

A Heuristic Search and its Roughness

Chun-Hung Tzeng
Computer Science Department
Ball State University
Muncie, Indiana 46301
Email: tzengf@bsu.edu

Fu-Shing Sun
Computer Science Department
Ball State University
Muncie, Indiana 46301
Email: fsunf@bsu.edu

Abstract—This paper introduces an abstract model of heuristic search for handling uncertainty. In the search, heuristic information and evaluation are precisely defined mathematically. As special cases, the model includes a previous probabilistic game-tree search and a pattern search. This paper also considers the relationship between the model and the rough set approach. Each rough set formulation can be reformulated as a special case of the abstract model.

Index Terms—Heuristic information, game-tree search, pattern recognition, tolerance relation, tolerance representative, rough set.

I. INTRODUCTION

This paper reports a preliminary result of a recent study on uncertainty based on probability theory. The study of uncertainty often includes search and evaluation. For the evaluation, the information returned by the search usually is incomplete. In this paper, the incomplete information is called heuristic information and the corresponding search is called a heuristic search. This paper introduces an abstract model of heuristic search. Special cases include a probabilistic game-tree search and a pattern search in a pattern recognition. This paper also considers the relationship between the abstract model and the rough set approach in handling uncertainty.

The rough set approach of handling uncertainty can be interpreted as a task of classifying some objects in a universe of discourse, denoted by U . The information for U is usually provided by a set of attributes. An indiscernibility relation R is an equivalent relation on U , which is usually defined by a subset of attributes. Each R -equivalent class $[x]_R$ is an R -exact elementary set. Any subset $Y \subseteq U$ is R -exact if it is a union of R -exact elementary sets. Any subset $X \subseteq U$ which is not R -exact is R -rough. The R -lower and R -upper approximation of X are defined, respectively, as follows:

$$\begin{aligned}\underline{R}X &= \{\omega \in \Omega \mid [\omega]_R \subseteq X\}. \\ \overline{R}X &= \{\omega \in \Omega \mid [\omega]_R \cap X \neq \emptyset\}.\end{aligned}$$

The rough set is extended by use of a tolerance relation instead of an equivalent relation. A tolerance relation is a reflexive and transitive binary relation on U .

For a game-tree search, the goal is to estimate the game values at certain nodes of the game tree. The information provided by a search is usually incomplete to determine the game values. Intuitively, a deeper search returns more precise information and the decision can be improved. However, there

is a pathological phenomenon, in which the minimax algorithm may degrade the information provided by a static evaluation in a deeper search.

A probabilistic formulation of heuristic information in game-tree search has been introduced. The goal of the game search is to estimate the game values at a certain nodes for a given game $\omega \in \Omega$. The set Ω of all possible games is a probability space. Let M be one of the game values, which is a random variable. A search returns a piece of information about ω , which is a subset $X \subseteq \Omega$ containing ω . The theoretical evaluation of M is the conditional expectation $E_X(M)$. A deeper search returns more precise information about ω , which corresponds to a more accurate estimation of M . The actual algorithm depends on each individual game structure. The minimax algorithm does not compute $M_X(M)$ in general, even if M is the minimax value of a node.

Uncertain information also appears in a pattern recognition, which is to classify certain objects. For a given element ω , a two-class recognition problem can be formulated as a determination of the predicate $\omega \in X$, where X is the set of all elements of one class.

The classical probabilistic formulation of pattern recognition on a Euclidean space $\mathbb{R}^n$ is extended to general probability spaces. This approach uses a tolerance relation as a similarity. The heuristic information is generated by a non-crispy representative data clustering which is defined by a set of representative elements. The cluster of a representative is the set of all similar elements to the representative. A conditional probability given a piece of heuristic information is used to estimate the pattern of the given element.

This paper introduces a general abstract model for heuristic search based on probability theory. A search task consists of a probability space $(\Omega, \mu, \mathcal{F})$ and a random variable M . The goal is to approximate the random variable M . A heuristic search is to find heuristic information I about a given element ω . The heuristic information I is a measurable set containing ω : $\omega \in I$. The set of all such possible I 's forms a partition of Ω . The partition generates a Borel subfield $\mathcal{F} \subseteq \mathcal{F}$. The evaluation of M is the conditional expectation $E_{\mathcal{F}}(M)$.

The heuristic game-tree search and the pattern recognition mentioned above are special cases of this abstract model. Each rough set formulation can also be formulated in a special heuristic search model.

The rest of this paper is organized as follows: Section 2 defines the abstract heuristic search model, using a game-tree search as an example. Section 3 discusses the relationship between this model and the rough set approach. Section 4 introduces a pattern recognition model by use of representative data clustering. Section 5 considers an example. The rough set properties in a representative data clustering are computed for different similarities in this section. Section 6 is the conclusion.

II. HEURISTIC INFORMATION AND EVALUATION

Based on probability theory, this section introduces a mathematical model for heuristic information and heuristic search. The heuristic information in a game-tree search (Fig. 1) and the heuristic search by use of representative coverings (Fig. 2) are special cases.

Definition 2.1. A *search model* is a probability space with a random variable, denoted by $(\Omega, \mathcal{F}, \mu, M)$, where Ω is a sample space of all possible elements, $\mathcal{F}$ is a Borel field on Ω , μ is a probability measure on $\mathcal{F}$, and M is a random variable of the probability space $(\Omega, \mathcal{F}, \mu)$.

For simplicity, we assume that Ω is a finite set. Let $F \subseteq \mathcal{F}$ be any Borel subfield. A non-empty set $A \in F$ is F -minimal if there is no non-empty F -measurable set properly contained in A . Suppose that $P_F = \{I_1, \dots, I_n\}$ consists of all F -minimal sets. Then P_F forms a partition of Ω and any F -measurable set is a union of F -minimal sets.

Definition 2.2. A *heuristic search* of a search model $(\Omega, \mathcal{F}, \mu, M)$, is a pair (F, f) where F is a Borel subfield $F \subseteq \mathcal{F}$ and f is a function on Ω . For any given element $\omega \in \Omega$, f returns the F -minimal set I containing ω : $\omega \in I$. The function f is the *heuristic search function* and I is the *heuristic information* about the element ω .

In a heuristic search, we consider an evaluation of the random variable M .

Definition 2.3. The estimator of M , based on the heuristic search $S = (F, f)$, is the conditional expectation M_S of M with respect to F :

$$M_S = E_F(M). \quad (1)$$

The estimator is also a random variable. For any $\omega \in \Omega$, let $I = f(\omega)$, the heuristic information about ω . Then $M_S(\omega)$ is the conditional expectation of M relative to I :

$$M_S(\omega) = E_I(M) = \frac{1}{\mu(I)} \int_I M d\mu. \quad (2)$$

That is, M_S has a constant value in I , which is the average of M . Here we assume I has a positive measure.

Let $P_F = \{I_i \mid i = 1, \dots, n\}$ be the partition corresponding to the heuristic search S . Then M_S is a linear combination:

$$M_S = \sum_i E_{I_i}(M) \mathbf{1}_{I_i}, \quad (3)$$

where $\mathbf{1}_{I_i}$ is the indicator function of I_i : $\mathbf{1}_{I_i}(\omega) = 1$ if $\omega \in I_i$, 0 otherwise.

Consider two heuristic searches $S_1 = (F_1, f_1)$ and $S_2 = (F_2, f_2)$. If F_1 is a subfield of F_2 : $F_1 \subseteq F_2$, then we say that S_2 is *more precise* than S_1 . In this case, each F_1 -minimal set is a union of F_2 -minimal sets. The corresponding estimators have the following property (Fig. 3).

Theorem 2.1. If S_2 is more precise than S_1 , then the estimator M_{S_1} is also an estimator of M_{S_2} :

$$M_{S_1} = (M_{S_2})_{S_1}. \quad (4)$$

In each F_1 -minimal set, the value of M_{S_1} is the average of M_{S_2} . That is, M_{S_2} is more accurate than M_{S_1} in estimating M .

To illustrate the idea of the heuristic information formulation, we consider a uniform binary game tree of four levels (Fig. 4), where each terminal node has independently the value 1 (a win for MAX) or 0 (a loss for MAX) with a probability p ($0 < p < 1$) and $1 - p$, respectively. The space $\Omega = \{0, 1\}^8$ of all leaf patterns forms a probability space. The game value at each internal node is the minimax value at that node, which is a random variable. We consider the task of estimating the minimax value at the root, denoted by M . Then the probability space Ω and M form a search model.

We consider two searches. First, consider the heuristic searches $S_1 = (F_1, f_1)$ where the function f_1 counts the number of 1's at the leaves under the root. The corresponding partition of F_1 consists of 9 members: $I_k = \{(x_1, \dots, x_8) \in \Omega \mid \sum_{i=1}^8 x_i = k\}$, $k = 0, 1, \dots, 8$. For a given game $x' = (x'_1, \dots, x'_8)$, the heuristic information is I_k , where $k = \sum_{i=1}^8 x'_i$, based on which the estimation of M is the conditional probability:

$$M_{F_1}(x') = P(M = 1 \mid x' \in I_k).$$

The second heuristic search is $S_2 = (F_2, f_2)$. The function f_2 counts the numbers of 1's at the leaves under each of the two nodes of level 2. The Borel field F_2 is generated by the partition consisting of following 25 sets:

$$I_{ij} = \{(x_1, \dots, x_8) \in \Omega \mid \sum_{k=1}^4 x_k = i, \sum_{k=5}^8 x_k = j\}, 0 \leq i, j \leq 4.$$

For the given game $x' = (x'_1, \dots, x'_8)$, the heuristic information is I_{ij} , where $i = \sum_{k=1}^4 x'_k$ and $j = \sum_{k=5}^8 x'_k$, based on which the estimation of M is

$$M_{F_2}(x') = P(M = 1 \mid x' \in I_{ij}).$$

The search S_2 is more precise than S_1 because $I_k = \cup_{i+j=k} I_{ij}$ for each I_k in F_1 . The both estimators have the following relation:

$$P(M = 1 \mid x' \in I_k) = \sum_{i+j=k} \mu(I_{ij}) P(M = 1 \mid x' \in I_{ij}).$$

III. HEURISTIC SEARCH AND ROUGH SET

This section shows that in a heuristic search any measurable set of probability space may be treated as a rough set $(\Omega, \mathcal{F}, \mu)$. And each rough set formulation can be reformulated as a heuristic search.

Suppose that X is a measurable set in a given a probability space $(\Omega, \mathcal{F}, \mu)$. Let $M = \mathbf{1}_X$, the indicator of X , which is a random variable because X is measurable. Let $S = (F, f)$ be a heuristic search for the search model $(\Omega, \mathcal{F}, \mu, M)$. Consider the corresponding partition of Ω : $P_F = \{I_1, \dots, I_n\}$, which consists of all F -minimal sets. The partition defines an indiscernibility relation (also denoted by F) in the rough set terminology $(\Omega, \mathcal{F}, \mu, M)$: two elements ω_1 and ω_2 are indiscernible if $f(\omega_1) = f(\omega_2)$ that is, for the two elements, the heuristic search returns the same heuristic information I_i : $\omega_1 \in I_i$ and $\omega_2 \in I_i$. Based on this indiscernibility relation, all F -exact elementary sets are those I_i 's in P_F . The F -lower and F -upper approximations of X are, respectively,

$$\underline{F}X = \{\omega \in \Omega \mid M_F(\omega) = 1\} \quad \text{ffi}$$

$$\overline{F}X = \{\omega \in \Omega \mid M_F(\omega) > 0\} \quad \text{ffi}$$

That is, $\underline{F}X$ consists of all F -minimal sets contained in X and $\overline{F}X$ consists of all F -minimal sets which has a non-empty intersection with X and the intersection has a positive measure. If $\underline{F}X = \overline{F}X$, then X is F -exact. Any member in the Borel field F is F -exact.

Conversely, each rough set formulation $(\Omega, \mathcal{F}, \mu, M)$ can be treated as a special heuristic search. In a rough set formulation, let Ω be the universe of objects and R the equivalence relation on Ω . The R -exact elementary sets are the R -equivalence classes $[x]_R$, $x \in \Omega$. Any union of R -exact elementary sets is R -exact. Any set $X \subseteq \Omega$ which is not R -exact is R -rough. The lower and upper approximation of X are defined, respectively, as follows:

$$\underline{R}X = \{\omega \in \Omega \mid [\omega]_R \subseteq X\} \quad \text{ffi}$$

$$\overline{R}X = \{\omega \in \Omega \mid [\omega]_R \cap X \neq \emptyset\} \quad \text{ffi}$$

Consider the universe of objects (i.e., Ω) as a probability space, where the Borel field $\mathcal{F}$ is the set of all subsets of Ω and each object $\omega \in \Omega$ is measurable and has the same probability measure $\mu(\omega) = \frac{1}{|\Omega|}$. Then the indicator $M = \mathbf{1}_X$ of X is a random variable. For the search model $(\omega, \mathcal{F}, \mu, M)$, we consider the following heuristic search. Let F be the Borel field generated by the R -exact elementary sets. That is, the F -minimal sets are the R -exact elementary sets and the members of F are R -exact. The function $f(\omega) = [\omega]_R$ is a heuristic function. The lower and upper approximation of X are, respectively, $\underline{F}X = \{\omega \in \Omega \mid M_F(\omega) = 1\}$ and $\overline{F}X = \{\omega \in \Omega \mid M_F(\omega) > 0\}$. The estimator M_F at an element ω is the conditional probability $M_F(\omega) = P(\omega \in X \mid \omega \in [\omega]_R)$ that is, it is the rough membership function defined in $(\Omega, \mathcal{F}, \mu, M)$:

$$M_F(\omega) = \mu(\omega, X) = \frac{|[\omega]_R \cap X|}{|[\omega]_R|}.$$

Therefore each rough set formulation is a special heuristic search for a search model.

III. HEURISTIC SEARCH USING REPRESENTATIVES

This section introduces a heuristic information computed by comparing an element with representatives. First, we consider a pattern recognition model on a probability space $(\Omega, \mathcal{F}, \mu)$, $(\Omega, \mathcal{F}, \mu, \eta)$, which is an extension of the classical model $(\Omega, \mathcal{F}, \mu)$.

A. Pattern Recognition and Classifiers

Pattern recognition is about guessing or predicting the unknown nature of an observation. Formally, we use ω to denote an observation and Ω to denote the space of all possible observations. For simplicity, we consider only two classes (e.g., normal and anomaly), denoted by 0 and 1.

We assume the observation is probabilistic. Consider the probability space $(\Omega, \mathcal{F}, \mu)$. Let (X, Y) be a random pair taking their respective values from Ω and $\{0, 1\}$. The random pair is defined by a pair (μ, η) , where μ is the probability measure of X and η is the posterior probability, given the value of X . That is, for any $A \in \mathcal{F}$, $P(X \in A) = \mu(A)$, and for any $\omega \in \Omega$, $\eta(\omega) = P(Y = 1 \mid X \in [\omega]_{\mathcal{F}})$, where $[\omega]_{\mathcal{F}}$ is the $\mathcal{F}$ -minimal set containing ω . Note that η is measurable (relative to $\mathcal{F}$) and the 4-tuple $(\Omega, \mathcal{F}, \mu, \eta)$ forms a search model defined in Section 1.

In pattern recognition, one creates a classifier $g : \Omega \rightarrow \{0, 1\}$ to represent one's guess of y , the class of a given observation ω . The classifier errs on ω when $g(\omega) \neq y$. We only consider measurable classifiers, which always assign indistinguishable observations (i.e., elements in the same $\mathcal{F}$ -minimal set) to a same class.

Definition 4.1. A classifier on the search model $(\Omega, \mathcal{F}, \mu, \eta)$ is an F -measurable function $g : \Omega \rightarrow \{0, 1\}$, where $\{0, 1\}$ is treated as a discrete space.

The probability of error $L(g) = P(g(X) \neq Y)$ is the *error rate* of g :

$$L(g) = P(g(X) \neq Y) = \int_{\Omega} P(g(X) \neq Y \mid X \in [\omega]_{\mathcal{F}}) d\mu(\omega). \quad \text{ffi}$$

The *false positive rate* $L^{(0)}(g)$ and the *false negative rate* $L^{(1)}(g)$ are

$$L^{(0)}(g) = P(g(X) = 1 \mid Y = 0), \quad \text{ffi0}$$

and

$$L^{(1)}(g) = P(g(X) = 0 \mid Y = 1). \quad \text{ffi1}$$

The relation among the three types of error is

$$L(g) = L^{(0)}(g)P(Y = 0) + L^{(1)}(g)P(Y = 1).$$

Let $0 \leq \theta < 1$. The classifier g_{θ} is defined as follows

$$g_{\theta}(\omega) = \begin{cases} 1 & \text{if } \eta(\omega) > \theta, \\ 0 & \text{otherwise.} \end{cases} \quad \text{ffi2}$$

Note that, if $\eta(\omega) \in \{0, 1\}$, then $L(g_{\theta}) = L^{(0)}(g_{\theta}) = L^{(1)}(g_{\theta}) = 0$ for $0 < \theta < 1$. The classifier $g_{0.5}$ is the *Bayes classifier* and $L(g_{0.5})$ is the *Bayes Error*. Similarly to the classical theory $(\Omega, \mathcal{F}, \mu)$ we have following theorems.

Theorem 4.1. (Bayes) For any classifier g , $L(g_{0.5}) \leq L(g)$.

Theorem 4.2. (Neyman-Pearson) For any classifier g and θ ($0 < \theta < 1$), if $L^{(0)}(g) \leq L^{(0)}(g_\theta)$, then $L^{(1)}(g) \geq L^{(1)}(g_\theta)$.

The Bayes classifier is an optimal classifier if we are required to minimize the error rate. If the false positive rate $L^{(0)}$ is required to be kept under a certain level, then g_θ minimizes the false negative rate $L^{(1)}$ for some θ .

B. Representative data clustering and heuristic information

First we consider a tolerance relation ξ on the search model $(\Omega, \mathcal{F}, \mu, \eta)$, which is a reflexive and symmetric binary relation that is, $(\omega, \omega) \in \xi$ for any $\omega \in \Omega$, and $(\omega_1, \omega_2) \in \xi$ implies $(\omega_2, \omega_1) \in \xi$. A space with a tolerance relation is called a tolerance space.

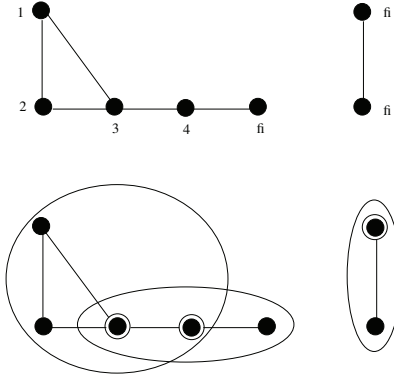


Fig. 1. A tolerance space and a minimal representative clustering.

We also use ξ as a predicate: $\xi(\omega_1, \omega_2)$ iff $(\omega_1, \omega_2) \in \xi$. If $\xi(\omega_1, \omega_2)$, we say that ω_1 is ξ -similar to ω_2 , or ω_1 and ω_2 are ξ -similar. We may omit ξ when there is no ambiguity. Any undirected graph is a tolerance space (and vice versa), where Ω is the set of all vertices, and two vertices are similar if they are the same vertex or they are adjacent (e.g., Fig. 1).

For each $\omega \in \Omega$, the set $\xi(\omega) = \{u \in \Omega \mid \xi(\omega, u)\}$ is called the neighborhood of ω . Here we assume that each neighborhood is measurable (i.e., an element of $\mathcal{F}$). We consider a data clustering of the search model by use of the neighborhoods (e.g., Fig. 1). Each $\omega \in \Omega$ is called a *representative* of its neighborhood $\xi(\omega)$. A set of elements $\{r_1, \dots, r_k\}$ is a representative system of Ω if the corresponding neighborhoods cover the whole space Ω . The representative system is minimal (if) if there is no other representative system with less than k members.

A representative system forms a clustering of Ω , in which the clusters are the corresponding neighborhoods. Note that each cluster is measurable. We call such a clustering a *representative clustering*. The clustering is in general not a partition (e.g., Fig. 1).

An element $\omega \in \Omega$ is called a *surprise* (if) if ω is not similar to any other element in Ω (that is, $\xi(\omega) = \{\omega\}$). Therefore, each surprise is always in any representative system. Let the numbers of the surprises be denoted by α_ξ , which is independent of the choice of representative systems. The

similarity ξ does not provide any pattern about surprises. In application, searching a similarity should eliminate or reduce the number of surprises.

Let $\mathbf{R}_\xi = \{r_1, \dots, r_k\}$ be a minimal representative system. Let the neighborhoods of r_i 's generate the Borel subfield D ($D \subseteq \mathcal{F}$), and the corresponding partition be $P_D = \{A_i \mid i = 1, \dots, n\}$. For example, in Fig. 1 there are three representatives and the corresponding partition is $\{\{1, 2\}, \{3, 4\}, \{5\}, \{6, 7\}\}$. For an $\omega \in \Omega$, the member A_i containing ω , denoted by $f(\omega)$, is computed as follows. We compare ω with the k representatives. Let S_ω be the set of all representatives similar to ω and $N_\omega = \mathbf{R}_\xi - S_\omega$. Then

$$f(\omega) = A_i = \cap_{r \in S_\omega} \xi(r) - \cup_{r \in N_\omega} \xi(r). \quad (3)$$

The pair $S = (D, f)$ is a heuristic search for the search model $(\Omega, \mathcal{F}, \mu, \eta)$. The estimator of η is $\eta_D = E(\eta|D)$. Given ω , $\eta_D(\omega)$ is the conditional probability:

$$\eta_D(\omega) = \frac{1}{\mu(A_i)} \int_{A_i} \eta(t) d\mu(t) = P(Y = 1 | X \in A_i). \quad (4)$$

We define classifiers based on the heuristic search $S = (D, f)$. A D -classifier is a D -measurable function $g^D : \Omega \rightarrow \{0, 1\}$. Note that g^D is constant on each A_i . As in Section 1, any D -classifier g^D is a linear combination of the indicators 1_{A_i} 's. Any D -classifier is also $\mathcal{F}$ -measurable and hence a classifier on the search model $(\Omega, \mathcal{F}, \mu, \eta)$, but not vice versa.

For D -classifiers, we have following similar definitions and properties. The D -classifier $g_\theta^D = 1$ iff $\eta_D(x) > \theta$, $0 \leq \theta < 1$. The Bayes D -classifier $g_{0.5}^D$ minimizes the error rate of all D -classifiers. The Neyman-Pearson Theorem is also true (that is, for any D -classifier g^D , if $L^{(0)}(g^D) \leq L^{(0)}(g_\theta^D)$, then $L^{(1)}(g^D) \geq L^{(1)}(g_\theta^D)$). Since the Bayes classifier $g_{0.5}$ minimizes the error rate, $L(g_{0.5}^D) \geq L(g_{0.5})$.

To search for a minimal representative system is intractable in general. A heuristic method to search for sub-minimal representative system has been introduced (Fig. 1, Fig. 4).

C. Application in Supervised Learning

This subsection briefly summarizes an application of heuristic search in a supervised learning, which has been introduced in Fig. 1. Suppose a supervised learning process consists of two sets: a training data set and a testing data set. The goal of the process is to classify each element in the testing data set. We assume that there are two classes, 0 and 1. We treat the training data set as the space Ω . We assign each element in Ω the same probability measure $\frac{1}{|\Omega|}$ if there is no previously given probability. From the given class label in the training data set, the posterior probability η can be computed. That is, we formulate the training data set into a search model $(\Omega, \mathcal{F}, \mu, \eta)$, where $\mathcal{F}$ is the Borel field corresponding to the probability measure.

The training process is to search for a suitable similarity ξ (i.e., a tolerance relation) and to compute a sub-minimal representative system $\mathbf{R}_\xi$, which defines a heuristic search $S = (D, f)$. Based on $\mathbf{R}_\xi$ the estimator η_D is computed. Then the classifier g_θ^D ($0 \leq \theta < 1$) is defined by use of η_D .

For each element ω in the testing data set, the heuristic information about ω in the training data set is computed by comparing ω with the representatives, as discussed above. Then we use the value of η_D on the searched heuristic information as an evaluation of the element: $\eta_D(\omega)$. Based on this evaluation, each classifier g_D^{θ} is extended to the testing data set. If ω is an unknown surprise (i.e., not similar to any representative), then the evaluation is not defined. In this case, ω may be evaluated by the distribution of surprises in the training data set.

The training result can be studied by use of different type of errors and the receiver operating characteristic analysis (fifth). In the training process, the Bayes error $L(g_{0.5}^D)$ should be as near to the original Bayes error $L(g_{0.5})$ as possible. To be a typical pattern, each representative should be similar to as many records as possible because surprising elements predict existence of unknown surprises in the testing data set which are not similar to any representatives. There is no heuristic information for the unknown surprise. One the other hand, for the computational purpose, the number of representatives should be small relative to the training data set size $|\Omega|$.

fi. EXAMPLES

The section uses the Breast Cancer $\mathbb{R}$ (Original) Data Set $\mathbb{R}$ as an example data set. The data set consists of 99 Instances. The instances have 10 attributes and a class attribute. Except the first attribute which is the sample code number, there are 9 attributes to consider for a pattern search task. The domain for the attributes are integers from 1 to 10. There are few instances that contain a single missing (i.e., unavailable) attribute value. For the class attribute, there are two classes: 2 for benign and 4 for malignant. There are 459 benign instances and 241 malignant instances. Let X be the set of malignant instances, we consider the approximations of X in the rough set terminology by use of representative clusterings.

Let $(x_1^\omega, \dots, x_9^\omega)$ represent the 9 attributes of an instance ω . We consider three different kinds of similarities. In the following we assume that $|x_i^{\omega_1} - x_i^{\omega_2}| = 5$ if one of the values is missing.

The first similarity is:

$$\xi_1(\omega_1, \omega_2) \quad \text{if} \quad |x_i^{\omega_1} - x_i^{\omega_2}| \leq 3, \quad i = 1, \dots, 9. \quad \text{fil fi)}$$

That is, for two instances if the difference of each attribute values is less than or equal to 3, then they are ξ_1 -similar.

The second one is:

$$\xi_2(\omega_1, \omega_2) \quad \text{if} \quad |x_i^{\omega_1} - x_i^{\omega_2}| \leq 2 \text{ for a least 8 attributes.}$$

That is, ω_1 and ω_2 are ξ_2 -similar if there are 8 attributes for which the differences are less than or equal to 2.

The third one is:

$$\xi_2(\omega_1, \omega_2) \text{ if } \sum_{i=1}^9 |x_i^{\omega_1} - x_i^{\omega_2}| \leq 15. \quad \text{fil fi})$$

That is, ω_1 and ω_2 are ξ_3 -similar if the sum of of all differences is less than or equal to 1fi .

TABLE I
SUMMARIES OF THE EXAMPLES

	$ \mathbf{R}_\xi $	α_ξ	$ \mathbf{P}_\xi $	$ \mathcal{L}_\xi $	$ \mathbf{B}_\xi $	b_ξ	m_ξ	B-error
ξ_1	97	36	229	1f5f	7	20	10	0.0114
ξ_2	66	12	224	1f5f	5	9	6	0.002f
ξ_3	32	3	176	1f3	10	59	24	0.01f2

For each similarity ξ , we compute a sub-minimal representative system $\mathbf{R}_\xi$ and the corresponding partitions $\mathbf{P}_\xi$. The number of surprises is α_ξ . The set of the exact elementary sets in the lower approximation of X is represented by L_ξ . The set of boundary exact elementary sets is denoted by $\mathbf{B}_\xi$. A boundary exact elementary set is an exact elementary set which is in the upper approximation but not in the lower approximation. Let b_ξ and m_ξ be the numbers of benign and malignant instances in the boundary, respectively.

The results are summarized in the Table 1, where B-error represents the Bayes error. For example, in the first similarity ξ_1 , there are 9fi representatives, 229 exact elementary sets. There are 10 malignant instances in the boundaryfi that is 231 malignant instances are in the lower approximation of X . The similarity ξ_2 gives a better result, where 23fi malignant instances are in the lower approximation of X and there are only 1fi fi.e., 9+6) instances in the boundary. The Bayes error is 0.002fi.

fi I. CONCLUSION

This paper introduces an abstract model of heuristic search based on probability. Both the previous probabilistic game-tree search and the pattern search are special cases of the abstract model. Each rough set formulation can also be reformulated as a special case of the abstract model. Further research on the abstract model is needed in order to understand more about the theoretical properties of heuristic search and heuristic information.

REFERENCES

- [1] Breast Cancer is inconsin+f(Original) data set, 1992
[http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+isconsin+f\(Original\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+isconsin+f(Original)).
- [2] K. L. Chung. *A Course in Probability Theory*. Academix Press, New fiork, 19fi4.
- [3] L. Devroye and fi. L. fiyoerffi. *A Probabilistic Theory of Pattern Recognition*. Springer, 199fi.
- [4] fi. Maak. *Fastperiodische Funktionen*. Springer-fierlag, 19fiifi.
- [5] M. A. E. Maloof. *Machine Learning and Data Mining for Computer Security*. Springer-fierlag, 200fi.
- [6] J. Pearl. On the nature of pathology in game tree searching. *Artificial Intelligence* 20, 19fi3, pages 42fi–4fi3.
- [7] J. Pearl. *Heuristics: intelligent search strategies for computer problem solving*. Addison-fi esley, Reading, MA, 19fi4.
- [8] L. Polkowski. *Rough Sets*. Springer-fierlag, 2002.
- [9] A. Skowron and J. Stepaniuk. Tolerance Approximation Spaces. *Fundamenta Informaticae*, 2fi, pages 24fi–2fi3, 199fi.
- [10] C.-H. Tzeng and O. C.-S. Tzeng. Tolerance Spaces and Almost Periodic Functions. *Bulletin of the Institute of Mathematics Academia Sinica. fiol. fi*, No. 1, pages 1fi9–1fi3, 19fiifi.
- [11] C.-H. Tzeng and P. fi. Purdom, Jr. A Theory of fi ame Tree. *Proceedings of the National Conference on Artificial Intelligence AAAI-83*, pages 41fi–419, fi illiam Kaufmann, Los Altos, CA. 19fi3.
- [12] C.-H. Tzeng. *A Theory of Heuristic Information in Game-Tree Search*. Springer-fierlag, 19fiifi.

- fil 3fi C.-H. Tzeng and F.-S. Sun. Data clustering in tolerance space. *Berthold, Lenz, Bradley, Kruse, and Borgelt, editors, Advances in Intelligent Data Analysis V*, pages 29fi–30fi. Springer-fierlag, 2003.
- fil 4fi C.-H. Tzeng. Similarity and pattern recognition. *Proc. Intern. Conf. on Data Mining and Appl.*, 200fi, pages fifi4–fi90
- fil fifi C.-H. Tzeng A Probabilistic Model of Pattern Recognition on Abstract Data. *Proceedings of the 2009 Intern. Conf. on Data Mining CSREA Press* 2009, pages 319-32fi
- fil fifi C.-H. Tzeng and F.-S. Sun. A Pattern Search in Data Analysis. *International Journal of Electronic Commerce Studies* fiol. 1, No. 2, 2010, pages 11fi-13fi.
- fil fifi fi . H. fi olberg and O. L. Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences, U.S.A.*, fiol fifi, pages 9193–919fi, 1990.
- fil fifi L. fiadeh. A New Direction in AI Toward a Computational Theory of Perceptions, *AI Magazine*, Spring 2001, pages fi3–fi4.
- fil 9fi E. C. fieman. The topology of the brain and visual perception. *Topology of 3-Manifolds and related Topics*, pages 240–2fifi. Proc. The Univ. of fi eorgia Institute, 19fi2.