Control of Nosocomial Infections by Data Mining

D. Benhaddouche
Laboratory Simpa Faculty of Science
Department of Computer Science,
University of Science and Technology of Oran “Mohammed Boudiaf” USTO

A. Benyettou
Laboratory Simpa Faculty of Science
Department of Computer Science,
University of Science and Technology of Oran “Mohammed Boudiaf” USTO

Abstract: These last 15 years have been rich in publishing high quality scientific studies evaluating the effectiveness of measures to prevent nosocomial infections, particularly in intensive care unit (ICU), comparison of the results of these studies and practices in intensive care units can now to better define a program for preventing nosocomial infections to develop in these services. Focused on managing the risk of infection and prevention of nosocomial infections, our study, using tools that use data mining methods, together proposals for how well resuscitation. Among the techniques we use in data mining classification, neural networks and decision trees that also use the description used for prevention or for the unsupervised classification and clustering, we estimate we have for the rules Association. These techniques are used with several algorithms that give different results and which are distinguished from each other.

Keywords: Data Mining, Medical Data Mining, Knowledge Extraction from Databases (KDD), Hospital Environment, The Intensive Care Unit (ICU), Nosocomial Infections, Decision Trees, Association Rules, Neural Networks.

I. INTRODUCTION

Currently, no tool can help clinicians to predict the occurrence of nosocomial complications or worsening of patients in ICU. Yet many data are collected daily at the bedside. Exploitation of bio-statistics databases can reveal the correlations between data; thereby specify tools for predicting the evolution of gravity by integrating risk nosocomial.

II. INTRODUCTION TO OUR BASE

The data we have come from the database created during a study conducted last year 2008 in an intensive car unit of University Hospital of Oran, to investigate hospital infections, this study has been prepared by a team of epidemiologists belonging to the epidemiology department of Oran in collaboration with staff of the ICU [1]. Part of the data was entered manually and another part deduced by calculations. Over 1500 patients were registered for this study. These data were received as Microsoft Access; we have converted into the format recognizable by our application and platform Weka what the format « Arff ».

After targeting the essential data to complete the various tasks of the monitoring program, we obtained a database and less charged with 17 attributes and 1520 records, with the attributes:
1. Sex of patient:
2. The age of the patient:
3. Nosocomial infections earlier:
4. Immunosuppression:
5. Antibioticothérapie:
6. Chemotherapy:
7. The B.M.I:
8. The reason for hospitalization:
9. Intubation - Extubation / reintubation:
10. KT Central:
11. Catheter:
III. IMPLEMENTATION AND RESULTS

The computer tool for surveillance of nosocomial infections that we have achieved, processes the data retrieved through epidemiology services to key stakeholders that we will develop in the following

A. Pre-treatment

B. Classification

In this part we can apply multiple classifiers to generate such rules to prevent future infections nosocomial, classify different patients present in the database into classes; here are some examples of these algorithms in this section

1 Decision trees

The algorithm AD Tree

The basic algorithm is founded by Freund Y. and L. Mason [2]. In this version, search heuristics have been implemented to accelerate learning and this algorithm works only on a database with an attribute class consists of only 2 values (classes). That's why we had to change the class attribute of our database to use it, in fact, the class attribute is "IN" can take 5 values (None, Lung, Bacteremia, Urinary, otherwise) the first value indicates that the patient has contracted any nosocomial infection and the other 4 values indicates the type of infection the patient contracted, as a solution: to have only two values in this attribute, we grouped these values into two groups "yes" if the patient has contracted an IN or "no" if it has not contracted.

The algorithm J48

Known as "C4.5" [3], this algorithm generates a decision tree, but unlike the previous algorithm AD Tree, that one can generate a tree with a class attribute which can take more than two values.

With this algorithm we can better identify the type of infection contracted by the patient - If a patient stays longer than 10 days in ICU without a urinary catheter and undergoes neither extubation / reintubation or an invasive central line but has an IMB (body mass index) that exceeds 16.9 then it will contract an HAI type bacteremia (blood infection).

We have all the information of the selected instance (Fig 1) and we notice that the attribute PréditIN gives us the type of nosocomial infection for which this proceeding was filed and the IN attribute gives us the true type of infection Nosocomial for this instance.

Figure 1. Classified information on a lawsuit with the C4.5 tree.

Neural Networks

The algorithm applied "MultilayerPerceptron" uses the backpropagation and our program enables a graphical interface will accept the break and the alteration of neural networks during the training model of classification.

Figure 2. Results of classification of instances by neural network
TABLE I. COMPARISON BETWEEN THE ALGORITHMS

<table>
<thead>
<tr>
<th></th>
<th>ADtree</th>
<th>C4.5 + J48</th>
<th>Id3</th>
<th>RN A</th>
<th>Decision Tables</th>
<th>NNg e « KP PV »</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of instances correctly classified on all data</td>
<td>1503 / 1520</td>
<td>1447 / 1520</td>
<td>1273 / 1520</td>
<td>150 / 1520</td>
<td>1209 / 1520</td>
<td>152 / 1520</td>
</tr>
<tr>
<td>%</td>
<td>98.88</td>
<td>95.2%</td>
<td>83.75%</td>
<td>98.82%</td>
<td>79.54%</td>
<td>100%</td>
</tr>
<tr>
<td>The number of instances misclassified on all data</td>
<td>17 / 1520</td>
<td>73 / 1520</td>
<td>247 / 1520</td>
<td>187 / 1520</td>
<td>311 / 1520</td>
<td>0 / 1520</td>
</tr>
<tr>
<td>%</td>
<td>1.11%</td>
<td>4.8%</td>
<td>16.25%</td>
<td>1.18%</td>
<td>20.46%</td>
<td>0%</td>
</tr>
<tr>
<td>Execution time</td>
<td>0.05 Sec</td>
<td>0.03 Sec</td>
<td>0.09 Sec</td>
<td>30.3 Sec</td>
<td>0.06 Sec</td>
<td>0.49 Sec</td>
</tr>
<tr>
<td>The number of instances correctly classified the data set to test</td>
<td>77 / 86</td>
<td>79 / 86</td>
<td>75 / 86</td>
<td>84 / 86</td>
<td>65 / 86</td>
<td>86 / 100</td>
</tr>
<tr>
<td>%</td>
<td>89.53%</td>
<td>91.86%</td>
<td>86%</td>
<td>87.2%</td>
<td>75.58%</td>
<td>100%</td>
</tr>
<tr>
<td>The number of instances correctly classified the data set to test</td>
<td>9 / 86</td>
<td>7 / 86</td>
<td>11 / 86</td>
<td>27 / 86</td>
<td>21 / 86</td>
<td>0 / 86</td>
</tr>
<tr>
<td>%</td>
<td>10.46%</td>
<td>12.79%</td>
<td>86%</td>
<td>12.79%</td>
<td>24.41%</td>
<td>0%</td>
</tr>
<tr>
<td>Execution time</td>
<td>0.11 Sec</td>
<td>0.05 Sec</td>
<td>0.08 Sec</td>
<td>31.84 Sec</td>
<td>0.01 Sec</td>
<td>0.49 Sec</td>
</tr>
</tbody>
</table>

The algorithm ID3

We applied the ID3 algorithm on a portion of our database; the attributes chosen for our case represent a fraction of all the attributes found in our database.

We obtained the following decision tree
IV. Conclusion

All results obtained by the application of these algorithms have been presented in the center of Epidemiology Oran and discussed with experts and policymakers who are doctors. These same doctors were very cooperative and directed us throughout this project.

Association rules have allowed us to extract risk factors obey the reported major risk factors such as the risks reported by the CDC Centers for Disease Control, new risk factors were found which correspond to cases Nosocomial infections present in the database, the error rate prediction of nosocomial infections were more or less high, this is due to the lack of information in the database, in fact, the presence of microbiological samples for the patient on admission improves the prevention of possible infections, the downside is that they are very costly in terms of time and resources, which does not make them systematically on admission of the patient and they do make that request by the laboratory or physician. Decision trees allow us to predict potential infections for example by placing a patient in the decision tree and follow the branch that corresponds to it, so having a probability on the contraction of a nosocomial infection.

The main themes for the establishment of a surveillance system of nosocomial infection are [4]:
1- Strategies and methods of treatment of nosocomial infection surveillance.
2- Antibiotics in hospital, evaluation of resistance of isolates (patient and environment)
3- Supports genetic resistance.

References