Intelligent Decision Support System based on Data Mining

Alberto Ochoa-Zezzatti[1], Sergio Enriquez[2], Arturo Elías[2], Rubén Jaramillo[3], Ricardo Pérez[3], Alfredo Jiménez[3], Ángeles Duhart[3] & Cyntia Araiza[1]

[1]Centro de Investigaciones Sociales, UACJ.
[2] Universidad Autónoma de Aguascalientes.
[3] CIATEC Centro Conacyt

Abstract. The making decision process involves a lots of information to reduce as much as possible the probability of errors. We propose an intelligent tool for decision making with ability to organize large amounts of data and represent them in many forms besides an easy interpreting for users. The tool allows the discovery of hidden patterns and predicts the tendencies through of data mining looking for improve of the making decision process.

Keywords: Intelligent tool, decision making, data mining.

1 Introduction

The decision making is a process common in the organizations whereby several alternatives are considered to provide a solution to many kinds of problems. The range of decision making process is so wide that the majority of the occasions involve several areas of knowledge. Most important to take a decision is the information related to the subject at issue. The amount of information is very relevant; more information implies enhancing in a successful decision making. The organizations are generating of great amounts of information, but a big problem is how to organize and summarize these amounts of information for be useful and easy to interpret. An intelligent tool for the decision making allows the user to analyze, to organize and to present/display the information of way summarized and easy to interpret, and thus the takers of decisions can understand better the context or situation and so the process of decision making improves.

2 Intelligent Decision Support System

2.1 Data Mining

The data mining techniques are based on great amounts of related data and allow discovering information hides as well as predict their tendencies. The data mining is a passage in the process of knowledge discovery. The mining of data involves the use of sophisticated tools of data analysis, which can include statistical models like the

© C. Delgado, C. Gutiérrez, R. Velázquez, H. Sossa (Eds.)
Advances in Computer Science and Electronic Systems
Research in Computing Science 52, 2011, pp. 123-131
Received 23/01/11
Accepted 20/02/11
Final version 07/03/11

time series seen in [3], linear regression [4], mathematical algorithms and methods of machine learning [1]. As much this technique as others of Artificial intelligence can be those that contribute more to the future of some Decision Support Systems (DSS's) [2]. The tool presented in this article, uses techniques of data mining to analyze the information and to present/display it to the user, so that it is evident recognize certain existing features in the data and then predict their tendencies, which is very useful in the decision making process.

2.2 Decision Support System

The decision making is a process that is daily carried out in all type of activities by all the people. Particularly in the enterprise surroundings, this process acquires great importance since, generally the success or failure of a company is linked to right or incorrect decision making. The process of decision making is carried out to different levels with the purpose of to achieve certain objectives and where the time also is a factor to consider, the decisions may varies depending if they are short term or long term decisions. Additionally it consists of several steps and exist different models to explain them. The Fig.1. shows some common steps in the process of decision making: Generally the systems of support for the decisions have like intention fundamental to support and to facilitate this process, through the opportune and reliable acquisition of excellent information [5].

2.3 Statistical Techniques

The statistical techniques allow realize a descriptive analysis of the data, make predictions and obtain a better decisions making. Different statistical techniques were applied from the data (formed through surveys, which are used to feed the system), like the linear regression, to define an equation or function that allows us to consider the average stature of one second generation by the sex of the individuals and the statures of their grandparents as paternal and maternal.

3 System Development

3.1 Methodology

The tool that we propose was developed in language JAVA by means of the IDE (Integrated Development Environment) of Borland, JBuilder. The handling and presentation of data were devised to diverse functions within the system for. It is possible to load different archives to work with the data, to modify them and to present/display them by regions, among others options. Software re-engineering was applied in some re-usable modules, to adapt these to the requirements, for it was begun with the design of the system, being the function it bases, the load of a data file from which the other operations of the system can be used.

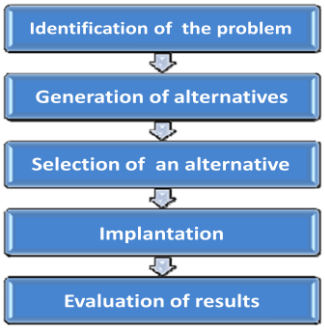

Fig. 1. *The process of Support Decisions*

4 Development Tool

This Intelligent Tool for Support Decisions (HITODE) displays to the user the information organized and summarized, in a clear way with certain characteristics of groups of people associated with properties in the collected data. Data set was obtained by surveys to a group of students of bachelor level. These data included questions on their musical, literary tastes, sports, purchases by Internet and origin municipality enters others. This data base was used to prove the functionality of the system. HITODE is composed by seven modules, which appear in the Fig. 2.

The following six modules are based on the input data that feed the system, the modules are: the file load, the graphic display of data, the mapping of regions, the reports in sequence hierarchic, the data modification and the parameters or faces of Chernoff [6]The two remaining modules, calculation of statures and migratory model were developed to divide the data of the survey and generate an equation that calculates its respective values, that are independent modules of the new data input. For the modules based on input data, the first step consists in load a data file that contains registries with n attributes. The text file requires a specific format to permit to be read by the system. The file must be in format CSV (Control On Value) and requires three lines of headed, first line has the names of the attributes, second has the amount of possible options for an attribute and finally third line contains the possible options for each attribute. When the data is loaded, the file load module qualifies its functions or buttons, areas of text, among others, allowing it to be used by the user.

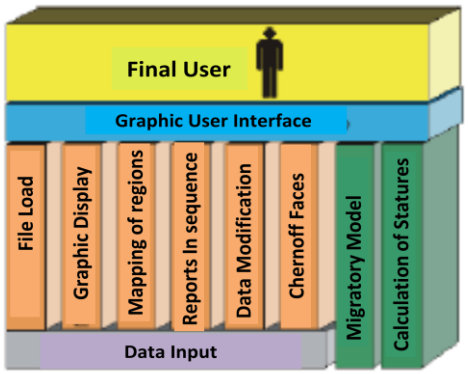

Fig. 2. Internal Structure of HITODE

The module of graphic display of data only supports attributes that are different to chain or string types. For example, the sex attribute (Fig. 3.) may be plotted because it has a defined dominion {Masculine, Feminine}. In this way, with the graphics module, it is possible show through a bar chart the amount of men and women registered in the data file of entrance.

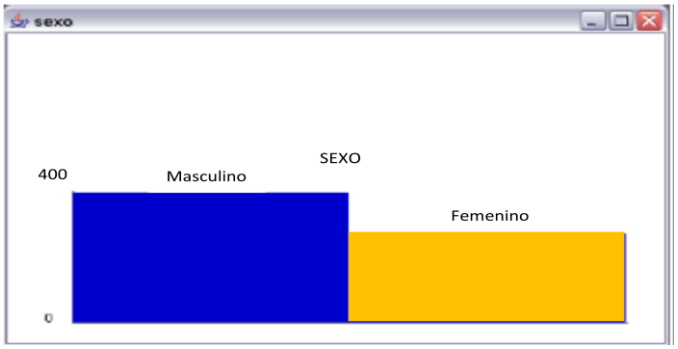

Fig. 3. HITODE Interface: Graphics Module

The module of mapping of regions, consists of presenting/displaying the information of certain attributes in a map divided by regions (Fig. 4.), where the color represents the highest percentage of occurrence of an attribute and a smaller circle in each region represents the importance of the second attribute of greater occurrence. For example when selecting the attribute "cinema", HITODE will present/display the preferences of cinematographic sort by region. Allowing the user visualize on the map predominant cinematographic sorts in each region. This function turns out particularly useful in a market study to know preferences the consumers, displayed by region. The module of faces or parameters of Chernoff displays the data by regions by means of the propose technique in [6]. It consists of using faces whose parts represent certain attributes of a certain region. For example, if a region is characterized to have a low index of unemployment, it would have a smiling mouth, and otherwise it would have a sad face. It is not possible represent

all the attributes by means of the parameters of Chernoff, only attributes whose dominion is numerical or dichotomizing, because another type of attributes would not show excellent information that can be represented through this parameters, that means if for example exists an attribute called Literature, where its dominion is {Poetry, Terror, Science fiction, Suspension, History} when trying to be presented with the size of the eyes of a face of Chernoff, would not be easy to interpret what they mean great eyes or small eyes. Although this module was realized of independent way, the original idea was taken from [6].

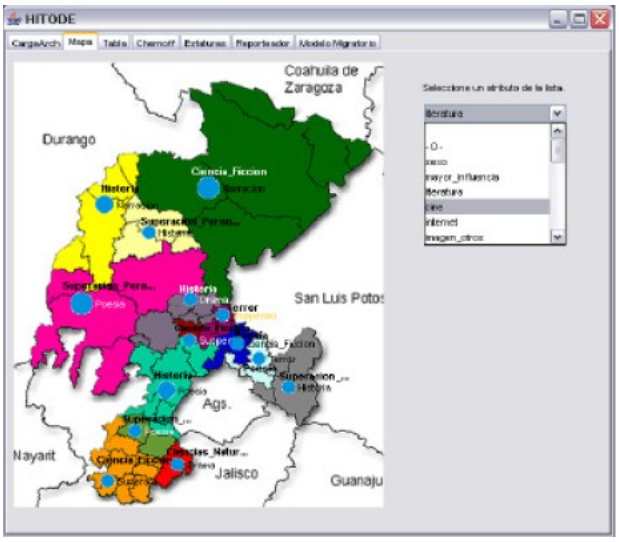

Fig. 4. HITODE Interfaces displays the attributes by regions.

The module reports in sequence hierarchic allows making a filtrate of information based on a combination of certain attributes. For example, to the first step for starting is select a region, then choice an attribute like for example the Literature type, immediately unfold options of the different types from Literature that exist in the data input together with sex, the module has the possibility of filtering by sex: Men or women solely and even both. An example of this combination of attributes could give the results of sex "masculine" that belong to "region 3" where its "literary preference" is the kind of "science fiction".

In the module of data modification, modifications from the data input file may be doing either to add new registries or to correct to the existing ones. The data appear in the form of table where each of attributes is organized by columns as it showing in Fig. 6.

The module of the migratory model is independent of the data input file of the system and makes the calculations of migration by regions, using a function generated from of data collect through surveys.

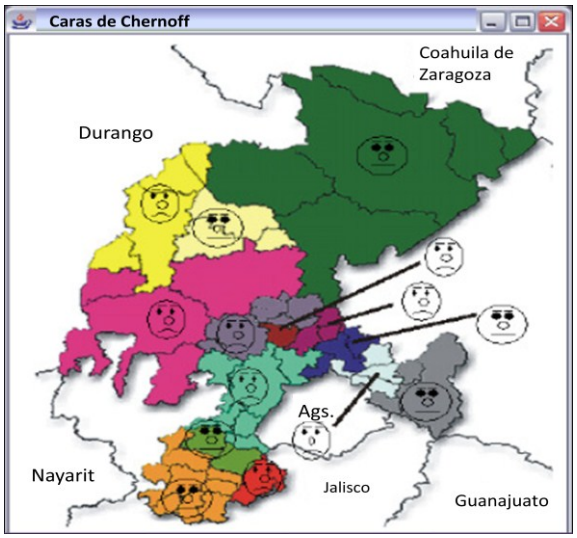

Fig. 5. HITODE Interface displays the attributes by means of Chernoff Faces

The module of statures calculation uses a function generated from statistic analysis and shows the stature average for the last two generations of the grandparents from both paternal and maternal families. Furthermore for the accomplishment of the calculation, sex of individual must be chosen. The generated equation (1) is showing follow:

$$\text{Statures}=129.5713+13.099(\text{Sex})+0.168(\text{Prom\_Abue\_Mat})+0.026(\text{Prom\_Abue\_Pat}) \quad \textbf{(1)}$$

Here, the Sex attribute takes values from 1 for masculine and 0 for feminine and Prom_abue_Pat and Prom_abue_Mat are the average of the statures of the maternal and paternal grandparents respectively. The confidence intervals and calculate of prediction from equations (2) and (3) that can to find in [4] (see this reference for more information). These values are displayed in the system followed of average stature and the symbol "±".

$$\hat{Y}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 X_0^T (X^T X)^{-1} X_0} \qquad (2)$$

$$\hat{Y}_0 \pm t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 (1 + X_0^T (X^T X)^{-1} X_0)} \qquad (3)$$

Although the modules of the migratory model and the statures calculation working on independently new data input, they can be implemented to work with different data files. These modules were developed with the purpose of display useful and simplified information of the input data file obtained on the survey realized over the tastes and preferences of the Zacatecans young people and that contain uncommon attributes in other data files.

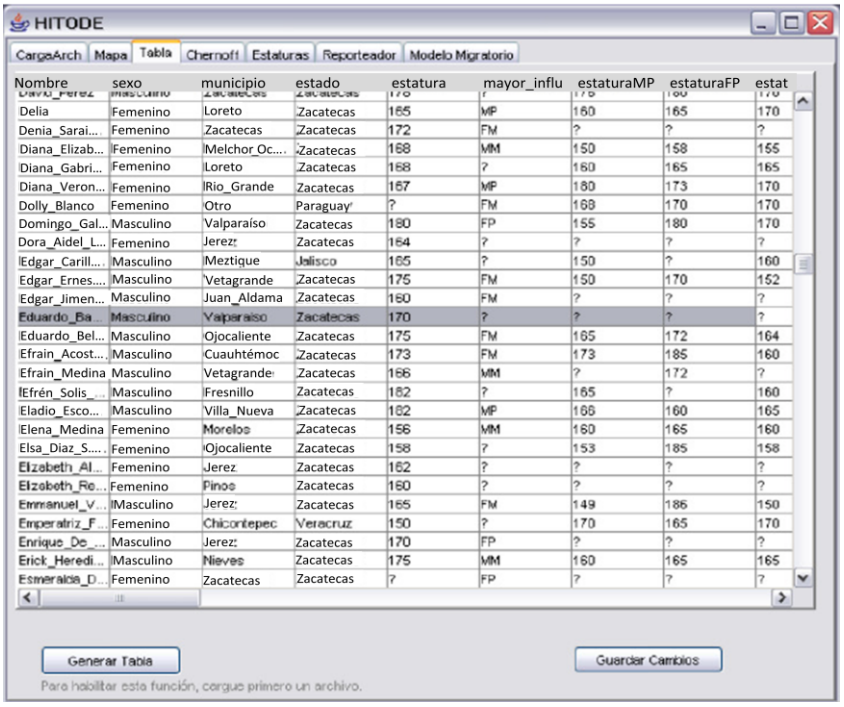

Fig. 6. HITODE Interface: Module of data modification

5 Results

The objective of this survey was to know information on Zacatecan youth to design strategies of businesses for this segment of market in specific. When using HITODE Tool with the input data file generated from this survey, was possible analyze the data of diverse forms and thus understand of simpler way the existing relations between the attributes including in the survey.

5.1 Future Research

When applying the statistical techniques, we were with the problem that high entropy in the collected data and due to this existed, was very complicated to apply some of these techniques, reason why it was necessary to make a cleaning of the data to be able to work with them. On the other hand, due to great amount of data lost percentage of prediction (security) that it was obtained with techniques as the simple linear regression were not very high when trying in particular to calculate the statures average of an individual from the statures of its grandparents, reason why it tries to realize another compilation of new data being looked for to improve the quality of these and this way to develop a function within HITODE that allows to make statistic analyses with other data. Finally HITODE is tried to migrate to Web, being implemented using XML

(Standard Markup language), to adapt the information of a comprehensible and easily analyzable way.

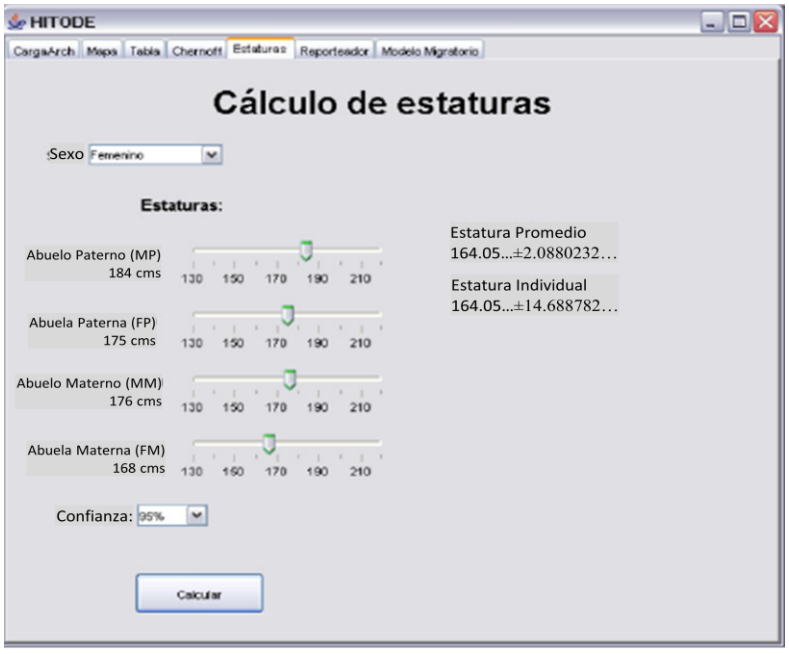

Fig. 7. HITODE Interface: Stature Calculation

6 Conclusions

An intelligent tool for the decision making allows us to compare and to corroborate information that is not so easy to visualize when great amounts of data are had. The system that we presented helps the ones in charge of decision making to visualize these data of way graphical, and divided by regions which is particularly useful at the time of making studies of market mention an example.

Acknowledges

The first author wishes to be thankful to the Drs. Hal, and Iztebegovic by its collaboration in the development of the module of the calculation of statures of the presented system as well as his collaboration for the revision of the article. Also it wishes to be thankful for to the Dr. Mendizhavili by his aid for the development and the revision of HITODE Tool and the present paper and also thank to A. Gutierrez for his support in the revision on this text.

References

[1] Seifert, Jeffrey W. "Data Mining: An overview". Congressional Research Service - The Library of Congress. 2004.

[2] Aiken, M. Liu Sheng, O. Vogel, D. "Integrating expert systems with group decision support systems". ACM Transactions on Information Systems (TOIS). 1991.

[3] Keogh, E. Pazzani, M. "Scaling up dynamic time warping for datamining applications". Conference on Knowledge Discovery in Data. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. 2000.

[4] Montgomery, D. Peck, E. Vining, G. *Introducción al Análisis de Regresión Lineal*. CECSA 3ra ed. México D.F. pp. 92 –98. 2004.

[5] Cohen Karen, D., Asín Lares, E. *Sistemas de Información para los negocios*. Mc Graw Hill, México D.F. pp. 194 – 199. 2005.

[6] Chernoff H. "The Use of Faces to Represent Points in KDimensional Space Graphically", in Journal of the American Statistical Association, Vol. 68, No. 342 (Jun., 1973), pp. 361-368.

[7] Hughes, D. *Mercadotecnia: planeación estratégica*. Ed. Addison-Wesley iberoameric. México, 1986. pp. 59 – 64