l_for_Investigating_Product_Attributes_and_Purchase_Decision.pdf

by

Submission date: 17-Apr-2023 06:08AM (UTC+0700)
Submission ID: 2066307773
File name: l_for_Investigating_Product_Attributes_and_Purchase_Decision.pdf (111.06K)
Word count: 2468
Character count: 14598

2018 7th International Conference on Computer and Communication Engineering (ICCCE)

Rough-Regression Model for Investigating Product Attributes and Purchase Decision

Rasyidah[1,2]
[1]Management System Information Department, State Polytehcnic of Padang
25164 Padang, West-Sumatra, Indonesia
rasyidah@pnp.ac.id

Nazri Mohd Nawi[2]
[2]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia
86400 Batu Pahat, Johor, Malaysia
nazri@uthm.edu.my

Riswan Efendi[3]
[3]Mathematics Department, State Islamic University of Sultan Syarif Kasim Riau
28294 Pekanbaru, Indonesia
riswan.efendi@uin-suska.ac.id

***Abstract*—Regression models have been widely applied to investigate the causal relationship between independent and dependent attributes with statistical assumptions. On the other hands, not easy to achieve all statistical assumptions using these models, especially for certain areas. This paper presents rough-regression model to handle the categorical data types with minimal assumptions. The proposed idea is address to solve the unclassified elements and decisive criteria in data sets. Moreover, the product attributes of Kompas newspaper are investigated and selected using rough-regression model. The result showed that three conditional attributes, namely, price, promotion, and location have positive effect to purchasing decision attribute. Proposed decisive criteria also may help decision makers or marketing management in providing information and planning precisely.**

***Keywords*—rough-regression, unclassified customer, decisive category, price, promotion, location, newspaper**

I. Introduction

Some model have been discussed to investigate the medical attributes, such as, blood pressure [1], sleeping hour [2],[3], weight or obesity [4] and calorie level [5] which affect the patient cholesterol level. However, these studies are still focusing with rough sets approximation or regression model only, no combination among both yet.

The rough sets theory as a new knowledge discovery tool with many advantages – inspired many scholars to do research in adapting this theory to the business domain [6]. So far they have achieved many promising results [16, 17]. For example, it is based on the original data only and does not need any external information, unlike probability in statistics or grade of membership in the fuzzy set theory [18]. The rough set theory is an appropriate tool for analyzing not only numerical attributes but also non-numerical (categorical) ones. It bridges important facts uncertainty in data and expresses them in the natural language of decision rules.

The set of decision rules derived by the rough set gives a Generalized description of the knowledge contained in the financial information tables, eliminating any redundancy typical of the original data. The decision rules obtained are based on data reduction and unclassified element elimination. In the above review of rough sets model for economic and financial prediction, the rough sets model is a promising alternative method to conventional methods [18].

Recently, rough-regression model has been implemented to medical data [15, 19]. However, we found some rooms can be improved. In this paper, we introduce the importance of data reduction in rough-regression model for better prediction values and it will be implemented in product attributes selection and purchase decision of Kompas newspaper. The proposed data reduction strategy is able to analyze the dependency among the categorical data. The detail of theories, proposed model, and implementation are discussed in Sections 2, 3 and 4, respectively.

II. Rough Set and Regression Models

A. Rough Sets and Applications

The rough set theory was initiated by Pawlak in 1982 [6], there are some components in this theory, such as, information systems, indiscernibility relation, set approximations, rough clustering, and others. An information system $S = (U, \Omega, V_q, f_q)$ consists of a nonempty finite set called the universe, finite set attributes, and others [7-11].

A cases, states, processes, patients and observations are interpreted as objects in rough sets. While, the features, variables, and characteristic information are denoted as attributes. A decision table is used for a special case of information systems or attribute-value table is applied in the following analysis. In a decision table, the row and column correspond to objects and attributes, respectively [7-11].

Based on [11], the indiscernibility, lower-upper approximations, and boundary region can be formulated by following Equations (1) – (5) accordingly.

$$\mathrm{IND}(B) = \{(x, y) \in U^2 : \forall a \in B, a(x) = a(y)\}, \qquad (1)$$

$$\underline{B}(X) = \{\{x \in U | [x]_B \subseteq X\}\}, \qquad (2)$$

This work is financially supported by Universiti Tun Hussein Onn Malaysia (UTHM) and Ministry of Higher Education (MOHE) Malaysia .

978-1-5386-6992-1/18/$31.00 ©2018 IEEE

$$\bar{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\}, \quad (3)$$

$$\text{BND}(X) = \underline{B}(X) - \bar{B}(X), \quad (4)$$

$$\alpha(X) = \frac{\underline{B}(X)}{\bar{B}(X)}. \quad (5)$$

While, the dependency attributes is formulated as [12]:

$$k = \frac{\sum_{x \in U/D} |\underline{C}(X)|}{|U|}; C, D \subseteq A \wedge C \cap D = \emptyset. \quad (6)$$

The maximum value of k can be interpreted as a dominant attribute or criteria.

3 The data reduction is a set of necessary minimum data, since the original properties of the system or information table are maintained. Therefore, the reduction must have the capacity to classify objects, without altering the form of representing the knowledge. The process of reduction of information is divided into some parts, such as [21]:

3 Verification inconclusive data.
b. Verification of equivalent information.
c. Given analysis of condition attributes.
d. Verification of equivalent (intersection) data.
e. Generate decision rules.

2 Majority, rough sets theories can be implemented to investigate the dependency between conditional attributes and decision attribute for various domains, such as, cognitive sciences and medical data.

B. Regression Models and Its Applications

A basic linear regression is denoted as [13]:

$$Y = \beta_0 + \beta_1 X + e, \quad (7)$$

where Y is a dependent variable, X is an independent variable, β_0 and $\beta_1$are constant and coefficient model, while e is error of model. Based on [15, 19], there are steps to derive Eq. (7) as follows:

Step 1: Verify the linear association between Y and X.
Step 2: Predict constant and coefficient regression using ordinary least square (OLS) method.
Step 3: Evaluate the contribution X to Y using ANOVA and F-test
Step 4: Check the significance constant and coefficient regression using t-test.
Step 5: Check the normality residual model.

1 In applications, regression models can be used to measure the causal effect between independent and dependent variables in various domains, such as, sciences, medical, education, economics, others.

III. Proposed Rough-Regression Model and Its Implementation

From Section 2.2, the Eq. (7) can be implemented and generalized to investigate the product attributes and purchasing decision as follows:

$$PD = a_0 + a_1 Pri + a_2 Pro + a_3 Loc, , \quad (8)$$

Based on Eq. (8), $Pri = price, Pro = promotion, Loc = location$ are independent (conditional) attributes which affect to purchase decision, PD. While, $a_0, \ldots, a_3$ are estimated parameters (coefficients). In the previous studies, the discussion of decisive criteria from each attribute is very limited. Motivated by previous studies [15, 19], the removing unclassified patients do not yet considered in building rough-regression model. In this paper, we are interested to implement rough set approximations into regression model for investigating the product attributes and purchase decision using data reduction and dominant criteria by following steps in Cross Industry Standard Process for Data Mining (CRISP-DM) which describes the life cycle of a data mining project in form of six different phases, such as, business (any area) understanding, data understanding, data preparation, modeling, evaluation and deployment [20].

Step 1: Transform numerical product attributes to criteria or categorical values o as presented in Table 1.

TABLE I. Transform Numerical into Categorical Data

News Code	Price		…	Purchasing Decision	
	Num.	Cat.		Num.	Cat.
1	1	In	…	0	No
2	2	Out	…	1	Yes
3	1	In	…	1	Yes
…	…	…	…	…	…
n	1	In	…	0	No

Step 2: Based on Table 1, determine regression equation (before data reduction) by using the steps given in Section 2.2.

Step 3: Based on Table 1 and Equations (2)-(4), determine lower approximation, upper approximation and boundary region (BR) between conditional attributes and decision attribute.

Step 4: Based on Step 3 and [21], eliminate unclassified elements or objects.

Step 5: Based on Step 4, construct new regression equation (after data reduction).

Step 6: Based on Table 1, provide data reduction in information system until equivalent (intersection) data can be obtained.

Step 7: Based on Step 3, generate decision support rules.

Step 8: Based on Steps 2 and 4, predict the actual decision.

Step 9: Based on Step 5, verify the prediction error from the proposed model with the existing models based on average of percentage error as follows:

$$Accuracy = \frac{number\ of\ corrected\ value}{total\ observation} \times 100\%. \quad (9)$$

Step 10: Determine decisive criteria as mentioned in [15] for each product attribute.

By following Steps 1-10 and using row data from 130 customers of Kompas newspaper, a linear relationship (regression model) between product (X_1), price (X_2), location (X_3), promotion (X_4) and purchase decision (PD) is shown mathematically in Fig. 1.

TABLE II. REGRESSION MODEL BEFORE DATA ELIMINATION

Regression model before data reduction is:

$$PD = 3.481 + 0.096\, x_1 + 0.112\, x_2 + 0.175\, x_3 + 0.143\, x_4$$

R-Sq = 27.6% R-Sq(adj) = 25.3%

Based on Table 2, there are three attributes or independent variables which contribute to purchase decision (PD) significantly with R-square 27.6%. However, this percentage can be increased by using data reduction strategy, namely, removing unclassified customers have to be eliminated from data set. Thus, a new regression model can be shown in Table 3.

TABLE III. REGRESSION MODEL AFTER

Regression model after data reduction is:

$$PD = 3.617 + 0.091\, x_1 - 0.099\, x_2 + 0.361\, x_3 + 0.125\, x_4$$

R-Sq = 37.6% R-Sq(adj) = 31.7%.

Based on Tables 2 and 3, the data elimination strategy is able to improve the performance of regression model, especially R-Square value as presented in Table 4.

TABLE IV. PERFOMANCE REGRESSION BEFORE-AFTER REDUCTION

Parameters	Regression Models		Remark
	Before reduction	After reduction	
Constant	3.481	3.617	Down
Coefficient X_1	0.096	0.091	Down
Coefficient X_2	0.112	-0.099	Down
Coefficient X_3	0.175	0.361	Up
Coefficient X_4	0.143	0.125	Down
R-Square	27.6%	37.6%	Up

Table 4 shows the significance of data elimination for unclassified customers of Kompas newspaper in improving the performance of regression model. R-square value is increased 10%. In this case, we proposed the new strategy to improve the contribution of independent variables or conditional attributes to purchase decision without removing independent variables, but elimination of unclassified elements. Moreover, these elements can be denoted as "vague information in data sets". The information may positively influence the model building and its accuracy. On the other hands, not many studies have been discussed the combination between rough sets approximation and regression models using data reduction. In previous regression models, the dominant category from each independent variables to dependent variable are not yet determined clearly. In this paper, we are also interested to determine the decisive category from each attributes as presented in Fig. 1.

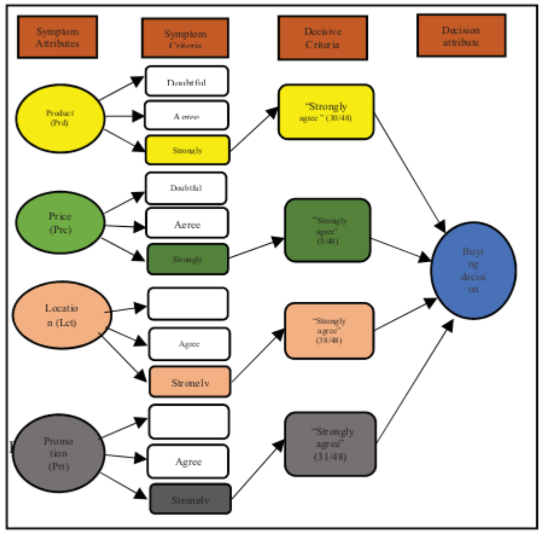

Fig. 1. Decisive category for product attributes of newspaper.

From Fig. 1, the decisive category for each product attribute is 30/48 for strongly agree (X_1), 5/48 for strongly agree (X_2), 38/48 for strongly agree (X_3), and 31/48 for strongly agree (X_4) which influence the purchase decision significantly. While, for other category are not dominant. In this case, most of purchase decision of customers are strongly agree for all product attributes (X_1, X_2, X_3, X_4). In this case, the investigating to product attributes and its category are very detail explained if compared with conventional regression models. The proposed rough-regression model is one of appropriate model in investigating of categorical data, especially for conditional attributes which have high uncertainty in measurement.

IV. CONCLUSION

In this paper, we suggest two strategies in improving the conventional regression models with rough sets approximation. First strategy, the data reduction is very important to remove the unclassified data or elements in the data sets. Its contribution is to increase the dependency between conditional attributes and decision attribute. A procedure to investigate the dominant category of each product attributes which influence the purchase decision for Kompas newspaper. This proposed model has been also examined to evaluate the dominant criteria and factors affecting customers purchase decision. Interestingly, the proposed rough-regression model is also help the decision makers in determining the dependency product attributes and the dominant criteria precisely which affecting decision attribute. Based on our perspective, the proposed model is one of alternative strategy to improve the performance of classical regression models.

Acknowledgment

The authors would like to thank Universiti Tun Hussein Onn Malaysia (UTHM) and Ministry of Higher Education (MOHE) Malaysia for financially supporting this Research under IGSP grants note U420 and under Trans-disciplinary Research Grant Scheme (TRGS) vote no. T003.

References

[1] Sakurai, M., Stamler, J., Miura, K., Brown, I.J., Nakagawa, H., Elliot, P., Ueshima, H., Chan, Q., Tzoulaki, I., Dyer, A.R., Okayama, A., Zhao, L.: Relationship of dietry cholesterol to blood pressure: The intermap study. J. Hypertens. **29**, 222-228 (2011)

[2] Wolk, R., Somers, V.K.: Sleep and the metabolic syndrome. Exp. Physiology. **92**, 67-78 (2007)

[3] Kaneita, Y., Uchiyama, M., Yoshiike, N., Ohida, T.: Associations of usual sleep duration with serum lipid and lipoprotein levels. Sleep. **31**, 645-652 (2008)

[4] Miettinen, T.A.: Cholesterol production in obesity. Circulation. **64**, 842-850 (1971)

[5] Ueshima, H., Iida, M., Shimamoto, T., Konishi, M., Tanigaki, M., Doi, M., Nakashini, N., Takayama, Y., Ozawa, H., Komachi, Y.: Dietary intake and serum total cholesterol level: their relationship to different lifestyles in several Japanese populations. Circulation. **66**, 519-526 (1982)

[6] Pawlak, Z.: Rough sets. Int. J. Compt. Inf. Science. **11**, 341-356 (1982)

[7] Pawlak, Z.: Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publisher Dordrecht (1991)

[8] Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance. **5**, 25-29 (1997)

[9] Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance. **6**, 40-42 (1998)

[10] Hampton, J.: Rough set theory: the basics (part 1). J. Compt. Intel. Finance. **6**, 35-37 (1998)

[11] Tay, F.E.H., Shen, L.: Economic and financial using rough sets model. European J. Operation Research. **141**, 641-659 (2002)

[12] Herawan, T., Deris, M.M., Abawajy, H.: A rough set approach for selecting clustering attribute. Knowledge-Based Systems. **23**, 220-231 (2010)

[13] Wooldridge, M.: Introductory econometrics a modem approach. Third Ed. Thomson, South Western, USA (2006)

[14] Ali, R., Hussain, J., Siddiqi, M.H., Hussain, M., Lee, S.: A hybrid rough set reasoning model for prediction and management of Diabetes Mellitus. Sensors. **15**, 15921-15951 (2015)

[15] Efendi, R., Deris, M.M.: Decision Support Model in Determining Factors and Its Dominant Criteria Affecting Cholesterol Level Based on Rough-Regression. Recent Advances on Soft Computing and Data Mining, 243-251 (2018)

[16] Polkowski, L., Skowron, A.: Rough Sets in Knowledge Discovery 1: Methodology and Applications; Rough Sets in Knowledge Discovery 2: Application, Case Studies, and Sofware Systems. Physica-Verlag, Wurzburg (1998)

[17] Shen, L., Tay, F.E.H., Qu, L., Shen, Y.: Fault Diagnosis using Rough Sets Theory. Computers in Industry (**43**), 61-72 (2000)

[18] Krusinska, E., Slowinski, R., Stefanowski, J.: Discriminant versus Rough Set Approach to Vague Data Analysis. Applied Stochastic Models and Data Analysis (**8**), 43-56 (1992).

[19] Efendi, R., Samsudin, N.A., Deris, M.M.: Medipre: Medical diagnosis prediction using rough-regression approximation. ACM Proceeding on High Compilation, Computing and Communications, 35-39 (2018)

[20] Chapman, P., Clinton, J., Khabaza, T., Reinartz, T and Wirth, R. The CRISP-DM Process Model. August 2000

[21] Rissino, S., Torres, G. L. Rough set theory-fundamental concepts, principals, data extraction, and applications, Julio Ponce and Adam Karahoca (Ed), Data Mining and Knowledge Discovery in Real Life App. Inform. (2009) 35-58.

l_for_Investigating_Product_Attributes_and_Purchase_Decisio...

ORIGINALITY REPORT

22%	%	22%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Riswan Efendi, Susnaningsih Mu'at, Voni Apriana Dewi, Nelsi Arisandy, Noor Azah Samsudin, Dadang Syarif Sihabudin Sahid. "Rough-Regression for Categorical Data Prediction based on Case Study", 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019 Publication	7%
2	Riswan Efendi, Novi Yanti, Alex Wenda, Susnaningsih Mu'at, Noor Azah Samsudin, Mustafa Mat Deris. "Dominant Criteria and Its Factor Affecting Student Achievement Based on Rough-Regression Model", 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018 Publication	3%
3	Silvia Rissino, Germano Lambert-Torres. "Chapter 3 Rough Set Theory — Fundamental Concepts, Principals, Data Extraction, and Applications", IntechOpen, 2009 Publication	2%

4	"Intelligent and Interactive Computing", Springer Science and Business Media LLC, 2019 Publication	2%
5	Laisheng Wang. "Fuzzy Comprehensive Evaluation Model based on Rough Set Theory", 2006 5th IEEE International Conference on Cognitive Informatics, 07/2006 Publication	2%
6	Riswan Efendi, Noor Azah Samsudin, Mustafa Mat Deris, Yip Guan Ting. "Flu Diagnosis System Using Jaccard Index and Rough Set Approaches", Journal of Physics: Conference Series, 2018 Publication	2%
7	ASHWANI KUMAR, D. P. AGRAWAL, S. D. JOSHI. "ADVERTISING DATA ANALYSIS USING ROUGH SETS MODEL", International Journal of Information Technology & Decision Making, 2011 Publication	2%
8	Kweku-Muata Osei-Bryson. "Organization-Ontology Based Framework for Implementing the Business Understanding Phase of Data Mining Projects", Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), 01/2008 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		