IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB12608

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2* edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

Hybrid Rough/Fuzzy Modeling of Advertising Effects on Consumer Preferences

Ashwani Kumar, Assistant Professor, Indian Institute of Management, Lucknow, India,
P: 91-522-2736647, F: 91-522-2734025, ashwanti@timl.ac.in,

ABSTRACT

Advances in computational methods have led, in the marketing domain, to huge databases of consumer and marketing information. In the past decade, various machine intelligence techniques have been applied in mining this data for obtaining knowledge and in-depth information about the consumers and the markets. This paper presents a hybrid rough/fuzzy model to discover knowledge on advertising effectiveness from time series databases on consumer preferences and advertising expenditures. The data set used for application example contains weekly investments in different media categories: TV, radio, cinema, morning press, evening press, popular press, special interest press, and outdoor posters; for seven makes of cars on Swedish market, and consumer awareness and preference indices.

HYBRID ROUGH/FUZZY MODELING

In the rough set/ fuzzy set approach, collection of if-then rules constitutes a model. In the review paper by Guillaume [1], soft computing methods have been compared and analyzed according to interpretability. What are the necessary conditions for a set of induced rules to be interpretable? First, the fuzzy/rough partitions must be readable, in the sense that fuzzy/rough sets can be interpreted as linguistic labels. These labels must be meaningful for experts of the problem under study, so as to allow the rules to be compared to each other, and to lead to knowledge discovery. Second, the set of rules must be as small as possible. The reduction of a set of rules may result in some loss of numerical performance, but a more compact set has a better interpretability. For large systems, a third condition is required: the rules should be incomplete rules. If the rule premises involve the whole set of variables, there is a loss of interpretability without a corresponding gain of performance, when the rules context can be defined by a subset of available variables only. The systematic presence of all variables in all rules can be considered as a drawback of most automatic rule induction methods, due to the techniques themselves. It is not an intrinsic characteristic of the problem.

Two kinds of fuzzy inference systems (FIS) are in use as rule-based models. The first kind of FIS – the Mamdani model [2], designed from expert knowledge, from data, or from both, possesses high transparency and explainability features. In the second kind – the Takagi-Sugeno model [3], designed from data, there is generally a loss of semantic. Both of these methods can be considered as rule generation techniques[1] comprising two main steps : 1) rule induction, and 2) rule-base optimization. Obviously, the rules can be easier to interpret if they are defined by the most influential variables, and the system behavior will be easier to understand as the number of rules are getting smaller. Variable selection and rule reduction through clustering of data are, thus, two important steps of the rule-optimization process. They are usually referred to as *structure optimization*. Apart from structure optimization, a FIS has many parameters that can also be optimized, *e.g.*, membership function parameters. Neuro-fuzzy models [4], have parameter optimization features: neural networks bring their learning algorithm and numerical accuracy to FIS; without paying much attention to the semantic. These tools prove useful for applications for which semantic is not a prime concern. The backpropagation tuning algorithms may generate an unreadable fuzzy partitioning. The only approach that deals with the third interpretability condition is the fuzzy decision tree[1]. They offer a compact description of a given context by using only the locally most significant variables. The incomplete rules generated by decision trees will be informative for experts provided that the initial partitioning was carefully defined [5,6].

Rough set theory, introduced by Pawlak in the early 1980s [7,8], is a new mathematical tool to deal with vagueness and uncertainty. Rough set theory does not compete with fuzzy set theory, with which it is frequently contrasted, but rather complements it [7]. Rough set theory and fuzzy set theory are independent approaches to imperfect knowledge. In the marketing domain applications, human-computer cooperation requirements dictate rule interpretability as the major criterion for modeling. With this in mind, we focus on rough set data analysis (RSDA), and Mamdani fuzzy models with structure optimization but without parameter optimization.

Models from volumes of data are typically constructed in the framework of rough sets and fuzzy sets using the following steps [9]:

- *Selection*: The study population is defined, a set of features is defined, and a modeling target is chosen. Together, this corresponds to assembling a decision system from a data source.
- *Preprocessing*: It covers the issue of data cleaning, which is generally application specific. Removing obvious outliers is one common preprocessing task. Another common task is that of completion, *i.e.*, the process of converting a decision system with missing values into a new system without any such "holes" [10].
- *Transformation*: The most common transformation procedure in rough sets based knowledge discovery in databases is that of discretization, which basically corresponds to defining a coarser view of the world through making the attributes' value sets smaller. For numerical attributes, we can introduce intervals which in turn may be given linguistic labels and can be treated as qualitative rather than quantitative entities. Several alternative discretization functions are currently implemented including methods based on discernibility preservation, entropy minimization, equal frequency binning and various naïve approaches [11]. Note that discretization may or may not be needed to carry out.

In fuzzy sets based knowledge discovery in databases, the coarser view of the world is obtained through fuzzification of numerical attributes. [12].

- *Data mining*: The purpose of the data mining step is to produce a model from our preprocessed and transformed database. Since we are looking for relationships that are presumably unknown beforehand, it might not be entirely straightforward to *a priori* specify what constitutes potentially valuable inputs. Letting a method loose on data material that has not been sufficiently

> "focused" by narrowing down the number of inputs may not only increase the computational burden in the data mining step, but is also likely to result in spurious and nonsensical correlations being found. A pre-requisite of a 'good' model construction is therefore feature selection step.

The issues of feature selection have been studied by statisticians and by machine-learning specialists for decades and many interesting techniques have been developed. The fuzzy-neural network proposed by Li *et.al.*[13] attempts to select important features from among the originally given plausible features; value of weight connection in trained network represents the degree of importance of a feature. Lin and Cunningham [14] introduce "fuzzy curves" and use them to identify input variables. These approaches, based on fuzzy inference systems, are in fact variants of sensitivity-based pruning for input variables selection studied by many researchers using neural networks [15-17]. Principal component analysis (PCA) is another classical technique to remove correlations between variables. [18]. In rough set data analysis (RSDA), entire data are classified into a small set of independent information granules, and effects of an input variable on all information granules are examined [19]. It is a semantics-preserving data analysis method, and does not suffer from the limitations of sensitivity-based pruning. The new information or decision system constructed from data after deleting insignificant condition variables leads to a rule-based model with optimum set of rules.The availability of PC-based commercial software systems for database mining has made this technology accessible to users [20, 21]. For a detailed review on the theory, readers may refer to Komorowski *et al.* [22] or Ashwani Kumar *et al.* [23].

An alternative to RSDA is fuzzy set approach for inducing rules. Users can control model reduction by varying the degree of coarseness of fuzzification of both the condition and decision variables independently (in RSDA, discretization of condition variables is constrained by that of decision variable). This gives a better control on rule-based model reduction, thereby giving better interpretability and implementation in expert systems. Defuzzification is the technique employed to realize classifiers/predictors from the ensembles of decision rules [12].

In our study, we obtain the structure of the model by means of Wang and Mendal [12] clustering technique and therefrom extract a set of relevant rules based on the input-output data samples. In our hybrid approach, we use RSDA for feature selection and fuzzy set approach for rule-based model construction.

EMPIRICAL RESULTS

The data used for the application example contains weakly investments in different media categories – TV, radio, cinema, morning press, evening press, popular press, special interest press, outdoor posters; for seven makes of cars on Swedish market.

Every week a number of individuals are interviewed to find out if they have seen and remember advertisements. From these interviews, the following percentages are produced for each make [17]:

- Top of Mind (TOM); the make is the first mentioned by the interviewee
- In Mind (IM); the interviewee mentions the make
- Preference (PREF); the make is the preferred choice
- Possibility (POSS); the make is a possible choice for the interviewee

Johansson and Niklasson [17] performed sensitivity analysis to find out the relative importance of the media categories. The sensitivity estimates produced by this procedure are however very approximate; variations in neural network weights have been ignored in computing the sensitivity measure. In RSDA, entire data are classified into a small set of independent information granules, and effects of input variable on all information granules are examined. Granules then induce variation on the decision variable.

Table 1. Significance factors of media on TOM

Brand	T.V.	Radio	Movie	Outdoor	Morning Press	Evening Press	Popular Press	Special Press
G	0.419355	0	0	0.090323	0.63871	0.212903	0.290323	0.16129
F	0.496815	0	0	0	0.738854	0.363057	0.165605	0.152866
E	0.261146	0	0	0	0.751592	0.471338	0.197452	0.261146
D	0.116129	0	0	0	0.948387	0.393548	0.083871	0.277419
C	0.825806	0	0	0	0.748387	0.632258	0	0
B	0.235669	0	0	0	0.719745	0.579618	0.076433	0.197452
A	0.235669	0	0	0	0.719745	0.579618	0.076433	0.197452

Table 2. Significance factors of media on IM

Brand	T.V.	Radio	Movie	Outdoor	Morning Press	Evening Press	Popular Press	Special Press
G	0.387097	0	0	0.058065	0.645161	0.187097	0.264516	0.141935
F	0.33121	0.050955	0	0	0.687898	0.305732	0.267516	0.095541
E	0.264516	0	0	0	0.722581	0.580645	0.16129	0.264516
D	0.141935	0	0	0	0.954839	0.341935	0.103226	0.225806
C	0.851613	0	0	0	0.858065	0.683871	0	0
B	0.197452	0	0	0.012739	0.713376	0.56051	0.146497	0.22293
A	0.229299	0	0	0	0.936306	0.375796	0.070064	0.101911

We have performed RSDA on the subset of the data (7 brands instead of 17) with the brand names hidden. The results are given in Tables 1 and 2 (Comparison with the results of Johansson and Niklasson [17] could not be made because of hidden brand names in the data available with us). The overall conclusions are: the most effective media were TV and morning press. For some makes, evening press/popular press/special press were effective. Radio, movie and outdoor advertisements were, in general, ineffective.

Detailed analysis of data for one brand (F) with TOM as the output variable was carried out using RSES (Rough Set Exploration System) software (http://logic.mimuw.edu.pl/~rses/). For three levels of output discretization, the following reduct was obtained:

{TV, Morning Press, Evening Press}

Number of cuts (levels) generated by RSES for TV variable= 8
Number of cuts generated by RSES for Morning Press variable = 20
Number of cuts generated by RSES for Evening Press variable = 4

The large number of levels for input variables given by RSES, for three levels of output (decided by the user) results in large number of rules; 125 in our example from 157 data points.

Analysis using fuzzy sets was done with trial values of K_i and K_0 taken as 3,5 and 7. Three fuzzy subsets are linguistically interpreted as {low, medium, high}, and five subsets interpreted as {very low, low, medium, high, very high}. Interpretation becomes poorer as the number of subsets becomes 7 or higher; however, the mean square error performance is expected to improve with increase in the number of fuzzy subsets. Input variable selection dictated by the reduct is {TV, Morning Press, Evening Press}. It only takes a single, noisy object to change the indiscernibility relation; reducts, therefore, are prone to incorporate noise and other peculiarities in the data set. Since our data is polluted with noise and other imperfections, we do a little fine tuning on the reduct. From Table 1, we observe that for the TOM variable of F-brand, {TV, Morning Press, Evening Press, Popular Press, Special Press} could be considered as significant. In our simulations, we have allowed a variation in condition variables: 3 to 5. Also a large variation was allowed in the number of lags of the condition variables for creating a FIS prediction model: from 1 to 7.

75% of the data was used for training and the remaining 25% for testing.

Large number of simulations were carried out. The best model of our empirical study (not necessarily the optimal) has the following parameters:

Figure 1. Comparison of predicted and actual values (test data) for three legs

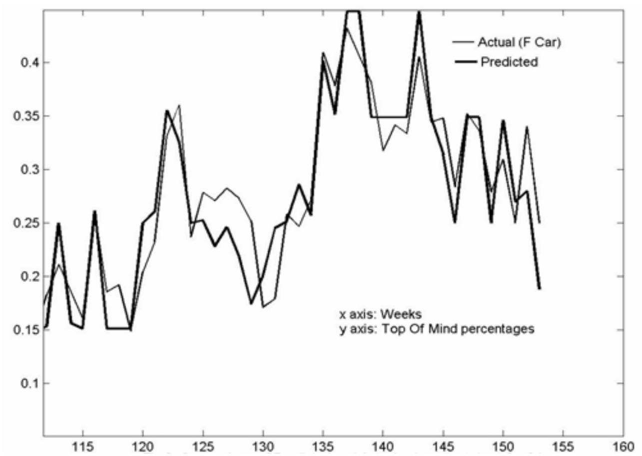
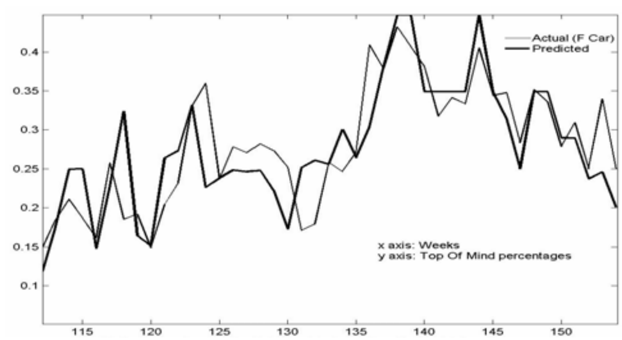


Figure 2. Comparison of predicted and actual values (best data) for one leg



Condition variables = 3

No. of lags = 3

$K_i = 5$

$K_0 = 5$

Mean square error (MSE) on test data = 0.0012

Number of fuzzy rules = 113

Simulation results on test data are shown in Fig.1.

To get better interpretability, we sacrifice a little the mean square error performance. For condition variables {TV, Morning Press, Evening Press}, number of lags = 1, K_i = 5 and K_0=5, the FIS predictor has $29´10^{-4}$ MSE on test data. Number of fuzzy rules = 54. Simulation results on test data are shown in Fig.2.

We propose the following simple strategy for generating incomplete fuzzy rules. Using "MIN" fuzzy operator, firing strength of rule r(r=1,2, …, R) may be defined as

$$\mu_{rf}^{(p)} = \underset{(i, j_i)}{MIN} \left\{ \mu_{ij_i}(x_i^{(p)}) \right\}; \begin{array}{l} i = 1,2,...,n \\ j_i = 1,2,...,K_i \\ p = 1,2,...,P \end{array} \tag{1}$$

The linkage of the IF-part fuzzy sets and the THEN-part fuzzy sets in a rule r is given by the degree of rule r defined as

Table 3.

Rule No.	TV(t)	TV(t-1)	MP(t)	MP(t-1)	EP(t)	EP(t-1)	
1	5	0	0	0	0	0	5 : very high
2	3	5	3	5	0	0	4 : high
3	2	2	2	2	0	0	3 : medium
4	1	2	2	3	1	1	2 : low
5	3	5	4	0	0	0	1 : very low
6	1	3	3	4	0	0	

$$\mu_{rd}^{(p)} = \underset{(i, j_i, k)}{MIN} \left\{ \mu_{ij_i}(x_i^{(p)}), \mu_k(y^{(p)}) \right\}; k = 1,2,..., K_0 \tag{2}$$

Certainty Factor (CF) of a rule r when inferring output values:

$$CF_r = \frac{\sum_{p=1}^{P} \mu_{rd}}{\sum_{p=1}^{P} \mu_{rf}} \tag{3}$$

For the case study under consideration, three present and three past values of condition variables yield a 6´1 vector **x**. Certainty factors of rule r for all possible (2^6) combinations of {x_1,x_2, …, x_6} variables are calculated, i.e., CF_r for premise parts consisting of {x_1}, .,{x_6}, {x_1x_2}, …, {$x_1x_2x_3$}, … {$x_1x_2x_3x_4x_5x_6$} are calculated. The combination with highest CF_r is selected to generate an incomplete fuzzy rule. A sample result for 'very high' output TOM is given in Table 3.

Interpretation follows:

1) If we invest large amount of money on TV in the present work, TOM will be very high next week.
2) If we have already spent very high amount in the previous week on TV, then in order to get very high TOM next week, it is required to invest medium amount on TV in the present week along with

a) high investment on MP in the present week if no investment was made on MP in the previous week (rule 5)

OR

b) medium investment on MP in the present week if very high investment was made on MP in the previous week (rule 2)

CONCLUSIONS

In this paper, rough sets and fuzzy sets were used cooperatively during the process of building a marketing decision support system for retail business. In particular, the significance of the input features has been analyzed using rough sets and subsequently rule-based fuzzy inference models based on these features have been built. The rough sets analysis of the input attributes allowed understanding of the relative importance of advertising in different media categories for the effects on consumer preferences. The derived rules allow the advertising system designers to understand the relationships between the various media categories contributing to consumer awareness and preferences.

In order to further improve the modeling of advertising effects on consumer preferences, the following issues need attention:

- We have assumed that advertising is essentially independent across competitors. A multi-state game-theoretic model of advertising competition may be constructed to see the validity of this assumption [26-29]
- Empirical support model may be constructed for a generalization concerning the expected length of advertising carry-over (as estimated from consumer preference response models of different levels of aggregation) [30-31].
- Use of decision trees may be explored to improve interpretability of the model [5, 6].

REFERENCES

[1] S.Guillaume, “Designing fuzzy inference systems from data: An interpretability-oriented review”, *IEEE Trans. Fuzzy Systems*, vol.9, no.3, June 2001.

[2] E.H.Mamdani, and S.Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller”, *Int. J. Man-Mach. Stud.*, vol.7, pp.1-13, 1975.

[3] T.Takagi, and M.Sugeno, “Fuzzy identification of systems and its applications to modeling and control”, *IEEE Trans. Syst., Man, Cybern.*, vol.SMC 15, pp.116-132, 1985.

[4] J-S.R. Jang, C-T. Sun, and E.Mizutani, *Neuro-Fuzzy and Soft Computing*, Englewood Cliffs, NJ: Prentice Hall, 1997.

[5] J.R.Quinlan, “Induction of decision trees”, *Machine Learning*, vol.1, pp.81-106. Aug. 1986.

[6] H.Ichihashi, T.Shirai, K.Nagasaka, and T.Miyoshi, “Neuro-fuzzy ID3: A method of inducing fuzzy decision trees with linear programming for maximizing entropy and an algebraic method of incremental learning”, *Fuzzy Sets Syst.*, vol.81, pp.157-167, 1996.

[7] Z.Pawlak, “Rough sets”, *International Journal of Computer and Information Sciences*, 11(5), 341-356, 1982.

[8] Z.Pawlak, “Rough sets”. *In Theoretical Aspects of Reasoning About Data*, Kluwer, Netherlands, 1991.

[9] U.Fayyad, G.Piatetsky-Shapiro, P.Smyth, and R.Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.

[10] Marzena Kryszkiewicz, “Rules in incomplete information systems”. In Paul P.Wang, editor, *Proc.Third International Joint Conference on Information Sciences*, vol.3, Durham, NC, pp.73-76, March 1997.

[11] Hung Son Nguyen and Sinh Hoa Nguyen, “Discretization methods in data mining”. In Lech Polkowski and Andrzej Skowron, editors, *Rough Sets in Knowledge Discovery 1 : Methodologies and Applications*, vol.18 of Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, Germany, pp.451-482, 1998.

[12] Li-Xin Wang and Jerry M.Mendel, “Generating fuzzy rules by leaning from examples”, *IEEE Trans. Systems, Man, and Cybernetics*, vol.22, no.6 pp.1414-1425, Nov/Dec 1992.

[13] Ashwani Kumar, D.P Agarwal and S.D. Joshi, “Advertisig Data Analysis Using Rough Sets Model ”, Intenrnational Journal of Information Technology and Decision Making, 4 (2), 2005.

[14] Y.Lin, and G.A. Cunningham-III, “A new approach to fuzzy-neural system modeling”, *IEEE Trans.Fuzzy Systems*, vol.3, no.2, May 1995.

[15] J.Utans, J.Moody, S.Rahfuss and H.Siegalmann, ”Input variable selection for neural networks : Application to predicting the U.S.business cycle”, *Proc.1995, IEEE/IAFE Conference on Computational Intelligence for Finance Engineering*, pp.118-122, 1995.

[16] H-L Poh, J.Yao and T.Jasic, “Neural networks for the analysis and forecasting of advertising and promotion impact”, *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7, 253-268, 1998.

[17] U.Johansson, and L.Niklasson, “Increased performance with neural nets – An example from the marketing domain”, *IEEE 2002, International Joint Conference on Neural Networks*, vol.2, pp.1684-1689.

[18] A.Hyvarinen, J.Karhunen, and E-Oja, *Independent Component Analysis*, John Wiley, 2001.

[19] R.W.Swiniarski, and A.Skowron, “Rough set methods in feature selection and recognition”, *Pattern Recognition Letters*, vol.24, issue 6, pp.833-849, March 2003.

[20] Z.Pawlak, J.Grzumala–Busse, R.Slowinski, and W.Ziarko, “Rough Sets”, *Communications of the ACM*, vol.38, No.11, pp.89-95, Nov.1995.

[21] J.W.Grzymala-Busse, “LERS-A system for learning from examples based on rough sets”. In *Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory*, R.Slowinski, Ed., Kluwer, pp.3-18, 1992.

[22] J.Komorowski, Z.Pawlak, L.Polkowski, and A.Skowron”, Rough Sets: A tutorial”, In Pal, S.K. and A.Skowron (eds), *Rough Fuzzy Hybridization*: *A new Trend in Decision-Making*, Springer, Singapore, pp.3-98, 1999.

[23] R.P.Paiva and A.Dourado, “Interpretability and learning in neuro-fuzzy systems”, *Fuzzy Sets and Systems*, In Press, Corrected Proof, Available online 2 March 2004, (www.ElsevierComputerScience.com).

[24] C.C., Lin, “Fuzzy logic in control systems: fuzzy logic controller-Part I and II”, *IEEE Trans. Systems, Man, and Cybernetics*, vol.20, no.2, 404-435, 1990.

[25] Lin-Xin Wang, “The WM method completed: A flexible fuzzy system approach to data mining”, *IEEE Trans.Fuzzy Systems*, vol.11, no.6, Dec.2003.

[26] M.L.Littman, “Markov games as a framework for multi-agent reinforcement learning”. In *Proceedings of the Eleventh International Conference on Machine Learning (ICML-94)*, New Brunswick, pp.157-163, 1994

[27] E.Kutschinski, T.Uthmann and D.Polani, “Learning competitive pricing strategies by multi-agent reinforcement learning”, *Journal of Economic Dynamics and Control* 27: 2207-2218, 2003.

[28] Ram C.Rao, Arjunji, V.Ramesh, and B.P.S.Murthi, “Game theory and empirical generalizations concerning competitive promotions”, *Marketing Science* 1995, vol.14, iss. 3, part 2, pg G89, 12 pgs.

[29] B.Banerjee and S.Bandyopadhyay, “Advertising competition under consumer inertia”, *Marketing Science*, 2003, vol.22, iss.1, pg 131, 14 pgs.

[30] Marnik G.Dekimpe and D.M.Hanssens, “The persistence of marketing effects on sales”, *Marketing Science*, 1995, vol.14, iss 1, pg 1, 21 pgs.

[31] Robert P.Leone, “Generalizing what is known about temporal aggregation and advertising carryover”, *Marketing Science*, 1995, vol.14, iss 3, part 2, pg G141, 10 pgs.

Related Content