Proceedings of ECML-95 Workshop on Statistics, Machine Learning and Knowledge Discovery in Databases, 1995, Heraklion, Greece, pp 199-203

Patient Ventilation Management Expert Rules derived from Ulm University Clinic Database Using PolyAnalyst- Knowledge Discovery System

Sergei Arseniev PhD
Mikhail Kiselev PhD

Computer Patient Monitoring Laboratory, National Research Centre for Surgery,
per. Abrikosovski, 2, Moscow, 119874, Russia
sarseniev@glas.apc.org

Bernd Classen MD

ATV Section, University Clinic for Anaesthesiology, Ulm University, Germany
bc@atvulm.atv.uni-ulm.de

A B S T R A C T

Symbolic KDD System "PolyAnalist v. 1.0R" was applyed for automated classification rules finding in the Respiratory Assessment Clinical database from University Clinic of ULM (Germany). At the first stage of this study four classification tasks have been solved. Only objective data collected from measuring monitors and results of blood tests were selected for the analysis. In all the tasks PolyAnalyst has found classification rules which provide very high accuracy of classification. Only in one task the classification error was slightly above 10% . The found rules also have clear physiological interpretation.

Keywords: Knowledge discovery in databases, artificial patient ventilation.

Introduction

Critically ill patients in Intensive Care often require artificial lung ventilation. This ventilation is performed using special apparatus-ventilator which provides gas exchange (oxygen supply and carbon dioxide removal) essential for patient life support [Spence 82]. The artificial ventilation can last from several hours to several weeks and its adequate management is one of the key factors of the whole treatment success.

At the University Clinic of Ulm, section ATV, a computerized system called "RICA" was developed for the purposes of objective assessment of patient ventilation status. The system colects data from monitoring devices, blood laboratory, adds to them doctor’s examination data, results of XR, tomorgaphy etc. Then medical experts make their evaluation of every aspect of ventilation management and patient status according to a predetermined classification scheme. In the output the system forms a database consisting of synchronized in time collected data and expert classification results.

In the present study we made an attempt to apply PolyAnalyst - symbolic knowledge discovery system [Kiselev 1994], [Kiselev, Arseniev & Flerov 1994] to derive classification rules from the RICA database automatically.

Data sets characterization

The RICA database which was used in this study included clinical data of 12 patients with the total number of training examples (ventilation status assessments) of about 200. As the number of patients is relatively small at the first stage of this study we decided to include into analysis only initial data collected from measuring devices and blood tests without doctor's examinations and other patient related information such as patient history, tomograpy, XR etc.

The following classification tasks have been selected for the analysis:

1. oxigenation diagnostics:

class "1" - excessive oxigenation;
class "0" - normal oxigenation;
class "-1"- hypoxia;

2. CO_2 diagnostics:

class "1" - hypercarbia;
class "0" - normocarbia;
class "-1"- hypocarbia;

3. acid-base status:

class "1" - alcalosis;
class "0" - normal acid-base status;
class "-1"- acidosis;

4. alveolar ventilation diagnostics:

class "1"- hyperventilation;
class "0" - adequate ventilation;
class "-1"- hypoventilation;

Analyzed data sets consisted of 31 parameters and 196 or 197 training cases. The basic independent parameters included haemodynamics parameters: heart rate, arterial and venous blood pressures; gases and blood tests: fraction of oxigen, arterial blood partial pressures of oxigen and carbon dioxide, arterial oxigen saturation, blood pH; ventilator settings: minute and tidal ventilation, respiratory rate, different ventilation pressures and also some additional parameters such as atmospheric pressure, ventilation mode, expiratory temperature. There was about 20 percent of missed values in the data.

Methods

In the present study we used PolyAnalyst v1.0R - a symbolic data mining system developed in the Laboratory of Computer Patient Monitoring (NRCS, Moscow). This version is already in practical use for more than a year and several real-world problems have been successfully solved with its assistance.

PolyAnalyst automatically derives knowledge from body of data in the form of algebraic fomulae and structural algorithms. The hypotheses are constructed as programs (procedures) in the internal functional language. The are considered as regression models. In general they are non-linear and are solved by means of hill climbing and other numerical methods.

Though PolyAnalyst presents knowledge as formulae which return continuous values it can be used also in classification tasks. In this case the system tryes to find a dependency(s) which approximates the discrete dependent variable with best possible accuracy. Then it is neccessary to determine a discrimination thresholds which correspond to a minimal classification errors.

There are several methodological schemes of PolyAnalyst usage in classification problems. Scheme choice depends on a particular task and the size of data. If the number of training examples is relatively small (as in our case, less then 200) it is worth trying to apply a strategy beased on calculation of residuals. After the first result is obtained the corresponing residuals are calculated by means of substracting the predicted valuses from the true discrete values of the dependent variable. After that the PolyAnalyst is started again this time seeking for a second term. Generally speaking this process can be continued recursively while new terms give the decrease of classification errors.

Obviously each new term should be checked for its significance. In the present study for this purpose we used randomized tests. In these tests we compared the standard errors reached for real data sets with standard errors obtained for randomized data sets. The randomized data sets were generated from real data by means of randomly changing the position of each variable value within the column not changing the value itself so that the statistical characteristics of the distribution remain the same. If the mean standard error for randomized tests was obviously greater than for a corresponding real data set we considered the new term significant.

Classification errors were calculated directly as a ratio of a number of false classifications to a total number of training examples.

Fig.1 illustrates the classification scheme used in this study.

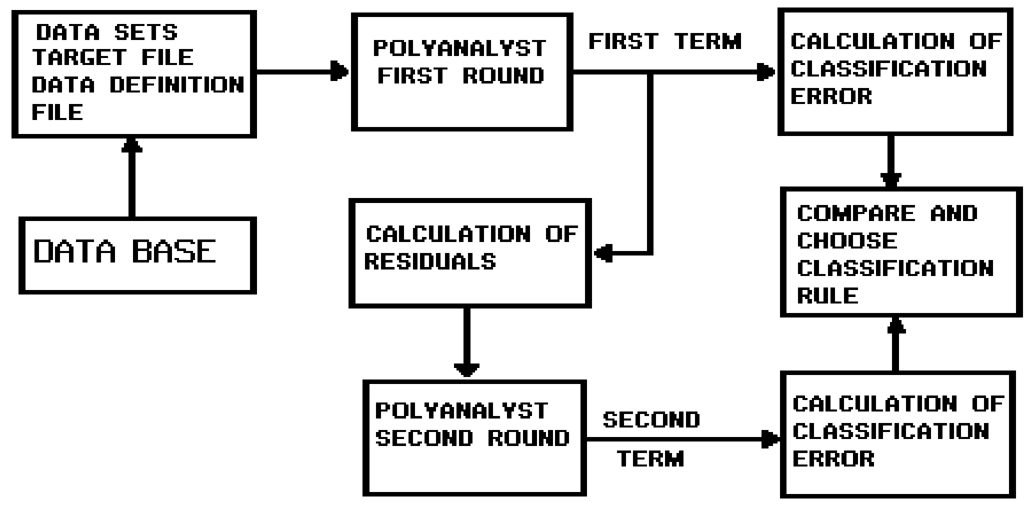

Fig 1. METHODOLOGICAL SCHEME APPLIED FOR ANALYSIS OF VENTILATION ASSESSMENT DATABASE

Results and discussion

Classification accuracy achieved by the found rules is summarized in Table 1.

Table 1

name of task	number of training examples	number of false classifications	%
1. oxigenation diagnostics	196	12	6.1
2. CO_2 diagnostics	197	15	7.6
3. acid-base status	196	10	5.1
4. alveolar vent. diagnostics	197	23	11.7

Following is an example of a derived rule (task "oxigenation diagnostics"):

```
if PaO2/ (FiO2*100)  ≥  1.46
        P=A;
else
        P=A+B;
        where:
                A=(369.361*PaO2 - 32262)/(PaO2*PaO2 + 6512);
                B=-0.0544532*SaO2 + 4.88387;
if P < -0.449332 - class "hypoxia";
if -0.449332 ≤  P < 0.542186 - class " normal oxigenation";
if P ≥  0.542186 - class " excessive oxigenation"
```

Here:
PaO2 - partial pressure of oxygen in the arterial blood;
FiO2 - fraction of inspired oxygen;
SaO2 - arterial blood oxygen saturation.

Figure 2 presents a graph of term A in the example rule as a function of oxygen partial pressure PaO2. One can see that the pattern of the graph reminds the known "oxyhemoglobin dissociation curve".

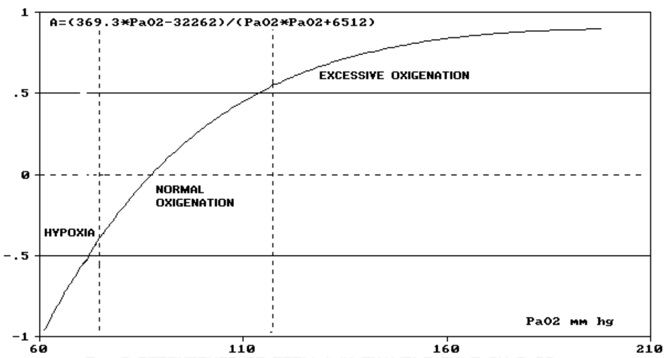

Fig. 2 DEPENDENCE OF TERM A IN EXAMPLE RULE ON PaO2

Table 1 shows that in all 4 tasks a high accuracy of classification was obtained. Only in the task "alveolar ventilation diagnostics" the error exceeded 10%. That proves the ability of the symbolic KDD System PolyAnalyst to solve classification tasks successfully. It is important to point out that the continuous value returned by a found rule is valuable itself. It can be interpreted as a continuous measure reflecting a probability that a given case belongs to certain class.

All classification rules derived in this study have clear physiological interpretation. We hope that more nontrivial rules can be derived in this important problem proving there will be more clinical cases included in the RICA database.

REFERENCES

Kiselev M.V. (1994)
PolyAnalyst - a Machine Discovery System Inferring Functional Programs, Proceedings of AAAI Workshop on Knowledge Discovery in Databases’94, Seattle, pp 237-249.

Kiselev M.V., Arseniev, S.B. & Flerov E.V. (1994)
PolyAnalyst- a Mashine Discovery System for Intelligent Analysis of Clinical Data, ESCTAIC-94 Abstracts (European Siciety for Computer Technology in Anaesthesiology and Intensive Care), Halkidiki, Greece, October 1994, p. H6.

Spence A.A. (1982)
Respiratory monitoring in intensive care, Churchill Livingstone, Ediburgh.