Soft Comput
DOI 10.1007/s00500-015-1735-0

METHODOLOGIES AND APPLICATION

Mining unexpected patterns using decision trees and interestingness measures: a case study of endometriosis

Ming-Yang Chang[1] · **Rui-Dong Chiang**[2] · **Shih-Jung Wu**[3] · **Chien-Hui Chan**[2]

Abstract Because clinical research is carried out in complex environments, prior domain knowledge, constraints, and expert knowledge can enhance the capabilities and performance of data mining. In this paper we propose an unexpected pattern mining model that uses decision trees to compare recovery rates of two different treatments, and to find patterns that contrast with the prior knowledge of domain users. In the proposed model we define interestingness measures to determine whether the patterns found are interesting to the domain. By applying the concept of domain-driven data mining, we repeatedly utilize decision trees and interestingness measures in a closed-loop, in-depth mining process to find unexpected and interesting patterns. We use retrospective data from transvaginal ultrasound-guided aspirations to show that the proposed model can successfully compare different treatments using a decision tree, which is a new usage of that tool. We believe that unexpected, interesting patterns may provide clinical researchers with different perspectives for future research.

Keywords Treatment comparison · Unexpected patterns · Domain-driven data mining · Interestingness measures

Abbreviations

ANOVA	Analysis of variance
D^3M	Domain-driven data mining
CA-125	Cancer antigen 125, carcinoma antigen 125, or carbohydrate antigen 125
BMI	Body mass index
CART	Classification and regression tree
EST	Ethanol sclerotherapy
ID3	Iterative Dichotomiser 3 algorithm
CHAID	Chi-Square Automatic Interaction Detector
tech()	Technical interestingness measures
biz()	Business interestingness measures
act()	Actionability of a pattern
tech_obj()	Technical objective interestingness measures
tech_sub()	Technical subjective interestingness measures
biz_obj()	Business objective interestingness measures
biz_sub()	Business subjective interestingness measures
RecoveryRate()	Probability patient recovers from illness
$IM_{\text{tech\_obj}}$	Technical objective interestingness measure
$IM_{\text{tech\_sub}}$	Technical subjective interestingness measure

Communicated by V. Loia.

✉ Rui-Dong Chiang
081863@mail.tku.edu.tw

Ming-Yang Chang
mychang@adm.cgmh.org.tw

Shih-Jung Wu
wushihjung@mail.tku.edu.tw

Chien-Hui Chan
emmacc@gmail.com

[1] Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, Taipei, Taiwan, ROC

[2] Department of Computer Science and Information Engineering, Tamkang University, New Taipei City, Taiwan, ROC

[3] Department of Innovative Information and Technology, Tamkang University, Yilan County, Taiwan, ROC

Published online: 07 June 2015

$IM_{\text{biz\_obj}}$	Business objective interestingness measure
$IM_{\text{biz\_sub}}$	Business subjective interestingness measure

1 Introduction

This research targeted patients whose endometriosis had recurred after undergoing related surgeries. In our experience medical experts usually feel more comfortable using statistical methods to conduct studies. To measure the difference between treatment regimens, researchers usually divide patients into two groups (Hsieh et al. 2009; Kafali et al. 2003; Noma and Yoshida 2001). For example, in Hsieh et al. (2009), all patients were divided into two groups: "ethanol irrigation" and "ethanol retention," respectively. Student's t tests were used to compare the means of the two treatments. The typical method is to first propose a null hypothesis, meaning that the two sets of data are statistically similar. Generally, a p value of less than 0.05 is regarded as statistically significant and small enough to justify rejection of the null hypothesis, which means that the two sets of data are essentially different. Whereas the t test is only suitable for comparing two treatment means, an analysis of variance (ANOVA) can be used both to compare several means and in more complex situations (Bolton and Bon 2009a).

In medical research it is important to identify and treat the cause of each problem correctly because different conditions may generate different recovery rates for patients. However, it is difficult to accurately describe the conditions of each group of patients using statistical methods in statements such as: when a patient's "*body weight* $<$ 56.3 kg" and "*age* $<$ 40," treatment A is better than treatment B. Neither t test nor ANOVA can directly partition continuous variables into different groups. Thus, using statistical methods, we can only describe the conditions of each group of patients with statements such as: there are significant differences in a patient's *body weight* and *age* between treatment A and treatment B.

Even though we can use regression analysis to estimate the relationships between variables (Bolton and Bon 2009b), we still cannot easily determine the appropriate cutoff points for continuous variables. In addition, inappropriately partitioned data may result in a finding of no statistically significant differences between groups, particularly when the medical conditions of two groups of patients are very similar. Therefore, when we use statistical methods to find the cutoff point for continuous variables, it is necessary to repeatedly examine each cutoff point. On the other hand, through a decision tree algorithm, we can easily divide patients into different groups and generate the different medical conditions of each group. In our previous research (Wang et al. 2013), we successful employed decision trees as a classification function to divide patients into different groups.

The aim of this paper was to find unexpected patterns which can describe scenarios that contradict domain experts' prior knowledge. Unexpectedness was first brought up by Silberschatz and Tuzhilin (1995). Unexpectedness is formalized with respect to background knowledge which either is explicitly defined by a user or represents common sense domain knowledge (Kontonasio et al. 2012). Unexpected patterns are interesting because they contradict a person's existing knowledge or expectations and may suggest an aspect of the data that needs further study. Bay and Pazzani (2001) proposed a search algorithm to mine contract sets and prune association rules, as well as a post-process to present a surprising subset to the user. According to a survey by Kontonasio et al. (2012), most studies that measure the unexpectedness of patterns focus on association rules. To the best of our knowledge, none of these studies use interestingness measures to find unexpected patterns in decision trees.

According to Lenca et al. (2008), each domain and problem has a different set of best measures. In this case, we define our interestingness measures based on the domain-driven data mining (D^3M) concept so as to detect unexpected patterns that contrast with domain knowledge. Generally, data-centered mining is done to identify interesting patterns. During the data mining process, environmental factors are usually filtered or simplified; pattern identification is often based on technical significance or interestingness. Individual user requirements and domain-related knowledge are less considered. Cao et al. (2010) advocated that the current algorithms, patterns, and produced models lack workability, actionability, and operable capability. Therefore, they proposed D^3M to solve these issues. Domain-driven data mining has several key components: constraint of the knowledge delivery environment, in-depth pattern mining, enhanced knowledge actionability, and closed-loop and iterative refinement (Cao et al. 2010; Zhu et al. 2009).

In-depth pattern mining can discover interesting and actionable knowledge from a domain-specific viewpoint, and uncover deep data intelligence and interior business rules that data-driven data mining cannot discover (Cao et al. 2010). A closed-loop process means that outputs of data mining can be fed back in to change relevant factors or parameters at particular stages (Cao et al. 2010). A pattern is actionable in a domain if it can be used to make decisions about future actions in the domain (Ling et al. 2002; Wang et al. 2002). Since clinical studies occur in complex environments, utilizing prior domain knowledge, constraints, and expert knowledge can enhance the capabilities and performance of data mining. Sebastian and Then (2011) also mention that, by focusing the mining process on domain-knowledge-compliant rules, the outputs will also be more acceptable to domain experts.

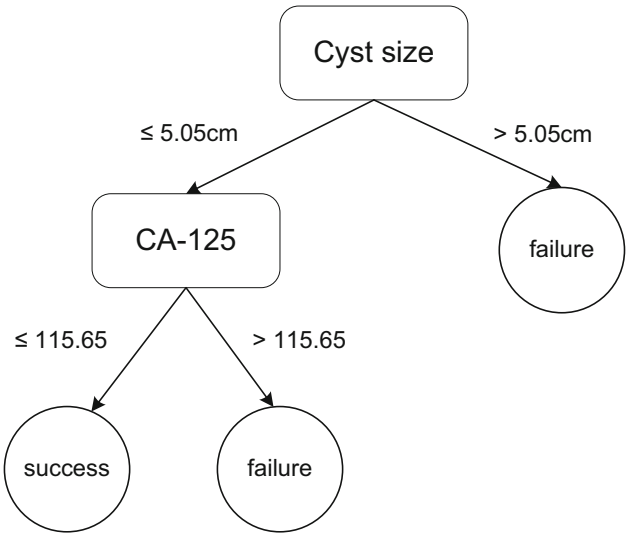

Fig. 1 Resulting tree built on data from patient group: treatment A

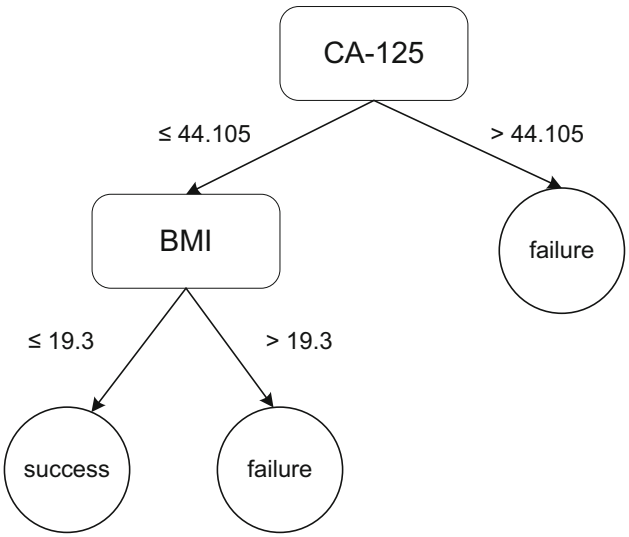

Fig. 2 Resulting tree built on data from patient group: treatment B

To find unexpected patterns, we used decision trees to compare the effectiveness of different treatments, at each individual node, for patients with the same physiological conditions. This allows users to make comparisons between treatments. For example, a rule could be extrapolated that patients who are under 40 years old with a *BMI* of less than 23 respond better to treatment *A* than treatment *B*. However, when we used different trees to analyze different treatments, we obtained the results shown in Figs. 1 and 2. We can thus retrieve statements such as: when a patient's "*Cyst size* $\leq$ 5.05 cm" and "*CA*-125 $\leq$ 115.65," treatment *A* is successful. And when a patient's "*CA*-125 $\leq$ 44.105" and "*BMI* $\leq$ 19.3," treatment *B* is successful.

Because the trees were built on data from different groups, the variables chosen as nodes and the cutoff points of continuous variables may differ. Even if we increase the weights so that the same variables are used in the nodes of two different trees, the cutoff points of the continuous variables for different trees may still be different. It is therefore difficult to directly compare our results. Moreover, since decision trees use binary targets to compare differences between sibling nodes, no studies use decision tree technology for comparisons at each individual node.

Lastly, we created a mining system architecture using closed-loop processing and in-depth mining, in order to uncover unexpected patterns with a decision tree. When using SPSS Clementine in our experiment, we found that the unexpected patterns of the C4.5/C5.0 and CART (Classification and Regression Tree) trees differed only slightly in terms of the cutoff points of the continuous variables. Thus, our method can be successfully applied to both C4.5/C5.0 and CART decision trees. Since we are unable to limit the depth of the C4.5/C5.0 tree, the resulting tree may include a lot of nodes. On the other hand, CART allows us to set a "prune level" which can limit the tree depth and reduce the number of nodes. Therefore, we used the CART algorithm in order to clearly demonstrate our method.

In this paper, we integrated CART (Breiman 1984) with a D^3M approach, and compared two different treatments at each individual node to find an unexpected pattern for endometriosis; for the entire decision tree, several unexpected patterns may result. Therefore, this study's contributions include the following:

1. Proposal of a new method to use decision tree techniques for comparison at individual nodes.
2. Definition of interestingness measures based on the D^3M concept to detect unexpected patterns.
3. Creation of a mining system architecture using decision trees to uncover unexpected patterns.

The rest of the paper is organized as follows. A brief summary of related work is presented in Sect. 2, and the proposed method is in Sect. 3. The experiments and an example using retrospective data from transvaginal ultrasound guided aspirations are presented in Sect. 4. Section 5 presents the conclusions.

2 Related works

2.1 Endometriosis

Endometriosis is one of the most common issues in gynecology (Nap et al. 2004). It is defined as the presence of endometrial-like tissue outside the uterus (Kennedy et al. 2005); this causes pain (e.g., pelvic pain, dysmenorrhea, and

dyspareunia) and infertility, although 20–25 % of patients are asymptomatic (Bulletti et al. 2010). The associated symptoms of endometriosis can impact a woman's quality of life in many ways. Endometriosis should be considered a chronic disease characterized by high recurrence rates (Bulletti et al. 2010). In addition, women who have their ovaries removed are still at risk of recurrence of endometriosis. The treatment for endometriosis should be chosen by each individual patient, depending on symptoms, age, and fertility (Bulletti et al. 2010). In other words, treatment depends on how severe the patient's symptoms are and whether the patient still has an intention to bear children. Traditional treatment approaches are medical (hormone therapy and anti-inflammatories) and surgical (laparoscopy and laparotomy); also, a combination of these approaches can be offered to patients (Bulletti et al. 2010; Kennedy et al. 2005). Many patients require a combination of treatments (Bulletti et al. 2010). Unfortunately, in terms of reproductive success, none of these approaches has an absolute advantage or disadvantage over the other (Zhu et al. 2011). Hormone therapies for endometriosis can cause side effects and pose certain health risks. Additionally, patients face repeated laparoscopy or laparotomy surgical interventions; they commonly face problems of dense pelvic adhesion, which may cause serious tissue damage and diminished ovarian reserve. Some researchers have reported that repeated, conservative surgery has the same efficacy and limitations as primary surgery on symptomatic endometriosis, but pregnancy rates are almost half those obtained after primary surgery (Berlanda et al. 2010; Vercellini et al. 2009).

Transvaginal ultrasound-guided aspiration with ethanol sclerotherapy (EST) is an alternative treatment that can minimize surgical risks and effectively decrease cyst size and related symptoms of cyst compression (Donnez et al. 2011; Ikuta et al. 2006). The main goal of EST is to prevent the suffering and possible complications of surgical procedures, and to treat illnesses with more patient-friendly strategies. Ideally, if treatment time is sufficient, this procedure may result in total regression of the inflammatory cyst (Kafali et al. 2003). Although Kafali et al. (2003) performed five minutes of irrigation on endometriosis patients with 70 % ethanol and erythromycin, results were not encouraging. Therefore, ten minutes of ethanol irrigation of each endometrioma as proposed by Noma and Yoshida (2001) is still the preferred therapeutic reference guide.

2.2 Decision trees

A decision tree uses a branch structure to produce easily understandable classification rules. In current practice the decision tree can be considered a fairly mature technique. Decision trees are popular due to their simplicity and transparency. They can be used to represent both classifier and regression models. When a decision tree is used for classification tasks, it is referred to as a classification tree; when it is used for regression tasks, it is called a regression tree (Rokach and Maimon 2008). The decision tree can produce results according to different variables through repetition, and can thus be used to analyze characteristics, similarities, and differences in data; this data is often represented graphically and in a visually effective manner. Figure 3 presents the tree that result from applying the CART procedure to the examples in Table 1. At each leaf is the class distribution, in the format of $[A, B]$.

A decision tree is built by selecting the best feature from the set of candidate features as the root of the decision tree. The same procedure is done on each branch to induce the decision tree until the growth stopping criteria is reached. In a decision tree, each internal node splits the instance

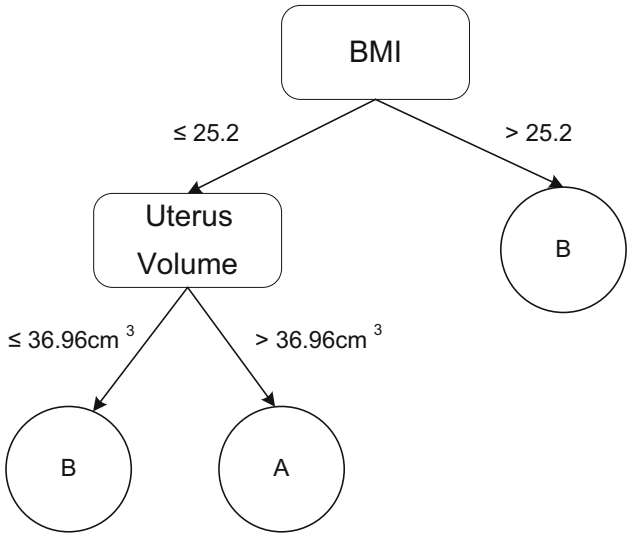

Fig. 3 Decision tree after CART has been applied to the dataset in Table 1

Table 1 Example dataset for endometriosis therapy

No.	Age	BMI	CA-125	Uterus length	Uterus volume	Cyst size	Treatment (target)
1	39	19.43	9.58	9.4	153.76	3.4	A
2	33	18.04	82.9	7.6	78.71	5.0	B
3	40	25.15	41.82	8.8	200.66	4.5	A
4	25	22.10	120.6	6.7	86.83	4.5	A
5	36	22.15	49.91	6.0	42.41	3.4	B
6	34	22.60	45.6	5.5	68.68	3.3	A
7	30	17.88	19.75	6.3	41.37	4.5	A
8	30	17.88	51.45	5.6	37.53	4.4	A
9	39	20.57	56.05	6.5	68.54	3.6	B
10	43	20.57	42	5.8	71.82	5	A
...							
208	41	19.95	145.86	7.23	149.53	3.4	B

space into two or more sub-spaces according to the values of the input attributes. Each path from the root of a decision tree to one of its leaves can be transformed into a rule. The most famous decision tree algorithms are ID3 (Quinlan 1986), C4.5/C5.0 (Quinlan 1993), CART (Breiman 1984), and CHAID (Kass 1980). Since the ID3 algorithm was only designed to handle categorical variables, continuous variables must be divided into discrete categorical values before the decision tree construction process. During the construction processes of the CART, C4.5/C5.0, and CHAID models, continuous variables are automatically divided into discrete categorical values.

The CART model constructs binary trees: each internal node has exactly two outgoing edges (Breiman 1984). A post-pruning process will sequentially collapse nodes that result in the smallest change in purity. Therefore, a significant difference between two branches of a node can be maintained. When dealing with continuous variables, a multi-branch decision tree might divide continuous variables into several ranges, which could make it difficult for domain experts (doctors) to interpret the decision tree. In this situation, the CART model is useful for producing binary cutoff points for continuous variables. The entire process can be automatically completed with mining tools. Thus, we can interpret the corresponding rules directly from the tree.

2.3 Interestingness measures

Discovering interesting patterns in data is an important objective of data mining (Padmanabhan and Tuzhilin 1999). By applying interestingness measures, experts can find interesting patterns (Piatetsky-Shapiro 1991). Baena-García and Morales-Bueno (2012) proposed a method using association rules with interestingness measures to detect interesting factors, rather than all possible factors. Most researchers divide interestingness measures into objective and subjective measures (Cao et al. 2007; Freitas 1999; Glass 2013; Kontonasio et al. 2012; Liu et al. 1999; McGarry 2005; Padmanabhan and Tuzhilin 1999; Shaharanee et al. 2011; Silberschatz and Tuzhilin 1995; Tsay and Raś 2005; Yao et al. 2006), although Geng and Hamilton (2006) categorize nine interestingness criteria as objective, subjective, or semantics-based. The objective interestingness measures depend only on raw data; they are data-driven and domain-independent (Tsay and Raś 2005; Yao et al. 2006). Most objective interestingness measures are based on statistical, probability, and information theory (Geng and Hamilton 2006). Subjective interestingness measures should consider both the data and users. To define subjective measures we acquire users' insights on data and their background knowledge. Consequently, the subjective measures are user-driven and domain-dependent (Geng and Hamilton 2006; Tsay and Raś 2005).

To date, based on diverse definitions, nine interestingness criteria have been proposed in previous studies (Geng and Hamilton 2006): conciseness, coverage, reliability, peculiarity, diversity, novelty, unexpectedness, utility, and actionability. From a subjective point of view, an interesting pattern is either unexpected or actionable. Therefore, as described by Silberschatz and Tuzhilin (1995), both unexpectedness and actionability are important for subjective interestingness measures. Unexpected patterns are interesting because they cannot be identified by previous knowledge and may suggest a particular status of the data that needs further study (Geng and Hamilton 2006). Furthermore, the patterns that contradict prior knowledge can be used to build theories about the domain (Padmanabhan and Tuzhilin 1999).

Based on domain-driven data mining, Cao and Zhang (2006) claimed that patterns extracted from a database must simultaneously satisfy technical and business interestingness. In other words, the patterns have to satisfy Formula 1.

$$\forall x \in I, \quad \exists P : x.tech(P) \cap X.biz(P) \rightarrow X.act(P) \qquad (1)$$

In Formula 1, I represents a set of items; x is an item-set in a database DB that consists of a set of transactions. A pattern P(composed of item-set x) is an interesting pattern discovered in DB through a modeling method. $Tech()$ is the technical interestingness measure that implies how interesting the pattern is from a technical viewpoint. It often utilizes specific technical metrics for data mining. $Biz()$ is the business interestingness measure, which indicates how interesting the pattern is from the users' point of view. It is determined by the domain-oriented criteria accepted by domain users. $Act()$ represents the actionability of a pattern. Thus, if a pattern $P()$ satisfies both $tech()$ and $biz()$, it is both interesting and actionable.

To be integrated with the subjective and objective concepts described above, interesting domain knowledge should satisfy $tech\_obj()$ (technical objective interestingness measures), $tech\_sub()$ (technical subjective interestingness measures), $biz\_obj()$ (business objective interestingness measures), and $biz\_sub()$ (business subjective interestingness measures) (Cao and Zhang 2007). Therefore, the output knowledge should satisfy Formula 2.

$$\begin{aligned}\forall x \in I, \quad & \exists P : X.tech\_obj(P) \cap x.tech\_sub(P) \\ & \cap x.biz.obj(P) \cap xbiz\_sub(P) \rightarrow x.act(P)\end{aligned} \qquad (2)$$

3 Proposed methods

3.1 Architecture of unexpected knowledge discovery

To detect whether the extracted patterns are both unexpected and actionable, we propose an unexpected knowledge discovery model based on interestingness measures, as shown in Fig. 4. During data preprocessing, domain experts input and select variables for decision tree induction. At the data decentralization stage, the CART algorithm is applied, and the cutoff points of the continuous variables will be generated during the decision tree construction, which can help to divide patients into different conditions. We use the Gini index as a split selection criterion for the CART algorithms; it selects the split results in the greatest increase in purity. After the decision tree is generated, we search for potentially interesting nodes in the decision tree. The corresponding rules of these nodes will be extracted and examined using interestingness measures.

We define interestingness measures based on the D^3M concept and use them to determine whether the node has domain interestingness and unexpectedness. We are looking for potentially interesting patterns that may pass the interestingness examination. If the patterns satisfy the technical objectives and subjective interestingness measures but not the business objective interestingness measure, then closed-loop and iterative refinement processes are initiated. We feed the initial results back into the data decentralization stage, readjust the input parameters, and conduct mining again to obtain in-depth patterns. Based on the results of the initial mining, we select those groups which require further analysis and induce a new decision tree. Because adding more constraints to data selection diminishes the sample pool, the input variables of the trees and the pruning level will be adjusted accordingly. This process is repeated until no more unexpected patterns are found.

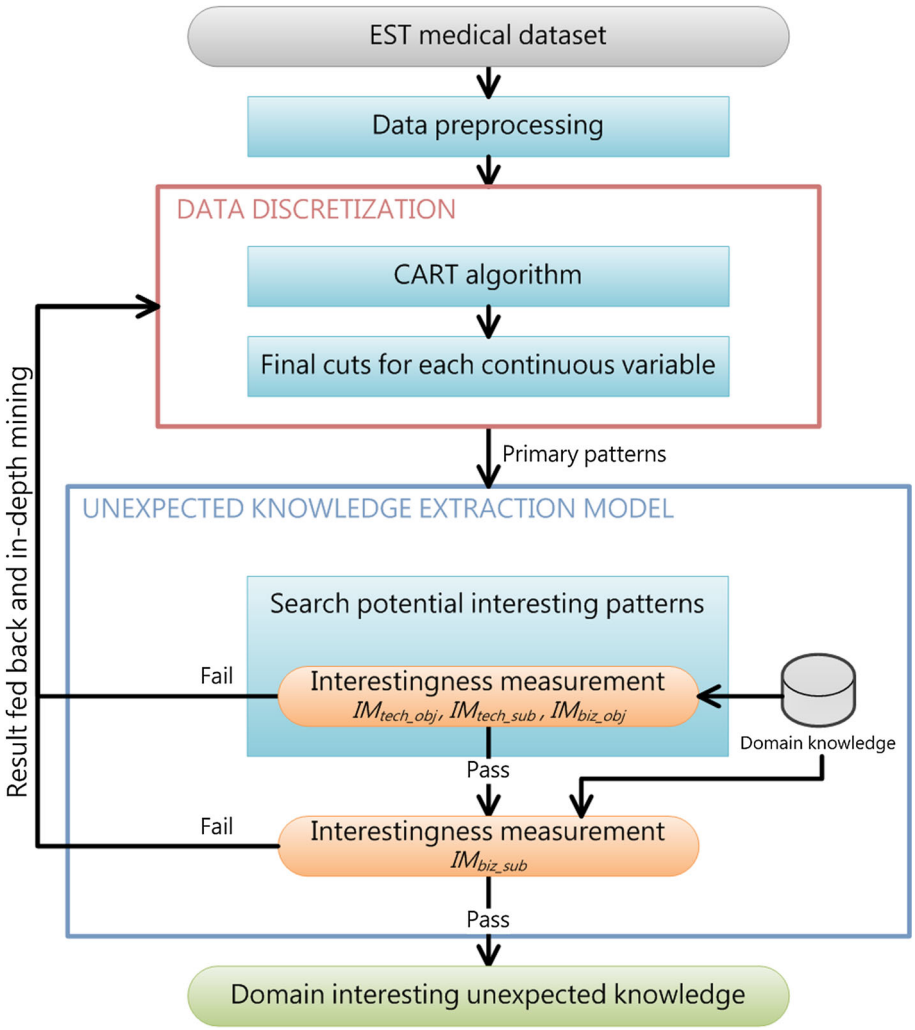

Fig. 4 Architecture of the unexpected knowledge discovery system

3.2 Unexpected nodes and rule interestingness examination

We assume that there are two different treatments: X and $\neg X$. For each node i in the decision tree, patients are categorized into four groups according to the success/failure of the treatments:

1. $X_i.success$: successful recovery with treatment X.
2. $X_i.failure$: recovery failure with treatment X.
3. $\neg X_i.success$: successful recovery with treatment $\neg X$.
4. $\neg X_i.failure$: recovery failure with treatment $\neg X$.

Subsequently, the recovery rates of treatments X and $\neg X$ are computed for each node using Formula 3. For each node i in the decision tree, the interestingness measures are formulated as Formulae 4, 5, 6, and 7.

$$RecoveryRate(Treatment_i) = \frac{|Treatment_i.success|}{|Treatment_i.success| + |Treatment_i.failure|}. \quad (3)$$

If, based on their prior knowledge, domain experts believe that the recovery rate of treatment X will be higher than that of $\neg X$, then $RecoveryRate(X_i) > RecoveryRate(\neg X_i)$ signifies that treatment X has a better curative effect than treatment $\neg X$ in node i; this is consistent with prior knowledge. In contrast, when $RecoveryRate(\neg X_i) > RecoveryRate(X_i)$, then treatment $\neg X_i$ has a better curative effect than treatment X in node i; this contrasts with prior knowledge. Therefore, this node may be a potentially unexpected node and need further investigation. We define our technical objective interestingness measure $IM_{tech\_obj}()$ as Formula 4.

$$IM_{tech\_obj}(i): RecoveryRate(\neg X_i) > RecoveryRate(X_i) \quad (4)$$

Formula 5 is our technical subjective interestingness measure, $IM_{tech\_sub}()$. We recognize that when there is only one $X(\neg X)$ treatment in the node of the decision tree, this node cannot be compared. Moreover, when there is merely one sample of X or $\neg X$ in the tree node, the recovery rate for X or $\neg X$ will either be 0 or 100 %. Due to reliance on the outcome of a single sample, the node is over fit. As a result, such exceptional conditions should not be included in our discussion. When a node satisfies IM_{tech_obj} and IM_{tech_sub}, then this node is a useful potentially unexpected node.

$$IM_{tech\_sub}(i): (|X_i.success| + |X_i.failure|) > 1 \quad \text{and} \quad (|\neg X_i.success| + |\neg X_i.failure|) > 1 \quad (5)$$

In Formula 6, $subnode_i$ is the node that immediately follows unexpected node i. $Subnode_{i-left}$ and $subnode_{i-right}$ represent the immediate left and right subnodes, respectively, of unexpected node i; p_i represents the p value of node i. In $subnode_{i-left}$ and $subnode_{i-right}$, when the recovery rate from treatment $\neg X$ is higher than that of treatment X, it is denoted as "$subnode_{i-left} \rightarrow \neg X$" or "$subnode_{i-right} \rightarrow \neg X$." When the immediate subnodes of the potentially unexpected node are both unexpected, $\neg X$ is the general case of these nodes. Thus, this potentially unexpected node can also be a terminal node.

Since domain experts are interested in what makes the recovery rate of one treatment higher than the other, we give more consideration to successful cases than to failures. Moreover, when the total amount of patients with treatment X is more than that of treatment $\neg X$, a low total number of patients with treatment $\neg X$ can distort $RecoveryRate(\neg X_i)$. In other words, even if the recovery rate of treatment $\neg X$ is better than that of treatment X, the patterns of the unexpected node may not be meaningful. Therefore, one still has to consider whether the number of recovering patients who received treatment $\neg X$ is also more than that of X. Our aim is to find patterns that contrast with domain experts' knowledge: $\neg X$. Therefore, the threshold in each node is that the sample number of $\neg X$ should greater than X.

Also, the rules should be validated to a certain extent by the existing body of knowledge (Sebastian and Then 2011), so that they are acceptable to domain experts. In medical research, we often use t tests to understand whether there is a significant difference among different groups. When $p_i < 0.05$, it implies that a remarkable variance exists among groups. Therefore, we use a t test to confirm whether there is significant difference between the recovery rate of treatment X and treatment $\neg X$ at each node.

$$IM_{biz\_obj}(i): (subnode_i = \emptyset \; or (subnode_{i-left} \rightarrow \neg X \quad \text{and} \quad subnode_{i-right} \rightarrow \neg X)) \quad \text{and} \quad |\neg X_i.success| \geq |X_i.success| \quad \text{and} \quad p_i < 0.05 \quad (6)$$

In Formula 7, *expertDefine* represents the threshold value determined by domain experts. The business subjective interestingness, $IM_{biz\_sub}(i)$, serves as the minimum threshold value in each node. According to the prevalence of a disease and the sample size, domain experts set a minimum amount of patients for each node so that a node can be meaningful. In other words, the amount of patients for an unexpected node must satisfy the threshold value, so it can pass IM_{biz_sub}.

$$IM_{biz\_sub}(i): (|X_i.success| + |X_i.failure| + |\neg X_i.success| + |\neg X_i.failure|) > expertDefine \quad (7)$$

In general cases, these four interestingness measures can be pre-defined. Then we can directly analyze data by applying the algorithm. However, in this study, we output $|X_i|$ and $|\neg X_i|$ without pre-defining the value of IM_{biz_sub} because doctors need to first understand the conditions of each node before setting IM_{biz_sub}.

3.3 Unexpected pattern detection algorithm

The pseudocode of the unexpected detection process is shown as Algorithm 1. In this algorithm, we use the Gini index as a split selec tion criterion for the CART algorithm. We will examine IM_{tech_obj} and IM_{tech_sub} first, then IM_{biz_obj}, and finally IM_{biz_sub}. When a rule only satisfies IM_{tech_obj} and IM_{tech_sub}, the rule will be reported as a potentially interesting pattern for conducting in-depth mining. Finally, when a rule satisfies IM_{tech_obj}, $IM_{tech\_sub,}$ and IM_{biz_obj}, the rule and the nodes X and $\neg X$ will then be output to the domain expert for a final determination. In this algorithm, we do not discuss the threshold of IM_{biz_sub}.

Based on their prior knowledge (from previous research or experience), experts believe the recovery rate of treatment X is better than treatment $\neg X$. So it would be considered unexpected if the recovery rate of treatment $\neg X$ is better than the recovery rate of treatment X. The class imbalance problem typically occurs when there are many more instances of some classes than others. In such cases, standard classifiers tend to be overwhelmed by large classes and small classes are ignored (Chawla et al. 2004). Because experts believe treatment X to be more effective based on their prior knowledge, the sample number for treatment X will usually be higher than that of treatment $\neg X$. In other words, there exists a proclivity to select treatment X.

When a decision tree algorithm is growing, it will keep splitting the data into branches until all the data in a single branch belongs to the same class (or until the stopping condition is reached). Therefore, an imbalance learning problem may result from using decision trees for classification. In this paper we use four target categories [*X.success*, *X.failure*, $\neg$*X.success*, $\neg$*X.failure*] to induce classification trees. Although the proposed method identifies classification rules that still emphasize the homogeneity of node data, in Algorithm 1, our method will further analyze the recovery rates of treatment X and $\neg X$ to identify any unexpected patterns by the following steps:

1. We used *X.success* and *X.failure* to calculate the recovery rate of treatment X at each node, and used $\neg$*X.success* and $\neg$*X.failure* to calculate the recovery rate of treatment $\neg X$. We then compared the recovery rates of each treatment X and $\neg X$ at each node. Even though the sample number for treatment X is generally higher than treatment $\neg X$ (i.e., there are more samples of *X.success* and *X.failure* than $\neg$*X.success* and $\neg$*X.failure*), the recovery rates will not be affected because they are expressed as ratios. When a node shows $RecoveryRate(X_i) > RecoveryRate(\neg X_i)$, this means that under these conditions, treatment X is more effective, therefore this is expected. On the other hand, when a node shows $RecoveryRate(\neg X_i) > RecoveryRate(X_i)$, this means that under the conditions of that node, treatment $\neg X$ is more effective. Therefore, this node may generate an unexpected pattern.
2. When a node contains the samples for only one treatment (treatment X or $\neg X$), then we cannot compare recovery rates of this node. Moreover, when there is only one sample of X or $\neg X$ in the tree node, the recovery rate for X or $\neg X$ will either be 0 or 100 %, therefore the node is over fit. As a result, such exceptional conditions were not included in our discussion.
3. If we find any potentially unexpected nodes after filtering in accordance with the two steps described above, we must still examine $IM_{biz\_obj,}$ and IM_{biz_sub} to determine whether this is an unexpected pattern.

Therefore, imbalanced learning problems do not affect our results.

Algorithm 1 Unexpected pattern detection

Input: A set of data partitions; an interested attribute set A; the CART attribute selection method; interestingness measure methods
Output: An unexpected rule of data partition

```
Create a decision tree T
for each node i ∈ T do
  Call interestingness measure IM_tech_obj, IM_tech_sub, IM_biz_obj to examination node i.
  if(IM_tech_obj = true && IM_tech_sub = true) then
    if(IM_biz_obj = true) then
      /* Return rule to physicians and surgeons for further examination IM_biz_sub.
      return rule_i, |X_i|, |¬X_i|
    else
      /* Report rule as useful, potentially interesting pattern for in-depth mining.
      return rule_i
    end if
  end if
end for
```

4 Experimental results and discussion

4.1 Materials

In this study, a retrospective review was done on 208 aspirations from 2001 through 2010 at Taipei Chang Gung Memorial Hospital. All of the records were collected from patient medical records (paper/electronic) and input by medical experts. The Institutional Review Board of Chang Gung Medical Foundation has approved this study. To preserve patient confidentiality, direct patient identifiers were not collected. Therefore, each registry contains therapy records and follow-up episodes, rather than the specific records of each patient. These endometriosis patients with postsurgical recurrence of pelvic cysts received transvaginal ultrasound aspirations with 95 % ethanol sclerotherapy at the outpatient gynecological department. Our patients randomly received ethanol instillations of short duration: 0 min (ethanol was injected, irrigated, and then removed) to 6 min; or long duration: 7–10 min (including total retention).

Further repeated surgical interventions, including repeated sclerotherapy and abdominal operations, were also recorded. The immediate preoperative data of patients undergoing repeat operations were characterized as the endpoint data for the patient. They were followed up with vaginal ultrasounds, CA-125 determinations, and pain score records every 3–6 months for at least one year. Twelve-month recovery was defined as: (a) pregnancy achieved; treatment successful, (b) no repeat surgery; treatment successful; and (c) no cyst development; treatment successful. If any cysts developed with a diameter of 3.0 cm or larger the treatment is considered a failure. In this research, domain experts recognize that the recovery rate for ethanol instillation time over 7 min (X) is higher than that for ethanol instillation time under 7 min ($\neg X$). Accordingly, the patients were categorized into four groups as the target categories of the decision tree:

1. $X_i.success$: successful recovery with ethanol instillation time over 7 min.
2. $X_i.failure$: recovery failure with ethanol instillation time over 7 min.
3. $\neg X_i.success$: successful recovery with ethanol instillation time under 7 min.
4. $\neg X_i.failure$: recovery failure with ethanol instillation time under 7 min.

Using preliminary statistics for the 12-month follow-up period, it was found that the recovery rate of Group $\neg X$ ($N = 64$) and Group X ($N = 144$) were 31.25 and 44.45 % ($p > 0.05$), respectively, which shows that the retention time of each group has no significant effect on curative outcomes. To further identify the factors that influence the curative outcomes, we adopted a decision tree to generate the cutoff points in the numeric data for grouping. The dataset includes preoperative patient characteristics, operation details, and pathological and laboratory findings. The original records contained more than 80 preoperative and postoperative examination follow-up fields; however, most were null values. Therefore, after discussion with domain experts, the variables that physicians consider influential on curative effects were selected to build the decision tree (for more details, please refer to Table 2). In this experiment, we used SPSS Clementine 12 to generate the decision tree.

Table 2 Selected variables

Variable	Definition
CA-125	Preoperative serum CA-125 level
Uterus length	Length of preoperative uterus, mm
Uterus volume	Volume of preoperative uterus, mm^3
Cyst size	Total size of preoperative cysts, cm
Cyst number	Total amount of preoperative cysts
Cyst content	Majority type of cysts, clear/bloody
Patient group (Target)	Successful recovery with ethanol instillation time over 7 min
	Recovery failure with ethanol instillation time over 7 min
	Successful recovery with ethanol instillation time under 7 min
	Recovery failure with ethanol instillation time under 7 min

4.2 Experiment discussion

Since decision trees use binary targets to compare the difference between sibling nodes, researchers generally use [*treatment X, treatment* $\neg X$] or [*success, failure*] as targets. In this way we can retrieve statements such as: when a patient's "$BMI \leq 25.2$" and "*uterus volume* $> 36.96\,cm^3$", they received treatment X. To further analyze under what conditions patients can be treated successfully we have to select the data of patients with one treatment only, e.g. treatment X, and use [*success, failure*] as targets. Then, we can describe patient conditions in statements such as: when a patient's "*cyst size* ≤ 5.05 cm" and "*CA-125* ≤ 115.65," treatment X is successful. Since each different treatment has its own tree, the rules of the conditions may not be the same. Therefore, we cannot directly use a single decision tree to compare different treatments, and retrieve statements such as: when a patients "$BMI \leq 25.2$" and "$age < 40$," treatment X is better than treatment $\neg X$. In this situation, if we want to analyze the recovery rate of different treatments under the same condition we need to review each node, and perform a manual calculation.

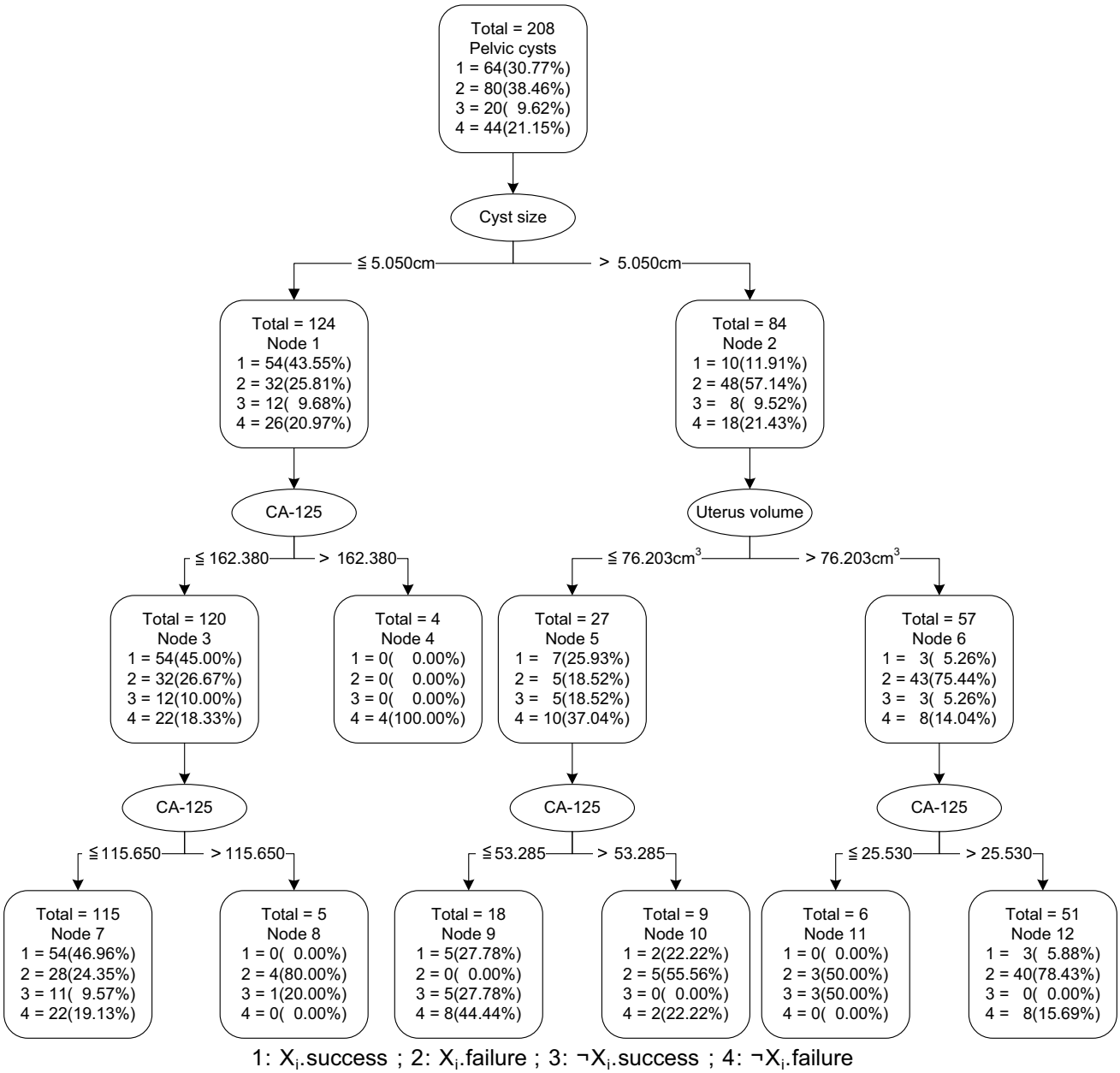

Fig. 5 Resulting decision tree constructed from data from the endometriosis dataset

By using the proposed method, we use four targets to build a decision tree, as shown in Fig. 5. With Formula 3, the resulting tree can directly compare the recovery rate of each treatment. For example, in node 1, $RecoveryRate(X) = 62.79\,\%$, $RecoveryRate(\neg X) = 31.58\,\%$. By using interestingness measures, we can directly find that nodes 2, 6, 8, and 11 may be unexpected. The results of the interestingness examination of these nodes are shown in Table 3.

As shown in Table 3, since node 8 represents only one sample in ¬X, it is an exceptional condition and does not satisfy IM_{tech_sub} thus it should not be included in our discussion. Moreover, in the left subtree of the root, all other nodes depict treatment X as having a better curative effect than treatment $\neg X$; this is consistent with prior knowledge. Therefore, the left subtree of the root does not need to have in-depth pattern mining carried out on it. On the other hand, in the right subtree of the root, node 11 satisfies IM_{tech_obj}, IM_{tech_sub} and IM_{biz_obj}, so the rule was output to domain experts. They considered the total patient number of the node was too low, and requested in-depth mining. Since nodes 2 and 6 satisfy IM_{tech_obj} and IM_{tech_sub}, but not IM_{biz_obj} therefore, the in-depth pattern mining strategy was used to further analyze the data. According to the closed-loop method, we utilized the feedback results to adjust the parameters.

Table 3 Interestingness check of potentially unexpected rules for Fig. 5

Node	Rule	IM_{tech_obj}	IM_{tech_sub}	IM_{biz_obj}	IM_{biz_sub}
2	*Cyst content* $=$ *bloody* with *cyst size* $>$ 5.05 cm $\rightarrow \neg X$	Pass	Pass	Fail $subnode_{i-left} \rightarrow \neg X, subnode_{i-right} \rightarrow \neg X$ $\lvert\neg X_i.success\rvert < \lvert X_i.success\rvert$ $p_i > 0.05$	
6	*Cyst content* $=$ *bloody* with *cyst size* $>$ 5.05 cm *and uterus volume* $>$ 76.203 $cm^3 \rightarrow \neg X$	Pass	Pass	Fail $subnode_{i-left} \rightarrow \neg X, subnode_{i-right} \rightarrow \neg X$ $\lvert\neg X_i.success\rvert = \lvert X_i.success\rvert$ $p_i < 0.05$	
8	*Cyst content* $=$ *bloody* with *cyst size* $\leq$ 5.05 cm, *CA*-125 $\leq$ 162.38 *and CA*-125 $>$ 115.65$\rightarrow \neg X$	Pass	Fail		
11	*Cyst content* $=$ *bloody* with *cyst size* $>$ 5.05 cm, *uterus volume* $>$ 76.203 cm^3,*and CA*-125 $\leq$ 25.53 $\rightarrow \neg X$	Pass	Pass	Pass $\text{subnode}_i = \emptyset$ $\lvert\neg X_i.success\rvert > \lvert X_i.success\rvert$ $p_i < 0.05$	$\lvert X_i\rvert = 3$ $\lvert\neg X_i\rvert = 3$

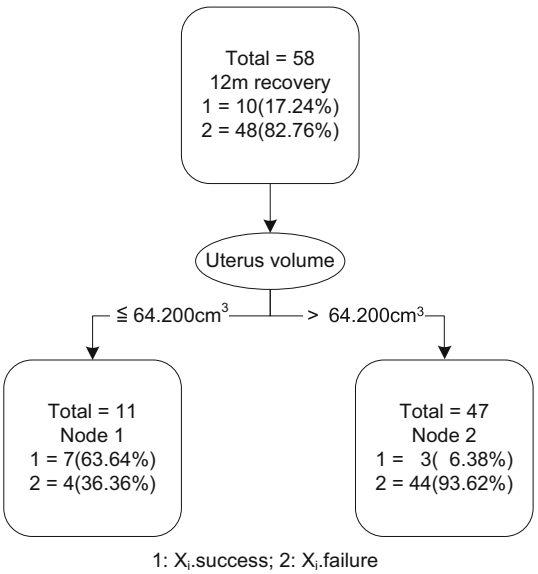

Fig. 6 Resulting tree of reselected data using the Node 2 conditions from Fig. 5

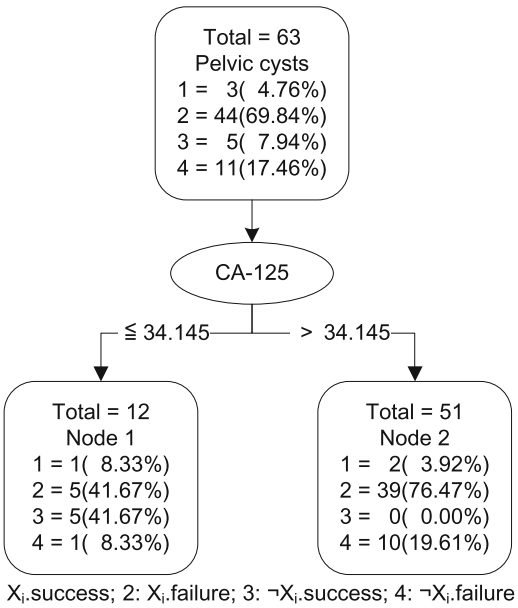

Fig. 7 Resulting tree of reselected data using the Node 2 conditions from Fig. 6

The decision tree makes locally optimal decisions for each node. Therefore, in order to prevent the final result of the decision tree from being affected by group $\neg X$, we only consider group X in this step. Thus, we reselect the data according to the conditions of node 2 in Fig. 5—*cyst content* $=$ *bloody, duration* $\geq$ 7 min, *and cyst size* $>$ 5.05 cm—to grow a new tree. The resulting tree is shown in Fig. 6. Without considering the effect of group $\neg X$, the decision tree algorithms select 64.2 cm^3 as the cutoff point of "*uterus volume*" for grouping, instead of 76.203 cm^3, as shown in Fig. 5. In this situation, $RecoveryRate(X)$ for node 1 in Fig. 6 (63.64 %, $N = 11$) is significantly different from that for node 2 in Fig. 6 (6.38 %, $N = 47$) ($p < 0.05$). Therefore, we recognize 64.2 cm^3 as an effective cutoff point. In addition, since node 5 in Fig. 5 is not a terminal node, we reselect the data set by the conditions of nodes 1 and 2 in Fig. 6, respectively, for further processing. Therefore, we return to the data decentralization stage to perform further in-depth mining with the decision tree, in order to further confirm whether any unexpected node exists.

Now, we consider both group X and $\neg X$, and reselect the dataset using the conditions of node 1 in Fig. 6 to reproduce the decision tree; the resulting tree has no unexpected nodes. On the other hand, when using the conditions of node 2 in Fig. 6, the resulting tree, as shown in Fig. 7, does have an unexpected node. As shown in Table 4, the corresponding rule of node 1 in Fig. 7 passes the interestingness examinations of IM_{tech_obj}, IM_{tech_sub}, and IM_{biz_obj}. Therefore,

Table 4 Interestingness check of potentially unexpected rules for Fig. 7

Node	Rule	IM_{tech_obj}	IM_{tech_sub}	IM_{biz_obj}	IM_{biz_sub}
2	*Cyst content* $=$ *bloody* with *cyst size* $>$ 5.05 cm, *uterus volume* $>$ 64.20 cm^3, and *CA*-125 $\leq$ 34.145 $\rightarrow \neg X$	Pass	Pass	Pass $\text{subnode}_i = \emptyset$ $\lvert\neg X_i.success\rvert > \lvert X_i.success\rvert$ $p_i < 0.05$	$\lvert X_i\rvert = 6$ $\lvert\neg X_i\rvert = 6$

the corresponding rule of node 1 is a potentially unexpected and interesting pattern. In the final examination of IM_{biz_sub}, experts approved the pattern that we found. In other word, "*cyst content* $=$ *bloody, cyst size* $>$ 5.05 cm, *uterus volume* $>$ 64.20 cm^3, *and CA*-125 $\leq$ 34.145 $\rightarrow \neg X$"—is the unexpected pattern we were looking for. The experts agreed that the pattern is interesting and unexpected. However, their attitude to this finding was reserved and they require more samples to support the threshold of business subjective interestingness. Therefore, further research needs to be conducted to define the appropriate threshold.

5 Conclusions

In this research, we used retrospective data on transvaginal ultrasound-guided aspirations to show that our proposed model can compare different treatments and retrieve unexpected patterns through use of a decision tree. Since decision trees compare differences between sibling nodes, they usually adopt binary targets to induce classification trees. In this paper, we adopted four target categories to induce classification trees. Our interestingness measures were designed to compare treatment at individual nodes, and retrieve unexpected patterns. Thus, we were able to conduct a comparison at each individual node automatically.

Using a pure decision tree without interestingness measures will produce the graphic in Fig. 5. In this situation, users need to calculate each individual node and search the whole tree manually to find unexpected patterns. However, in this research, we can directly compare different treatments of a group with the same medical condition through our algorithm. Drawing on domain knowledge to formulate interestingness measures, we can automatically retrieve patterns that domain experts may be interested in.

References

Baena-García M, Morales-Bueno R (2012) Mining interestingness measures for string pattern mining. Knowl-Based Syst 25:45–50. doi:10.1016/j.knosys.2011.01.013

Bay SD, Pazzani MJ (2001) Detecting group differences: mining contrast sets. Data Min Knowl Disc 5:213–246. doi:10.1023/a:1011429418057

Berlanda N, Vercellini P, Fedele L (2010) The outcomes of repeat surgery for recurrent symptomatic endometriosis. Curr Opin Obstet Gynecol 22:320–325

Bolton S, Bon C (2009a) Analysis of variance. Pharmaceutical statistics: practical and clinical applications, 5th edn. Informa Healthcare, New York, pp 182–221

Bolton S, Bon C (2009b) Linear regression and correlation. Pharmaceutical statistics: practical and clinical applications, 5th edn. Informa Healthcare, New York, pp 147–181

Breiman L (1984) Classification and regression trees. In: The Wadsworth statistics/probability series. Wadsworth International Group, Belmont

Bulletti C, Coccia M, Battistoni S, Borini A (2010) Endometriosis and infertility. J Assist Reprod Genet 27:441–447

Cao L, Zhang C (2007) Domain-driven, actionable knowledge discovery. IEEE Intell Syst 22:78–88

Cao L, Luo D, Zhang C (2007) Knowledge actionability: satisfying technical and business interestingness. Int J Bus Intell Data Min 2:496–514. doi:10.1504/ijbidm.2007.016385

Cao L, Zhang C, Yu PS, Zhao Y (2010a) Challenges and trends. Domain driven data mining. Springer, US, pp 1–25

Cao L, Zhang C, Yu PS, Zhao Y (2010b) D^3M methodology. Domain driven data mining. Springer, US, pp 27–47

Cao L, Zhang C (2006) Domain-driven actionable knowledge discovery in the real world. In: Ng W-K, Kitsuregawa M, Li J, Chang K (eds) Advances in knowledge discovery and data mining, Lecture notes in computer science, vol 3918. Springer, Berlin, pp 821–830. doi:10.1007/11731139_96

Chawla NV, Japkowicz N, Kotcz A (2004) Editorial: special issue on learning from imbalanced data sets. SIGKDD Explor Newsl 6:1–6. doi:10.1145/1007730.1007733

Donnez J, Squiffle J, Donnez O (2011) Minimally invasive gynecologic procedures. Curr Opin Obstet Gynecol 23:289–295. doi:10.1097/GCO.0b013e328348a283

Freitas AA (1999) On rule interestingness measures. Knowl-Based Syst 12:309–315

Geng L, Hamilton HJ (2006) Interestingness measures for data mining: a survey. ACM Comput Surv 38:1–31. doi:10.1145/1132960.1132963

Glass DH (2013) Confirmation measures of association rule interestingness. Knowl-Based Syst 44:65–77. doi:10.1016/j.knosys.2013.01.021

Hsieh C-L, Shiau C-S, Lo L-M, Hsieh Ts-Ta, Chang M-Y (2009) Effectiveness of ultrasound-guided aspiration and sclerotherapy with 95 % ethanol for treatment of recurrent ovarian endometriomas. Fertil Steril 91:2709–2713

Ikuta A et al (2006) Management of transvaginal ultrasound-guided absolute ethanol sclerotherapy for ovarian endometriotic cysts. J Med Ultrason 33:99–103

Kafali H, Yurtseven S, Atmaca F, Ozardali I (2003) Management of non-neoplastic ovarian cysts with sclerotherapy. Int J Gynaecol Obstet 81:41–45

Kass GV (1980) An exploratory technique for investigating large quantities of categorical data. Appl Stat 29(2):119–127

Kennedy S et al (2005) ESHRE guideline for the diagnosis and treatment of endometriosis. Hum Reprod 20:2698–2704

Kontonasios K-N, Spyropoulou E, De Bie T (2012) Knowledge discovery interestingness measures based on unexpectedness. Wiley Interdiscip Rev Data Min Knowl Discov 2:386–399

Lenca P, Meyer P, Vaillant B, Lallich S (2008) On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. Eur J Oper Res 184:610–626. doi:10.1016/j.ejor.2006.10.059

Ling CX, Tielin C, Qiang Y, Jie C (2002) Mining optimal actions for profitable CRM. In: Paper presented at the proceedings of the 2002 IEEE international conference on data mining, 2002

Liu B, Hsu W, Mun L-F, Lee H-Y (1999) Finding interesting patterns using user expectations. IEEE Trans Knowl Data Eng 11:817–832. doi:10.1109/69.824588

McGarry K (2005) A survey of interestingness measures for knowledge discovery. Knowl Eng Rev 20:39–61. doi:10.1017/s0269888905000408

Nap AW, Groothuis PG, Demir AY, Evers JLH, Dunselman GAJ (2004) Pathogenesis of endometriosis. Best Pract Res Clin Obstet Gynaecol 18:233–244

Noma J, Yoshida N (2001) Efficacy of ethanol sclerotherapy for ovarian endometriomas. Int J Gynaecol Obstet 72:35–39

Padmanabhan B, Tuzhilin A (1999) Unexpectedness as a measure of interestingness in knowledge discovery. Decis Support Syst 27:303–318

Piatetsky-Shapiro G (1991) Discovery, analysis, and presentation of strong rules. In: Piatetsky-Shapiro G, Frawley W (eds) Knowledge discovery in databases. AAAI/MIT Press, Cambridge, pp 229–248

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc, San Francisco

Quinlan JR (1986) Induction of decision trees. Mach Learn 1:81–106. doi:10.1007/bf00116251

Rokach L, Maimon O (2008) Data mining with decision trees: theory and applications. World Scientific Publishing Company, MA

Sebastian Y, Then PHH (2011) Domain-driven KDD for mining functionally novel rules and linking disjoint medical hypotheses. Knowl-Based Syst 24:609–620

Shaharanee INM, Hadzic F, Dillon TS (2011) Interestingness measures for association rules based on statistical validity. Knowl-Based Syst 24:386–392. doi:10.1016/j.knosys.2010.11.005

Silberschatz A, Tuzhilin A (1995) On subjective measures of interestingness in knowledge discovery. In: Paper presented at the proceedings of the 1st international conference on knowledge discovery and data mining (KDD' 95)

Tsay L-S, Raś ZW (2005) Action rules discovery: system DEAR2, method and experiments. J Exp Theory Artif Intell 17:119–128

Vercellini P, Somigliana E, ViganÒ P, De Matteis S, Barbara G, Fedele L (2009) The effect of second-line surgery on reproductive performance of women with recurrent endometriosis: a systematic review. Acta Obstet Gynecol Scand 88:1074–1082. doi:10.1080/00016340903214973

Wang YF, Chang MY, Chiang RD, Hwang LJ, Lee CM, Wang YH (2013) Mining medical data: a case study of endometriosis. J Med Syst 37:1–7. doi:10.1007/s10916-012-9899-y

Wang K, Zhou S, Han J (2002) Profit mining: from patterns to actions. In: Paper presented at the proceedings of the 8th international conference on extending database technology: advances in database technology

Yao Y, Chen Y, Yang X (2006) A measurement-theoretic foundation of rule interestingness evaluation. In: Young Lin T, Ohsuga S, Liau C-J, Hu X (eds) Foundations and novel approaches in data mining, Studies in computational intelligence, vol 9. Springer, Berlin, pp 41–59. doi:10.1007/11539827_3

Zhu Z, Gu J, Zhang L, Song W, Gao R (2009) Research on domain-driven actionable knowledge discovery. In: Shi Y, Wang S, Peng Y, Li J, Zeng Y (eds) Cutting-edge research topics on multiple criteria decision making, Communications in computer and information science, vol 35. Springer, Berlin, pp 176–183. doi:10.1007/978-3-642-02298-2_27

Zhu W, Tan Z, Fu Z, Li X, Chen X, Zhou Y (2011) Repeat transvaginal ultrasound-guided aspiration of ovarian endometrioma in infertile women with endometriosis. Am J Obstet Gynecol 204:61.e61–61.e66