two groups of objects sheds a new light upon the forming of information granule [5].

In passing, fourth, Pedrycz mentioned a problem of how to define the dimension of information granule as an open question in granule computing [3], and our formal approach suggests a possible candidate of the notion of the dimension of information granule. In this respect, and other comments and remarks on forming of information granules, see [7].

5. CONCLUSION

The growth of instrumentation technology has enabled us to obtain fine and precise information on our surroundings. However, we do not always need precise information such as long sequences of numerical data. We need information with a suitable degree of accuracy according to our intension of measurement: sometimes coarse observation meets our need. Balancing the fineness of observation with the required accuracy of information is a subject of interest in design of measurement systems. Regarding clustering task as a form of `coarse-grained' observation of objects, here we have discussed a formal approach to the interdependence analysis in the information obtained by `coarse-grained' observation. Starting from a formal framework of clustering, we examined the basic properties of new entropic measure of the degree of interdependence, and pointed out its potential significance in analysis of the relationship among various sequences in a stochastic chain. Our result suggests that when we examine the relationship between sequences of different length, or groups of different size, we should take the average entropy as a measure of interdependence. Our formal framework is a one possible approach to a theory of coarse-grained information, whose scope will embrace various subjects, such as mathematical theory of stochastic chains, error and coding problems in communication theory, algorithms of pattern recognition and clustering, and some fundamental topics in theoretical physics. The field of measurement science seems to be rich in examples where the notion of coarse-grained information plays a crucial role..

REFERENCES

[1] D. Blackwell, "The entropy of functions of finite-state Markov chains", *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pp.13--20, 1956.

[2] T. E. Harris, "On chains of infinite order", *Pacific Journal of Mathematics*, vol.5, pp.707—724, 1955.

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

[4] H.Watanabe, "Theory of classification: a new aspect of the nominal scale", *IMEKO-XV*, II, pp.147—154, 1999.

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

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