Data Mining Using Neural Fuzzy for Student Relationship Management

K.W. Wong*, C.C. Fung**, and T.D. Gedeon*

*School of Information Technology
Murdoch University
South St, Murdoch
Western Australia 6155
Email: {kwong | tgedeon} @ murdoch.edu.au

**School of Electrical and Computer Engineering
Curtin University of Technology
Kent St, Bentley
Western Australia 6102
Email: lanceccfung@ieee.org

***Abstract*: - Customer Relationship Management (CRM) initiatives have attracted much attention in recent years. The technology behind CRM is data mining, which require human interaction with the technology so as to achieve a successful CRM strategy. In this paper, we look at the implementation of CRM in the environment of an educational institution, known as Student Relationship Management (SRM). This paper examines the difference between SRM and CRM, and proposes a model for performing part of the SRM role. Intelligent techniques mainly Artificial Neural Network (ANN) and Fuzzy theory are used as intelligent tools to interact with the human analyst in formulating a SRM strategy.**

I. INTRODUCTION

Customer Relationship Management (CRM) initiatives have gained much attention in the business world over the last few years. Although CRM is based on information and computational technologies, the most important factor that leads to the success of a business requires an effective strategy. In other words, CRM is used to develop strategies that originate from a desire to build business around the customer [1]. With the aid of the proposed data mining techniques, it is designed to assist the decision makers in formulating their CRM strategy in an effective manner.

Recently, in the education sector and in some higher education literature, the treatment of students as "customers" has brought forth a new insight to the relationship between students and the education provider. In Spanbauer's *A Quality System for Education* [2], it was stated "Education is a service with customers like any other business and those customers do, indeed, express satisfaction and dissatisfaction about school services and instruction." However, there is a school of thought that the higher education sector is different from most business sectors. At least, the educators cannot meet all the customers' demands such as "give me a Distinction in all of my units". On the other hand, it cannot be denied that students are the most important reasons for the existence of an institution. Hence, we have to examine the relationship between students and an institution in order to establish the Student Relationship Management (SRM) model.

In this paper, the relationship between an institution and a student is considered to be equivalent or similar to the medical or legal professions. From the education institution's perspective, a number of important objectives are:

- to provide good educational services which are based on sound and well-balanced curriculum,
- to provide infrastructure to support the students in order to improve the student's employability,
- to enhance the student's creativity and analytical skills, and
- to broaden their perspectives.

In many countries, the above objectives also form the assessment criteria by the governing bodies such the Department of Education and professional accreditation bodies. The importance of the relationship is therefore to ensure that the students are able to fulfil their objectives satisfactorily under the guidance and support from the educator. The satisfaction factors from the students can in turn increase the credibility and reputation of the institution. This will then be a powerful tool used for the marketing of educational services and maintenance of the quality of the services.

In order to understand a customer, one has to analyse all the relevant data and attributes about the customer. Data mining definitely has a place in contributing to the success of a CRM strategy [3]. In SRM, there is no difference - understanding of a student is the *key* success factor. This technology of data mining is to transform data into useful information for the institution to focus on the needs and objectives of the students.

There are basically two main types of data mining: descriptive and predictive. Descriptive data mining generates information about the data so that we can realise useful and interesting underlying information. Predictive data mining makes use of past patterns and information in predicting issues, such as whether a student will continue with further studies, or is there a potential for referring other students etc. The objective of the data mining algorithm is to automate the detection of relevant patterns in a large database. The commonly used techniques in data mining are artificial neural networks

[4], decision trees [5], genetic algorithms [6], nearest neighbour method [7], and rule induction [8].

This paper will focus on the development of the SRM model, and an investigation into areas of the SRM model which may incorporate the use of intelligent data mining techniques.

II. Data Mining for SRM

In order to understand the needs of students, one has to start from analysing all the relevant data belonging to the students, thus data mining will be the intelligence behind a successful SRM strategy. This technology is to transform data into useful information for the institution to focus on the needs of the students. With most of the present techniques as mentioned in the introduction and those presented in [9], it is difficult to simultaneously perform these two analyses in the same model. All these methods are also insufficient to address uncertainties in the data. It is therefore the attempt of this paper to look for an intelligent model to address these factors.

There could be five main steps in the process of implementing a successful data mining solution for SRM: setting goals, data collection, data preparation, analysis and prediction, and, measurement and feedback. When setting the goals, identifying the student segmentation model is important. It allows reasonable goals under each category of student to precisely address issues like retention, success rate as well as possible cross-school major degrees. In data collection and preparation, it is important to address issues like feature selection, parameter identification and handling of incomplete data. When building analysis and prediction models, different methods may have to be used in each different segment to meet the intended goals. In this paper, we only cover the analysis and prediction stage in the data mining process.

III. SRM Model

In this section, we propose a SRM model. After the propose SRM model has been discussed, intelligent data mining techniques which can aid the formulation of the SRM strategy will then be presented in the next few sections. The purpose of the model is to assist the institution in developing some forms of SRM policy and strategy. We have to highlight that this model is strictly developed for this research, and the authors have no intention of copying any existing model. In business intelligence, data mining and campaign management technologies normally followed different objectives, thus it is difficult to integrate them into a single system [10]. It is not much different in the higher education environment. In this paper, we not only integrate the intelligent data mining model to the campaign management model, but we also propose models to handle the management of institutional effectiveness as well as management of enrolment.

The main objectives of the integration can be divided into the three main categories: Institutional Effectiveness, Campaign Marketing and Enrolment Management [11].

When accessing Institutional Effectiveness, the main objective is to assess the effectiveness of learning. This includes areas such as "how well does a student learn and the factors that determine effectiveness of the learning experience that can contribute to overall learning outcomes?" In Campaign Marketing, the main objective is to attract more students into the institution and build reputation. This includes areas like identifying who is more likely to select courses in the institution and who is more likely to return for life long learning experience. When dealing with Enrolment Management, the most important objective from SRM point of view is to increase the retention rate. This will reflect that the institution has good academic program and eventually increase revenue and reputation.

In a SRM data mining model, the first step is to collect knowledge from the past and existing students so that their wants and needs are known. This first stage of the SRM model is to perform segmentation in the large pool of student profiles within the data warehouse. This data warehouse normally contains information about the student, information gathered via surveys, information from the alumni regarding the graduates, and information about graduates continuing their life long training by revisiting courses from the institution. In this paper, the details of preparing this data warehouse are not presented.

After segmentation of the large data warehouse, a few smaller sub sets will be generated. Hopefully, after the initial segmentation, the sub sets data pool will be more homogeneous within each segment, which is more suitable for target analysis. A human interaction and analysis is required at this stage to identify which sub data sets could be used for different purposes. Normally, in this step, the human analyst will also take away any input variables that are irrelevant to the objectives. For example, if one wants to know how well a student learns, one will not need to include input variables such as Post Codes, Address etc.

After the human selection step of determining which sub data sets are suitable for the different purposes, the second phrase of the knowledge discovery is then carried out. The objective of this phrase of the data mining process is to find answers to questions that are required to understand the issues involved in Institutional Effectiveness, Campaign Marketing and Enrolment Management.

In each sub data set, further segmentation is carried out and score is given to each individual. The score is used to indicate the likelihood of a student who will exhibit a particular behaviour or to indicate how close the behaviour of a particular student with respect to the majority of the students in that segment. This further segmentation has the purpose of subdividing the population in the sub data sets to facilitate specific target analysis. For example, the campaign manager may just want to know the behaviour of the student group age between 18 to 25 years living in the south of the city.

The next stage in this SRM data mining model is to extract underlying meaning from the segments discovered in the previous step. Based on the scores, human analyst can then determine whether to include or exclude certain students when extracting the meaning. After this stage, the underlying meaning has two purposes: it can be used to understand the relationship of the variables in the segment and it can be used to act as a predictive model for any future unknown data. With this descriptive or predictive model, the human analyst can then use it to enhance their objectives in the SRM strategy.

IV. ARTIFICIAL NEURAL NETWORKS

In the last decade, Artificial Neural Networks (ANN) have emerged as an useful option for inferential data analysis and solving complex data analysis problem [12]. The observation sample that is used to derive the predictive model is known as training data in ANN development. The independent variables, or the predictor variables, are known as the input variables and the dependent variables, or the responses, are known as the output variables.

In supervised learning [13], an ANN makes use of the input variables and their corresponding output variables to learn the relationship between them. Once the relationship is found, the trained ANN is then used to predict new output variables given new input data set.

For unsupervised learning [13], an ANN will only make use of the input variables and attempts to arrange them in a way that is meaningful to the analyst. Self-organising Map (SOM) is a popular unsupervised neural network technique mainly because it is a fast, easy and reliable unsupervised clustering technique [14]. SOM is designed to simulate the organisation found in various brain structures and is related to brain maps. Its main feature is the ability to visualise high dimensional input spaces onto a smaller dimensional display, usually two-dimensional as shown in Figure 1. In this discussion, only two-dimensional arrays will be of interest. Let the input data space $\hat{A}^n$ be mapped by the SOM onto a two-dimensional array with i nodes. Associated with each i node is a parametric reference vector $m_i=[m_{i1},m_{i2},\ldots,m_{in}]^T$ Î $\hat{A}^n$, where m_{ij} is the connection weight between node i and input j. Therefore, the input data space $\hat{A}^n$ consisting of input vectors $X=[x_1,x_2,..,x_n]^T$, i.e. X Î $\hat{A}^n$, can be visualized as being connected to all nodes in parallel via a scalar weight m_{ij}. The aim of the learning is to map all the n input vectors X_n onto m_i by adjusting weights m_{ij} such that the SOM gives the best match response locations.

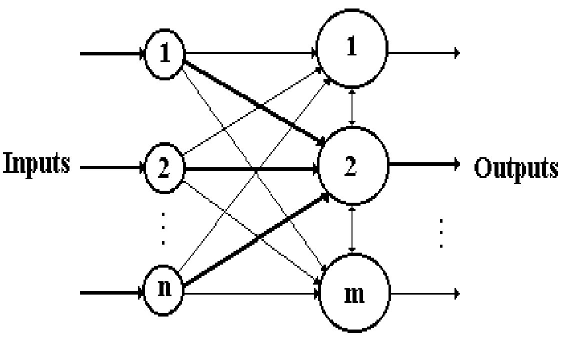

Figure 1(a): Self Organising Map

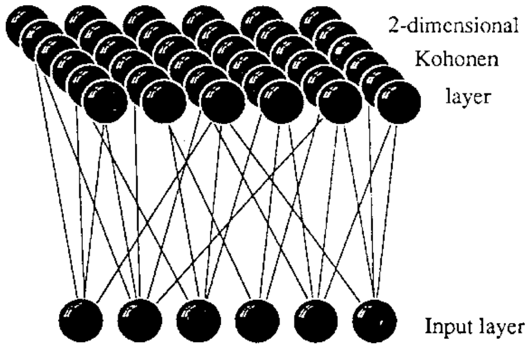

Figure 1(b): Visualisation for Self Organising Map

SOM can also be said that it is a nonlinear projection of the probability density function $p(X)$ of the high dimensional input vector space onto the two-dimensional display map. Normally, to find the best matching node i, the input vector X is compared to all reference vector m_i by searching for the smallest Euclidean distance:

$\| X - m_i \|$, indexed by c,

that is, $\| x - m_c \| = \min_i \| x - m_i \|$.

During the learning process, the node that best matches the input vector X is allowed to learn. Those nodes that are close to the node up to a certain distance will also be allowed to learn. The learning process is expressed as:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[X(t) - m_i(t)] \qquad (1)$$

where t is a discrete time coordinate, and $h_{ci}(t)$ is the neighbourhood function.

After the learning process has converged, the map will display the probability density function $p(X)$ that best describes all the input vectors. At the end of the learning process, an average quantisation error of the map will be generated to indicate how well the map matches the entire input vectors X_n. The average quantisation error is defined as:

$$E = \int \| X - m_c \|^2 \; p(X)dX \qquad (2)$$

ANN analysis is quite similar to statistical approaches in that both have learning algorithms to help them realise the data analysis model. However, an ANN has the advantage of being robust with the ability to handle large amounts of data. Novice users can also easily understand the use of an ANN.

V. FUZZY THEORY

Fuzzy theory works on the basis derived from fuzzy sets [15,16]. A fuzzy set allows for the degree of membership of an item in a set to be any real number between 0 and 1, this allows human observations, expressions and expertise to be modelled. The membership function of a fuzzy set A is denoted by:

$$A : X \rightarrow [0,1] \tag{3}$$

Once the fuzzy sets have been defined, it is possible to use them in constructing rules for fuzzy expert systems and in performing fuzzy inference. Fuzzy systems can produce more accurate results based on the basic idea of defuzzification. A defuzzification technique is used to calculate the conclusion by evaluating the degree of matches from the observation that triggered one or several rules in the model. This will lead to a better result by handling the fuzziness in the decision making process. Thus, fuzzy technique can improve on statistical prediction in certain cases.

Fuzzy sets allow human expertise and decisions to be modelled more closely. It is suggested that Fuzzy sets will play an important role in the SRM model. In this paper, with the availability of vast amounts of data in each sub set, it will be useful to extract knowledge from the data directly. This has the advantage of discovery of new knowledge or relations underlying the data. In extracting fuzzy rules from the data, the first step is to translate all the available data into linguistic fuzzy rules using linguistic labels. The following algorithm outlines the steps in extracting the fuzzy linguistic rules from the available data. For k inputs, the given input-output data pairs with n patterns:

$$\begin{gathered} (x_1^1, \ldots, x_k^1; y^1) \\ (x_1^2, \ldots, x_k^2; y^2) \\ \vdots \\ (x_1^n, \ldots, x_k^n; y^n) \end{gathered}$$

The number of linguistics terms T and the distribution of data in the regions of the whole domain are first determined. For ease of interpretation and computational simplicity, the shape of the membership function used in this algorithm is triangular. In this case, we will obtain for every $x \in X$,

$$A_t \in F(x) \rightarrow [0,1] \qquad \text{for all } t \in T \tag{4}$$

After the fuzzy regions and membership functions have been set up, the available data set will be mapped accordingly. If the value cuts on more than one membership function, the one with the maximum membership grade will be assigned to the value:

$$R_n \Rightarrow [x_1^n(A_{t1}, \max), \ldots, x_k^n(A_{tk}, \max) : y^n(B_t, \max)] \tag{5}$$

After all the data sets have been assigned with a fuzzy linguistic label, Mamdani type fuzzy rules are then formed and centroid defuzzification is used. After fuzzy rules have been generated from each data point, repeated rules are removed. In the event that there are repeated fuzzy rules, the number of repetitions of the fuzzy rules and the firing strengths of the rules will be examined to resolve conflicts.

Besides using fuzzy theory in the data mining process of the SRM model, fuzzy clustering can also be used. Fuzzy clustering provides a more precise measure to the institution in delivering value to the student. Given a set of data, clustering techniques partition the data into several groups such that the degree of association is strong within one group and weak for data in different groups. Classical clustering techniques result in crisp partitions where each data can belong to only one partition. Fuzzy clustering extends this idea to allow data to belong to more than one group. The resulting partitions are therefore fuzzy partitions. Each cluster is associated with a membership function that expresses the degree to which individual data belongs to the cluster. Technique such as Fuzzy C-Means (FCM) clustering is very reliable and popular in performing fuzzy clustering [17].

Given a set of data, FCM clustering iteratively searches for a set of fuzzy partitions and the associated cluster centres that represent the structure of the data. The FCM clustering algorithm relies on the user to specify the number of clusters present in the set of data to be clustered. Given the number of cluster c, FCM clustering partitions the data $X = \{x_1, x_2, \ldots, x_n\}$ into c fuzzy partitions by minimizing the within group sum of squared error objective function using the following equation:

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c}(U_{ik})^m \, \| x_k - v_i \|^2, \qquad 1 \le m \le \infty \tag{6}$$

where $J_m(U,V)$ is the sum of squared error for the set of fuzzy clusters represented by the membership matrix U, and the associated set of cluster centres V. $\|.\|$ is some inner product-induced norm. In the formula, $\|x_k - v_i\|^2$ represents the distance between the data x_k and the cluster centre v_i. The squared error is used as a performance index that measures the weighted sum of distances between cluster centres and elements in the corresponding fuzzy clusters. The number m governs the influence of membership grades in the performance index. The partition becomes fuzzier with increasing m and it is proven that the FCM clustering converges for any $m \in (1,\infty)$.

VI. INTELLIGENT DATA MINING

Figure 2 shows the block diagram of the data mining process described in Section III together with the intelligent techniques used in each stage. This SRM process is divided into two main phases. In the first phase, the main purpose is to break the data warehouse into smaller segments such that the nature of the data is more homogeneous within each segment. The algorithm of using SOM is as follows:

Step 1. Input all the available data from the data warehouse to SOM in identifying some meaningful clusters:-

- In this case, only two-dimensional map is used.
- When determining the number of clusters, the average quantisation error is observed. The number of output neuron is increased in steps of 1. When the difference between the two quantisation errors is small, the previous step is used as the final number of clusters.

Step 2. After the clusters are identified from the whole data warehouse, the huge database is then separated into smaller modules based on the clusters obtained from SOM. Each of these modules is known as a segment after this step.

In the second phase, depending on the number of objectives of the knowledge discovery, the number of data mining models required is directly related to the number of objectives required. In this paper, it is three: one for Institutional Effectiveness, one for Campaign Marketing and another one for Enrolment Management. Depending on the situation, sometime it could be more than one data mining model under one objective to answer multiple related questions. For example, under Campaign Marketing, the user could have asked the following two questions:

1. Who is more likely to continue their life long learning experience?
2. Who is more likely to be affected by the priority areas set by the government when selecting a course to study?

In this case, under the Campaign Marketing module, two data mining models need to be constructed based on the requirements needed to answer the above questions. The input variables selected in each model have to be decided by the human analyst. Each of the intelligent data mining model used in answering the above questions is then constructed as follows:

1. Fuzzy C-means algorithm is used in each cluster to perform scoring or ranking. In this paper, we used three cluster groups for each major segment identified by the SOM. They are named as "highly true", "moderately true", and "not so true".
2. The fuzzy membership grade acts as a kind of fuzzy score for a particular student belonging to that fuzzy cluster.
3. The alpha-cut that is selected to determine whether a particular student belongs to the fuzzy cluster is set at 0.2. In this case, students who fall in the fuzzy region could be handled and examined in both clusters with membership grades attached to it. The human analyst can then decide whether anything needs to be done to those students.
4. The last stage is to create a predictive model such that when new or potential student information is available, the data mining model can confidently tell the human analyst which category they are in:-

    - Using the fuzzy rule extraction algorithm to extract the fuzzy rule base that best describes the relationship between the input variables and the fuzzy scores are carried out.
    - In this stage, it also acts as a knowledge discovery technique to allow human analyst in gaining in depth understanding of the group of students for the objectives.

The human analyst can then make use of the information generated by the above data mining models to find answers to any questions put forward. Appropriate SRM strategy and policy can then be outlined to improve the management and marketing strategies.

VII. Conclusion

In this paper, a SRM model has been proposed. Areas in this SRM model that can incorporate the use of intelligent data mining techniques, in particular, ANN and fuzzy theory have also been examined. The interaction between the intelligent data mining and the human analyst has been examined. This proposed intelligent data mining model allows the human analyst to generate answers to the objectives in an semi-automatic fashion. In our future work, examination of the above model with intelligent data mining techniques will be implemented in a simulated environment.

VIII. Acknowledgement

The authors thank the supports provided to this research by the Murdoch University Early Career Research Grant.

References

[1] M. Ward. Expectations are High for CRM, but so are the Challenges, *Information Systems Control Journal*, vol. 4, pp. 41-44, 2001.

[2] S.J. Spanbauer. A Quality System for Education: Using Quality and Productivity Techniques to Save Our Schools, *ASQC Quality Press*, 1994.

[3] G. Linoff. Data Mining, *Inform*, vol. 13(9), pp. 18-24, 1999.

[4] J.P. Bigus. *Data Mining with Neural Networks*, New York, McGraw Hill, 1996.

[5] K.J. Cherkauer, and J.W. Shavlik. Growing Simpler Decision Trees to Facilitate Knowledge Discovery, *Proceedings of Second International Conference on Knowledge Discovery and Data Mining*, pp. 316-318, 1996.

[6] P.D. Scott, A.P.M. Coxon, M.H. Hobbs, and R.J. Williams. An Intelligent Assistant for Exploratory Data Analysis, *Principles of Data Mining and*

Knowledge Discovery, Lecture Notes in Computer Science number 1263, pp.11, Springer Verlag, 1997.

[7] T. Hastie, and R. Tibshirani. Discriminant adaptive nearest neighbour classification, *KDD-95: Proceedings First International Conference on Knowledge Discovery and Data Mining*, 1995.

[8] R. Agrawal, and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases, *Proceedings of International Conference on Very Large Data Bases*, pp. 478-499, 1994.

[9] A. Berson, S. Smith, and K. Thearling. *Building Data Mining Applications for CRM*, McGraw-Hill, pp. 123-199, 2000.

[10] K. Thearling. Increasing Customer Value by Integrating Data Mining and Campaign Management Software, *Direct Marketing Magazine*, February 1999.

[11] J. Luan. Data Mining and Knowledge Management in Higher Education – Potential Applications, *Workshop and Presentation at 42nd Associate of Institutional Research International Conference*, Toronto, Canada, 2002.

[12] J.H. Friedman. An Overview of Predictive Learning and Function Approximation, in *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*, Springer-Verlag, pp. 1-61, 1994.

[13] S.V. Kartalopulos. *Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications*, IEEE Press, 1996.

[14] T. Kohonen. *Self-Organising Map*, Springer-Verlag, 1995.

[15] L.A. Zadeh. Outline a new approach to the analysis of complex systems, *IEEE Trans. on Sys. Man and Cybernetics*, vol. 1, pp. 28-44, 1973.

[16] G.J. Klir, and B. Yuan. *Fuzzy Sets and Fuzzy Logic*, New Jersey, Prentice Hall, 1995.

[17] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithm*, New York, Plenum Press, 1981.

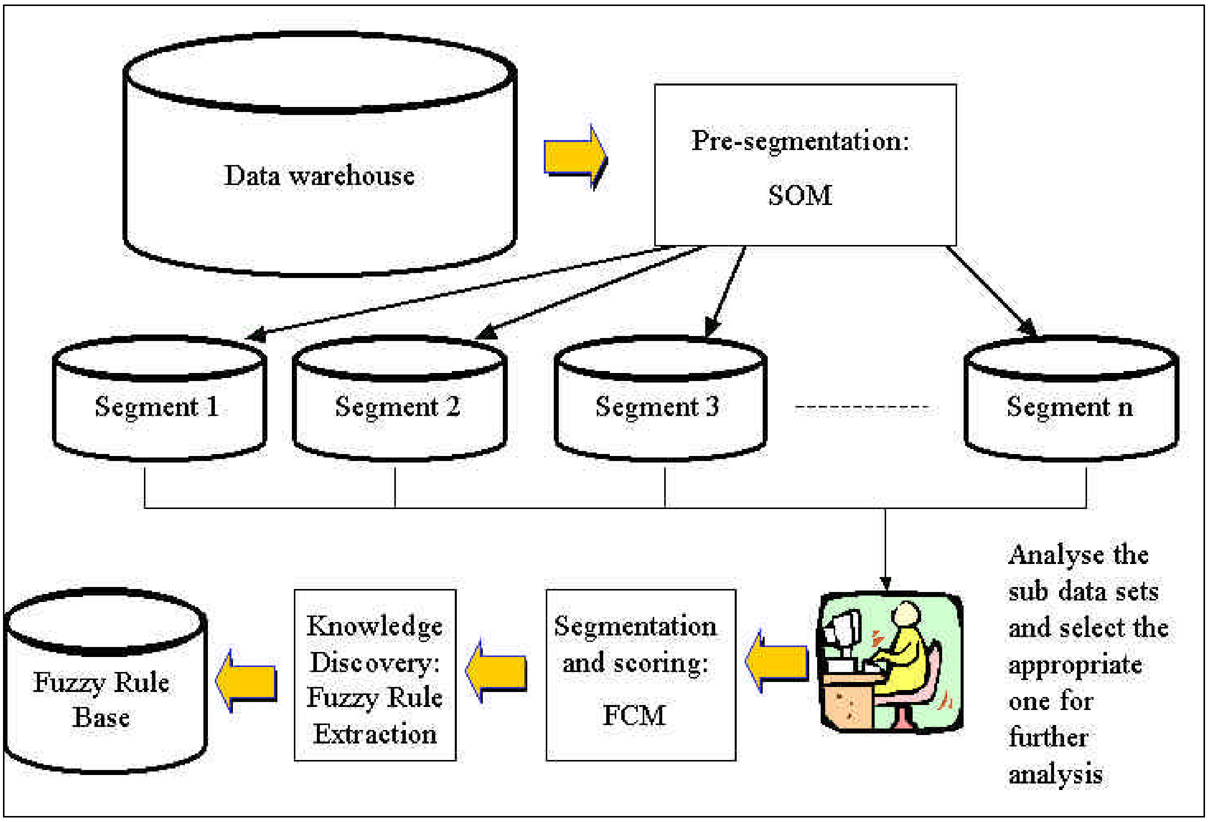

Figure 2: Block diagram of the proposed intelligent SRM model