# Mark*i*S: A Marketing Intelligent System Software Application for Causal Modeling

Francisco J. Marín[1], Jorge Casillas[2], and Francisco J. Martínez-López[3]

[1] Dept. Computer Science and Artificial Intelligence, Univesity of Granada, 18071 Granada, Spain
xmarin@correo.ugr.es

[2] Dept. Computer Science and Artificial Intelligence and Research Center on Information and Communication Technology (CITIC-UGR), University of Granada, 18071 Granada, Spain
casillas@decsai.ugr.es

[3] Dept. Management, University of Granada, 18071 Granada, Spain
fjmlopez@ugr.es

**Abstract.** Mark*i*S is a software platform that uses intelligent systems for knowledge extraction from large marketing databases. Mark*i*S allows the marketing expert to model, learn and analyze marketing models using two different genetic algorithms for learning fuzzy systems with multi-item variables. Using these intelligent systems the expert can obtain more valuable information about the model and improve his/her decisions. Mark*i*S functioning is divided into five steps: the creation of the model, the presentation of the dataset, the edition of the variables and items of the model, the learning of the fuzzy rule-based systems that explain the model, and the analysis of these fuzzy systems.

**Keywords:** causal modeling, fuzzy systems, genetic algorithms, marketing, software.

## 1 Introduction

Understanding customers is a core element in businesses' success. For this reason, it is necessary to take into account their opinions before designing marketing strategies and actions. A practical way to obtain these opinions is through questionnaires compounded by a collection of items used to gather variables of interest for practitioners and researchers. These questionnaires usually represent the articulation of a set of variables that need to be measured to solve certain decisional and/or research problems. Also, it is habitual, especially in academic surveys, which said variables are the result of a subjacent theoretical model that relations all of them.

When marketing practitioners have to process and analyze large amounts of information in a short time, the major problem that companies have to deal with is the lack of appropriate levels of knowledge to take the right decisions. Conventional analysis systems, based on statistical methods, are valid to explain the models but

J. Casillas et al. (Eds.): Management Intelligent Systems, AISC 171, pp. 165–174.
springerlink.com 

are not accurate enough, taking into account the current competitive requirements. For this reason, marketing practitioners are increasingly demanding new accurate and illustrative techniques that facilitate their decision processes. With this aim, new artificial intelligence techniques for knowledge extraction (KDD) in large marketing databases have been developed.

The use of knowledge extraction algorithms allows (automatically) finding accurate information about the relational structure among the database's variables, taking a theoretical model of base as an a priori information. The outputs provided by said algorithms improve the decision of the experts, as they are able to provide detailed information on the relations about variables along all their value range. Since the first expert systems [1, 2], algorithms based on different branches of the artificial intelligence have been tested on this field: neural networks [3, 4], case-based reasoning [5], fuzzy systems [6, 7], etc.

Casillas and Martínez-López [6, 7] developed two machine learning methods based on two genetic algorithms for learning fuzzy systems with multi-item variables, obtaining interesting results. The former tries to get complete fuzzy systems with a good precision degree while maintaining a good legibility, while the latter tries to find individual interesting patterns with a good description level. These algorithms have several options to modify their behavior and need a previous step to preprocess the dataset and the initial fuzzy systems. Also, to visualize the output results, the user needs some external programs. This may be a problem for a user without knowledge about the algorithms. For this reason, to promote their use in the professional (also in the academic) field, it has been necessary to create a software that includes these algorithms and facilitates their use to the experts: Mark*i*S.

Mark*i*S is a cross-platform application that gathers all the steps in a friendly way, from the creation of the model to the learning and analysis of the fuzzy systems that explain it, in order to facilitate the adaptation to the learning algorithms presented in [6, 7]. This software allows the experts easily using these algorithms, without worrying about the complexities of the preprocessing step, since it is internally handled by the program. However, Mark*i*S maintains the configurability options offered by the algorithms, so a user can apply it in all its potential, also with the support of a complete user guide.

In Section 2, the functioning of Mark*i*S is detailed, while Section 3 presents some concluding remarks and further work.

## 2 Mark*i*S Description

Mark*i*S covers step by step the entire process of generation, learning and analysis of the marketing models, structured in five steps including: the creation of the model, the presentation of the data set, edition of the items and variables of the model, the learning of the fuzzy models and their subsequent analysis.

- **Model creation**: The first step is performed importing the questionnaire data in `csv` format (comma-separated values) that can be created with spreadsheet programs like Excel or LibreOffice Calc, creating the variable set and assigning the different items to the variable set. Once the variable set has been created, the user can add the variables to the model and set relations between them in the model tab (Section 2.1).

- **Dataset viewer**: After that, the expert can analyze the data set with several ordering and filtering options in the dataset viewer tab (Section 2.2).
- **Variable editing**: The expert can check and eventually modify some properties of the variables and the items before starting the learning process. The variable editing tab presents useful information and several options for the variables of the model and the items assigned to the variables (Section 2.3).
- **Learning**: The next step is the learning algorithm that will derive the fuzzy rule-based systems (FRBSs) that explain the relations between variables. Mark*i*S includes the algorithms and their options in the learning tab (Section 2.4).
- **Fuzzy System Analysis**: After the learning is performed, the analysis tab (Section 2.5) presents information about the FRBSs (both data base and rule base) and several quality measures. The expert can also modify the fuzzy sets of the data base and adding or removing rules from the fuzzy rule base in order to analyze the effects on the learned FRBSs. Plots of the predicted surfaces can be also analyzed for a better understanding of the fuzzy model.

## *2.1 Model Creation*

The first step is the creation of the model. This step starts in Mark*i*S importing a `csv` file with the values of the itemset, being each column a single item of the questionnaire. The `csv` file can also include a header row with short names for the items. After that, the expert has to create the set of variables of the model and assign each item to its variable. However, Mark*i*S offers an option to generate the variable set and assign automatically the items to their respective variable, including this information in the header row of the `csv` file with the format “variable name#item name.” Although is not necessary to use the entire set of items in the causal model, Mark*i*S will warn the expert if some items are unassigned.

After the set of variables has been created, the user can add these variables to the causal model and operate with them. The construction of the causal model is made in a graphical way, allowing the user adding or removing variables from the causal model, editing the variable properties (Section 2.3), grouping several first order variables into a second order variable (or ungrouping a second order variable into a set of first order variables), and creating or deleting links between variables. An example of a complete model is presented in Figure 1 (which represents the model proposed by Novak, Hoffman, and Yung for web consumer behavior analysis [8]):

Relations among variables in the causal model are analyzed with FRBSs. This FRBS list is updated whenever a new link is added or removed. After the expert has finished the building of the causal model, each fuzzy system of the FRBS list can be automatically obtained in the learning step.

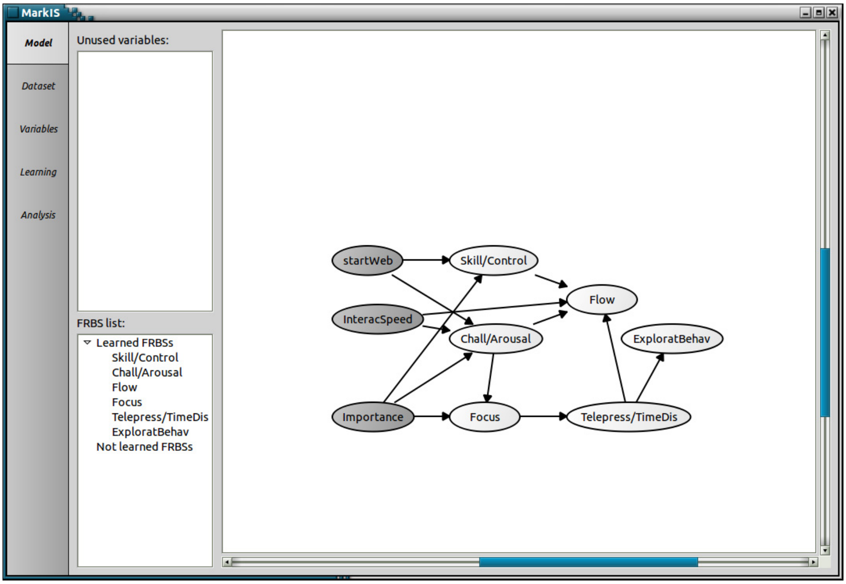



**Fig. 1** Model screenshot

## 2.2 Dataset Viewer

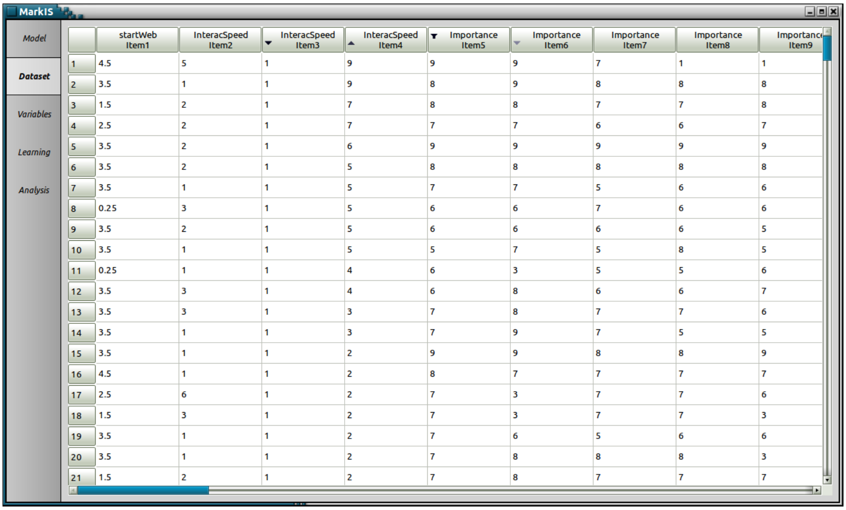
The dataset viewer tab (see Figure 2) collects the values of the itemset. Items are presented as columns of a table, and are grouped by its parent variable. The table offers some ordering and filtering options for facilitating an a priori analysis of the data:

- **Ordering options**: Rows can be ordered in ascending or descending order for a single item or for a set of items. The set of items is sorted by order of selection: the rows are sorted using the first item selected and, for the equal values, the ordering is performed with the next selected item.
- **Filtering options**: Like in Excel, the dataset viewer offers some filters to present only interesting values to the expert. For the selected item the user can filter a range of values so these values will not appear in the table; or filtering all values except the values in the range.

## 2.3 Variable Editing

The variable editing tab presents information about the variables that form the causal model and their set of items. A screen capture is shown in Figure 3.

Variable properties are presented in the *variable details* group and collects information like the variable name, group and type. The variable group property identifies the variable as a simple, first order variable or a second order group of variables.




| | startWeb Item1 | InteracSpeed Item2 | InteracSpeed Item3 | InteracSpeed Item4 | Importance Item5 | Importance Item6 | Importance Item7 | Importance Item8 | Importance Item9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.5 | 5 | 1 | 9 | 9 | 9 | 7 | 1 | 1 |
| 2 | 3.5 | 1 | 1 | 9 | 8 | 9 | 8 | 8 | 8 |
| 3 | 1.5 | 2 | 1 | 7 | 8 | 8 | 7 | 7 | 8 |
| 4 | 2.5 | 2 | 1 | 7 | 7 | 7 | 6 | 6 | 7 |
| 5 | 3.5 | 2 | 1 | 6 | 9 | 9 | 9 | 9 | 9 |
| 6 | 3.5 | 2 | 1 | 5 | 8 | 8 | 8 | 8 | 8 |
| 7 | 3.5 | 1 | 1 | 5 | 7 | 7 | 5 | 6 | 6 |
| 8 | 0.25 | 3 | 1 | 5 | 6 | 6 | 7 | 6 | 6 |
| 9 | 3.5 | 2 | 1 | 5 | 6 | 6 | 6 | 6 | 5 |
| 10 | 3.5 | 1 | 1 | 5 | 5 | 7 | 5 | 8 | 5 |
| 11 | 0.25 | 1 | 1 | 4 | 6 | 3 | 5 | 5 | 6 |
| 12 | 3.5 | 3 | 1 | 4 | 6 | 8 | 6 | 6 | 7 |
| 13 | 3.5 | 3 | 1 | 3 | 7 | 8 | 7 | 7 | 6 |
| 14 | 3.5 | 1 | 1 | 3 | 7 | 9 | 7 | 5 | 5 |
| 15 | 3.5 | 1 | 1 | 2 | 9 | 9 | 8 | 8 | 9 |
| 16 | 4.5 | 1 | 1 | 2 | 8 | 7 | 7 | 7 | 7 |
| 17 | 2.5 | 6 | 1 | 2 | 7 | 3 | 7 | 7 | 6 |
| 18 | 1.5 | 3 | 1 | 2 | 7 | 3 | 7 | 7 | 3 |
| 19 | 3.5 | 1 | 1 | 2 | 7 | 6 | 5 | 6 | 6 |
| 20 | 3.5 | 1 | 1 | 2 | 7 | 8 | 8 | 8 | 3 |
| 21 | 1.5 | 2 | 1 | 2 | 7 | 8 | 7 | 7 | 7 |

**Fig. 2** Dataset editing screenshot

Likewise, the type of the variable can be endogenous or exogenous whether it is or is not determined, respectively, by other/s variable/s of the model. The items of the variable are also listed on this group and can be activated or deactivated by the expert. Only the active items will be used in the learning and analysis processes; said items will be used to calculate the Cronbach's alpha of the multi-item variable, also presented in the variable details group. The last parameters of the variable details group are fuzzy parameters for the learning step: the number of fuzzy sets and the universe of discourse of the fuzzy variable. The universe of discourse is the $[min, max]$ range where the fuzzy variable is applied, while the number of fuzzy sets is the number of labels that form the linguistic variable.

For each item of the variable, the expert can find in the *item details* group statistical information as minimum, maximum, mean and standard deviation values and the histogram of the item values. Also, the user can reverse the values of the items. See that Likert-type (and differential semantic) scales have a subjacent bipolar assessment approach (e.g. totally disagree-totally agree) so, to avoid a response inertia when respondent is filling the questionnaire, it is usual introducing some reversed statements/items. The invert option is included to set, when necessary, all the items' scales of a variable in the same direction.

## 2.4 Learning

Once the causal model has been created, the fuzzy systems that explain the relations between variables have to be learned. As we said in Section 1, Mark*i*S includes two genetic algorithms for fuzzy learning with multi-item variables.

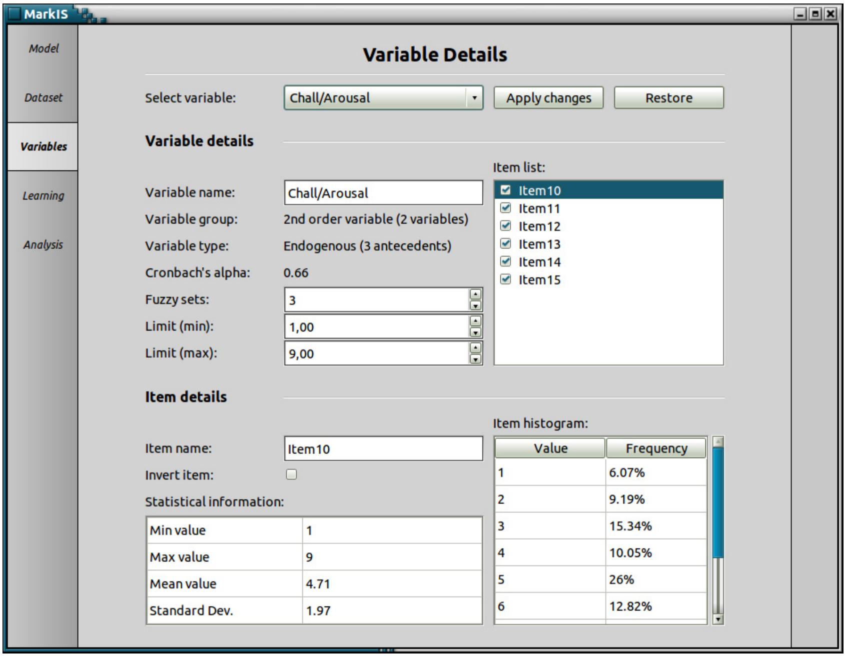



**Fig. 3** Variable editing screenshot

The former is a Pittsburgh-style genetic algorithm for predictive induction [6]. This algorithm encodes the entire set of rules into each individual of the population, in order to obtain precise and compact sets of rules. Thus, the learned rules can predict the model reliably while the description of the model is sufficiently legible.

The latter is a Michigan-style genetic algorithm for descriptive induction [7]. This algorithm tries to obtain descriptive rules with a high quality individually, instead of a precise set of rules. In this case, the model description is the main objective instead of the prediction of the variables, explaining the complexities of the model using simple, understandable rules with great precision. Michigan-style genetic algorithms encode one rule into each individual of the population, generating the set of rules using the entire population.

Both algorithms have the same learning parameters. The set of parameters can be divided into two groups: *general parameters* and *genetic parameters*. In the first group the user can set options like the type of algorithm (Pittsburgh or Michigan), a seed value for the random algorithm, the division of the dataset in training examples and test examples, and the number of iterations of the genetic algorithm. Internal parameters of the genetic algorithm are collected in the second group. These parameters define the behavior of the genetic process. Among these options the user can find the crossover and mutation probabilities, the number of individuals of the population and two options for operating with fuzzy sets: fusion (fusion of two sets

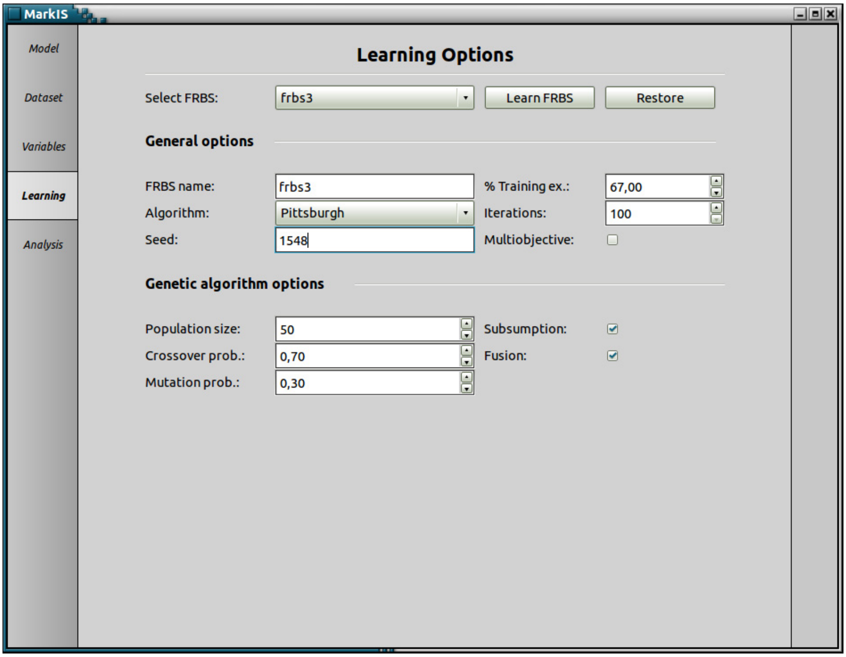



**Fig. 4** Learning options

into one) and subsumption (addition of one fuzzy set in the fuzzy variable). When these options are activated, the initial number of fuzzy sets can vary.

Mark*i*S includes all these options in the learning tab (See Figure 4). Parameters can be set for a single fuzzy system or for the entire set of FRBSs of the model.

## *2.5 Fuzzy Rule-Based System Analysis*

In the last step, the analysis tab (Figure 5) presents to the expert information about the performance and behavior of the learned FRBSs. The number of antecedents and rules and the mean square error obtained over the dataset is collected in the quality summary of the fuzzy system. Better fuzzy systems present lower values in mean square error and number of rules, specially when the Pittsburgh learning algorithm is used. Also, information about the quality of each rule of the fuzzy rule set is presented in the summary table of the rules. This information is very useful in general, but it is specially useful for the FRBSs learned with the Michigan algorithm. The quality of the rules is measured through three values: support, confidence and #Cases. Support (in the range $[0,1]$) measures the representation degree of the rule in the dataset, while confidence (also in the range $[0,1]$) measures the accuracy of the fuzzy rules. High values for support mean that the rule is more general and represents a higher portion of the sample; high values for confidence mean that the

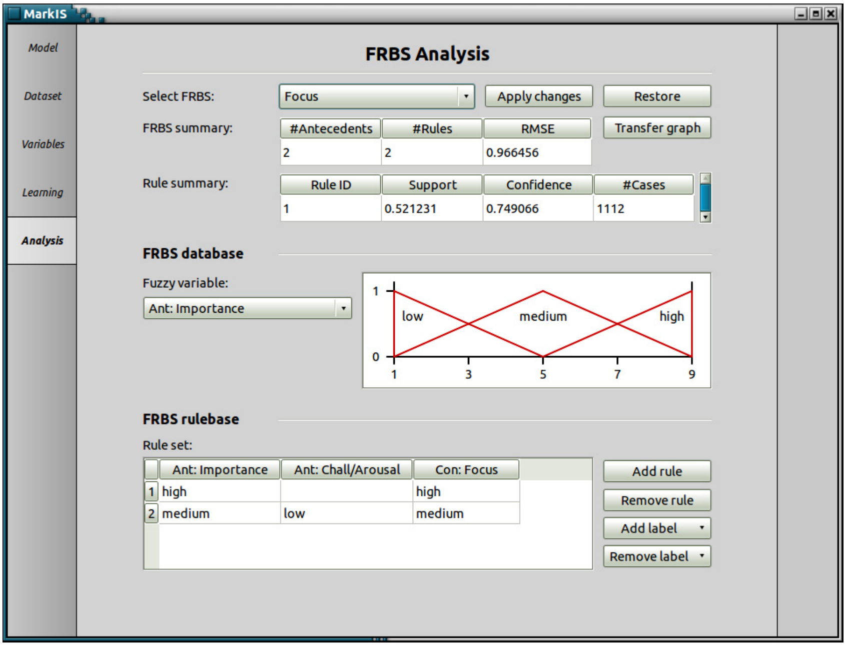



**Fig. 5** Analysis screenshot

rule is reliable. The last value (#Cases) is the number of examples matched by the antecedent of the rule.

Information about the FRBS components is also collected, divided into two groups: *data base information* and *rule base information*. The former presents graphical information about the fuzzy sets that form each variable of the FRBS (antecedents and consequents): the number of fuzzy sets and their shapes. The latter collects the rules of the rule base, presenting the labels of the antecedents and consequents that form each rule (if there is more than one label for a variable, each label is separated with the disjunction “O”).

A plot of the transference function of the fuzzy systems is also presented in this tab. It represents the predicted value of the consequent according to the values of the antecedents. Mark*i*S shows plots for fuzzy systems with one and two antecedents (2D and 3D plots, Figure 6), three antecedents (an array of density plots) and four antecedents (a matrix of density plots).

The expert can edit both the data base and the rule base of the FRBSs in order to tune them, adding the expert knowledge about the model. The data base can be edited adding new sets in a variable, removing sets or editing the shape of a existing set by changing it from a triangular shape to a trapezoidal shape or viceversa, or adjusting any of the vertices of the shape. On the other hand, the rule base can be edited adding new rules or removing an existing one. The labels of a rule can also be edited, adding a label or removing one. New values for mean square error, support,

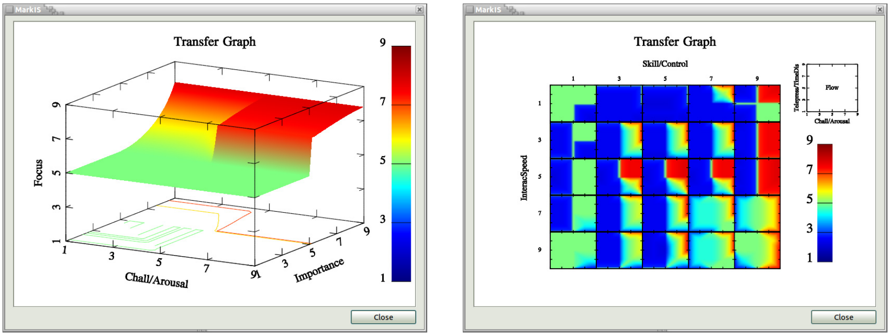



**Fig. 6** Transfer graphs for FRBSs with two and four antecedents

confidence and #Cases will be calculated for the edited FRBS. The transference function will also vary with the changes made in the FRBS.

## 3 Conclusion and Further Work

A software for analyzing causal models in marketing has been presented. This program allows an expert with little knowledge about machine learning to work with several soft computing algorithms for fuzzy learning with multi-item variables. These methods are rarely used in the professional field, dominated by conventional statistical methods, since they are quite recent and need some previous knowledge about machine learning. However, their use in the professional field can lead to a better knowledge on the customers' opinions and consequently better market decisions taken by the marketing expert/manager.

Mark*i*S reduces the complexities of these algorithms and the needs of previous knowledge, since the expert starts with a dataset with the answers to the questionnaire, creating the whole model and the relations between variables easily. Mark*i*S handles internally the complexities of the preprocessing, while offering the expert all the configurability of the algorithms. Finally, the user can analyze the quality of the learned fuzzy systems and edit them to add his/her expert knowledge about the marketing problem uner analysis (articulated with a model) in a simple way.

Finally, it should be highlighted that, though Mark*i*S has been developed to be applied in a marketing context, it is equally valid in modeling contexts that use theoretical models to structure the database, and also with a similar measurement philosophy to the one used here as a base for the experimentation. This means that Mark*i*S could be also used by management and business modelers, in general. Likewise, it could be applied in non-business areas where these kind of theoretical models are also used, as psychology or sociology, for instance.

As further work, new algorithms recently developed will be included in the software as well as new features, among them, translation to other languages (Mark*i*S is currently available only in English).

## References


1. Sisodia, R.S.: Marketing information and decision support systems for services. The Journal of Services Marketing 6(1), 51–64 (1992)
2. Wierenga, B., Van Bruggen, G.T.: Marketing management support systems: Principles, tools and implementation. Kluwer Academic Publishers (2000)
3. Fish, K.E., Johnson, J.D., Dorsey, R.E., Blodgett, J.G.: Using an artificial neural network trained with a genetic algorithm to model brand share. Journal of Business research 57(1), 79–85 (2004)
4. Boone, D.S., Roehm, M.: Retail segmentation using artificial neural networks. International Journal of Research in Marketing 19, 287–301 (2002)
5. Burke, R.R.: Reasoning with empirical marketing knowledge. International Journal of Research in Marketing 8, 75–90 (1991)
6. Casillas, J., Martínez-López, F.J.: Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. Expert Systems with Applications 36, 1645–1659 (2009)
7. Martínez-López, F.J., Casillas, J.: Marketing Intelligent Systems for consumer behaviour modelling by a descriptive induction approach based on Genetic Fuzzy Systems. Industrial Marketing Management 38, 714–731 (2009)
8. Novak, T., Hoffman, D., Yung, Y.: Measuring the customer experience in online environments: A structural modelling approach. Marketing Science 19(1), 22–42 (2000)