The Expert Explorer: a tool for hospital data visualization and adverse drug event rules validation

Adrian BĂCEANU [a1], Ionuț ATASIEI [a], Emmanuel CHAZARD [b], Nicolas LEROY [b], the PSIP consortium[2]

[a] Ideea Advertising, Bucharest, Romania
[b] Lille University Hospital, France

Abstract. An important part of adverse drug events (ADEs) detection is the validation of the clinical cases and the assessment of the decision rules to detect ADEs. For that purpose, a software called "Expert Explorer" has been designed by Ideea Advertising. Anonymized datasets have been extracted from hospitals into a common repository. The tool has 3 main features. (1) It can display hospital stays in a visual and comprehensive way (diagnoses, drugs, lab results, etc.) using tables and pretty charts. (2) It allows designing and executing dashboards in order to generate knowledge about ADEs. (3) It finally allows uploading decision rules obtained from data mining. Experts can then review the rules, the hospital stays that match the rules, and finally give their advice thanks to specialized forms. Then the rules can be validated, invalidated, or improved (knowledge elicitation phase).

Keywords. Data Mining, Association rules, Database, Knowledge Elicitation, Web application.

1. Introduction

Adverse Drug Events (ADEs) endanger patients' safety and instigate considerable extra hospital costs [1]. Therefore, a significant reduction of preventable ADEs is a challenging issue in Public Health. Electronic health records (EHRs) seem to be the best information source in the field of ADEs detection and prevention [2, 3].

The PSIP project (Patient Safety through Intelligent Procedures in medication) [4] follows two main objectives:

- To produce epidemiological knowledge on Adverse Drug Events
- To design a clinical decision support system (CDSS) implementing some ADE detection rules, those rules being deduced from data mining [5] of the structured hospital data bases, and semantic mining of free text collections (e.g. discharge letters) [6].

---

[1] Corresponding author: Adrian Băceanu, Ideea Advertising, P-ta Mihail Kogalniceanu nr 1 bl 1 sc C ap 5, 050064, Bucuresti, Romania; Email: adrian.baceanu@ideea.biz

[2] CHRU Lille F, CHU Rouen F, CH Denain F, Region H Hospitals of Copenhagen DK, Oracle® F, IBM-Acure® DK, Medasys® F, Vidal® F, Ideea-Advertising® Ro, Kite® I, Aalborg University DK, Aristotle University of Thessaloniki GR, Umit Innsbruck A

As a part of the project, an EHR visualization tool has been required. A software called "Expert Explorer" has been designed to meet the following requirements:

- **Hospital stay display:** the tool must be able to display a given hospital stay thanks to visual and comprehensive pages. This visualization tool must display medical and administrative information including diagnoses and procedures, drug prescriptions, lab results, and reports (e.g. discharge letters).
- **Dashboards:** the tool must allow for dashboards design, execution and display. These dashboards will be used to provide on the fly epidemiological information about ADEs in each medical department.
- **Rules review and validation:** the tool must allow implementing the decision rules provided by the data-mining team. Once a rule is implemented, the Experts must be able to review all the hospital stays that match the rule and to fill an evaluation form. The results will be used to validate or not the rule. The rules-review tool uses the hospital-stay-display feature.

Moreover, the tool must respect some common requirements:

- The tool must be usable for every member of the research project, some users being able to connect from another country to a hospital database.
- Anonymity and confidentiality of the datasets are mandatory.

2. Available data

In a first phase of the PSIP project, the available databases were identified and a generic access mechanism to the different data sources was developed. This was done by proposing and validating a common data model to extract data from the hospital databases, organize them under the same format, send them through FTP protocol, and prepare them for the next data (or semantic) mining phase.

A common data scheme (Figure 1) was agreed between partners to conduct data extraction from the various existing data [7].
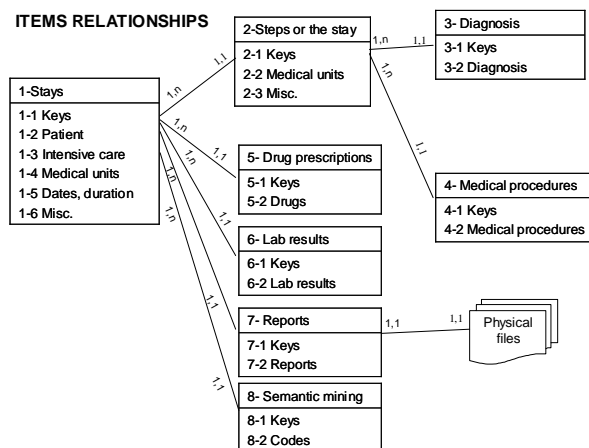

Figure 1. Simplified data scheme

The first extractions and exports of data were realized. The four data sources available to PSIP provided data as following:

- Some complete datasets, including patient description diagnoses and procedures (EHR), lab results (laboratory database) and drug prescriptions (Computer physician order entry CPOE):
  - o The Capital Region of Denmark Hospitals (RegionH hospitals, Dk) provided 2,600 hospital stays of year 2007 from cardiology and gerontology departments.
  - o The Denain hospital provided 10,000 hospital stays of year 2007 from surgery and medicine units
- Some incomplete datasets coming from hospitals without any CPOE. In those hospitals, information about drugs has been extracted from discharge letters thanks to semantic-mining:
  - o The Rouen university hospital provided 800 complete records. Its EHRs contain 1.4 million hospital stays and 5.3 million reports.
  - o The Lille university hospital has provided 10,000 records. Its EHR contains 2 millions patients and 20 millions reports and letters.

All the medical data have been anonymized and collected in a common repository. An important data management phase has been performed to ensure the quality of the data. This repository is completely anonymous and no personal information can be retrieved. Scripts have been written to make easier the data extraction. They can be re-utilised as frequently as necessary during the next steps of the project.

3. The Expert Explorer

3.1. General presentation

One of the available solutions to have a data visualization tool, was the Oracle Business Intelligence Suite (OBI) [8], but the requirements that were formulated for this tool determined us to develop a custom solution: we needed a tool to work specially with the PSIP data model, we needed a more operational view, a tool that could generate laboratory results charts, drug charts, a tool that could allow different users to access the data and fill questionnaires to evaluate the hospital stays, a tool easy to use and focused on the goal it was designed for.

The Expert Explorer is a web-based data visualization tool. It allows representing several data from a given hospital stay: medical and administrative information, diagnoses, medical procedures, lab results and drug prescriptions. The application also offers a general overview of the medical department. The medical personnel can identify particular cases in the medical unit they have in charge. This is done by some primary statistics such as death rate, distribution by patient sex or the percent of hospital stays that implied an ICU. The user can generate reports based on the available details of the hospital stay. He can target specific values of the variables, and see the distribution of the stays in percent or mean values. The application is also a tool used as a support for evaluating the rules defined in the data and semantic mining process.

Using Expert Explorer one can apply primary statistics on the existing data sets: generate reports, update data sets used in the application, define rules and load rules from files, see the details of the hospital stays corresponding to the rules.

The Expert Explorer allows for several tasks:
- report generation: reports design using basic statistics, report publication

- visualization of the hospital stays from various datasets
- rules import: the rules can be then executed as SQL queries
- validation of the hospital stays obtained from data-mining: the expert can view the rules and the related stays, and validate them (or not) using a pre defined questionnaire.

3.2. Implementation and availability

Expert Explorer is hosted on a web server and is available to anyone that knows the login credentials.

Data is stored using a MySQL [9] database and it uses the structure of the 8 tables structure defined during the first phase of the project.

The users that access the application can fit three different profiles: (1) general users, or guests that don't need a password, (2) medical experts that should login with a username and password and (3) administrators.

3.3. Data sets management

Before any activity, the user should choose the data set that he wants to work on. Also, on the same page (Figure 2) with data set selection, depending on his rights the users can update the data set, by uploading to the server the corresponding files.
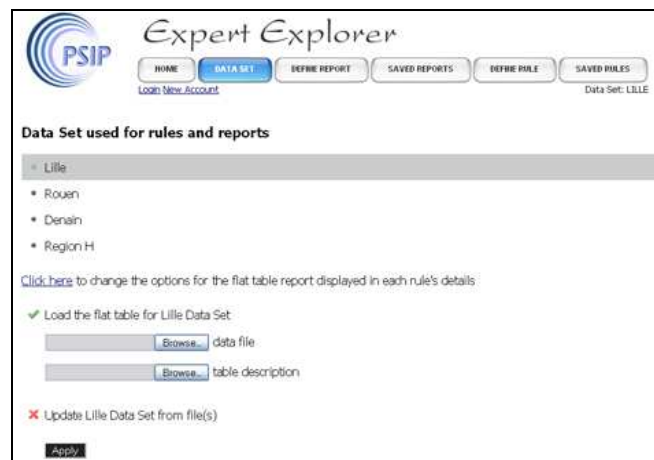
Figure 2. The page containing Data Set options

3.4. Reports management

Reports definition and visualization was one of the firsts task designed in the Expert Explorer. The reports embed some basic statistical operations (count, mean…). The aim is to provide the physicians a comprehensive look over the patients they take care of.

The report has a table structure. The fields to use as lines labels are in fact selection fields for only the hospital stays that fit the selected value (WHERE SQL clauses) and aggregation fields (GROUP BY SQL clauses). The fields to use as columns are the fields to be aggregated (sums, counts, means…). As an example, the Death=1 can be

used as a line label to restrict the dataset to hospital stays with death. At the opposite, the Mean (Death) can be used as a column label in order to compute the proportion of death in each subgroup. For this example, we have chosen sex and age as column labels and so there will a report for the hospital stays where the patient died, with details on the distribution in percent by sex and age.

Figure 3. Defining a new report

Figure 4. Visualisation of an already defined report

The report is displayed (Figure 4) using the table constructed when it was defined, with options to sort the table with a click on the columns names. When the option "Last column used to list patients" is checked, a link to a page that is listing all the hospital stays that match the report's criteria appears in the last column.

3.5. ADE detection rules management

This section is used to define ADE detection rules. The rules can be defined manually or any file meeting the data-mining-team XML format can be directly uploaded.

Figure 5. Correcting fields that weren't automatically matched in database

The rules use a structured formalism looking like "IF A1 AND A2 AND ... THEN B". This is not a natural language form. Only the fields existing in the tables can be used in a rule. The logical operator "and" and the usual comparison operators can be used. To define a rule, we select the field, the operator and the value on which the operator will be used. After defining a rule, it is kept in the database.

Using the "Import rule(s) from file" option the user can load rules from an external file. If there are two fields in different tables with the same name, the user will be asked to choose the correct field. (e.g. In Figure 5 the fields "kind" and "stay" are present in two different tables).

Figure 6. Details of a rule and the related hospital stays

The rule details page will display all the hospital stays that match this rule (Figure 6). Below the list of the stays returned by the rule, is a report based on an aggregated flat table. The columns used are defined on the Data Set page. The first line is an average of all the values from the flat table, for each column, the second line compute the averages only for the values associated to the group of hospital stays returned by the rule. A click on a hospital-stay identifier will bring up the page with all the details for that stay.

3.6. Visualization of a hospital stay

A hospital stay can be visualized in several contexts:
- As a standalone feature, the users are allowed to review every stays.
- Every implemented rule is associated to a report which displays links to the hospital stays that match the rule (Figure 6).

On the stay details page are displayed all the information available for that stay. The user only has to click on the desired tab. The following tabs are available:
- The first tab displays a description of the hospital stay (demographics, principal diagnosis, length of stay, etc.) (Figure 7)
- The second tab displays the different steps of the hospital stay, each step corresponding to a medical unit visited during the hospital stay.
- The third tab displays the medical procedures performed during the step of the hospital stay.
- The fourth tab shows the ICD10 diagnoses
- Another tab displays the drug administrations on a tabular form, and another tab presents that information in a comprehensive way thanks to a specific chart (Figure 9).
- The lab results can be displayed in a tabular form or thanks to a specific chart (Figure 8)

- Finally, it is possible to read all the anonymized reports free text documents (eg : discharge letter).

Figure 7. Hospital stay details page
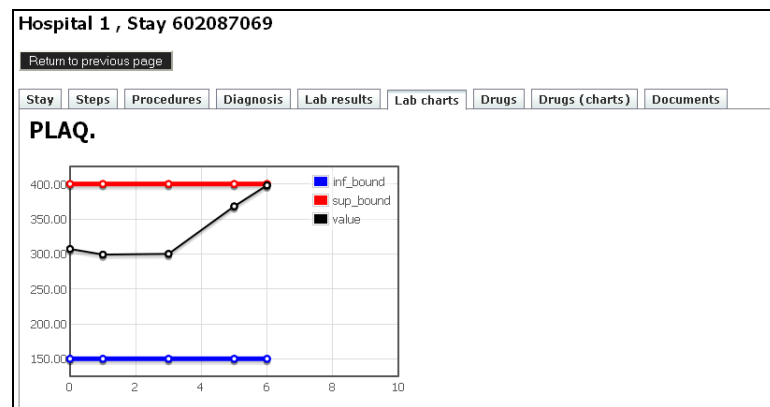
Figure 8. Lab-result-chart tab

Figure 9. Drug-chart tab

3.7. User account

The program is structured on three user levels of access, depending on the rule:
- Users (everybody) – can see/visualize data without having the power to modify them; can design and visualize dashboards
- Experts – can see/visualize and interpret the data; as a result to this he may register his advice and validate or invalidate the rules

- Administrator – has total access, the main role being that of implementing new rules and assign them for review; the administrator would also be in charge of updating the database's content

3.8. Experts' rule validation task

When an expert logs in, he is presented with a personal control panel (Figure 10) that displays him the hospital stays he has to review.

Figure 10. Expert's home page after login

Figure 11. The pre defined questionnaire that the experts have to complete

The expert's home page contains two tabs: one where are listed the hospital stays that the expert need to review, and another with the hospital stays already reviewed. The expert can review the details of each particular stay and he can go to the report page where the pre defined questionnaire (Figure 11) can be completed, in order to determine if it was or wasn't an ADE.

The expert is asked, one at the time, several questions. Depending on the answer, the next question is displayed. He only has to select the answer, confidence score and sometimes he'll need to write personal comments. The expert can review a hospital stay, even if it is not associated with a rule. If the hospital stay has a rule associated, then two additional questions about the rule are displayed. If the expert already reviewed that hospital stay, then the questionnaire is automatically filled with his previous answers, and he has the option to modify his answers, or cancel and go to previous page.

Additional interactive functions allow supporting the reviewing process itself, allowing the experts to display the list of cases to be reviewed, to navigate through the data and to document the ADE Analysis form simultaneously.

The administrator can review all the expert accounts. The administrator is the one that assign rules for review to the experts. This can be done either by uploading a text file containing the hospital stay ids or by going through the previously defined rules and assign the hospital stay that the rules have returned.

4. Conclusion

Expert Explorer is developed exclusively based on the requirements from partners from the extraction phase, the data-mining phase, and finally the knowledge elicitation phase. It's currently subject to continuous updates.

Regarding the future evolution of Expert Explorer, we plan to improve it by implementing a module capable to automatically generate statistic rules. These automatically generated rules will be available to users only after they are approved by medical personnel. But this software main interest will be data visualization, data and semantic produced rules integration and data validation. Positive feedback from some of our partners and also from medical personnel that used the tool showed that it is a useful tool. We are convinced that, if it is used correctly, it may contribute in a significant way to the quality of medical services.

Some of the weak points of the tool include right now the use of a MySQL database and the server it is hosted on. When the volume of data will get bigger, a problem of updating the database can be issued. But to solve it, we aim towards the use of Oracle database. Oracle has many years of development and releases. Their tools (pl/sql, analytical functions, bulk loading, imports, exports, etc.) are far more advanced than any other database system out there [10].

A server hosting centralized data repository, managed by or just running Oracle solutions, could be of great use, because the tool could then only connect to that specific server and select the needed data.

The tool also lacks some evaluation: it is developed while integrating the point of view of ergonomic specialists, but regularly reviews of Expert Explorer take place.

5. Acknowledgments

6. References

[1]   Kohn LT, Corrigan J, Donaldson MS. *To err is human : building a safer health system*. Washington, D.C.: National Academy Press; 2000.

[2]   Gurwitz JH, Field TS, Harrold LR, Rothschild J, Debellis K, Seger AC, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA*. 2003 Mar 5;**289**(9):1107-16.

[3]   Jalloh OB, Waitman LR. Improving Computerized Provider Order Entry (CPOE) usability by data mining users' queries from access logs. *AMIA Annu Symp Proc*. 2006:379-83.

[4]   Patient Safety by Intelligent Procedures in medication.   [cited 2009 february 24]; Available from: http://www.psip-project.eu.

[5]   Adriaans P, Zantinge D, Syllogic (Firm). *Data mining*. Harlow, England ; Reading, Mass.: Addison-Wesley; 1996.

[6]   Beuscart R, Beuscart-Zephir M-C, Brender J, Chazard E, Darmoni S, Jensen S, et al. PSIP Periodic report2009 February 26.

[7]   Bernard O, Koncar M, Sarfati J-C, Chazard E, Niès J. Structures and Data Models of the Data repositories available in the PSIP project2008 May 21.

[8]   Oracle Business Intelligence.   [cited 2009 April 21]; Available from: http://www.oracle.com/solutions/business_intelligence/index.html.

[9]   MySQL.   [cited 2009 April 22]; Available from: http://www.mysql.com/.

[10] Oracle database.   [cited 2009 April 22]; Available from: http://www.oracle.com/database/index.html.

[11] European Research Council.   [cited 2009 february 24]; Available from: http://erc.europa.eu.

[12] Seventh Framework programme.   [cited 2009 february 24]; Available from: http://cordis.europa.eu/fp7/home_en.html.