The Automatic Discovery of Alarm Rules for the Validation of Microbiological Data

E. Lamma[a], M. Manservigi[b], P. Mello[c], A. Nanetti[d], F. Riguzzi[a], S. Storari[a]

[a] *Dipartimento di Ingegneria, Università di Ferrara, Ferrara, Italy*
[b] *Dianoema S.p.A., Bologna, Italy*
[c] *D.E.I.S., Università di Bologna, Bologna, Italy*
[d] *Clinical, Specialist and Experimental Medicine Department, Microbiology section, Università di Bologna, Bologna, Italy*

Abstract
In this work, we describe a project, jointly started by University of Bologna and Dianoema S.p.A. in order to build a system which is able to validate microbiological data. Within the project we have experimented data mining techniques in order to automatically discover association rules from microbiological data, and obtain from them alarm rules to be used for data validation. To this purpose, we have exploited the WEKA system and applied it to a database containing data about bacterial antibiograms. Discovered association rules are then transformed into alarm rules, to be used for data validation within an expert system named ESMIS. Among automatically produced alarm rules, we have identified some already considered in ESMIS and suggested by experts according to the NCCLS compendium, and new rules which were not present in that report, but were recommended by interviewed microbiologists.

Keywords:
Data mining, Knowledge Based System, Microbiology, Medical Informatics

Microbiological Data validation

Today, in a modern microbiological laboratory of a hospital the process of analysis result production is similar to an assembly line where both efficiency and quality are fundamental. With respect to efficiency, in Italy, a great number of hospitals manages microbiological analysis results by means of a software system named Italab C/S, developed by Dianoema S.p.A., an Italian information technology company operating in the Health Care market. Italab C/S is a Laboratory Information System based on a Client/Server architecture, which manages all the activities of the various analysis laboratories of the hospital. Italab C/S stores all the information concerning patients, the analysis requests and the analysis results. In particular, for bacterial infections, data includes:

- information about the patient: sex, age, hospital unit where the patient has been admitted;
- the kind of material (specimen) to be analysed (e.g., blood, urine, saliva, pus, etc.) and its origin (the body part where the specimen was collected);
- the date when the specimen was collected (often substituted with the analysis request date);
- for every different bacterium identified, its species and its antibiogram.

For each isolated bacterium, the antibiogram represents its resistance to a series of antibiotics. The set of antibiotics used to test bacterial resistance can be defined by the user, and the antibiogram is a vector of couples (antibiotics, resistance), where four types of resistance are possibly recorded: R when resistant, I when intermediate, S when susceptible, and null when unknown.

The antibiogram is not uniquely identified given the bacterium species but it can vary significantly for bacteria of the same species. This is due to the fact that bacteria of the same species may have evolved differently and have developed different resistances to antibiotics. However, very often groups of antibiotics have similar answer when tested on a given bacterium species, despite its strains.

With respect to quality of the results produced through microbiological analysis, an important step of the entire process is validation. Some instruments already execute intelligent controls on performed antibiotic test results but these controls are limited because they haven't information about specimen, patient characteristics and infection history. A system, capable of using all available information, may represent a better support for laboratory personnel in the validation task. This system should also control the application of standard antibiotic testing guidelines: these guidelines, used by almost all microbiological laboratories, suggest antibiotic test execution methods and result interpretation. Examples of problems that this system should manage are: automatic correction of antibiotic results for particular species that present in vitro susceptibility but in vivo resistance, controls on the list of tested antibiotics, predictions of test results for a group of antibiotics using some representative antibiotic (e.g., Tetracycline is representative for all Tetracyclines).

In the validation task, one would like the system to control the results reported in antibiograms in order to verify the presence of inconsistencies and alarming situations (e.g., some results for given antibiotics should be in accordance with one another or the result with respect to an antibiotic is not the expected and usual one, but some unexpected resistance has occurred).

To guide this task, NCCLS [1], an international standard organization recognised by almost all laboratories as reference in routinely work, writes an annual compendium, titled "Performance Standards for Antimicrobial Susceptibility Testing" [2], regarding testing guidelines for microbiological laboratory. NCCLS guidelines, for each species, are basically composed of a table that specifies the antibiotics to be tested, a table that specifies how to interpret the test of antibiotics and a list of exceptions regarding particular antibiotic test results. Nonetheless, the validation task, when performed manually can be long and difficult, and some laboratory management system helping microbiologists in this task should be very useful.

During the last few years, many surveillance systems have been developed in order to validate and monitor microbiological analysis results, and to early identify infective and epidemiological events.

Within a joint project between University of Bologna and Dianoema S.p.A., we have implemented an expert system (named ESMIS [3]) for validating microbiological data and generating alarms for critical situations. ESMIS has been built by following a knowledge-base approach. One of the main and well-known problems in building expert systems is knowledge acquisition. In general, this is a very time consuming and hard task. With respect to ESMIS knowledge-base building, we were interested in extracting knowledge about anomalous situations of resistance to antibiotics by isolated bacterium, in order to generate suitable alarm rules. This kind of knowledge can be extracted by hand in accordance with NCCLS documents and by intensive colloquia with experts on microbiology (and this approach has been followed in building the first ESMIS prototype (see [3]).

Another approach could be the use of the existing database where a large number of antibiograms is stored, in order to automatically extract "rules" representing anomalous situations. This latter approach, described in this paper, not antithetic, but complementary to the former one, can be very effective in validating ESMIS's knowledge-base, and also in extending this knowledge base by "discovering" new rules not yet considered by official documents. Last but not least, these new discovered rules, taking into account the history of the specific laboratory, are better tailored to the considered hospital situation, and this is very important since some resistances to antibiotics are specific to particular, local hospital environments. In this work, we report on the application of data mining techniques in ESMIS. In particular, we have experimented these techniques in order to automatically discover association rules to be used for the validation of microbiological data and for the generation of alarming situations. Other details about ESMIS architecture and knowledge base can be found in [3].

The paper is organised as follows. Section 2 describes the discovery of association rules by exploiting the APRIORI algorithm and the WEKA system. Section 3 shows how alarm rules are generated from the discovered association rules. Section 4 describes the experiments done. Related work is surveyed in section 5. We conclude and mention future work in section 6.

Discovery of Association Rules

Association rules describe correlation of events and can be regarded as probabilistic rules. "Correlation of events" means that events are frequently observed together. A good example from real life is databases of sales transactions, which are very frequently used by the marketing department of many companies because knowledge about sets of items frequently bought together is useful to develop successful marketing strategies.

The problem of discovering association rules can be formally stated as follows:

Let $I = \{i_1, i_2, ..., i_m\}$ be a set of literals, called *items*.

A *transaction* T is a set of items such that $T \subseteq I$. A database of transactions D is a set of transactions and is usually stored as a table of the form:

Table 1: Schema of a database of transactions

Transaction ID	Item

Let an *itemset* X be a set of items such $X \subseteq I$. We say that a transaction T contains an itemset X if $X \subseteq T$.

An *association rule* is an implication of the form $X \Rightarrow Y$, where X and Y are itemsets and $X \cap Y \neq \emptyset$.

- The rule $X \Rightarrow Y$ holds with confidence c in database D, if and only if c% of transactions in D that contain X also contain Y.
- The rule $X \Rightarrow Y$ has *support* s in transaction set D, if and only if s% of transactions in D contain $X \cup Y$.

Given a set of transactions D, the task of mining association rules can be reformulated as finding all association rules with at least a minimum support (called *minsup*) and a minimum confidence (called *minconf*), where *minsup* and *minconf* are user-specified values.

The task of discovering association rules can be decomposed into two subproblems:

1. Find all itemsest that have transaction support above minimum support. The *support* for an itemset is the number of transactions that contain the itemset. Itemsets with minimum support are called *large* itemsets, all others are called *small* itemsets. This subtask is addressed by the algorithm APRIORI [4].

2. Generate all association rules with minimum support and confidence from the set of all large itemsets. This subtask can be addressed by a straightforward algorithm:
    - For each large itemset *l*, find all non-empty subsets.
    - For each such subset *a* of *l*, output the rule *l* $\Rightarrow$ *(l-a)*, iff the ratio of support(*l*) to support(*a*) is at least *minconf*.

The APRIORI Algorithm

The APRIORI algorithm discovers large itemsets by means of multiple passes over the data:

- In the first pass, APRIORI counts the support of individual items and determine which of them are large, i.e., have minimum support.
- Each subsequent pass starts with a seed set represented by the itemsets found to be large in the previous pass. From this set it generates the new potentially large itemsets, called candidate itemsets.
- The actual support for these candidate itemsets is counted during a new pass over the data.
- At the end of the pass, we determine which of the candidate itemsets are actually large. These itemsets become the seed for the next pass. This process continues until no large itemsets are found.

For the sake of completeness, the algorithm is reported in the following.

Algorithm APRIORI

Notation:
k-itemset: An itemset having k items.
L_k: Set of large k-itemsets (those with minimum support).
C_k: Set of candidate k-itemsets (potentially large itemsets).

1.L_1={large 1-itemsets};
2.for (k=2; $L_{k-1}\neq\emptyset$; k++) do begin
3. C_k=apriori-gen(L_{k-1}); // New candidates
4. forall transactions t $\in$ D do begin
5. C_t = subset(C_k,t); //Candidates contained in t
6. forall candidates c $\in$ C_t do
7. c.count++;
8. end
9. L_k = {c $\in$ C_k | c.count > minsup}
10. end
11.Answer=$\cup_k L_k$

The apriori-gen function takes L_{k-1}, the set of all large (k-1)-itemsets, as an argument, and returns a set of candidates for being large k-itemsets. It exploits the fact, that expanding an itemset will reduce its support. A k-itemset can be large only if all of its (k-1)-subsets are large. So apriori-gen generates only candidates with this property, which can easily be achieved given the set L_{k-1}.

Learning Association Rules by WEKA

In order to learn association rules for validating microbiological data, we have exploited the WEKA system [5], a collection of machine learning algorithms for solving real-world data mining problems. WEKA is written in Java and runs on almost any platform. WEKA is open source software issued under the GNU General Public License. WEKA contains algorithms for performing classification, numeric prediction, clustering and learning association rules.

As regards association rule learning, WEKA employs a version of the APRIORI algorithm that is able to learn association rules from a generic table (like Table 2) with n attributes and m records.

Table 2: Example of a table for knowledge extraction

$Attribute_1$	**$Attrbute_2$**	**…**	**$Attribute_n$**
$Value_{1,1}$	$Value_{1,2}$	…	$Value_{1,n}$
…	…	…	…
$Value_{m,1}$	$Value_{m,2}$	…	$Value_{m,n}$

In this case, an association rule is a rule of the form

$A_1=v_{A1}, A_2=v_{A2},\ldots,A_j=v_{Aj} \Rightarrow B_1=v_{B1}, B_2= v_{B2},\ldots,B_k=v_{Bk}$

where $A_1, A_2,\ldots, A_j, B_1, B_2,\ldots,B_k$ are attribute names and $v_{A1}, v_{A2},\ldots,v_{Aj}, v_{B1}, v_{B2},\ldots,v_{Bk}$ are values such that v_{Al} (v_{Bh}) belongs to the domain of the attribute A_l (B_h).

In practice, each record is considered as a transaction and each possible equivalence Attribute=Value an item. WEKA's version of the APRIORI algorithm works as if Table 2 is first transformed into a transaction database with the schema of Table 3:

Table 3: Example of a table from which WEKA extracts knowledge

Transaction ID	**Item**
1	$Attribute_1= Value_{1,1}$
1	$Attribute_2= Value_{1,2}$
1	….
1	$Attribute_n= Value_{1,n}$
…	…
m	$Attribute_1= Value_{m,1}$
m	$Attribute_2= Value_{m,2}$
m	….
m	$Attribute_n= Value_{m,n}$

and the standard version of the APRIORI algorithm is then applied.

The algorithm in WEKA takes into account two numbers: the number of records verifying the rule antecendent (NA), and the number of records verifying both the antecedent and consequent of the rule (NR). Starting from these two values, confidence and support are assigned to the rule as ratio NR/NA, and NR/N (where N is the total number of

record considered) respectively. Rules are generated and presented by decreasing value for the confidence.

Generation of Alarm Rules

Discovered association rules can be transformed into alarm rules, to be used for data validation, as follows.

We have first applied filtering to discovered rules, in order to consider the most general ones among them. A rule, R1, is *more general* than a second rule, R2, if they have the same consequent, but conditions in R1's antecedent are a (proper) subset of those in R2's antecedent. For instance among the four rules below:

```
1. Amoxicillin+ClavulanicAcid=S
Clindamycin=S 3187
             ==> Oxacillin=S 3187 (1)
2. Amoxicillin+ClavulanicAcid=S
Clindamycin=S
Trimethoprim+Sulfamethoxazole=S 3131
             ==> Oxacillin=S 3131 (1)
3. Amoxicillin+ClavulanicAcid=S
Clindamycin=S Penicillin=R 2641
             ==> Oxacillin=S 2641 (1)
4. Amoxicillin+ClavulanicAcid=S
Clindamycin=S Penicillin=R
Trimethoprim+Sulfamethoxazole=S 2590
             ==> Oxacillin=S 2590 (1)
```

rule 1 is the most general, rule 4 is the most specific, and rule 2 and 3 are intermediate (and not comparable WITH each other).

To the selected most general rules, we have then applied syntactic transformations in order to produce alarm rules, to be used in ESMIS [3]. Alarm rules have been obtained by considering that an association rule of the kind:

$$X \Rightarrow Y$$

represents a *regular* (and usually quite frequent) situation, whereas the rule:

$$X\ not Y \Rightarrow \text{alarm}(Y)$$

where the consequent is complemented and moved to the antecedent, represents an *abnormality* situation. When X and *not* Y simultaneously occur, and alarm has to be raised because the usual value for Y should be true instead of false, when X is true.

In order to apply this kind of transformation, when Y is a singleton condition, we have considered the result for an antibiotic in an antibiogram as two-valued, where R is the complementary value of S and vice-versa. For instance, the alarm rule produced from rule 1 is:

```
1'. Amoxicillin+ClavulanicAcid=S
Clindamycin=S Oxacillin=R
             ==> alarm(Oxacillin=S )
```

(for the sake of simplicity, we omitted support and confidence in the reported alarm rules).

Otherwise, when Y is a composed condition, e.g.:

```
537. Oxacillin=R
       ==>[Amoxicillin+ClavulanicAcid=R,
           Penicillin=R]
```

we just move its negation to the body of the alarm rule. In this case, for the sample rule above n. 537, we obtain the following alarm rule:

```
537'. Oxacillin=R,
      not([Amoxicillin+ClavulanicAcid=R,
      Penicillin=R])]
  ==>alarm([Amoxicillin+ClavulanicAcid=R,
            Penicillin=R]
```

Experimental Results

We have applied WEKA to an Italab C/S database containing data about bacterial antibiograms. We have considered all the bacteria belonging to the species Staphilococcus Aureus, Escherichia Coli and four species belonging to Enterobacteriaceae. All the data have been collected from the Clinical, Specialist and Experimental medicine Department of the University of Bologna, in Bologna, Italy. We report about the experiments in the following.

Staphilococcus Aureus

The considered dataset for Staphilococcus Aureus contains 7009 records having as attributes 41 different antibiotics, plus the site of the considered sample, patient sex, hospital department hosting the patient and information about the therapy for the patient.

First experiments have been done by running the system with decreasing values for minimal support and confidence. In particular, we first run the system with minimal support equal to 0.5, 0.4, 0.3 e 0.2. and confidence equal to 0.9. These experiments have not produced any known rule or discovered new rules confirmed by experts. Then, we choose to further diminish the requested *minsup*, and run the system with minimum support equal to 0.1 and *minconf* equal to 0.9. With this experiment, among produced alarm rules, we have identified some rules already suggested by the NCCLS report, and already considered in the ESMIS knowledge base. In particular, we have discovered those rules which relate to each other the results of two classes of antibiotics, i.e., Oxacillin and Penicillin (when a bacterium is resistant to Oxacillin it must also be resistant to any kind of Penicillin), and the resistance result for Oxacillin and Penicillin with β-lactamase inhibition (when a bacterium is resistant to Oxacillina it must also be resistant to any Penicillin with β-lactamase inhibition). For instance, the following two istances of these general rules were found:

```
537. Oxacillin=R,
not([Amoxicillin+ClavulanicAcid=R,
Penicillin=R])]
==>alarm([Amoxicillin+ClavulanicAcid=R,
Penicillin=R]

2071. Clarithromycin =R, Oxacillin=R,
not([Amoxicillin+ClavulanicAcid=R,
Ceftriaxone=R, Penicillin=R]
==>
alarm([Amoxicillin+ClavulanicAcid=R,
ceftriaxone=R, Penicillin=R])
```

The discovery of this set of rules both confirms part of the content of the NCCLS compendium and of rules elicitated by the experts and already considered in ESMIS.

Furthermore, the experiment has also discovered new rules which were not present in the NCCLS report and in ESMIS knowledge base, but have been validated and recommended by the interviewed microbiologists, and in particular, among them, the following two:

```
1080'. Teicoplanin =S, Vancomycin =R
==>alarm(Vancomycin =S)
1539'. Vancomycin =S, Teicoplanin =R
==>alarm(Teicoplanin =S)
```

which relate to each other the results in an antibiogram of two (last-generation) antibiotics (i.e., Teicoplanin and Vancomycin).

Further experiments for the Staphilococcus Aureus have been done by filtering data and removing from the dataset unuseful anitibiograms, i.e., all those for which the bacterium was always susceptible to each antibiotic in the antibiogram (a part from Penicillin, to which the Staphilococcus Aureus can be sometimes susceptible and sometimes resistant). This filtering has been suggested by interviewed microbiologists, and has reduced the dataset to 3734 records. With this last experiment (done with decreasing minimum support, till 0.1, and minimum confidence equal to 0.9) we have newly discovered the mentioned above rule 537', rule 1080' and rule 1539', but with a higher minimum support, since noisy and unuseful data have been removed from the database.

Escherichia Coli

The considered dataset for Escherichia Coli contains 7165 records having as attributes 25 different antibiotics, plus the site of the considered sample, patient sex and information on the hospital department hosting the patient

The most significant experiments were done for this bacterium by deleting from the dataset unuseful anitibiograms, i.e., all those for which the bacterium was always susceptible to each antibiotic in the antibiogram. From the remaining data (3285 records), with a minimum support equal to 0.8 (and confidence equal to 1) a new couple of rules was discovered, and confirmed by interviewed microbiologists:

```
Cefotaxime=S, Ceftazidime =R
==>alarm(Ceftazidime =S)
Ceftazidime =S  , Cefotaxime=R
==>alarm(Cefotaxime=S)
```

This couple of rules relates to each other the results of two classes of antibiotics, i.e., Cefotaxime and Ceftazidime (when a bacterium is susceptible to Cefotaxidime it must also be susceptible to Ceftazidime, and vice-versa).

With lower support, but with confidence equal to 1, we have also discovered rules already considered in ESMIS in accordance with the NCCLS compendium, e.g. those relating, when the bacterium was isolated from the urinary tract, the resistance to Piperacillin with the resistance to Ampicillin.

Enterobacteriaceae

We have also done further experiments by considering four different bacteria belonging to the same family (Enterobacteriaceae, in particular).The considered dataset contains 3387 records having as attributes the bacteria species, 28 different antibiotics, plus patient sex and information about the therapy for the patient

Also for Enterobacteriaceae, the most significant experiments were done by deleting from the dataset unuseful anitibiograms, i.e., all those for which the considered bacteria were always susceptible to each antibiotic in the antibiogram. From the remaining data (2656 records), with support equal to 0.68 (and confidence equal to 1) we have rediscovered the couple of rules relating to each other the results of Cefotaxime and Ceftazidime (previously discovered for Escherichia Coli).

With lower support, but with confidence still equal to 1, we have also discovered rules already considered in ESMIS in accordance with the NCCLS compendium, e.g. those relating the resistance to Cefotaxime with the resistance to Cephalotin.

Related Work

During the last few years, many surveillance systems have been developed in order to monitor microbiological analysis results and to early identify infection and epidemiological events. Some of them also encompassed data validation according to NCCLS compendium. We survey the most significant among them.

WHONET 5 [6] is a database software for the management of microbiology laboratory test results. The software was developed for the management of routine laboratory results but has also been used for research studies. Software development has focused on data analysis, particularly of the results of antimicrobial susceptibility testing.

GermWatcher [7] is an expert system, which applies both local and international culture-based criteria for detecting potential nosocomial infections. Its knowledge base was obtained by the analysis of some documents, written by CDC's NNIS [8] (Center for Disease Control, National

Nosocomial Infection Surveillance), providing explicit culture-based and clinical-based definition for the most significant nosocomial infections.

TheraTrac 2 [9] is a system for microbiological data validation and real-time alarming. It directly interacts with Vitek, an expert system for test results validation, that is integrated in particular analytical instruments.

All the systems mentioned above use international standard guidelines in order to define controls to be executed on laboratory test results.

Our data mining approach is deeply integrated with the expert system ESMIS [3], under development within a joint project between the University of Bologna and Dianoema S.p.A. ESMIS is able to validate microbiological data, according to the NCCLS document. In particular, given a newly isolated bacterium, ESMIS performs five main tasks: (i) Validates the culture results; (ii) Identifies the most suitable antibiotics list; (iii) Issues alarms regarding the newly isolated bacterium; (iv) Issues alarms regarding patient clinical situation; and (v) Identifies epidemic events inside the hospital. Furthermore, ESMIS it is also able to consider alarm rules discovered through the application of data mining techniques when confirmed by the microbiologist experts. In this respect, ESMIS is able both to consider standard validation rules as they are stated in the NCCLS documents, but it is also able to extend itself and embrace new rules once they have been discovered starting from data that are peculiar of a given hospital (or region).

In the past, the University of Bologna and Dianoema S.p.A. have designed and implemented an expert system for the validation of clinical analysis [10] named DNSEV (Expert System for clinical result Validation). DNSEV has been developed in order to improve the quality of the validation process performed by a specific Laboratory Information System, which is an Italab C/S database. Quality improvement of the validation process has led to a decrease in the time required by medical doctors in the validation task of clinical analysis data, permitting them to direct their energies toward other important tasks. In DNSEV the medical laboratory expertise on the validation process is translated into rules that perform all the necessary checks on analysis results. The reasoning made by the new system is documented in order to explain it to the medical team. The type of reasoning and the rules used are clearly shown and easy to change by a laboratory expert manager.

Previous work on the detection of data inconsistencies at the level of every patient record has been done by considering the application of inductive learning on a database of atherosclerotic coronary heart disease patients [11]. In particular, confirmation rules for the detection of outliers are discovered in that work by exploiting inductive methods. The authors also consider the application of descriptor based classifiers.

In [12], data mining techniques are applied to patient data from several hospitals and along three years in order to discover associations, e.g., within diagnoses and medical treatment, with the purpose of enhancing medical quality management.

Finally, as concerns the application of data mining techniques to microbiological data, two previous works have considered the analysis of microbiological data ([3] and [13]).

In [14] the system PTAH is presented that was developed for the analysis of antibiogram data in order to help medical doctors in the prescription of antibiotics for the cure of nosocomial infections. PTAH performs four types of analysis:

- resistance level over time,
- hierarchical clustering of antibiograms,
- similarity of antibiograms
- effectiveness of antibiotics over time

In [9] the demographic clustering algorithm that is enclosed in Intelligent Miner [15,16] is applied in order find interesting cluster of antibiograms.

We differ from these works because we consider the problem of discovering potential correlations among the tests of different antibiotics, to be used later on for result validation and alarm generation.

Conclusions

In this paper we have described the application of data mining techniques in order to automatically discover association rules from microbiological data, and obtain from them alarm rules for data validation. This has been done within a project, supported by MURST, jointly started by the University of Bologna and Dianoema S.p.A. Among automatically discovered alarm rules, we have identified some already considered in the knowledge base of the expert system ESMIS – to be used for monitoring microbiological data - and suggested by experts according to the NCCLS compendium. Furthermore, we have also discovered new rules which were not present in that report, but were recommended by interviewed microbiologists.

We are currently extending ESMIS knowledge base by considering other bacterium species, by interviewing experts and by applying, in parallel, the WEKA system to a database containing data about various bacteria.

Acknowledgements

This work has been partially supported by Dianoema S.p.A. under MURST (Ministero dell'università e della ricerca scientifica e tecnologica) Project n. 23204/DSPAR/99. The authors would like to thank Giovanni Pizzi of Dianoema S.p.A. Authors are in debt with Massimo Perelli for his help in doing the experiments.

References

[1] NCCLS, National Committee for Clinical Laboratory Standards. http://www.nccls.org.

[2] Mary Jane Ferraro et. al., *Performance Standards for Antimicrobial Susceptibility Testing; Eleventh Informational Supplement*, NCCLS document M100-S11 Vol. 21 No. 1, 2001.

[3] E. Lamma, P. Mello, A. Nanetti, G. Poli, F. Riguzzi, S.Storari, *An Expert System for Microbiological Data Validation and Surveillance*, to appear in Proceedings of ISMDA 2001, Lecture Notes in Computer Science, Springer-Verlag.

[4] Agrawal Rakesh, Srikant Ramakrishnan *Fast Algorithms for Mining Association Rules*, Proceedings of the 20th International Conference on Very Large Databases, 1994.

[5] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, October 1999.

[6] World Health Organization - Department of communicable disease surveillance and response, *WHONET 5 - Microbiology Laboratory Database Software*, WHO/CDS/CSR/DRS/99.1

[7] M.G.Kahn, S.A.Steib, V.J.Fraser, W.C.Dunghan, *An Expert System for Culture-Based Infection control Surveillance*, Washington University, 1992.

[8] Center for Disease Control National Nosocomial Infection Surveillance, CDC NNIS, www.cdc.org, 9 July 2001.

[9] Theratrac, Biomerieux, see at web site: http://www.theratrac.com, 9 July 2001.

[10] M.Boari, E.Lamma, P.Mello, S.Storari, S.Monesi, *An Expert System Approach for Clinical Analysis Result Validation*, Proceedings of ICAI2000, Las Vegas, Nevada,CSREA Press, USA, 2000..

[11] D.Gamberger,N. Lavrac, G. Krstacic, T. Smuc Inconsistency tests for patient records in a coronary heart disease database, Proceedings of IDAMAP2000.

[12] W.Stuhlinger, O.Hogl, H.Stoyan, M.Muller, *Intelligent data mining for medical quality management,* Proceedings of IDAMAP2000.

[13] E.Lamma, M.Manservigi, P.Mello, F.Riguzzi, R.Serra, S.Storari, *A System for Monitoring Nosocomial Infections,* Proceedings of IDAMAP2000.

[14] M. Bohanec, M. Rems, S. Slavec, B. Urh, *PTAH: A system for supporting nosocomial infection therapy*, in N. Lavrac, E. Keravnou, B. Zupan (eds) "Intelligent Data Analysis in Medicine and Pharmacology", Kluwer Academic Publishers, 1997.

[15] Intelligent Miner, see at web site: http://www.software.ibm.com/ data/iminer/fordata, 9 July 2001.

[16] Cabena, Hadjinian, Stadler, Verhees, Zanasi, *Discovering Data Mining – from concept to implementation*, Prentice Hall – IBM.

Address for correspondence

Paola Mello

DEIS, Università di Bologna

Viale Risorgimento, 2

40136 Bologna, Italy

Tel. ++39 051 2093818 Fax. ++39 051 2093073

E-mail: pmello@deis.unibo.it