Feature Selection Through Dm in Electronic Commerce

Dr. Dinesh Kumar Singh

Assistant Professor, Department of Information Technology, DSMNRU Lucknow
Email: dineshsingh025@gmail.com

Abstract

This study article employs several data mining approaches to assess the feature selection of distinct factors. The data mining methods encompass RBFN, Artificial Neural Network, Decision tree, and others. The multilayer perceptron (MLP) is the most used neural network model. This neural network is categorized as a supervised network since it necessitates a desired output for the purpose of learning. A decision tree is a graphical representation that illustrates the logical steps of the decision-making process. The Radial Basis Function Network (RBFN) employs a methodology akin to k-means clustering to divide the data into segments based on the values of the target field.

Keywords: *ANN, DT, DT-ANN, RBFM, SA, Negotiation*

I. Introduction:

This study paper explores the utilization of different data mining techniques, including Artificial Neural Networks (ANN) and Decision Trees, to determine the significance and selection of features.Data mining is commonly characterized as the process of exploring, analyzing, and sifting through vast quantities of data in order to discover linkages, patterns, or any noteworthy statistical correlations. Thanks to the emergence of computers, extensive databases, and the internet, it has become increasingly convenient to amass vast amounts of data, ranging from millions to trillions of data points. This data may then be methodically evaluated to identify correlations and find answers to complex issues. In addition to its governmental applications, data mining is often employed by marketers to identify robust consumer trends and linkages. Large corporations and educational institutions also engage in data mining to identify meaningful relationships that can improve our society. A neural network is an effective data modeling tool that can capture and depict intricate input/output relationships. The impetus for the advancement of neural network technology arose from the aspiration to create an artificial system capable of executing "intelligent" tasks akin to those carried out by the human brain. The multilayer perceptron (MLP) is the most used neural network model. This neural network is categorized as a supervised network as it necessitates a desired output for the purpose of learning. The objective of this network is to construct a model that accurately links the input to the output based on past data. This model may then be utilized to generate the output in situations when the desired outcome is not known. A decision tree is a graphical representation that illustrates the logical steps involved in making a choice. It serves the purpose of providing an explanation for the reason behind asking a question. The decision tree operates on the assumption that questions can only be answered with a definitive "yes" or "no". A tree that permits responding with a partial affirmation or negation would possess a significantly greater quantity of terminal nodes. When dealing with real-world situations, the expertise of a human expert or expert system software is essential in order to identify the most probable outcome. Every terminal node symbolizes a scenario characterized by well-established and productive leadership approaches.

II. ANN Model implementation Using Radial Basis Function Network

The Radial Basis Function Network (RBFN) employs a methodology akin to k-means clustering to divide the data into segments based on the values of the target field. The RBFN method utilizes 23 variables as input, consisting of 1 boolean parameter, 1 integer parameter, and 21 categorical parameters. Out of the 21 parameters, 3 parameters have four stages, 6 parameters have just two stages, and the remaining 12 parameters have three phases. The input layer of the Artificial Neural Network (ANN) consists of 1 neuron for each boolean variable, 1 neuron for each integer variable, 6 neurons for each categorical variable with two stages, and 3 neurons for each of the 12 category variables with three stages. Additionally, there are 4 neurons for the 3 parameters. Hence, the total number of neurons in the input layer of the Artificial Neural Network (ANN) is 60, calculated as the sum of 1, 1, 4 multiplied by 3, 14 multiplied by 3, and 4. The ANN consists of a single hidden layer comprising twenty neurons, along with an output layer including four neurons. Out of the total of 90 test cases in this model, 85 instances are accurately classified while 5 cases are

inaccurately classified. The final accuracy of the mining results obtained using the artificial neural network model is 94.44%, with an error rate of 5.56%.

III.   Feature Selection through DM

We utilize the feature selection node of Clementine 11.1 software in this model. The feature selection node facilitates the identification of the most crucial fields for predicting a specific outcome. The feature selection node evaluates and prioritizes a large number of predictors to identify the ones that are perhaps the most significant. In the end, it could result in a faster and more efficient model, which utilizes fewer predictors, runs more swiftly, and is potentially easier to comprehend. Table 1 displays the outcomes of feature selection.

		Rank	Field	Type	Importance	Value
1	true	1	**DESIRE (X8)**	set	**Important**	1.0
2	true	2	**INTENTION (X9)**	set	**Important**	1.0
3	true	3	**BELIEF (X7)**	set	**Important**	1.0
4	true	4	**DELIVERY TIME(X19)**	set	**Important**	1.0
5	true	5	**ADDRESS MODE(X20)**	set	**Important**	1.0
6	true	6	**REPUTATION(X13)**	set	**Important**	1.0
7	true	7	**CAPABILITY(X12)**	set	**Important**	1.0
8	true	8	**COMMITMENT(X11)**	set	**Important**	1.0
9	true	9	**RELIABILITY(X14)**	set	**Important**	1.0
10	true	10	**PAYMENT TYPE(X17)**	flag	**Important**	1.0
11	true	11	**QUALITY (X3)**	flag	**Important**	1.0
12	true	12	**START TIME(X15)**	flag	**Important**	1.0
13	true	13	**RESPONSE TO MARKETING STIMULI (X23)**	flag	**Important**	1.0
14	true	14	**PREFERENCE (X10)**	set	**Important**	1.0
15	true	15	**BRAND (X5)**	flag	**Important**	1.0
16	true	16	**SERVICE TIME (X6)**	range	**Important**	1.0
17	true	17	**PROFIT (X2)**	set	**Important**	0.999
18	true	18	**PERFORMANCE (X1)**	set	**Important**	0.999
19	true	19	**PAYMENT MODE(X16)**	set	**Important**	0.999
20	true	20	**LOYALTY(X18)**	set	**Marginal**	0.935
21	true	21	**SELF MONITORING(X22)**	set	**Marginal**	0.935
22	true	22	**QUANTITY (X4)**	set	**Unimportant**	0.894
23	true	23	**SUPPORT(X21)**	set	**Unimportant**	0.833

IV.   Conclusion:

The Ann model has been implemented using Radial Basis Function Networks (RBFN) in this case. The RBFN employs methodologies akin to k-means clustering to divide the data into segments based on the values of the target field. In the k-means algorithm, it is not necessary to save all of the training data, allowing for online implementation. The feature selection node is used in sensitivity analysis to discover the key variables that have the greatest impact on a specific outcome. It could result in a faster and more efficient approach that utilizes fewer predictors, has faster execution, and is potentially easier to comprehend. We have endeavored to emphasize the significance of qualities in negotiation through the application of data mining techniques. The decision tree method serves as a visually intuitive and comprehensible alternative to the numerical charts and statistical probabilities employed in other decision analysis techniques, such as spreadsheets. Decision trees are flexible, allowing for modifications as new decisions arise or when new information is obtained, leading to changes in the scenarios.

Reference

[1].   Carter, J., Bitting, E. & Ghorbani, A. (2002). Reputation Formalization for an Information-Sharing Multi-Agent Sytem. Computational Intelligence 18(2): pp. 515- 534.

[2].   Chee, C. N. (2004). Three Critical Steps to Customer-centric Business Orientation. Meta Core Asia Pte Ltd. Retrieved October 2004, from www.metacore-asia.biz.

[3].   Faratin, P., Sierra, C. and Jennings, N., (2003). Using Similarity Criteria to Make Issue Trade-offs in Automated Negotiations. Journal of Artificial Intel. vol. 142 (2), pp. 205-237.

[4].   Kang, N., Han, S., (2002). Agent-based e-marketplace system for more fair and efficient transaction, Decision Support Systems 34 pp. 157– 165.

[5].   Russell S. Winer, (2001). Customer Relationship Management: A Framework, Research Directions, and the Future.

[6].   Sim, K. M. & Chan, R. (November 2000). A Brokering Protocol for Agent-Based E-Commerce, IEEE Transactions on Systems, Man, and Cybernetics part c: applications and reviews, vol.30, no. 4.