we are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists



122,000

135M



Our authors are among the

TOP 1%





WEB OF SCIENCE

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected. For more information visit www.intechopen.com



Interactive Knowledge Discovery for Temporal Lobe Epilepsy

Mostafa Ghannad-Rezaie¹ and Hamid Soltanian-Zadeh² ¹University of Michigan, Ann Arbor, MI ²Henry Ford Hospital, Detroit, MI USA

1. Introduction

Medical data mining and knowledge discovery can benefit from the experience and knowledge of clinicians, however, the implementation of this data mining system is challenging. Unlike traditional data mining methods, in this class of applications we process data with some posterior knowledge and the target function is more complex and even may include the opinion of user. Despite the success of the classical reasoning algorithms in many common data mining applications, they failed to address medical record processing where we need to extract information from incomplete, small samples along with an external rulebase to generate 'meaningful' interpretation of biological phenomenon. Swarm intelligence is an alternative class of flexible approaches that is promising in data mining. With full control over the rule extraction target function, particle swarm optimization (PSO) is a suitable approach for data mining subject to a rulebase which defines the quality of rules and constancy with previous observations. In this chapter we describe a complex clinical problem that has been addressed using PSO data mining. A large group of temporal lobe epilepsy patients are studied to find the best surgery candidates. Since there are many parameters involved in the decision process, the problem is not tractable from traditional data mining point of view, while the new approach that uses the field knowledge could extract valuable information.

The proposed method allows expert to interaction with data mining process by offering manual manipulation of generated rules. The algorithm adjusts the rule set with regard to manipulations. Each rule has a reasoning which is based on the provided rulebase and similar observed cases.

Support vector machine (SVM) classifier and swarm data miner are integrated to handle joint processing of raw data and rules. This approach is used to establish the limits of observations and build decision boundaries based on these critical observations.

2. Background

2.1 Data mining in medicine

The overall process of knowledge discovery from database (KDD) is a multistage process. The main step in KDD, Data Mining (DM), is the most commonly used name to describe the computational efforts meant to process feature space information, in order to obtain

Source: Data Mining in Medical and Biological Research, Book edited by: Eugenia G. Giannopoulou, ISBN 978-953-7619-30-5, pp. 320, December 2008, I-Tech, Vienna, Austria

valuable high level knowledge, which must conform to three main requisites: accuracy, comprehensibility and interest for the user (Apte et la. 1997). DM discovers patterns among database-stored information to support special user interest such as data classification. Computer-aided diagnosis (CAD) systems are one the primary areas of interest in data mining (Lavarac 1999, Toivonen et la 2000).

Designers of computer-based diagnosis systems often view the physician's primary decision-making task as a differential diagnosis. This term refers to a type of analytical task wherein the decision maker is confronted with a fixed set of diagnostic alternatives.

Over the past two decades, a large number of specialized procedures have been developed to assist physician in differential diagnosis of a variety of well defined clinical problems. These have been extensively reported in the medical and computing literature. In addition, algorithms to deal with a host of common medical problems, expressed by means of detailed flowcharts, have increasingly found their way into the clinical application.

Many different techniques have been used in structuring these clinical algorithms. In some cases, special programs have been formulated to capture the logic involved in the workup of particular classes of clinical problems. In other cases, generalized procedures have been adopted that are tailored to a particular application by specification of certain parameters; for example, many diagnostic programs have been developed to use the normative models of statistical decision theory. In some complex diagnosis tasks too many parameters such as cancer staging and neurological disease, medical diagnosis systems evolve rapidly. Evaluation studies frequently show that these programs, whatever their basis, generally perform as well as experienced clinicians in their respective domains, and somewhat better than the non-specialist. It is interesting, therefore, to speculate on the reason that such programs have not had greater impact on the practice of medicine.

Resistance in the medical community is sometimes attributed to the natural conservatism of physicians or to their sense of being threatened by the prospect of replacement by machines. Some have argued that this can be resolved only on the basis of education and training, and that the next generation will be more comfortable with computer-based decision aids as these become routinely introduced into the medical community. Some clinicians argue rather forcefully that the real reason that they have not adopted computer-based decision aids is that these systems have often been based on unrealistic models, which fail to deal with the physicians' real problems.

Unlike most of data mining approaches, we propose to optimize the rule-discovery process by giving clinician flexibility of incorporating domain knowledge, in the form of desire rule formats, into the rule search.

There are many reasons why a physician might experience difficulty in formulating an appropriate differential diagnosis. It may be that the case involves a rare disease or unusual presentation. Often, such difficulties arise in clinical problems where two or more disease processes are at work, generating a complex sequence of abnormal findings that can be interpreted in a variety of ways. Supporting the results with human understandable evidences, rule based CAD systems may be able to eventually convince clinician to accept their proposed results.

2.2 Interactive data mining

Mining medical information is to discover useful and interesting information from raw patient's clinical and non-clinical data (Apte et la. 1997, Kim 1997, Chen 2007). Nowadays, huge amount of features are collected, thanks to recent progresses in medical information

systems and novel medical instrument (Lavarac 1999). This wide range of knowledge is confusing. Traditional manual analysis methods can not discover complex relationships and knowledge potentially may embed inside raw data. Also, the more complexity is added to expertise's diagnosis process, the more time is required to make a reliable judgment. Therefore, automated data mining applications in medical domain has become a very active field during recent years.

In this section, we discuss on knowledge discovery from raw data in medical diagnosis systems. The primary application of our interest is surgery candidate selection in temporal lobe epilepsy. In some aspects, this problem is a prototype of many complex medical diagnosis problems. Low number of samples, large number of features, and missing data are common problems we are facing. Thus, the method can be extended to similar medical diagnosis problems. Medical knowledge extraction process is divided into five steps: data collection, data pre-processing, modelling, rule extraction and evaluation. Some primary processing is required to extract feature vector presentation of the data. Data mining block is the main part of the system that generates rules. Finally, comprehensibility and completeness of the generated rules are evaluated. The main contribution of most of knowledge discovery works is to find appropriate data classification and rule extraction algorithms. The result is a more intelligent and more robust system providing a human-interpretable, low cost, approximate solution, compared to the conventional techniques. This article focuses on advantages of association support vector machine (SVM) and partial swarm optimization (PSO) approaches in medical data mining (Botee et la1998).

In Medical information mining, comprehensibility has a special value. In a nutshell, data mining comprehends the actions of automatically seeking out, identifying, validating and is used for prediction, structural patterns in data that might be grouped into five categories: decision trees, classification rules, association rules, clusters and numeric prediction.

This article proposes a data discovery algorithm for a small data set with high dimensionality. Vector fusion algorithm is used to construct a single feature vector of noncommensurate data sources. Support vector machine is applied to classify the feature vectors. Finally, particle swarm agents are used to discover the SVM classification rules (Barakat et la 2005). It has been shown that this algorithm can manage the rule extraction task efficiently. Fig 1 shows the general structure of proposed algorithm. Raw data and external rules both can contribute in the final rule generation.

2.3 Terminology

A rule-base is a finite set of rules. There is no agreed on definition of rules for medical information, however, to have a consistence terminology we use a simple rule language based on rules used by the well known Jena framework. The rules are in the form of:

$(?A,b,c) \leftarrow (?A,x,y) (?A,v,w) \dots (?B,p,q)$

The left hand side of (\leftarrow) is the goal of the rule and the right hand side is the body of the rule. In this language, each triples represents a relationship, for example (?A,b,c) represent relationship b between variable A and value c.

Each rule has an active region in the feature space where the rule is valid. This area is defined by another rule, called primitive rule.

In data mining subject to a rulebase a set of rules derive from database where derived rules should be consistent with the given rulebase. Each rule and its primitive in the rulebase have a degree of preciseness is a positive value. An established fact has a large degree of preciseness, while a weak rule may have a degree of preciseness around 0. The cost function of a data mining algorithm increases when invalidating a rule proportional to the preciseness of the rule.

Prooftree represent the interaction between rules and data that lead to the result. The prooftree is a tool to indentify the reason for unexpected results.

3. Method

In recent years, various soft computing methodologies have been applied to address data mining (Lan et la. 2002; et la. Susa 2004, Xu et la 2007). Generally, there is no universal best data mining algorithm. Choosing appropriate data mining algorithm utterly depends on individual applications. Social algorithms and Swarm intelligence (SI) algorithms are well-known alternatives of soft computing tools that can be used for retrieving information from raw data (Galea 2002). Distributed Genetic Algorithm, Ant Colony Optimizer (ACO) and Particle Swarm Optimizer (PSO) are the most commonly used evolutionary algorithms in this domain. The most interesting contribution of these methods is in flexible rule extraction where we are facing to Incomplete and Inaccurate Measured dataset (Galea 2002).

3.1 Rule extraction

According to previous discussions, finding a rule-based classifier and reasons behind the decision making process are essential parts of medical computer aided diagnosis systems. Also they are key parts of knowledge discovery from databases (KDD). This section discusses the mathematical modeling of rule extraction process and application of particle swarm optimization to find the rule set describing a support vector machine classifier.

Assume S is the search space and θ is a data point inside S and yi = V(θ i) is the data point class. A classifier is define by yi¹ = U (θ _i). The target of rule extraction is to find a rule set R^U_{1..n} that describes the U classifier.

Each rule is a "IF-THEN" statement with two clauses. In the simplest case, the former clause is a condition on the search space and the latter clause is the target class. This rule may express as $(?A,x,c) \leftarrow (?A,x,y)$.

The structure of these two clauses is called rule set grammar limiting phrases in the rule clauses. The simpler the grammar, the more comprehensive statements can be retrieved. Rectangular grammar limits the IF-clause to intervals of each individual feature. Fig.2 shows an example of the rectangular rules. Decision tree is another alternative of rule set topology which is quite common in medical applications because of better interpretability and higher searching speed. In this structure the IF-part may also include another rule in addition to intervals but number of intervals is limited.

3.2 Rule-set evaluation

The value of a rule is evaluated using different parameters:

- *Accuracy*: this term stands for the percentage of data points correctly classified by the rule.

$$A_{R} = \frac{\#\theta : V(\theta) = R(\theta)}{\#\theta} = \frac{TP + TN}{TP + TN + FP + FN}$$
(1)

- *Quality*: this term is the popular classification quality measure for the rule set and can be obtained from the ROC curve:

$$Q_R = \text{sensitivity} \cdot \text{specificity} = \frac{TP \cdot TN}{(TP + FN) \cdot (TP + FN)}$$
 (2)

- *Coverage*: this term describes the percentage of specific class data points that are covered by the rule set from total available data points. This percentage is represented by C_R.
- *Simplicity*: the number of terms in rule the clauses and the number of intervals on each term condition representation (S_R). Generally, a very complex rule can describe any classifier and achieve very high Coverage and Accuracy spontaneously. Comprehensibility as a critical parameter in medical data mining is measured by this term. The number of rules in a rule set, and the number of terms in a rule represent the complexity phenomena. Simplicity is defined as *1/Complexity*.
- *Rule interference*: Adding to the individual rule evaluating parameters, the final rule set is admirable when it can cover the entire search space while the conflict among rules is kept as low as possible. Interference parameter (I_R) is particularly considered for reasoning process and reliable decision making.

Finally, the rule set evaluation is $\text{Eval}(R) = \alpha A_R + \beta C_R + \gamma S_R - \delta I_R$. The accuracy measure can be replaced by the quality when the trade-off between sensitivity and specificity is highly interested. Finding the best rule set is a complex multi-objective process.

3.3 Swarm intelligence rule extraction

Previous work in the literature shows power of PSO in solving rule extraction problems in medical diagnosis systems. Many recommended modifications of PSO for providing a flexible approach to address common difficulties of medical information retrieval from databases (Jaganathan 2007). Here, we present hybrid approaches to overcome problems presented in earlier sections.

Ant Colony Optimizer (ACO) and ant miners is the first swarm intelligent data mining algorithm presented by Parpinelli (Parpinelli et al. 2002, Omkara 2008). The purpose of their algorithm, Ant-Miner is to use ants to create rules describing an underlying data set. The overall approach of Ant-Miner is a separate-and-conquer one as same as C4.5 (Quinlan 1993). It starts with a full training set, creates a best rule that covers a subset of the training data, adds the best rule to its discovered rule list, removes the instances covered by the rule from the training data, and starts again with a reduced training set. This goes on until only a few instances are left in the training data (fewer than max number allowed) or the fitness function meets the target, at which point a default rule is created to cover remaining instances. Once an ant has stopped building a rule antecedent a rule consequent is chosen. This is done by assigning to the rule consequent the class label of the majority class among the instances covered by the built rule antecedent. The rule is then pruned in order to improve its quality and comprehensibility. The basic idea is to iteratively remove one term at a time from the rule while this process improves the rule quality as defined by a fitness function. In the iteration, each term in turn is temporarily removed from the rule antecedent, a new rule consequent is assigned, and the rule quality is evaluated. At the end of the iteration, the term that has actually been removed is the one that improves the rule quality the most.



Fig. 1. General schematic of proposed rule extraction system.

4. Particle swarm intelligence

Rule discovery process can be done using a Particle Swarm Intelligence Algorithm. PSO imitates the intelligent behaviour of beings as part of a group to experience sharing in a society. In contrast to conventional learning algorithms with individual reaction to environment or searching space, PSO is based on adaptive social behaviours. The basic idea of the PSO model is constructed on three ideas: evaluation, comparison and imitation. Evaluation is the kernel part of any intelligent algorithm measuring quality of the result in the environment and usefulness inside the community. Some metrics are defined to represent the particle superiority. This evaluation is pointless without well-defined comparison process which is a sequential relationship in the particle space. The improvement of particles is made by imitating best solution up to now. Looking to best solution in the neighbourhood, a particle decides where to move in the next step. There are many alternatives for implementation of neighbourhood and distance between particle concepts.

4.1 PSO Algorithm

PSO is a set of individual agents simply searching for an optimal point in their neighbourhood. The movement of agents depends on behaviour of other agents in the vicinity and the best visited nodes. During PSO training particle's best met position (BPi) and the best solution met by neighbours (BPNi) is updated. A position vector and a velocity vector in the feature space are assigned to each particle. The standard PSO parameters update formulation is:

$$\begin{cases} v_i(t) = v_i(t-1) + \varphi_{1i}[BP_i - x_i(t-1)] + \varphi_{2i}[BP_{Ni} - x_i(t-1)] \\ x_i(t) = x_i(t-1) + v_i(t) \end{cases}$$
(3)

In the new version of PSO, a weight term is applied to prevent divergence of the velocity vector:

$$v_i(t) = \alpha (v_i(t-1) + \varphi_{1i}[BP_i - x_i(t-1)] + \varphi_{2i}[BP_{Ni} - x_i(t-1)])$$
(4)

For a more general definition of the distance aspect, a more general form of position update equation is:

$$v_i(t) = \alpha(v_i(t-1) + \varphi_{1i}[Dis(BP_i, x_i(t-1))] + \varphi_{2i}[Dis(BP_{Ni}, x_i(t-1))])$$
(5)

General optimization algorithm is summarized in Table 1.



Table 1. PSO pseudo-code.

4.2 PSO for database rule extraction

Rule extraction process usually contains two stages: rule set generation and pruning. Rule generation is a forward selection algorithm adding new rules to current rule-set. In contract, pruning or cleaning process is a backward elimination algorithm omitting extra rules form rule-set. PSO could efficiently apply in the rule generation process. While PSO is a strong optimization algorithm in the large search space, it could obtain rule-set with maximum fitness function, no mater how complex it is. Sousa (Sousa et al. 1999) proposed a particle swarm based data mining algorithm. Between different rule representation approaches, they followed Michigan rule where each particle encodes a single rule. Comparing different PSO implementations with C4.5 and other evolutionary algorithms, Sousa concluded PSO can obtain competitive results against other alternative data mining algorithms, although it need a bit more computational effort.

4.3 Neighbourhood structure effect on rule set

Neighbourhood and social networks phenomena are a new issues proposed in swarm intelligence by PSO. In (Clerc et al. 2002) Kennedy studies how different social network structures can influence the performance of the PSO algorithm, arguing that a manipulation affecting the topological distance among particles might affect the rate and degree to which individuals are attracted towards a particular solution. Four different types of

neighbourhood topologies have been proposed in the previous works. In circles structure each individual is connected to number of its immediate neighbours. In wheels structure one individual is connected to all others, and these are connected only to that one. In star neighbourhood every individual is connected to every other individual. Finally random topology for some individuals, random symmetrical connections are assigned between pairs of individuals.

Structure of neighbourhood could affect data mining process. Rule-set fitness and convergence depend on neighbourhood structure. Experimental results have shown that the neighbourhood topology of the particle swarm has a significant effect on its ability in finding optima. Best pattern of connectivity among particles depends on fitness function. In single rule extraction such as Sousa (Sousa et al. 1999) wheel structure shows improvements in data mining convergence speed. However, random start neighbourhood is the better choice in multi-rule extraction approaches.

4.4 Decision tree rule indication using structured PSO

Decision trees are powerful classification structures. Standard techniques such as C4.5 can produce structured rules for decision tree. These techniques follow divide and conquer strategy on the data set to obtain two subsets. The algorithm is applied to subsets recursively. Each intermediate node tests a feature. Following path between leafs and root could take as a simple rectangular rule. PSO neighbourhood concept could design to extract tree based rules directly. In this structure each agent has a single decision node. Neighbourhood definition could force tree rule indication. Adjacent agents are defining as similar limitations on all features but one. The best point in the vicinity is the solution that satisfies neighbourhood condition while having lowest limitation. From decision tree point of view, best node in the neighbourhood is the root of its sub-tree and of course the best solution is decision tree root. The final solution presents by all agents together. Also the fitness function in this application is different from original target function and depends on the size of rule-set as well as accuracy parameters as described in the previous chapter.

4.5 Rule injection and rejection

Clinician involves in the rule-set instruction with rejecting an existed rule inside database or injection of new fact into the dataset. After injection or rejection process, other rules inside database may be affected. In the PSO algorithm there are two absorption points, local best solution and global best solution. New injection rule could model as a new absorption point. Injected rule could affect other solutions in the vicinity; however, training process is not applying to this rule. On the other hand, rejected rules are modelling as penalty term in the fitness function of the neighbourhood solutions. By adjusting mandatory conditions on the rules the target function changes to produce more realistic rule-set.

5. Extreme decision points and support vector machines

Classification is a valuable tool in data mining. Reasoning rules could extract from a classifier where we have more control on data analysis. Many successful decision support algorithm use well know pattern recognition algorithm. In this part we present the idea of extreme point from functional analysis. Through this part, we seem how its concept could be used as a valuable tool to make decision support structure. Also we propose support vector machine as a practical way to find critical points.

5.1 Extreme observation points for decision making

A point in a set is extreme if it could not express as an affine combination of other points. On the other hand, it is easy to show each member of a set could express as an affine combination of extreme points. Formally, the extreme points p of a decision boundary of R1 is defined as:

$$\exists w \,|\, wp + b > 0 \Longrightarrow p \in R1 \tag{6}$$

Separation theory, which is the key theory in convex optimization, could express in term of extreme points: if two sets are linearly separable, their extreme points are separable or extreme points are sufficient to check linear separation of two set. In convex set case, this condition is sufficient for non-linear separation too. In other word, the nearest points of two decision boundary are always extreme points.

In practice the extreme theory is not very useful because of inaccuracy of data and limited observations. Since real-world data is quite noisy, identifying real extreme points are difficult. Also with limited observations of a set, there is not guarantee to have all extreme points in the observation set. Support vector machine is a powerful statistical learning tool that could use to find approximation of extreme points in the noisy observation.

5.2 Support vector machines and finding extreme points

Support vector machines (SVMs) have been successfully applied to a wide range of pattern recognition problems, including handwriting recognition, object recognition, speaker identification, face detection and text categorization (Zheng 1999; Majumder et la. , 2005, Valentini et la. 2005, Smach 2008). SVMs are attractive because they are based on an extremely well developed theory based on statistical learning. A support vector machine finds an optimal separating hyper-plane between members and non-members of a given class in the feature space. One can build a classifier capable of discriminating between members and non-members of a given class, such as feature vectors for a particular disease. Such a classifier would be useful in recognizing new members of the class. Furthermore, the classifier can be applied to the original set of training data to identify outliers that may have not been previously recognized.

In this section, we briefly introduce SVM. For detailed information, see (Cortes et la. 1995, Smach 2008).

For pattern recognition, we estimate a function using training data, where N is the feature space dimension (Webb et la. 2007).

To design learning algorithms, we need a class of functions. SV classifiers are based on the class of hyper-planes

$$(\vec{w} \cdot \vec{\theta}) - b = 0 \quad \vec{w} \in \mathbb{R}^N, \ b \in \mathbb{R}$$

$$\tag{7}$$

They are corresponding to the decision functions

$$V(\bar{\theta}) = sign(\bar{w} \cdot \bar{\theta} - b). \tag{8}$$

Rescaling and such that the point(s) close to the hyper-plane satisfy implies:

$$y_i(\vec{w} \cdot \theta_i - b) \ge 1, \ i = 1, 2, \dots N$$
 (9)

The margin, measured perpendicularly to the hyper-plane, equals. To maximize the margin, we thus have to minimize subject to Equation 9. We then minimize the function:

$$\Phi(\vec{w}) = \frac{1}{2} \|\vec{w}\|^2 = \frac{1}{2} (\vec{w}, \vec{w})$$
(10)

The optimal hyper-plane can be uniquely constructed by solving a constrained quadratic programming problem whose solution is in terms of a subset of training patterns that lies in the margin. These training patterns are called support vectors.

To construct the optimal hyper-plane in the case when the data are linearly non-separable, we introduce nonnegative variables and the function

$$\Phi(\xi) = (\vec{w}, \vec{w}) + C \sum_{i} \xi_{i}$$
⁽¹¹⁾

Here is a regularization parameter used to allow a trade-off between the training error and the margin. We will minimize the above function subject to the constraints

$$y_i((\vec{w} \cdot \vec{\theta}_i) - b) \ge 1 - \xi_i \tag{12}$$

It can also be solved by quadratic optimization.



6. Data mining for temporal lobe epilepsy

This section describes the main target problem of this article, epilepsy surgery candidate selection. In the following, we will describe the importance of the problem and the challenges we faces to find a solution. The problem is new in the area of soft computing but still can be considered as a prototype of common medical diagnosis problems such as breast cancer staging or leukaemia genome expression.

6.1 Problem statement

Epilepsy is recognized as an organic process of the brain. More formally, epilepsy is an occasional, excessive, and disorderly discharge of nerve tissue, seizure, which sometimes

can be detected by electroencephalographic (EEG) recording. It is a complex symptom caused by a variety of pathological processes that result in treatment selection difficulties. Pharmacotherapy or surgical treatments are the neurologist alternatives. Optimal treatment selection in the first step may change the patient's life. Temporal lobe epilepsy is one of the most common types of known epilepsy. The main origin of seizures in this type is located in the hippocampus.

Despite optimal pharmacotherapy, about 20–30% of the patients do not become seizure-free (Sisodiya 1995). For some of these patients, surgery is a therapeutic option. Success of resective epilepsy surgery increased from 43% to 85% during the period 1986–1999 (Nei Et la 2000). Data from multiple sources suggest that 55–70% of patients undergoing temporal resection and 30–50% of patient undergoing extra-temporal resection become completely seizure-free (Tonini et la 2004). A recent prospective randomized controlled trial of surgery for temporal lobe epilepsy showed that 58% of patients randomized to surgery became seizure-free compared to 8% of the medical group (Wiebe 2001).

Surgery is considered a valuable option for medically intractable epilepsy even in the absence of a proven drug resistance; in addition, surgical outcome may be greatly influenced by the presence of selected prognostic indicators (Jeha et la. 2007). However, there are still uncertainties on who are the best surgical candidates, i.e., those who most likely will present good surgical outcome.

In a recent narrative literature review of temporal resections, good surgical outcome was associated with a number of factors (hippocampal sclerosis, anterior temporal localization of interictal epileptiform activity, absence of preoperative generalized seizures, and absence of seizures in the first post-operative week) (Sisodiya 1995). However, the published results were frequently confusing and contradictory, thus preventing inferences for clinical practice. Methodological issues (e.g., sample size, selection criteria, and methods of analysis) were indicated by the authors as the most likely explanation of the conflicting literature reports (Jeha et la. 2007).

For this reason, a quantitative review of the available literature has been undertaken in (Jeha et la. 2007) to assess the overall outcome of the epilepsy surgery and to identify the factors better correlating to seizure outcome. The aim of the study was to perform a meta-analysis of the results of published observational studies and assess the prognostic significance of the selected variables outlining the characteristics of the clinical condition, the correlations between the epileptogenic and functional lesion, and the type of surgical procedure.

6.2 Database for epilepsy patients

Human brain image database system (HBIDS) is under development for epilepsy patients at Henry Ford Health System, Detroit, MI (Ghannad-Rezaie et la. 2005; 2006). The proposed HBIDS will examine surgical candidacy among temporal lobe epilepsy patients based on their brain images and other data modalities. Moreover, it can discover relatively weak correlations between symptoms, medical history, treatment planning, outcome of the epilepsy surgery, and brain images. The HBIDS data include modalities such as MRI and SPECT along with patient's personal and medical information and EEG study (Ghannad-Rezaie et la. 2005, Siadat et la. 2005, 2003). The data has been de-identified according to HIPPA regulations (Ghannad-Rezaie et la. 2006).

For the first phase of the EEG study, the non-visual feature extractor is an expert or specialist. The experts do this routinely in the clinic based on well-defined standards. For

un-structured text information, the wrapper is the expert or trained nurse. The structured data such as patient's personal information do not need to be analyzed by the wrapper, so they are directly stored in the database (Ghannad-Rezaie et la. 2006).

6.3 Candidate selection problem

Most of data mining methods are designed to work on a huge amount of data; thus KDD problem with small sample size does not broadly browse in the data mining literature. The most successful approach is to add classification or modelling stage before the rule extraction. A smart classifier can recover the patterns inside dataset and minder can recover the classification rules. In this approach, thoughtful selection of both steps is critical.



Fig. 3. a) Extracted rules in decision, b) rule injection, c) rule rejection.

Candidate selection in epilepsy, more generally in medical diagnosis, is a hard pattern recognition problem. As well as many current bioinformatics problems, the main challenge in candidate selection problem is to find an optimal point in a very large-dimensional data-space with few samples. As an example of other problems with the same challenge, functionally gene classification problem (Wallace 2006) has a reduced feature space with 200 dimensions while usually less than 50 samples are available in each case. Epilepsy problem has a 40-dimensions space and around 55 samples. Common soft computing tools such as neural networks are efficiently applicable only on large datasets. The longer feature vector, the larger database is required. Overtraining problem is always a threat for small samples machine learning. On the other hand, conventional feature space dimension reduction algorithms such as principle component analysis (PCA) are based on statistical computations that can not be applied to small number of samples. Other difficulties such as missing data, large variety of medical data types, feature disturbances, and prior knowledge make the problem more complicated. Furthermore, knowledge recovery in this problem is not straightforward.

7. Experimental results

7.1 Classifier evaluation

Medical classification accuracy studies often yield continuous data based on predictive models for treatment outcomes. Evaluation of the classifier efficiency is computed with regard to true or false classification results. True positive (TP), true negative (TN), false positive (FP) and false negative (FN) values are the basic evaluation measures for a classifier. The sensitivity and specificity of a diagnostic test depends on more than just the "quality" of the test---they also depend on the definition of what constitutes an abnormal test. A popular method for evaluating the performance of a diagnostic test is the receiver operating characteristic (ROC) curve analysis (Zhou 2005). ROC is a plot of the true positive rate against the false positive rate for the different possible cut-points of the classifier. Each point of the ROC curve is obtained by finding the true positive rate when the decision threshold is selected based on a specific false alarm rate.

The area under ROC curve represents accuracy of a classifier. In medical problems, false alarm rate as well as false rejection rate should be lower than pre-specified limits. The trade off between false alarm rate and false rejection rate is problem specific. In surgery decisionmaking problem, both rates must be considered; however, false alarm rate (doing surgery for a patient who does not need it) is more likely to be of concern.

7.2 Cross validation training

Because of a very low number of samples, complete separation of the test and train sets is not economical. Cross-validation is used to reuse train information in test process to measure the generalization error [20]. Assume is a set with cardinality of l and an algorithm maps F to VF in the results space. We would like to measure the generalization error. Crossvalidation uses l-p samples to find the function of Vl-p where the generalization error is measured by:

$$e_1 = \sum_{x_i \in F_p} Eval(V_{l-p}(\vec{\theta}_i), y_i)$$
(13)

This process repeats M times and the final error expectation is

$$\hat{e} = \frac{1}{M} \sum_{i=1}^{M} e_i \tag{14}$$

which is expected to be the generalization error of VI. When p is 1, it can be shown that the generalization error estimation is un-biased. Although this validation is time consuming, significantly increases the power of the training process. For most efficient use of the data, training and test sets are not separated. In each training epoch, 4/5 of the patients are randomly selected to train the classifier. The rest of the patients (1/5) are used to test. The final classifier is the average of many training processes. This training strategy provides maximum database usage efficiency at the cost of higher computational complexity. In this experiment, more than 50 train-test sets are used. The training process terminates when the classifier's mean squared error of the test-set increases in the last two epochs. The train and test vectors are the normalized classifications.

7.3 Performance

Here we present experimental result of comparison of four proposed algorithms: Structured decision tree training as the representative of classic mining algorithms, Ant colony miner as an evolutionary algorithm pioneer in medical data mining, previously proposed PSO database miner, hybrid approach. Common train and test dataset has been used for all data miners. Table 2 and 3 compare performance of different algorithms.

The performance of data mining algorithms is compared from different points of view. Performance of the generated rule sets has been compared using evaluation functions proposed in the previous parts. Relatively, C4.5 generates the most accurate solution. Actually it misses very few test cases but the overall score of this approach is quit low. Having a close look on simplicity metric and number of rules, it is obvious that the high accuracy of C4.5 is the result of a more complex rule set and the loss of generalization.

C4.5 is very fast algorithms compared with SI algorithm due to its iterative and divide/conquer strategy, thus can not be fairly compared with evolutionary data miners (Figure 4). It is especially designed to handle huge amount of data so obviously we expect very fast convergence. Among evolutionary algorithms, PSO shows a very good convergence speed. Simulation shows that even with an additional classification learning process, PSO is faster that conventional ACO miner while having the same performance.

Altogether, C4.5 shows to be a powerful method but the resulting rule is too complex to use. Fast convergence of C4.5 is impressive but for small databases it is not recommended. PSO after classification process obviously outperforms PSO direct knowledge recovery from database but it is comparable with ACO in some aspects. Generally, simulation results show that the proposed hybrid process has the best overall evaluation while is still somewhat faster than the previous evolutionary algorithms. Also a bit more memory usage can count as a drawback of the new approach compared to ACO and simple PSO. The effect of rule injection and rejection process has been shown Figure 6.



Fig. 4. New approach for different rule fitness factors.



Fig. 5. Compare ACO and PSO rule extaction to new approach.

	Number of Tota	al Accuracy	Overall	
	Terms in Rules	s (%)	Evaluation (%)	
C4.5 (J48)	9	92.9	81.7	
ACO	8	76.5	88.3	
PSO on Database	10	87.7	84.2	
PSO on SVM Classifier	6	89.1	91.7	
	Sensitivity (%)	Specificity (%)	Simplicity (%)	
C4.5 (J48)	0.8421	0.9911	64.7	
ACO	0.9521	0.9821	74.2	
PSO on Database 0.8192		0.9541	53.8	
PSO on SVM Classifier	0.9821	0.9781	87.5	

Table 2. Acuracy parametters for different apparoaches. The new approach gives the simplest rule expression while keeping good specificity.

Database	Percentage of dataset	Percentage of rule set	Accuracy on train set	Accuracy on test set	Traditional Method Performance
	100%	0%	94.50	92.18	92.18
Breast	50%	10%	90.31	89.51	89.52
Cancer	30%	15%	89.21	84.64	82.91
	25%	20%	86.69	84.88	80.94
	100%	0%	76.31	75.11	75.11
Pima	50%	10%	74.12	72,05	70.82
Diabetes	30%	15%	71.81	69.86	68.21
	25%	20%	72.92	70.45	62.94
Sonar	100%	0%	99.12	97.91	97.91
	50%	10%	93.21	91.96	91.21
	30%	15%	94.12	89.76	87.91
	25%	20%	91.49	88.29	85.63
	_100%	0%	88.21	85.41	85.41
Votes	50%	10%	77.62	76.91	75.42
	30%	15%	79.21	78.81	76.65
	25%	20%	76.83	74.21	73.12

Table 3. UCI database (Hettich et la.1999) used to verified the approach.

	Accuracy	Quality	Coverage	Simplicity	Interference	Overall	Overall
	(A %)	(Q %)	(C %)	(S %)	(I %)	(using A)	(using Q)
C4.5	89	81	94	74	11	81	79
ACO	84	73	90	86	25	78	74
PSO	88	76	85	81	35	72	68
Hybrid	91	82	89	89	16	83	81

Table 4. Overall performance of different data miners on Wisconsin Breast Cancer data (*a*, *b*, *c*, *d* are set to 0.33).

146

8. Conclusion

We will develop and evaluate a new approach for interactive data mining based on swarm intelligence. The proposed method will process external rules along with the raw data to do reasoning. The proposed method is designed to work in low sample and high dimensional feature space conditions where statistical power of the raw data is not sufficient for a reliable decision. The proposed method will have both of the injection and rejection of rules to allow interactive and effective contributions provided by an expert user. The well-known support vector machine (SVM) classifier and swarm data miner will be integrated to handle joint processing of the raw data and the rules. The primary idea will be used to establish limits of observations and build decision boundaries based on these critical observations.

9. References

- Apte C.; Weiss S. (1997) Data mining with decision trees and decision rules, *Future Generation Computer Systems*, Vol. 13, No 2-3, pp 197-210
- Barakat N; Diederich, J;Eclectic rule-extraction from support vector machines, International Journal of Computational Intelligence
- Botee, H. M.; Bonabeau, E.(1998) Evolving Ant Colony Optimization, ADVANCES IN COMPLEX SYSTEMS, Vol 1; No 2/3, pp 149-160
- Clerc M.; Kennedy J. (2002) The particle swarm-explosion, stability and convergence in a multidimensional complex space, *IEEE Transactions on Evolutionary Computation*, Vol. 6, No 1, pp 58-73
- Cortes C., Vapnik V. (1995) Support-vector networks, Machine learning, Vol. 20, Issue 3, 273-297
- Chien C; Wanga W; Chenga J (2007) Data mining for yield enhancement in semiconductor manufacturing and an empirical study, Expert Systems with Applications, Vol 33, No 1,pp 192-198
- Hettich S.; Bay S. D. (1999) The UCI KDD Archive [http://kdd.ics.uci.edu]. Irvine, CA: University of California, Department of Information and Computer Science
- Galea M. (2002) Applying swarm intelligence to rule induction, MS Thesis, University of Edinburgh, Scotland
- Ghannad-Rezaie M.; Soltanian-Zadeh H.; Siada M.; Elisevich K. V.(2005) Soft computing approaches to computer aided decision making for temporal lobe epilepsy, *Proc. of IEEE International Conf. on Fuzzy Systems (NAFIPS)*, Ann Arbor, Michigan, USA
- Ghannad-Rezaie M.; Soltanian-Zadeh H.; Siadat M.-R.; K.V. Elisevich (2006) Medical Data Mining using Particle Swarm Optimization for Temporal Lobe Epilepsy, *Proceedings of the IEEE World Congress on Computational Intelligence*, Vancouver, Canada, July 15-21
- Jaganathan P.; Thangavel K., Pethalakshmi A.; Karnan M. (2007) Classification rule discovery with ant colony optimization and improved quick reduce algorithm, *IAENG International Journal of Computer Science*, Vol. 33, No. 1 pp 50-55
- Jeha, L. E. ;Najm, I.; Bingaman, W. ;Dinner, D. ;Widdess-Walsh, P. Luders, H.(2007) Surgical outcome and prognostic factors of frontal lobe epilepsy surgery, BRAIN, Vol 130; No 2, pp 574-584
- Kim, S. H.; Hyun Ju Noh (1997) Predictability of Interest Rates Using Data Mining Tools, EXPERT SYSTEMS WITH APPLICATIONS, Vol 13; No 2, pp 85-96
- Lan Y.; Zhang L.; Liu L. (2002) A method for extracting rules from incomplete information system, *Notes in Theoretical Computer Science*, Vol. 82, No. 4, pp. 312-315
- Lavrac M. (1999) Selected techniques for data mining in medicine, Artificial Intelligence in medicine, Vol 6, No.1, pp 3-23.

- Majumder S. K. ; Ghosh N.; Gupta P.K.(2005) Support vector machine for optical diagnosis of cancer, Journal of Biomedical Optics
- Nei, M.; Ho, R. T. ;Sperling, M. R. (2000) EKG Abnormalities During Partial Seizures in Refractory Epilepsy, EPILEPSIA, VOL 41, No 5, pp 542-548
- Omkara, S.N.;Karanth R (2008) Rule extraction for classification of acoustic emission signals using Ant Colony Optimisation, Engineering Applications of Artificial Intelligence, Vol 4, No 1, 320-324
- Parpinelli R.S.; Lopes H. S.; Freitas A. A. (2002) Data mining with an ant colony optimization algorithm, IEEE Transactions on Evolutionary Computation,

Quinlan J. R. (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann Inc.

- Siadat M;Soltanian-Zadeh H; Fotouhi F;Elisevich K (2003) Multimodality medical image database for temporal lobe epilepsy, Proceedings of SPIE Vol. 5033, pp 487-491.
- Siadat M;Soltanian-Zadeh H; Fotouhi F;Elisevich K (2005) Content-based image database system for epilepsy .Computer Methods and Programs in Biomedicine , Vol 79 , No 3 , pp 209 – 226
- Sisodiya, S. M.; Free, S. L. ;Stevens, J. M. ;Fish, D. R. (1995) Widespread cerebral structural changes in patients with cortical dysgenesis and epilepsy, Brain, Vol 118, No4, pp 1039
- Sousa T.; Silva A., Neves A. (2004) Particle Swarm based data mining algorithms for classification tasks, *Parallel Computing J.*, Vol. 30, pp. 767-783
- Smach, F. (2008) Generalized Fourier Descriptors with Applications to Objects Recognition in SVM Context, JOURNAL OF MATHEMATICAL IMAGING AND VISION, Vol 30, No 1, pp 43-71
- Tonini C; Beghi E;Berg A.;Bogliun G;Giordano L;Newton R; Tetto A;Vitelli E;Vitezic D;Wiebe S(2004)Predictors of epilepsy surgery outcome: a meta-analysis, Epilepsy Research , Vol. 62 , No 1 , pp 75 87
- Toivonen, H. T. T. Onkamo, P. Vasko, K. Ollikainen, V. Sevon, P. Mannila, H. Herr, M. Kere, J. (2000) Data Mining Applied to Linkage Disequilibrium Mapping, AMERICAN JOURNAL OF HUMAN GENETICS, Vol 67; No 1, pp 133-145
- Wallace M.; Ioannou S., Karpouzis K.; Kollias S. (2006) Possibility rule evaluation: a case study in facial expression analysis, *International Journal of Fuzzy Systems*, Vol. 8, No 4,pp 219-23
- Webb-Robertson B. J. M. ; Oehmen C. S. ; Cannon R.W.(2007) Support Vector Machine Classification of Probability Models and Peptide Features for Improved Peptide Identification from Shotgun Proteomics, Machine Learning and Applications, Vol 1,No 1, pp 500-505
- Wiebe, S.; Blume, W. T.; Girvin, J. P.; Eliasziw, M.(2001)A Randomized, Controlled Trial of Surgery for Temporal-Lobe Epilepsy, NEW ENGLAND JOURNAL OF MEDICINE, Vol 345; No 5, pp 311-318
- Valentini G. (2005) An experimental bias-variance analysis of SVM ensembles based on resampling techniques, IEEE Transactions on Systems, Man and Cybernetics, Part B,
- Xu, J. Huang, Y. (2007) Using SVM to Extract Acronyms from Text, SOFT COMPUTING, Vol 11; No 4, pp 369-373
- Zheng Z., Low BT, (1999) Classifying unseen cases with many missing values," Proc. of Methodologies for Knowledge Discovery and Data Mining Conf., Vol. 2, pp. 370-372
- Zou K. H.; Resnic F. S.; Talos I. (2005) A global goodness-of-fit test for receiver operating characteristic curve analysis via the bootstrap method, *J. of Biomedical Informatics*, Vol. 38, Issue 5, pp 395-403



Data Mining in Medical and Biological Research

Edited by Eugenia G. Giannopoulou

ISBN 978-953-7619-30-5 Hard cover, 320 pages Publisher InTech Published online 01, November, 2008 Published in print edition November, 2008

This book intends to bring together the most recent advances and applications of data mining research in the promising areas of medicine and biology from around the world. It consists of seventeen chapters, twelve related to medical research and five focused on the biological domain, which describe interesting applications, motivating progress and worthwhile results. We hope that the readers will benefit from this book and consider it as an excellent way to keep pace with the vast and diverse advances of new research efforts.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mostafa Ghannad-Rezaie and Hamid Soltanian-Zadeh (2008). Interactive Knowledge Discovery for Temporal Lobe Epilepsy, Data Mining in Medical and Biological Research, Eugenia G. Giannopoulou (Ed.), ISBN: 978-953-7619-30-5, InTech, Available from:

http://www.intechopen.com/books/data_mining_in_medical_and_biological_research/interactive_knowledge_di scovery_for_temporal_lobe_epilepsy



InTech Europe

University Campus STeP Ri Slavka Krautzeka 83/A 51000 Rijeka, Croatia Phone: +385 (51) 770 447 Fax: +385 (51) 686 166 www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai No.65, Yan An Road (West), Shanghai, 200040, China 中国上海市延安西路65号上海国际贵都大饭店办公楼405单元 Phone: +86-21-62489820 Fax: +86-21-62489821 © 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the <u>Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License</u>, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.



