## PAPER

# A Randomness Based Analysis on the Data Size Needed for Removing Deceptive Patterns 

Kazuya HARAGUCHI ${ }^{\dagger \text { a }}$, Mutsunori YAGIURA ${ }^{\dagger \dagger b}$, Members, , and Toshihide IBARAKI ${ }^{\dagger \dagger \dagger \mathrm{j})}$, Fellow


#### Abstract

SUMMARY We consider a data set in which each example is an $n$ dimensional Boolean vector labeled as true or false. A pattern is a cooccurrence of a particular value combination of a given subset of the variables. If a pattern appears frequently in the true examples and infrequently in the false examples, we consider it a good pattern. In this paper, we discuss the problem of determining the data size needed for removing "deceptive" good patterns; in a data set of a small size, many good patterns may appear superficially, simply by chance, independently of the underlying structure. Our hypothesis is that, in order to remove such deceptive good patterns, the data set should contain a greater number of examples than that at which a random data set contains few good patterns. We justify this hypothesis by computational studies. We also derive a theoretical upper bound on the needed data size in view of our hypothesis. key words: frequent/infrequent item sets, association rules, knowledge discovery, probabilistic analysis


## 1. Introduction

### 1.1 Background

Development of computer hardware technology enables us to save massive data at a low cost. In order to discover hidden meaningful knowledge from such data, various methodologies have been studied so far under the name of knowledge discovery, data mining, and so on.

A data set consists of examples drawn from the population of the considered phenomenon. One of the most challenging problems in the literature is to generate (or to enumerate) all patterns, substructures of examples, appearing frequently/infrequently in a given data set. This problem is often formulated as frequent/infrequent pattern mining, an important issue in data mining and bioinformatics (e.g.,

[^0]knowledge discovery from genome databases) [1], [6], [14]. (The term "frequent/infrequent set" is widely used in the literature, but in order to avoid the confusion with a simple set of elements, we use the term "pattern" in this paper.)

A "good" pattern in some sense may carry us useful information on decision making. However, its reliability as knowledge must heavily depend on the size of the data set; if the data set is too small, a pattern may be deceptive and thus may not serve as meaningful knowledge. While we can store a massive data set cheaply these days, data collection is still expensive in many application areas (e.g., weather data) [10]. In such areas, it is difficult to collect enough examples, and in this paper, we analyze the size of a data set needed for removing deceptive patterns, as an attempt to establish a criterion on the amount of examples needed for efficient knowledge discovery.

### 1.2 Preliminaries

Let us introduce the notations and terminologies used throughout this paper. Let $\mathbf{B}=\{0,1\}$. We denote a data set by $X$. Each element in $X$, an example, is represented by a vector in $\mathbf{B}^{n}$, and is labeled either by 1 (true) or by 0 (false). We denote by $X_{1}$ (resp., $X_{0}$ ) the set of true (resp., false) examples in $X$, i.e., $X=X_{1} \cup X_{0}$ with disjoint $X_{1}$ and $X_{0}$. We call the cardinality $|X|$ the size of a data set $X$. If $m_{1}=\left|X_{1}\right|$ and $m_{0}=\left|X_{0}\right|$, then we call $X$ an $\left(m_{1}, m_{0}\right)$-data set.

For a vector $x \in \mathbf{B}^{n}$ and a subset $J \subseteq\{1, \ldots, n\}$, we denote by $\left.x\right|_{J}=\left(x_{j} \mid j \in J\right)$ the sub-vector of $x$ corresponding to the index set $J$. A pattern $r=(J, b)$ is defined by a subset $J \subseteq\{1, \ldots, n\}$ and a Boolean vector $b \in \mathbf{B}^{|J|}$. Given a pattern $r=(J, b)$ and a Boolean vector $x \in \mathbf{B}^{n}$, we say that $r$ appears in $x$ if $\left.x\right|_{J}=b$ holds. Let us denote by $X(r)$ the set of examples in $X$ in which $r$ appears; i.e., $X(r)=\left\{x \in X|x|_{J}=b\right\}$. In particular, $\mathbf{B}^{n}(r)$ denotes the set of all binary vectors in which $r$ appears. We define the frequency of $r$ by $f(r, X)=|X(r)| /|X|$, i.e., by the proportion of examples of $X$ in which $r$ appears. Given a constant $a(0 \leq a \leq 1)$, we call a pattern $r a$-frequent (resp., $a$-infrequent) in $X$, if $f(r, X) \geq a$ (resp., $f(r, X) \leq a$ ).

For given constants $a_{1}, a_{0}\left(0 \leq a_{1}, a_{0} \leq 1\right)$, we call a pattern $r$ an $\left(a_{1}, a_{0}\right)$-pattern in $X$, if $f\left(r, X_{1}\right) \geq a_{1}$ and $f\left(r, X_{0}\right) \leq a_{0}$. If $a_{1}$ is "large enough" and $a_{0}$ is "small enough" such a pattern describes a feature of the true examples in $X$. One could also consider ( $a_{1}, a_{0}$ )-patterns capturing the features of false examples (i.e., patterns $r$ that
satisfy $f\left(r, X_{1}\right) \leq a_{1}$ and $f\left(r, X_{0}\right) \geq a_{0}$ for "small" $a_{1}$ and "large" $a_{0}$ ). Since those could be obtained by interchanging the roles of true and false examples, we focus on "good" patterns for true examples in this paper.

It is well-known that one can find frequent/infrequent patterns in incrementally polynomial time [1], and many fast algorithms for this task have been proposed so far (e.g., [12]). By applying the previous algorithms to generate frequent/infrequent patterns, we can generate all $\left(a_{1}, a_{0}\right)$ patterns from the data set $X$ in incrementally polynomial time; e.g., by taking the intersection of the set of $a_{1}$-frequent patterns in $X_{1}$ and that of $a_{0}$-infrequent patterns in $X_{0}$.

### 1.3 Description of Problems

We consider the problem of determining the data size needed to remove deceptive ( $a_{1}, a_{0}$ )-patterns. Let us define a domain by $D=\left(n, \rho, P_{1}, P_{0}\right)$, where $n$ denotes the number of Boolean variables, $\rho \in[0,1]$ denotes a probability, and $P_{1}, P_{0}: \mathbf{B}^{n} \rightarrow[0,1]$ denote probability distributions. Since $P_{1}$ and $P_{0}$ are probability distributions, it holds that

$$
\begin{equation*}
\sum_{x \in \mathbf{B}^{n}} P_{1}(x)=\sum_{x \in \mathbf{B}^{n}} P_{0}(x)=1, \tag{1}
\end{equation*}
$$

and $P_{1}(x), P_{0}(x) \geq 0$ for any $x \in \mathbf{B}^{n}$. We make an assumption on the distribution of examples as follows;

Assumption 1: Given a domain $D=\left(n, \rho, P_{1}, P_{0}\right)$, an example $(x, \omega)$ is independently generated by the following steps:
Step 1: The label $\omega$ is set to 1 with probability $\rho$, and to 0 otherwise (i.e., with probability $1-\rho$ ).
Step 2: A vector $x$ with label $\omega$ is drawn according to the distribution $P_{\omega}$.

Now, given a data set $X$ generated from a domain $D$, we expect that ( $a_{1}, a_{0}$ )-patterns in $X$ carry important information about $D$. It is possible, however, that some of them are deceptive; they might be present as $\left(a_{1}, a_{0}\right)$ patterns in $X$ only by chance, independently of the underlying structure of $D$. Obviously, such deceptive ( $a_{1}, a_{0}$ )patterns would exist with high probability if $m_{1}$ and $m_{0}$ of an ( $m_{1}, m_{0}$ )-data set $X$ are small, but the probability will diminish if $m_{1}$ and $m_{0}$ are sufficiently large. More precisely, let $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ denote the expected number of $\left(a_{1}, a_{0}\right)$ patterns found in an $\left(m_{1}, m_{0}\right)$-data set $X$ generated from $D$. If $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \gg E_{D}\left(m_{1}^{\prime}, m_{0}^{\prime} ; a_{1}, a_{0}\right)$ holds for sufficiently large $m_{1}^{\prime}$ and $m_{0}^{\prime}$ with $m_{1} / m_{0}=m_{1}^{\prime} / m_{0}^{\prime}$, then we conclude that $X$ is not large enough and that it contains a lot of deceptive ( $a_{1}, a_{0}$ )-patterns. This will be experimentally studied in Sect. 3 .

We then would like to estimate the sizes of $m_{1}$ and $m_{0}$, which guarantee that most of the ( $a_{1}, a_{0}$ )-patterns found in an ( $m_{1}, m_{0}$ )-data set are not deceptive. For this purpose, we introduce the random domain $R=(n, 1 / 2, U, U)$, where $U$ denotes the uniform distribution with $U(x)=1 / 2^{n}$ for all $x \in \mathbf{B}^{n}$. A data set generated from $R$ is called a random
data set. We consider that $R$ has no particular structure, and any ( $a_{1}, a_{0}$ )-pattern found in a random data set is deceptive. By using the random domain, our hypothesis for the needed data size is summarized as follows.

Hypothesis 1: Let $X$ be an ( $m_{1}, m_{0}$ )-data set generated from a given domain $D$. Then the probability of deceptive ( $a_{1}, a_{0}$ )-patterns found in $X$ is (approximately) the same as the probability that an ( $m_{1}, m_{0}$ )-random data set $Y$ contains ( $a_{1}, a_{0}$ )-patterns.
Therefore, we estimate experimentally and theoretically the sizes of $m_{1}$ and $m_{0}$, at which an ( $m_{1}, m_{0}$ )-random data set $Y$ contains approximately no ( $a_{1}, a_{0}$ )-patterns; they will be used as the sizes of an ( $m_{1}, m_{0}$ )-data set $X$ which contains no deceptive $\left(a_{1}, a_{0}\right)$-patterns.

The composition of this paper is as follows. After describing the related works in Sect.2, we study such data sizes experimentally in Sect. 3, and derive their theoretical upper bounds by some probabilistic analysis in Sect. 4. Then in Sect. 5, we give the concluding remarks.

## 2. Related Works

If a pattern $r=(J, b)$ is a frequent pattern in $X$ and there is no frequent pattern $r^{\prime}=\left(J^{\prime}, b^{\prime}\right)$ with $J^{\prime} \supset J$ and $\left.b^{\prime}\right|_{J}=b$, then we call $r$ a maximal frequent pattern. If $r$ is an infrequent pattern in $X$ and there is no infrequent pattern $r^{\prime}=\left(J^{\prime}, b^{\prime}\right)$ with $J^{\prime} \subset J$ and $\left.b\right|_{J^{\prime}}=b^{\prime}$, then we call $r$ a minimal infrequent pattern. Boros et al. [4] showed that, given a family of $O\left(n^{\varepsilon}\right)$ maximal frequent patterns, it is NP-complete to decide whether $X$ has any further maximal frequent patterns (for arbitrarily small fixed $\varepsilon>0$ ), and that all minimal infrequent patterns can be enumerated in incremental quasipolynomial time.

The problem of finding frequent patterns is closely related to that of association rules. An association rule is generally defined by two patterns $\left(r, r^{\prime}\right)=\left((J, b),\left(J^{\prime}, b^{\prime}\right)\right)$ with $J \cap J^{\prime}=\emptyset$; it represents that an example $x$ with $\left.x\right|_{J}=b$ is likely to attain $\left.x\right|_{J^{\prime}}=b^{\prime}$.

An association rule $\left(r, r^{\prime}\right)$ is usually evaluated by its support (which is the proportion of examples in $X$ where both $r$ and $r^{\prime}$ appear) and confidence (which is the frequency of $r^{\prime}$ in $X(r)$ ), while we evaluate a pattern $r$ by its frequency in $X_{1}$ and infrequency in $X_{0}$. Thus, the generation of frequent patterns is a basic operation in finding association rules.

As the task of finding association rules from a huge data set is very time-consuming, Li et al. [9] and Toivonen [11] discussed the proper size of a randomly drawn subset $X^{\prime}$ of the original data set $X$ such that $f\left(r, X^{\prime}\right)$ is close enough to $f(r, X)$ with a high probability, for all patterns $r$. While they consider the random sampling of a manageable size from the given huge data set, we consider the situation in which the size of the given data set is small, and discuss whether the extracted $\left(a_{1}, a_{0}\right)$-patterns are deceptive or not. This is the main difference between our approach and the existing ones.

## 3. Experimental Studies

### 3.1 Expected Number of $\left(a_{1}, a_{0}\right)$-Patterns

We derive the expected number of ( $a_{1}, a_{0}$ )-patterns in an $\left(m_{1}, m_{0}\right)$-data set generated from a domain $D=$ $\left(n, \rho, P_{1}, P_{0}\right)$.

Consider a pattern $r$. Under the condition that a generated example is labeled 1 (resp., 0) in Step 1 of Assumption 1 , the probability $c_{1}(r ; D)$ (resp., $c_{0}(r ; D)$ ) that $r$ appears in this new example is:

$$
\begin{align*}
& c_{1}(r ; D)=\sum_{x \in \mathbf{B}^{n}(r)} P_{1}(x), \\
& c_{0}(r ; D)=\sum_{x \in \mathbf{B}^{n}(r)} P_{0}(x) . \tag{2}
\end{align*}
$$

More generally, under the condition that $m_{1}$ true examples and $m_{0}$ false examples are generated, the probability that a pattern $r$ is $a_{1}$-frequent in the $m_{1}$ true examples is:

$$
\begin{align*}
b_{+} & \left(m_{1}, a_{1}, c_{1}(r ; D)\right) \\
& =\sum_{s=\left\lceil a_{1} m_{1}\right\rceil}^{m_{1}}\binom{m_{1}}{s} c_{1}(r ; D)^{s}\left(1-c_{1}(r ; D)\right)^{m_{1}-s}, \tag{3}
\end{align*}
$$

and the probability that $r$ is $a_{0}$-infrequent in the $m_{0}$ false examples is:

$$
\begin{align*}
& b_{-}\left(m_{0}, a_{0}, c_{0}(r ; D)\right) \\
& =\sum_{s=0}^{s=\left\lfloor a_{0} m_{0}\right\rfloor}\binom{m_{0}}{s} c_{0}(r ; D)^{s}\left(1-c_{0}(r ; D)\right)^{m_{0}-s} . \tag{4}
\end{align*}
$$

Note that the product $b_{+} b_{-}$gives the expectation that $r$ is an $\left(a_{1}, a_{0}\right)$-pattern in an ( $m_{1}, m_{0}$ )-data set generated from the domain $D$.

For a pattern $r=(J, b)$, let us call the cardinality $|J|$ the level of $r$. We denote by $R_{k}$ the set of all possible patterns of level $k(1 \leq k \leq n)$. Note that $\left|R_{k}\right|=2^{k}\binom{n}{k}$ holds and that $\left|\mathbf{B}^{n}(r)\right|=2^{n-k}$ holds for any $r \in R_{k}$. Let $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ be the expected number of ( $a_{1}, a_{0}$ )-patterns in an ( $m_{1}, m_{0}$ )data set from the domain $D$, and $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ be the same number when their levels are restricted to $k$. From the linearity of expectations, they are formulated as follows:

$$
\begin{align*}
E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)= & \sum_{k=1}^{n} E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \\
= & \sum_{k=1}^{n} \sum_{r \in R_{k}} b_{+}\left(m_{1}, a_{1}, c_{1}(r ; D)\right) \\
& \times b_{-}\left(m_{0}, a_{0}, c_{0}(r ; D)\right) \tag{5}
\end{align*}
$$

### 3.2 Real Data Sets

We take ten real data sets from UCI Repository [3]; i.e.,

Table 1 Summary of ten data sets from UCI repository.

| Data | $n$ | $m_{1}^{*}$ | $m_{0}^{*}$ | $m_{1}^{*}+m_{0}^{*}$ |
| :---: | ---: | ---: | ---: | ---: |
| AUS | 14 | 307 | 383 | 690 |
| BCW | 13 | 239 | 444 | 683 |
| BUPA | 21 | 200 | 145 | 345 |
| CAR | 12 | 518 | 1210 | 1728 |
| CRX | 13 | 296 | 357 | 653 |
| HABER | 18 | 75 | 219 | 294 |
| HEART | 10 | 120 | 150 | 270 |
| IONO | 9 | 225 | 126 | 351 |
| PIMA | 15 | 268 | 500 | 768 |
| TTT | 12 | 626 | 332 | 958 |



Fig. $1 E_{R}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ with $n=13$ and $m_{1} / m_{0}=239 / 444$ corresponding to data set BCW . (Lines with points $\times, \square$, $\square$ represent $a_{0}=$ $0.00,0.01,0.02$, respectively.)

AUS, BCW, BUPA, CAR, CRX, HABER, HEART, IONO, PIMA, TTT. In order to handle these data sets in our scheme, we modify them as follows:

- CAR is a four-labeled data set, and we modify it to a two-labeled data set; an example in CAR is labeled one of the four labels (i.e., unacc, acc, good, v-good). We treat those labeled unacc as false examples, and the rest as true examples.
- Some data sets contain examples with missing values or contradiction, and we exclude such examples.
- Finally, since the examples in some data sets are numerical and/or categorical vectors, we transform them into binary examples by the method used in [7].

For each real data set, let us denote by $X_{1}^{*}$ and $X_{0}^{*}$ the sets of true and false examples, respectively. We denote $X^{*}=X_{1}^{*} \cup X_{0}^{*}, m_{1}^{*}=\left|X_{1}^{*}\right|$ and $m_{0}^{*}=\left|X_{0}^{*}\right|$. Table 1 shows a summary of the binary data sets transformed from the ten real data sets.

## 3.3 $E_{R}$ on Random Data Sets

We first compute the expected number of $\left(a_{1}, a_{0}\right)$-patterns in an ( $m_{1}, m_{0}$ )-random data set $E_{R}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ by (2) to (5) with $D=R$. In order to compare this $E_{R}$ with the expected number $E_{D}$ on a real data set (where we write its domain by $D$ ) later, we adopt the number $n$ of Boolean variables and the ratio $m_{1} / m_{0}=m_{1}^{*} / m_{0}^{*}$ corresponding to the real data set, and test various $m_{1}$ and $m_{0}$ for all combinations of $a_{1} \in$ $\{0.10,0.20\}$ and $a_{0} \in\{0.00,0.01,0.02\}$.

Figures. 1 and 2 show the computed $E_{R}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ with the parameters corresponding to BCW and BUPA, respectively; i.e., $n=13$ and $m_{1}+m_{0}$ is changed with keeping $m_{1} / m_{0}=239 / 444$ for BCW, and $n=21$ and $m_{1} / m_{0}=$


Fig. $2 \quad E_{R}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ with $n=21$ and $m_{1} / m_{0}=200 / 145$ corresponding to data set BUPA. (Lines with points $\times, \square$, $\square$ represent $a_{0}=$ $0.00,0.01,0.02$, respectively.)

Table 2 The $\left(M_{1}^{*}, M_{0}^{*}\right)$ for real data sets with $a_{1}=0.10$. (A figure with an underline indicates that $M_{1}^{*}+M_{0}^{*} \leq m_{1}^{*}+m_{0}^{*}$ holds.)

| Data | $a_{0}$ | $M_{1}^{*}$ | $M_{0}^{*}$ | $M_{1}^{*}+M_{0}^{*}$ |
| :---: | ---: | ---: | ---: | ---: |
| AUS | 0.00 | 101 | 124 | $\underline{225}$ |
|  | 0.01 | 121 | 149 | $\underline{\underline{270}}$ |
|  | 0.02 | 161 | 199 | $\underline{360}$ |
| BCW | 0.00 | 81 | 149 | $\underline{230}$ |
|  | 0.01 | 91 | 169 | $\underline{260}$ |
|  | 0.02 | 132 | 243 | $\underline{375}$ |
| BUPA | 0.00 | 172 | 123 | $\underline{295}$ |
|  | 0.01 | 212 | 153 | $\underline{365}$ |
|  | 0.02 | 273 | 197 | $\underline{470}$ |
| CAR | 0.00 | 71 | 164 | $\underline{235}$ |
|  | 0.01 | 81 | 189 | $\underline{\underline{270}}$ |
|  | 0.02 | 122 | 283 | $\underline{405}$ |
| CRX | 0.00 | 102 | 123 | $\underline{225}$ |
|  | 0.01 | 121 | 144 | $\underline{265}$ |
|  | 0.02 | 161 | 194 | $\underline{355}$ |
| HABER | 0.00 | 81 | 234 | $\underline{315}$ |
|  | 0.01 | 101 | 294 | $\cdots 395$ |
|  | 0.02 | 151 | 439 | $\underline{590}$ |
| HEART | 0.00 | 83 | 102 | $\underline{185}$ |
|  | 0.01 | 103 | 127 | $\underline{230}$ |
|  | 0.02 | 143 | 177 | $\underline{320}$ |
| IONO | 0.00 | 132 | 73 | $\underline{205}$ |
|  | 0.01 | 132 | 73 | $\underline{205}$ |
|  | 0.02 | 174 | 96 | $\underline{270}$ |
| PIMA | 0.00 | 91 | 169 | $\underline{260}$ |
|  | 0.01 | 102 | 188 | $\underline{290}$ |
|  | 0.02 | 151 | 279 | $\underline{430}$ |
| TTT | 0.00 | 161 | 84 | $\underline{245}$ |
|  | 0.01 | 161 | 84 | $\underline{245}$ |
|  | 0.02 | 242 | 128 | $\underline{370}$ |

200/145 for BUPA. Each figure contains two cases corresponding to $a_{1}=0.10$ and 0.20 , where the horizontal (resp., vertical) axis represents $m_{1}+m_{0}$ (resp., $E_{R}$ ) and three curves correspond to different values of $a_{0}$. Note that the vertical axis is in the logarithmic scale. The $E_{R}$ appears to be monotonically decreasing with $m_{1}+m_{0}$ if we neglect small irregularities, and becomes less than 1 as $m_{1}+m_{0}$ becomes larger than a certain point.

Among the examined values of $m_{1}$ (resp., $m_{0}$ ), let us denote by $M_{1}^{*}$ (resp., $M_{0}^{*}$ ) the smallest value that attains $E_{R} \leq$ 1. Table 2 shows the observed $\left(M_{1}^{*}, M_{0}^{*}\right)$ for the parameter values $m_{1}$ and $m_{0}$, which correspond to the ten real data sets, where we always use $a_{1}=0.10$. In this table, a real data set whose $M_{1}^{*}+M_{0}^{*}$ is underlined indicates that $M_{1}^{*}+M_{0}^{*} \leq$ $m_{1}^{*}+m_{0}^{*}$ holds; i.e., the data set contains a sufficient number of examples in view of our hypothesis.


Fig. 3 Expected number $E_{D}$ of $\left(a_{1}, a_{0}\right)$-patterns on real data sets with $a_{1}=0.10$. (Lines with points $\times, \square$, $■$ represent $a_{0}=0.00,0.01,0.02$, respectively. A broken line parallel to the vertical axis represents $M_{1}^{*}+M_{0}^{*}$.)

## 3.4 $E_{D}$ on Real Data Sets

Now, for the ten real data sets, we would like to know how the expected number of $\left(a_{1}, a_{0}\right)$-patterns $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ changes as $m_{1}$ and $m_{0}$ increase relative to the $M_{1}^{*}$ and $M_{0}^{*}$. However, we cannot compute $E_{D}$ exactly by (5) since we do not know the domain $D$ of a real data set. (Furthermore, we do not know even whether the examples are generated according to Assumption 1 or not. In this experiment, however, we regard that they are.) In order to estimate $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ experimentally for $m_{1} \leq m_{1}^{*}$ and $m_{0} \leq m_{0}^{*}$, we randomly sample subsets $X_{1} \subseteq X_{1}^{*}$ and $X_{0} \subseteq X_{0}^{*}$ with $\left|X_{1}\right|=m_{1}$ and $\left|X_{0}\right|=m_{0}$, satisfying $m_{1} / m_{0}=m_{1}^{*} / m_{0}^{*}$, and enumerate all ( $a_{1}, a_{0}$ )-patterns in $X=X_{1} \cup X_{0}$. For each
tested values of $m_{1}$ and $m_{0}$, we repeat this process $\tau$ times and estimate $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ by the average of the numbers of ( $a_{1}, a_{0}$ )-patterns. In this experiment, we use $\tau=100$.

The results are shown in Fig. 3, where the vertical axis indicates $E_{D}$ and the horizontal axis indicates the size $m_{1}+m_{0}$. Note that the vertical axes in these figures are not in the logarithmic scale in contrast to Figs. 1 and 2. In this experiment, we always use $a_{1}=0.10$, and $a_{0}$ is set to 0.00 , 0.01 and 0.02 , respectively. Three curves correspond to different values of $a_{0}$, and a broken line parallel to the vertical axis represents the value that corresponds to $M_{1}^{*}+M_{0}^{*}$.

As shown in these figures, when the size of $m_{1}+m_{0}$ is small, the ( $m_{1}, m_{0}$ )-data set contains ( $a_{1}, a_{0}$ )-patterns much more than the $\left(m_{1}^{*}, m_{0}^{*}\right)$-data set. In such cases, we conclude that the expected number $E_{D}$ contains many deceptive ( $a_{1}, a_{0}$ )-patterns. Hypothesis 1 states that we need to satisfy $m_{1} \geq M_{1}^{*}$ and $m_{0} \geq M_{0}^{*}$ in order to remove deceptive ( $a_{1}, a_{0}$ )patterns, and it surely holds in the results; all the curves are stabilized after passing the point of $m_{1}+m_{0}=M_{1}^{*}+M_{0}^{*}$. For AUS, for example, if $m_{1}+m_{0}$ is small (e.g., less than $100)$, then more than $5.0 \times 10^{3}\left(a_{1}, a_{0}\right)$-patterns exist, while there are substantially smaller number of ( $a_{1}, a_{0}$ )-patterns if $m_{1}+m_{0} \geq M_{1}^{*}+M_{0}^{*}$ holds, for all three cases of $a_{0}=$ $0.00,0.01,0.02$. Furthermore, the expected numbers $E_{D}$ do not change to a great extent in the range $m_{1}+m_{0} \geq M_{1}^{*}+M_{0}^{*}$.

For BUPA and HABER, $\left|X^{*}\right|=m_{1}^{*}+m_{0}^{*}<M_{1}^{*}+M_{0}^{*}$ holds for almost all tested parameter combinations. According to our hypothesis, the size of $X^{*}$ is not large enough. In fact, $E_{D}$ is still making a rapid change even if $m_{1}+m_{0}$ is increased to the limit of $m_{1}^{*}+m_{0}^{*}$, and thus the given data set $X^{*}$ appears to contain many deceptive ( $a_{1}, a_{0}$ )-patterns.

## 4. Upper Bounds on the Needed Data Size

### 4.1 Preliminaries

The determination of $M_{1}^{*}$ and $M_{0}^{*}$ by using (5) requires a nontrivial computational cost. To alleviate this, we derive an upper bound on $M_{1}^{*}+M_{0}^{*}$ in this section. For the derivation, we assume that any domain $D=\left(n, \rho, P_{1}, P_{0}\right)$ satisfies the following assumption.

Assumption 2: For any $x \in \mathbf{B}^{n}, P_{1}(x) \leq p$ and $P_{0}(x) \geq q$ hold for some constants $p$ and $q$.
From (1), it is implied that $p \geq 1 / 2^{n}$ and $q \leq 1 / 2^{n}$. Note that the random domain $R$ is realized by setting $p=q=1 / 2^{n}$.

For a domain $D$ satisfying Assumption 2, we show that an upper bound on $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ becomes sufficiently small (i.e., not more than $\varepsilon$, a small positive value) if $k$ is in some range, either $m_{1}$ or $m_{0}$ is larger than some threshold, and a few other conditions hold. If an upper bound on $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ becomes sufficiently small for all $k=1, \ldots, n$, then their sum $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)=\sum_{k} E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ also becomes small; thus such thresholds on $m_{1}$ and $m_{0}$ can respectively be used as upper bounds on the needed numbers of true and false examples, $M_{1}^{*}$ and $M_{0}^{*}$.

Note that $E_{D, k}$ with "large" $k$ or "small" $k$ cannot be large for the following reason. Consider a pattern $r$ with level $k$ in an ( $m_{1}, m_{0}$ )-data set $X$. If $k$ is large (resp., small), then $\left|\mathbf{B}^{n}(r)\right|=2^{n-k}$ tells that the $r$ appears in a small (resp., large) portion of vectors in $\mathbf{B}^{n}$. Thus the $r$ is unlikely to be $a_{1}$-frequent in the $m_{1}$ true examples (resp., $a_{0}$-infrequent in the $m_{0}$ false examples), and thus unlikely to be an $\left(a_{1}, a_{0}\right)$ pattern in $X$. Our analysis in the following is to refine this observation.

### 4.2 Probabilistic Analyses on $E_{D, k}$ and Bounds on $M_{1}^{*}$ and $M_{0}^{*}$

We first introduce some well-known bounds in the probability theory.
Theorem 1: (Chernoff [5]) Given a positive integer $m$ and $0 \leq \mu \leq 1$, let $Q_{i}$ be a random variable taking the value as follows:

$$
Q_{i}= \begin{cases}1-\mu & \text { with probability } \mu  \tag{6}\\ -\mu & \text { with probability } 1-\mu\end{cases}
$$

and let $Q=\sum_{i=1}^{m} Q_{i}$. Then, for any $\beta>1$,

$$
\begin{equation*}
\operatorname{Pr}(Q \geq(\beta-1) \mu m)<\left(\exp (\beta-1) \beta^{-\beta}\right)^{\mu m} \tag{7}
\end{equation*}
$$

holds.
Theorem 2: (Hoeffding [8]) For a positive integer $m$ and $0 \leq a \leq 1$, if $0 \leq \mu \leq a$, then the $b_{+}$in (3) satisfies

$$
\begin{equation*}
b_{+}(m, a, \mu) \leq \exp \left(-2 m(a-\mu)^{2}\right) \tag{8}
\end{equation*}
$$

Similarly, if $a \leq \mu \leq 1$, then the $b_{-}$in (4) satisfies

$$
\begin{equation*}
b_{-}(m, a, \mu) \leq \exp \left(-2 m(\mu-a)^{2}\right) \tag{9}
\end{equation*}
$$

Variations of Theorem 1 are found in [2], for example.
Now we derive two types of upper bounds on $E_{D, k}$ for "large" $k$.

Theorem 3: Suppose that $D, m_{1}, m_{0}, a_{1}, a_{0}, k$ and $\varepsilon \in$ $(0,1]$ are given. If $k \geq K_{+}$and $m_{1} \geq M_{1}$, then $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \leq \varepsilon$ holds, where

$$
\begin{align*}
& K_{+}=n-\log _{2} \frac{a_{1}}{e^{2} p}  \tag{10}\\
& M_{1}=\frac{n \ln (2 n)-\ln \varepsilon}{a_{1}} \tag{11}
\end{align*}
$$

and $e$ denotes the base of the natural logarithm.
Proof : Let $r$ be a pattern of level $k \geq K_{+}$. From Assumption 2 and $\left|\mathbf{B}^{n}(r)\right|=2^{n-k}$, we have $c_{1}(r ; D) \leq \min \left\{1,2^{n-k} p\right\}$, and since $2^{n-k} \leq 2^{n-K_{+}}=a_{1} /\left(e^{2} p\right)$, we have $c_{1}(r ; D) \leq$ $2^{n-k} p \leq a_{1} / e^{2}<1$. Let $Z_{i}$ be a random variable taking the value as follows:

$$
Z_{i}= \begin{cases}1 & \text { with probability } 2^{n-k} p  \tag{12}\\ 0 & \text { with probability } 1-2^{n-k} p\end{cases}
$$

and let $Z=\sum_{i=1}^{m_{1}} Z_{i}$. Let $Q_{i}=Z_{i}-2^{n-k} p$ and $Q=\sum_{i=1}^{m_{1}} Q_{i}=$ $Z-2^{n-k} p m_{1}$. Then, we have

$$
\begin{align*}
& E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \\
& \quad=\sum_{r \in R_{k}} b_{+}\left(m_{1}, a_{1}, c_{1}(r ; D)\right) b_{-}\left(m_{0}, a_{0}, c_{0}(r ; D)\right) \\
& \quad \leq b_{+}\left(m_{1}, a_{1}, 2^{n-k} p\right) \times\left|R_{k}\right| \\
& \quad=\operatorname{Pr}\left(Z \geq a_{1} m_{1}\right) \times 2^{k}\binom{n}{k} \\
& \quad=\operatorname{Pr}\left(Q \geq 2^{n-k} p m_{1}\left(\frac{a_{1}}{2^{n-k} p}-1\right)\right) \times 2^{k}\binom{n}{k} \tag{13}
\end{align*}
$$

From $k \geq K_{+}$, it holds $a_{1} /\left(2^{n-k} p\right) \geq e^{2}>1$. By applying Theorem 1 with $m=m_{1}, \mu=2^{n-k} p$ and $\beta=a_{1} /\left(2^{n-k} p\right)$, we have

$$
\begin{align*}
E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) & <\left(\frac{2^{n-k} p e}{a_{1}}\right)^{a_{1} m_{1}} \times 2^{k}\binom{n}{k} \\
& \leq\left(\frac{2^{n-k} p e}{a_{1}}\right)^{a_{1} m_{1}} \times(2 n)^{k} \\
& \leq e^{-a_{1} m_{1}} \times(2 n)^{n} . \tag{14}
\end{align*}
$$

The right hand side of (14) is not more than $\varepsilon$ if and only if

$$
\begin{equation*}
m_{1} \geq \frac{n \ln (2 n)-\ln \varepsilon}{a_{1}}=M_{1} \tag{15}
\end{equation*}
$$

Another upper bound on $E_{D, k}$ for large $k$ is given below. It depends on a parameter $t$ and can bound $E_{D, k}$ for $k$ with $k>K_{+}-3$.

Theorem 4: Suppose that $D, m_{1}, m_{0}, a_{1}, a_{0}, k$ and $\varepsilon \in$ $(0,1]$ are given. If $k \geq K_{+}(t)$ and $m_{1} \geq M_{1}(t)$ for some $t \in\left(0, a_{1}\right)$, then $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \leq \varepsilon$ holds, where

$$
\begin{align*}
& K_{+}(t)=n-\log _{2} \frac{a_{1}-t}{p}  \tag{16}\\
& M_{1}(t)=\frac{n \ln (2 n)-\ln \varepsilon}{2 t^{2}} \tag{17}
\end{align*}
$$

Proof : For an arbitrary $t \in\left(0, a_{1}\right)$, let $r$ be a pattern of level $k \geq K_{+}(t)$. From Assumption 2 and $\left|\mathbf{B}^{n}(r)\right|=2^{n-k}$, we have $c_{1}(r ; D) \leq \min \left\{1,2^{n-k} p\right\}$, and since $k \geq K_{+}(t)$, we have $2^{n-k} p \leq a_{1}-t<a_{1} \leq 1$. Thus, $c_{1}(r ; D) \leq 2^{n-k} p$ and

$$
\begin{equation*}
b_{+}\left(m_{1}, a_{1}, c_{1}(r ; D)\right) \leq b_{+}\left(m_{1}, a_{1}, 2^{n-k} p\right) \tag{18}
\end{equation*}
$$

By applying (8) of Theorem 2 with $m=m_{1}, a=a_{1}$ and $\mu=2^{n-k} p$, we have

$$
\begin{align*}
& b_{+}\left(m_{1}, a_{1}, 2^{n-k} p\right) \\
& \quad \leq \exp \left(-2 m_{1}\left(a_{1}-2^{n-k} p\right)^{2}\right) \tag{19}
\end{align*}
$$

and hence

$$
\begin{align*}
& E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \\
& \quad \leq \exp \left(-2 m_{1}\left(a_{1}-2^{n-k} p\right)^{2}\right) \times 2^{k}\binom{n}{k} \\
& \quad \leq \exp \left(-2 m_{1} t^{2}\right) \times(2 n)^{n} . \tag{20}
\end{align*}
$$

The right hand side of (20) is not more than $\varepsilon$ if and only if

$$
\begin{equation*}
m_{1} \geq \frac{n \ln (2 n)-\ln \varepsilon}{2 t^{2}}=M_{1}(t) \tag{21}
\end{equation*}
$$

Given $D$ and $a_{1}$, the $K_{+}$in Theorem 3 is a constant while $K_{+}(t)$ in Theorem 4 depends on the parameter $t$. The following corollary about the range of $t$ is useful in obtaining an upper bound $E_{D, k} \leq \varepsilon$ with $K_{+}(t) \leq k \leq K_{+}$from Theorem 4 .
Corollary 1: If we set $t=a_{1}\left(1-\ell / e^{2}\right)$ for a constant $1 \leq$ $\ell<e^{2}$, then $K_{+}-K_{+}(t)=\log _{2} \ell$.

Proof : It directly comes from the definition of $K_{+}$and $K_{+}(t)$.

Note that $K_{+}-K_{+}(t)<\log _{2} e^{2}<3$ holds.
Now we turn to an upper bound on $E_{D, k}$ for "small" $k$.
Theorem 5: Suppose that $D, m_{1}, m_{0}, a_{1}, a_{0}, k$ and $\varepsilon \in$ $(0,1]$ are given. If $k \leq K_{-}(s)$ and $m_{0} \geq M_{0}(s)$ hold for some $s \in(0,1)$, then $E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \leq \varepsilon$ holds, where

$$
\begin{align*}
& K_{-}(s)=n-\log _{2} \frac{a_{0}+s}{q}  \tag{22}\\
& M_{0}(s)=\frac{K_{-}(s) \ln (2 n)-\ln \varepsilon}{2 s^{2}} \tag{23}
\end{align*}
$$

Proof : The proof is similar to that of Theorem 4. For an arbitrary $s \in(0,1)$, let $r$ be a pattern of level $k \leq K_{-}(s)$. From Assumption 2, $\left|\mathbf{B}^{n}(r)\right|=2^{n-k}$ and $k \leq K_{-}(s)$, we have $c_{0}(r ; D) \geq 2^{n-k} q \geq a_{0}+s>a_{0}$. By applying (9) of Theorem 2 with $m=m_{0}, a=a_{0}$ and $\mu=2^{n-k} q$,

$$
\begin{align*}
& b_{-}\left(m_{0}, a_{0}, c_{0}(r ; D)\right) \\
& \quad \leq b_{-}\left(m_{0}, a_{0}, 2^{n-k} q\right) \\
& \quad \leq \exp \left(-2 m_{0}\left(2^{n-k} q-a_{0}\right)^{2}\right) \tag{24}
\end{align*}
$$

holds and hence we have

$$
\begin{align*}
& E_{D, k}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \\
& \quad \leq \exp \left(-2 m_{0}\left(2^{n-k} q-a_{0}\right)^{2}\right) \times 2^{k}\binom{n}{k} \\
& \quad \leq \exp \left(-2 m_{0} s^{2}\right) \times(2 n)^{K_{-}(s)} \tag{25}
\end{align*}
$$

The right hand side of (25) is not more than $\varepsilon$ if and only if

$$
\begin{equation*}
m_{0} \geq \frac{K_{-}(s) \ln (2 n)-\ln \varepsilon}{2 s^{2}}=M_{0}(s) \tag{26}
\end{equation*}
$$

Recall that Theorems 3 and 4 hold for large $k$ and Theorem 5 holds for small $k$. Then, if one of these theorems holds for every $k=1, \ldots, n$, then we will have $E_{D, k} \leq \varepsilon$ for all $k=1, \ldots, n$ and hence, $E_{D}=\sum_{k} E_{D, k} \leq n \varepsilon$. More precisely, if we choose parameters $t$ and $s$ so that $K_{+}(t) \leq K_{-}(s)$ holds, and we have $m_{1} \geq \max \left\{M_{1}, M_{1}(t)\right\}$ and $m_{0} \geq M_{0}(s)$, then one of these theorems holds for every $k=1, \ldots, n$. A sufficient condition for $K_{+}(t) \leq K_{-}(s)$ to hold is given in the following corollary.

Corollary 2: If $t \in\left(0, a_{1}\left(1-1 / e^{2}\right)\right]$ and $s \in\left(0, q\left(a_{1}-t\right) / p-\right.$ $\left.a_{0}\right]$, then $K_{-}(s) \geq K_{+}(t)$ holds.

Proof : It directly comes from the definitions of $K_{+}(t)$ and $K_{-}(s)$.

Finally, $E_{D}=\sum_{k} E_{D, k}$ becomes sufficiently small under the conditions given in the following theorem.

Theorem 6: Suppose that $D, m_{1}, m_{0}, a_{1}, a_{0}$ and $\varepsilon \in(0,1]$ are given. If $t \in\left(0, a_{1}\left(1-1 / e^{2}\right)\right]$ and $s \in(0,1)$ satisfy $s \leq q\left(a_{1}-t\right) / p-a_{0}, m_{1} \geq \max \left\{M_{1}, M_{1}(t)\right\}$ and $m_{0} \geq M_{0}(s)$, then $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right) \leq n \varepsilon$ holds.

Corollary 3: For appropriate values of $p, q, a_{1}$ and $a_{0}$ (e.g., $p \simeq q$ and $a_{1} \gg a_{0}$ ), there exist $t$ and $s$ that satisfy the above condition $s \leq q\left(a_{1}-t\right) / p-a_{0}$. Then, if we take $\varepsilon$ sufficiently small (e.g., $\varepsilon=2^{-n}$ ), $E_{D}\left(m_{1}, m_{0} ; a_{1}, a_{0}\right)$ converges to 0 .

Corollary 4: The $\max \left\{M_{1}, M_{1}(t)\right\}$ and $M_{0}(s)$ in Theorem 6 are upper bounds on $M_{1}^{*}$ and $M_{0}^{*}$ in Sect. 3, respectively.

Let us consider the possibility of using $\max \left\{M_{1}, M_{1}(t)\right\}$ and $M_{0}(s)$ as estimates on $M_{1}^{*}$ and $M_{0}^{*}$ in Sect. 3, respectively. To see how close they are, we computed $M_{1}, M_{1}(t)$ and $M_{0}(s)$ on the random domain $R$ for some combinations of ( $n, a_{1}, a_{0}$ ), where we set $t$ and $s$ to the values that minimize $\max \left\{M_{1}, M_{1}(t), M_{0}(s)\right\}$ among all $t=\ell \times 10^{-3} \in$ $\left(0, a_{1}\left(1-1 / e^{2}\right)\right]$ and $s=\ell^{\prime} \times 10^{-3} \in\left(0, a_{1}-a_{0}-t\right]$ with natural numbers $\ell$ and $\ell^{\prime}$. The obtained upper bounds, however, are not very tight; e.g., for $\left(n, a_{1}, a_{0}\right)=(13,0.10,0.00)$, $M_{1}=449.20, M_{1}(t)=6036.04$ and $M_{0}(s)=6029.60$, while $\left(M_{1}^{*}, M_{0}^{*}\right)=(81,149)$ and $(102,123)$ from the results for BCW and CRX in Table 2, respectively. It may indicate that the bounds $\max \left\{M_{1}, M_{1}(t)\right\}$ and $M_{0}(s)$ are not very accurate indicators of $M_{1}^{*}$ and $M_{0}^{*}$.

It is left as our future work to derive tighter theoretical estimates of $M_{1}^{*}$ and $M_{0}^{*}$.

## 5. Conclusion

In this paper, we considered how many examples are needed in a given data set in order to remove deceptive $\left(a_{1}, a_{0}\right)$ patterns. Our hypothesis is that the data set should contain a greater number of examples than that at which the probability of having ( $a_{1}, a_{0}$ )-patterns vanishes for the random data set. We justified the hypothesis by computational experiments in Sect. 3, and derived estimates of such number of examples by probabilistic analysis in Sect. 4.

Our future work includes the theoretical study of the hypothesis based on such theories as randomness and VC dimension [13] from learning theory, and an application of the hypothesis to other enumeration problems (e.g., graph mining).

## Acknowledgments

We gratefully acknowledge very careful and detailed comments given by anonymous reviewers. We are indebted to Prof. Hiroshi Nagamochi and Prof. Liang Zhao of Kyoto

University, Japan, for a number of helpful comments and suggestions. This work is partially supported by Grant-inAids for Scientific Research on Priority Areas "Comparative Genomics" and "New Horizons in Computing" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," Proc. ACM SIGMOD Conf. Management of Data, pp.207-216, Washington, USA, May 1993.
[2] N. Alon, J.H. Spencer, and P. Erdös, eds., The Probabilistic Method, John Wiley \& Sons, 1992.
[3] A. Asuncion and D.J. Newman, UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 2007.
[4] E. Boros, V. Gurvich, L. Khachiyan, and K. Makino, "On the complexity of generating maximal frequent and minimal infrequent sets," Proc. STACS 2002, pp.133-141, Antibes Juan-les-Pins, France, March 2002.
[5] H. Chernoff, "A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations," Annals of Mathematical Statistics, vol.23, pp.493-509, 1952.
[6] U. Fayyad, G. Piatetsky-Shapiro, and S. Padhraic, "From data mining to knowledge discovery in databases," AI Magazine, vol.17, no.3, pp.37-54, 1996.
[7] K. Haraguchi, T. Ibaraki, and E. Boros, "Classifiers based on iterative compositions of features," Proc. ICKEDS 2004, pp.143-150, Porto, Portugal, June 2004.
[8] W. Hoeffding, "Probability inequalities for sums of bounded random variables," J. American Statistical Association, vol.58, pp.1330, 1963.
[9] Y. Li and R.P. Gopalan, "Effective sampling for mining association rules," Proc. 17th Australian Joint Conference on Artificial Intelligence, pp.391-401, Cairns, Australia, Dec. 2004.
[10] N. Ramakrishnan and C. Bailey-Kellogg, "Gaussian process models of spatial aggregation algorithms," Proc. IJCAI 2003, pp.10451051, Acapulco, Mexico, Aug. 2003.
[11] H. Toivonen, "Sampling large databases for association rules," Proc. 22nd VLDB Conference, pp.134-145, Bombay, India, Sept. 1996.
[12] T. Uno, M. Kiyomi, and H. Arimura, "LCM ver 2: Efficient mining algorithms for frequent/closed/maximal itemsets," Proc. IEEE ICDM'04 Workshop FIMI'04, Brighton, UK, Nov. 2004.
[13] V. Vapnik, ed., The Nature of Statistical Learning Theory, SpringerVerlag, New York, 1995.
[14] G. Yang, "The complexity of mining maximal frequent itemsets and maximal frequent patterns," Proc. 10th ACM SIGKDD Intl. Conf. Knowledge Discovery and Data Mining, pp.344-353, Seattle, USA, Aug. 2004.

Kazuya Haraguchi received the B.E., the Master of Informatics and the Ph.D. degree in Informatics from Kyoto University, in 2000, 2002 and 2006, respectively. He is currently with the Department of Information Technology and Electronics, Faculty of Science and Engineering, Ishinomaki Senshu University. His interest includes algorithms, optimization, machine learning and their applications.


Mutsunori Yagiura received the B.E., M.E. and Ph.D. degrees in Engineering from Kyoto University, in 1991, 1993 and 1999, respectively. He is currently with the Department of Computer Science and Mathematical Informatics, Graduate School of Information Science, Nagoya University. His research interest includes metaheuristics, algorithms, optimization, computational complexity and their applications.


Endre Boros is a Professor of Operations Research at Rutgers University, studied mathematics and operations research at the Eötvös Lorád University, Budapest Ilungary, and received his Doctorate in 1985. His main areas of research include combinatorial optimization, game and graph theory, logical analysis of data (LAD), and the theory of boolean functions and data mining.


Toshihide Ibaraki received the B.E., M.E., and Ph.D. degrees in Engineering from Kyoto University, in 1963, 1965 and 1970, respectively. After retiring from Kyoto University in March 2004, he is currently with Department of Informatics, School of Science and Technology, Kwansei Gakuin University. His interest includes algorithms, optimization, computational complexity and their applications.


[^0]:    Manuscript received June 19, 2006.
    Manuscript revised October 29, 2007.
    ${ }^{\dagger}$ The author is with the Department of Information Technology and Electronics, Faculty of Science and Engineering, Ishinomaki Senshu University, Ishinomaki-shi, 986-8580 Japan.
    ${ }^{\dagger \dagger}$ The author is with the Department of Computer Science and Mathematical Informatics, Graduate School of Information Science, Nagoya University, Nagoya-shi, 464-8603 Japan.
    ${ }^{\dagger \dagger \dagger}$ The author is with the Rutgers Center for Operations Research, Rutgers University, NJ, 08854-8003 USA.
    ${ }^{\dagger+\dagger}+\mathrm{T}$, Science and Technology, Kwansei Gakuin University, Sanda-shi, 669-1337 Japan.
    a) E-mail: kazuyah@ isenshu-u.ac.jp
    b) E-mail: yagiura@nagoya-u.jp
    c) E-mail: boros@rutcor.rutgers.edu
    d) E-mail: ibaraki@ksc.kwansei.ac.jp

    DOI: 10.1093/ietisy/e91-d.3.781

