Using DSS in an Industrial Context

Authors: Juan Pablo Musella, Gustavo Janezic, Diego Branca, Daniela López De Luise, James Stuart Milne, Germán Ricchinni, Francisco Milano, Santiago Bosio
Universidad de Palermo, Engineering School, Mario Bravo 1050, C1188AAB, Argentine, aigroup@palermo.edu

Abstract

The purpose of this paper is to present the administration of socio-political aspects in the context of a help decision-making prototype, commonly known as DSS (Decision Support System). It was set to be consulted by industrial enterprises. It consists of a base of the company's own parameters that directly affect the business, an input of parameters associated industrial sectors and macro economic factors that directly affect the activity. In its architecture, there are two expert systems (one feeds the other) and a predefined set of data mining tools. The aim of the prototype presented here is to provide those responsible for decision-making information to minimize the risks faced in a given business. The main contribution of this study is the inclusion of socio-political variables as part of the information entered in the Knowledge Base and how these are derived from existing data in the community.

Keywords: Industry, Decisions, DSS, Expert System, Data Mining.

1. Introduction

DSS are specific information systems oriented to organizational decision making, and business. Their main purpose is to answer questions about things being difficult to evaluate, speculations about future events (typically where the context of the company is complex and mutable). They are commonly used for decision making of low and middle management. They usually provide information transactional and across departments. Data is typically collected from operational or or transactional data bases. One DSS definition describe it as computer-based interactive information systems whose models describe and predict production processes. It is also defined as systems designed to help decision makers to use data and models to identify and solve problems and make decisions. [1]

DSS has four major characteristics:
- integrate data and models
- the design is intended to assist managers in decision-making processes
- they do not replace (but support) human decision-making
- their purpose is to improve the effectiveness of decisions and the decision efficiency.

In the late seventies, DSS through computers was encouraged. According to Power [2] [3] they can process several input types and provide alternate decision. Therefore they can be classified in four categories as follows:

- Communication driven DSS: Allow concurrent support for many users at a specific shared task. Ex.: chats, instant messengers, collaborative systems, etc.
- Data driven DSS: focused on temporal series access and manipulation for certain organization. They are usually used for specific queries in databases or time mutable data Warehousing. In this category lies Geographic information systems, that can be used for smart managing of geographical data from maps.
- Document driven DSS: arrange, retrieve and manage unstructured information into a variety of electronic formats. They are intended to a big number of users. Their main purpose is to crawler web pages to find documents. As example there are several text analysis software, named as Wikis (Web sites freely editable by volunteers).
- Knowledge driven DSS: provide accumulated experience from facts, rules, procedures, from the actual enterprise or imported similar structures. They intend to be applied for problem resolution. They are mainly used to provide advice for management about certain products or services. As example the processing of vast data volumes, hidden pattern recognition and smart representation of such discovered patterns.
- Model driven DSS: centered in statistical, financial model access and administration. It could be used for optimization or simulation. As case is the business process prediction using past event information in order to answer what if questions.

Some of the above described models exist in the market, as proprietary or open-source solutions as described in Table 3.

[1] Database with specific content, usually oriented to an specific field (mainly for certain enterprise, organization, or topic)
As can be seen, both commercial and open source systems have several application fields. Any case, there is a strong tendency to business for commercial DSS.

This paper depicts main architecture of the HERCULES DSS prototype of the AIGroup research Lab. It handles business specific operative and account transactional information. It also has a statistical database from the business sector and one more database for macroeconomic of the country it belongs. From the logical perspective, all the data stored represents a large variable set classified as: business sector (sector for short), macro and microeconomic.

Selected macroeconomic variables arise from macroeconomic models rules. Among them GDP (Gross Domestic Product), CPI (Consumer Price Index), and occupation index. Even tough several variables are constantly used by the system, many other will depend on the specific query to be answered by the DSS.

Sector variables are mainly focused in the business activity, covering topics ranging from holding issues to competitors characteristics. Wherefore, actual variable set changes with the specific analysis the DSS has to perform.

Among variables for microeconomic study, are balance information, since it provides measure of required topics such as enterprise Economic Activity normalized by country or region.

In HERCULES, decision maker perspectives are included as part of the solution search through the process of variable set definition. A good selection enables more or less solution alternatives even when some variables represent conflicting information. Also expertise endow better definitions and as a consequence tuned answers.

From industrial point of view it could be stated that HERCULES constitutes a useful tool for a wide range of questions. In order to delimit implementation complexity, the type of possible questions was analyzed, defined and restricted. They are classified into two main types: open and restricted queries. As questions are reusable for many studies, there are meta variables to act as binded wild-cards (see fig. 1).

Questions are stated as a set of meta variables connected each other with certain specific words that force a special relationship between them. They are noted in squared brackets. In the figure, the question has two meta-variables: parameter1 and parameter2. These two meta-variables will eventually be replaced with specific variables (pre-defined by the system operator, of course). For instance parameter1 could be binded with total-screw-consumption variable, and parameter2 with screw-total-production-time.

What happens with [parámetro1] upon [parámetro2]

Fig. 1. Sample of query with two meta-variables

In the following, this papers organizes as follows: architecture (section II), interaction and query management approach (section III), economic model and specifications (section IV), socio-economic topics administration (section V), actual state of the implementation (section VI), conclusion and future work (section VI).

2. HERCULES architecture

The prototype has two user interfaces (see fig.1). The first for administrative duties over the problem domain. It enables data universe configuration and import activities through a special module named Parser. The Parser makes internal files to be used by an Expert System named Internal Expert System (IES). These files are processed automatically with alternate Data Mining techniques and the results formatted by the user GUI to be properly presented. Parser follows three steps: select files and variables, explicit relationships between selected files (using foreign keys and primary keys, also with record filtering based on a set of variable values), export data to a plain text file able to be processed by any traditional Data Mining tool.

The second user interface is aimed to provide business consultant interaction and it is directly connected to a second Expert System named User Expert System (UES) fed by the IES upon user commands to show DSS results. UES is complemented with other minor modules to conform the Advisory System. The user interface provides a formalized way to input variable declarations and to ask for specific output formatting. Figure 1 shows the connection between configuration interface and data conversion and processing. It builds one or more CSV files with processed data. Such files will be data mined for a set of tools as indicated by the IES. To do that, IES assess the best previously known DM approaches for subsets of data it recommends. This way the decision of which data is used and how it is processed remains on IES and its knowledge about the kind of query and the problem.

Results are recorded into Historical DB for further usage by the Advisory System and for self tuning of internal metrics.
 Regarding man-machine interaction, this prototype provides a sophisticated data capture and conversion toolkit embedded in the user interfaces. It is important to note that one of the critic functionalities in HERCULES is the ability to upload information from a variety of sources (document files, plain files, worksheets, databases, etc.) into its internal database, and the possibility to process it successfully with a varying subset of tools. The user can define main variable subsets for the present business problem and afterwards import them considering standard units (set ISO/IEC 80000 is included) or custom ones.

The main menu (see Fig. 2), makes it possible to determine the type of study to be performed, define elementary or composed variables, configure Database connection and settle DM restrictions. It permits also to write a textual description of the test for future reference. For elementary variable creation (see Fig. 3 and 4) there are a set of fields to fulfill regarding domain, units, banned and allowed values. It is also possible to create variable groupings by selecting one or more elementary variables (see Fig. 5).

Data generator subsystem is another important module. Its main goal is information unification for later DM. It generates some output files from all inputs. Figure 6 depicts the three modules that implement the three main steps previously described here:

1) Importer – its main goal is to import from every source into a common repository. It reads a configuration file with the type of formats to be processed. It can perform temporal transformation over data during exportation, in order to keep temporary datafile information synchronized.

2) Processor and filterer – it merges data inputted into an unified database, joining, filtering and projecting sub-tables as needed. It performs intensive usage of dynamic SQL generated upon user choices and definitions.

3) Exporter – builds a set of files usable for internal DM toolkits.
3. User query administration

Due man-machine complexity, it is required a generic and flexible mechanism for querying. It is acquired by a set of meta-variable usage, composing a large number of predefined customizable queries named here as open queries. Meta-variables can be assigned to specific variables during consulting process. Table 1 shows some open queries, classified by its coverage.

<table>
<thead>
<tr>
<th>Type</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>(1) What follows in sales?</td>
</tr>
<tr>
<td>Survey</td>
<td>(2) What is [producer] making, how fast, and why?</td>
</tr>
<tr>
<td>Human Resources</td>
<td>(3) What factors influence sales?</td>
</tr>
<tr>
<td></td>
<td>(4) How much [increase / decrease] if [expected profit] decreases?</td>
</tr>
</tbody>
</table>

It can be seen that some cases replaces meta-variables with a specific variable (for instance sales-rate in sample 1). There is also a special type of question such as case 3. It does not has an explicit meta-variable but, in order to answer it, it is required a task called expansion before processing. During expansion the query is related and replaced by a set of expanded queries where there are no meta-variables. Both meta-variables and expansion procedure are parametric and can be defined by system administrator.

Although open queries are very flexible, there are many other questions that are frequently used and thus easily foreseen because they are within the set of classical questions. For such cases there is a set of restricted queries without any meta-variable nor variable. Table 2 shows the complete list designed for HERCULES. Some of them have been considered for future implementation.

4. Economic models and specifications

Generally speaking, models and specifications have been extracted for: macro-economy, micro-economy and socio-politics perspectives. In this section the first two will be briefly described and the next section is focused in socio-politics concerns.

Macro-economy deals with Economics aggregates, monetary expansion and Recession, total goods and services, Economic growth, inflation rate, unemployment, Balance of Payments and Exchange Rate. It also covers production growth and employment along time periods (since it reflects economy growth) and short term variations which constitutes business cycle. [10] [11] [12]

In spite of the big contrast between macro and micro-economy, there is no conflict between them. Macroeconomy is just the aggregation of markets. The difference stays mainly in the focus and presentation. The scientific study of how to define, explain, anticipate economic phenomena using formalized tools to assess, model and structure information is named Econometrics.

Among main projects in the area, it worth to mention LINK project. It actually includes almost 80 models representing a total of 73 national economies and 7 regional aggregation. One of them is the Wharton-UAM model. It defines about 30000 study variables. [13] [14] [15]
Some other tools and models are used in macro-economics field, mainly related to the evaluation of politics effects and some future projections. There are also certain regional models used by public entities from government or not, to make projections. Among them are XS21 model family, EDGE models (Stochastic General Equilibrium), ARIMA (Auto-Regressive Integrated Moving Average), REM (Run time Execution Monitoring), ARMA (Autoregressive moving average), etc. [16][17][18].

Macro-economic variables are defined in HERCULES to cover several kinds of analysis. All of them feed an internal module that selects important variables and models for the actual query to be solved. Typically selected variables are the ones that represent basic national equations, such as: GDP (Gross Domestic Product), C (Private consumption), G (public consumption), I (Internal Gross Investment, private or public), X (Goods and services export), M (Goods and services Import). Other equally important are accounts related to Balance of Payments, Money Aggregates, Price variations, and Commercial exchange rates.

Due mentioned variables and their components are usually used for temporal series extraction. HERCULES design includes them along with additional ones as they result related with the problem. It is important to note that the result is derived from a set of previously known rules that model knowledge about database content and usage in every similar case. Data is updated periodically to improve precision. [19][20][21]

Other complementary perspective is the micro-economy. In this context specific metallurgy enterprises that are quoted companies (in Argentina Acindar, Siderar and Aluar). They were elicited because they are the most important at the present. Related information was extracted from trustworthy sources. Sales, production amount, statement of financial position, stock are variables considered here. All this information is required to be able to perform a deep analysis of the enterprise. All these variables derive in a big set of data useful to cross information, infer external changes and business cycle. From the resulting analysis a numeric and nominal set of values are fed to an Expert System.

5. Socio-Politics perspective variable management

The relationship between culture and socio-politics has been studied by sociology since its firsts stage as a discipline derived from industrial, economics, sociology and politics changes in the last century [22][23][24]. Society structures reflect consolidation of such relation, making them relevant for socio-economics assesment of a region, country or sector. There is a deep link between culture (values, beliefs, attitudes, regulations, etc.) and socio-economics development. (1) It has been observed also that social matter is crucial in determining enterprise performance [25].

Henceforward, to relate economy with social sciences is a subjective, relative, complex and comparative concept that becomes one of the key recurrent social science topics. (2)

next there is a brief description of the considered variables in this field:

a) demographic variables: a preliminary study was performed with classic demographic variables. The provide information about regional context for the enterprise. To betray specific data for every case study, public well known source are referred. For instance, every State has a statistic center that promulgates official evaluations. In several cases there are also private entities that are good alternatives.

b) Socio-economics variables: this kind of variables are defined for addressing coverage issues. The first obstacle is to provide data faithfulness and credibility. Most of the studies and papers are written by sociologist and epidemiologist (3) (3). That biases the focus on, shifting the approaches in a way not totally compatible with an automatic processing software device.

c) Socio-politics variables: every variable is analyzed to split it into several measurable indicators. A good number of these indicators are covered by macro-economics, micro-economics and socio-economics variables. This point of view also focuses on the governability and legal security or each region, considering precedents.

In a general sense, variable are a demonstration of models built over reality. For instance, the human development index has been extended to cover several welfare extra topics. It can be measured as a collection of economics and sociologics indexes developed by United Nations Development Program(PNUD)(4). Particularly, human development index (HDI) is based on the individual income, health (life expectation) and culture (literacy and elementary, high school and university student attendance rates).

Variables are collected, protocoled and uploaded into the prototype. Data are organized into a set of tables and are pre-processed according to actual requirements to get them able to be processed by DM. This module also may perform conversions to different formats and a set of precoded adaptations.

Once prepared and converted, exploratory analysis is performed and the resulting data is used to feed the IES and get the best interaction and behavioral model. It is important to note that typically the classical variables in this field lack of a probed model for integrate them into an equation.

6. Actual implementation status
The user interface for input data and configure variables are implemented as long with adaptation modules, database and the IES with a few rules. Actually the full set of rules are being deployed to model DM initial results. Rules have been developed for modeling part of macro-economy, micro-economy and socio-politics.

7. Conclusion and future work

HERCULES prototype has been presented, its overall design and main architectural characteristics. It remains to study self tuning approach of internal working parameters and further open and restricted queries. Alternates pending to be covered are problem automatic break down into several subproblems, Expert systems full adjustment and results GUI.

8. References

[19]. Grupo De Monitoreo Macroeconómico de la República Argentina (Argentinean macroeconomic monitor Group); GMM (2009). URL: gmm.mecon.gov.ar
[22]. La cultura empresarial en Andalucia. Un estudio sociológico de la pequeña empresa. Pablo Galindo Calvo, Universidad de Granada. Departamento de Sociología, Campus de Cartuja. 18071 Granada gailindo@ugr.es
[24]. Noemí Guillamón Cano. Variables socioeconómicas y problemas interiorizados y exteriorizados en niños y adolescentes
[25]. MANUEL ALFONSO DE LA VEGA TOLEDO. Análisis geográfico de la relación entre variables biofísicas y socioeconómicas y su influencia en categorías de pobreza en el territorio nacional. Universidad Rafael Landivar. Facultad de Ciencias Ambientales y Agrícolas