Performing Variable Selection by Multiobjective Criterion: An Application to Mobile Payment

Alberto Guillén[1](✉), Luis-Javier Herrera[1], Francisco Liébana[2], Oresti Baños[3], and Ignacio Rojas[1]

[1] Department of Computer Architecture and Technology, University of Granada, Granada, Spain
aguillen@ugr.es
[2] Department of Marketing and Market Research, University of Granada, Granada, Spain
[3] Ubiquitous Computing Lab (UCLab), Kyung Hee University, Seoul, Korea

Abstract. The rapid growth social networks have led many companies to use mobile payment systems as business sales tools. As these platforms have an increasing acceptance among the consumers, the main goal of this research is to analyze the individuals' use intention of these systems in a social network environment. The problem of variable selection arises in this context as key to understand user's behaviour. This paper compares several non-parametric criteria to perform variable selection and combines them in a multiobjective manner showing a good performance in the experiments carried out and validated by experts.

1 Introduction

Social commerce presents two advantages over existing forms of commerce [25]. First of all, it facilitates the interaction between Internet users, allowing direct interaction and exchange of opinions, purchase advice and experiences (participatory environment). Secondly, it makes web surfing and knowledge about a variety of products possible, both of which are limited in an offline context (unlimited access).Therefore, given the growth that Virtual Social Networks (VSNs) are experiencing in our society, it is necessary for both companies and users to analyze the intention to use the new tools implemented on social networks, to establish the level of general acceptance of these tools, as well as the factors that determine user behavior regarding payment systems that can generate profit for a company.

There are numerous behavioral decision-making theories and intention models in scientific literature, analyzing individuals' behavior when facing innovation, the majority of which are based on social psychology studies [20]. However, this problem can be reformulated so machine learning and data mining algorithms can be applied. The definition of the problem from these disciplines point of view is that given a set of input vectors $X = [x_1^1 x_1^2 x_1^d; x_2^1 x_2^2 x_2^d; x_m^1 x_m^2 x_m^d]$ and their corresponding output $Y = [y_m]$, it is necessary to find out which subset

I. Rojas et al. (Eds.): IWANN 2015, Part II, LNCS 9095, pp. 333–340, 2015.
DOI: 10.1007/978-3-319-19222-2_28

of variables are the most relevant to determine the output. Therefore, there are (2^d)-1 possible solutions only considering if the variable i is selected or not.

There are several ways to determine which subset of variables is the most adequate, but two classes can be defined: parametric versus non-parametric. Parametric criteria are the ones that design a model that adjusts the output Y, so the more accurate the model obtained using a particular subset of variables, the better this subset is. Examples of previous algorithms performing this task include models built simultaneously using a genetic algorithm[9]. The main problem with these approaches is that they are very expensive in terms of computation so very few solutions can be evaluated. Another problem that this approach presents is that the value obtained for each subset may change depending on the parameters chosen for the model, (ex. the kernel function, number of neurons, weights in the hidden layer, etc.) so the evaluation is not robust.

In order to overcome these drawbacks, several papers apply non-parametric values so the result is independent of the choices made beforehand. This paper proposes a new approach to tackle the variable selection problem by using multiobjective criteria that consider opposite non-parametric criteria as could be mutual information and delta test [10],[16], [8], [14], [11], [12].

The rest of the paper is organised as follows: Section 2 explines how the data were collected. Then, Section 3 introduces the procedures to rank variable selections. In Section 4, experiments are carried out and, afterwards in Section 5, conclusions are drawn.

2 Data Compilation

The research was carried out on the worlds best-known social network, Facebook, because of its widespread diffusion, the large number of users and the usefulness it confers. These users were exposed to an experimental scenario on a Facebook profile, in which they watched a video explaining the proposed new payment system. This system, called Zong, facilitates the purchase of both physical goods and virtual content using a mobile terminal. This very simple process accepts different payment methods for purchases via the Internet, social networks, television and even the point of sale itself). Payment can be charged to a users phone bill or to a credit card authorized during service activation.

The size of the final sample was 1840, using a quota sampling method based on the characteristics of users who participated in the Survey about Equipment and Use of Information and Communication Technologies in Households by the National Statistics Institute [17].

The variables considered in this research to define intention to use the new payment system were structured in five groups: behavioral variables, socio-demographic variables, user technological and social media comfort level and users previous experience with similar payment tools. First, social influence or social image and subjective norms , since in the context of this research there is a very close relationship between them due both to desired social value, and to the influence the social environment on the social network exerts on users of the new payment system [13]; secondly, ease of use based on individuals perception that using a certain system will be effortless or simply easy to use [2]; thirdly, perceived usefulness, defined as the potential users subjective probability that using a specific system will improve their job performance in an organizational context [3]; fourthly, attitude based on favorable or unfavorable feelings people express towards a given behavior [22]; fifthly, trust based on the psychological state that leads one to accept the vulnerability of those who trust and is based on favorable expectations about the intentions and behaviors of others [23]; sixthly, perceived risk considered in view of a potential unsafe use of the innovation [7]; seventhly, perceived quality based on users subjective comparison between the desired and the received quality of service [6]; and finally, perceived satisfaction, defined as the difference between the expectations and the feelings generated by the use experience [19].

Finally, the level of experience shows that a users positive experience with a similar tool in the past will have a decisive impact on their behavior [5].

Once the data was collected, in order to analyze it, preprocessing was required to obtain concrete values for each type of question. In order to do this, an average value was computed considering all the answers for a particular type of question. Afterwards, these values were normalized with a mean equal to zero and a standard deviation equal to one.

3 Criteria for Variable Selection

In order to determine the relevance of each variable measured, two criteria were selected: the Delta Test (DT), and Mutual Information (MI). Four different algorithmic approaches have been considered and compared: the minimum value of Delta Test; the maximum value of the estimation of the MI and a novel approach that combines the former two criteria in a multi-objective approach described later.

3.1 Mutual Information

Given a single-output multiple input function approximation or classification problem, with input variables $X = [x_1, x_2, \ldots, x_d]$ and output variable $Y = y$, the main goal of a modeling problem is to reduce the uncertainty on the dependent variable Y. According to the formulation of Shannon, and in the continuous case, the uncertainty on Y is given by its entropy defined as:

$$H(Y) = -\int \mu_Y(y) \log \mu_Y(y) dy, \tag{1}$$

considering that the marginal density function $\mu_Y(y)$ can be defined using the joint probability density function $\mu_{X,Y}$ of X and Y as $\mu_Y(y) = \int \mu_{X,Y}(x,y)dx$.

Given that we know X, the resulting uncertainty of Y conditioned to known X is given by the conditional entropy, defined by:

$$H(Y,X) = -\int \mu_X(x) \int \mu_Y(y|X = x) \log \mu_Y(y|X = x) dydx. \tag{2}$$

The mutual information (also called cross-entropy) between X and Y can be defined as the amount of information that the group of variables X provide about Y, and can be expressed as $I(X,Y) = H(Y) - H(Y|X)$. In other words, the mutual information $I(X,Y)$ is the decrease of the uncertainty on Y once we know X. Due to the mutual information and entropy properties, the mutual information can also be defined as:

$$I(X,Y) = H(X) + H(Y) - H(X|Y), \tag{3}$$

leading to:

$$I(X,Y) = \int \mu_{X,Y}(x,y) \log \frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)} dxdy. \tag{4}$$

Thus, only the estimate of the joint PDF between X and Y is needed to estimate the mutual information between two groups of variables. Estimating the joint probability distribution can be performed using a number of techniques like histograms and kernel density estimators have been used for this purpose. Although there exists a variety of algorithms to calculate the mutual information between variables, this paper uses the approach presented in [26] which is based on the k-nearest neighbors.

3.2 Delta Test

The Delta Test (DT), introduced by Pi and Peterson for time series and proposed for variable selection in [4], is a technique to estimate the variance of the noise, or the mean squared error (MSE), that can be achieved without overfitting. Given N input-output pairs $(\mathbf{x}_i, y_i) \in \mathbb{R}R^d \times \mathbb{R}R$, the relationship between $\mathbf{x}_i$ and y_i can be expressed as

$$y_i = f(\mathbf{x}_i) + r_i, \qquad i = 1, ..., N \tag{5}$$

where f is an unknown function and r is the noise. The DT estimates the variance of the noise r.

The DT is useful for evaluating the nonlinear correlation between two random variables, namely, input and output pairs. The DT can also be applied to input variable selection: the set of input variables that minimizes the DT is the one that is selected. Indeed, according to the DT, the selected set of input variables is the one that represents the relationship between input variables and the output variable in the most deterministic way.

This test based on a hypothesis coming from the continuity of the regression function. If two points $\mathbf{x}$ and $\mathbf{x}'$ are close in the input space, the continuity of the regression function implies that the outputs $f(\mathbf{x})$ and $f(\mathbf{x}')$ are also close enough in the output space. Alternatively, if the corresponding output values are not close in the output space, this is due to the influence of the noise.

3.3 MultiObjetive Criteria: MOmaxMI or MOminDT

This paper presents a novel approach that is able to combine the two criteria considered. Very little research has previously been done in this direction. In [21] two ways of computing MI are optimized simultaneously. However, the concept being computed is the same. Although the evaluation of a good concrete subset of variables should be similar for both DT and MI, the reality is that, for many solutions, these two values are opposites. The reason is because the MI measures the entropy and the DT measures the variance of the noise in the output so, if there is a high variance of the noise, there is high entropy. Therefore, it is interesting to optimize both values at the same time in order to obtain a compromise solution.

Multi-objective optimization is a well-known problem that is usually solved by using heuristics such as genetic algorithms. However, in this case, we can afford to compute the values for all the possible solutions and select the solutions from the complete Pareto front [1].

4 Experiments

In order to validate the experimental results a set of experts (with more than 10 years of experience) related to the different research topics were consulted.

The evaluations of the methods from the subsets of experts are shown in Table 1 using a Likert (1-7) scale. This table presents 4 criteria to perform the variable selection. The first one considers the criterion of maximizing the mutual information. The second one is one of the novel contributions of the paper as it tries to maximize the mutual information while also, selecting a solution that has a relatively small Delta test value. The third criterion is also a multi-objective approach that minimizes the Delta test value but it considers it as a solution that provides a high value for the MI. The forth one consists of the minimization of the Delta test. The two multiobjective solutions were taken from the Pareto front shown in Figure 1.

Table 1. Punctuation results from expert's opinion on the variables selected by the different criteria. Group A are marketing and market research experts and Group B corresponds to Financial and payment experts.

A	maxMI	MO-maxMI	MO-minDT	minDT
Expert 1	3	3	4	3
Expert 2	2	4	4	2
Expert 3	3	3	5	3
Mean(std)	2,6	3,3	2,6	2,6
B	**maxMI**	**MO-maxMI**	**MO-minDT**	**minDT**
Expert 1	3	7	3	3
Expert 2	2	6	4	2
Expert 3	4	7	5	2
Mean(std)	3	6,6	4	2,3
Global Mean (std)	2,8	4,95	3,3	2,45

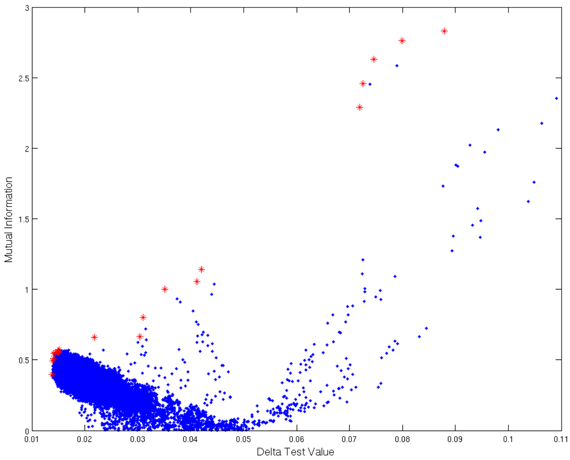

Fig. 1. Blue dots represent all posible solutions according to their DT and MI values. Red stars represent the Pareto front from the whole solution set.

As the table shows, the best total score is obtained by the second criterion. The other criteria are not highly rated in comparison with the first two although the second one is another MO approach.

Once all the results were collected, a second interview was carried out and the experts justified the scores given to each method. The reasons for the scores were: Regarding Max MI criterium (option 1): It only selects one variable (users experience with Internet). The experts argue that, although the variable is important, it is not the only one that should be taken into account. Regarding the MO approach that maximizes MI (option 2) and keeps a small value for the DT: It includes several variables such as the ones related to the users experience using Internet, mobile and social networks. Some experts comment that the method does not select variables that represent the behavior of the potential user as it does not consider the trajectory of the user with similar situations. On the other hand, the fact that it selects the elements related to the new payment platform it is highly scored by the experts in the marketing field. Regarding the MO approach that minimizes the DT (option 3) and tries to keep a high MI: It selects nearly all the variables presented as inputs so the total number of variables is very high and this is punished by the experts (in both the marketing and the financial fields) who could not make a decision based on so many factors. Regarding the minimization of the DT (option 4), it has the same problem as the previous method. The number of variables selected is too high and experts agree that it is not practical to make a decision.

From the machine learning point of view, the results are very interesting. In the first place, they confirm that the multi-objective way of thinking is more practical and provides better results than considering only one criterion. Specifically, in this case, maximizing the entropy but keeping in mind that the noise should be reduced, by considering a small value of DT, was a very successful approach.

5 Conclusions and Further Work

The developments that are taking place in the Information and Communication Technologies (ICT) sector in our society in recent years have had repercussions in terms of profitability, productivity, competitiveness and economic growth, at the company level. Consumers are increasingly using social networks to obtain recommendations and opinions from friends, relatives, experts and the entire social community. It is very important to be able to identify the key elements that drive the user behaviour. The paper has presented a fusion between two well-known criteria to perform variable selection and optimize them in a multi-objective way. The results obtained were validated by several experts showing that it is interesting to select the variables with high entropy, maintaining variables with small variance of the noise in the output.

Acknowledgments. This work has been supported by the GENIL-PYR-2014-12 project from the GENIL Program of the CEI BioTic, Granada, and the Junta de Andalucia Excellence Project P12-TIC-2082.

References

1. Coello, C.A.C., Lamont, G.B., Van Veldhuisen, D.A.: Evolutionary algorithms for solving multi-objective problems. Springer (2007)
2. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly **13**(3), 319–340 (1989)
3. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science **35**, 982–1003 (1989)
4. Eirola, E., Liitiinen, E., Lendasse, A., Corona, F., Verleysen, M.: Using the delta test for variable selection. In: ESANN, pp. 25–30 (2008)
5. Fishbein, M., Ajzen, I.: Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading (1975)
6. Gefen, D.: E-commerce: The role of familiarity and trust. The International Journal of Management Science **28**, 725–737 (2000)
7. Gerrard, P., Cunningham, J.B.: The diffusion of internet banking among Singapore consumers. International Journal of Bank Marketing **21**(1), 16–28 (2003)
8. Guillén, A., Del Moral, F.G., Herrera, L.J., Rubio, G., Rojas, I., Valenzuela, O., Pomares, H.: Using near-infrared spectroscopy in the classification of white and iberian pork with neural networks. Neural Computing and Applications **19**(3), 465–470 (2010)
9. Guillén, A., Pomares, H., González, J., Rojas, I., Valenzuela, O., Prieto, B.: Parallel multiobjective memetic rbfnns design and feature selection for function approximation problems. Neurocomputing **72**(16), 3541–3555 (2009)

10. Guillén, A., Sovilj, D., Lendasse, A., Mateo, F.: Minimising the delta test for variable selection in regression problems. International Journal of High Performance Systems Architecture **1**(4), 269–281 (2008)
11. Guillén, A., Sovilj, D., van Heeswijk, M., Herrera, L.J., Lendasse, A., Pomares, H., Rojas, I.: Evolutive approaches for variable selection using a non-parametric noise estimator. In: de Vega, F.F., Pérez, J.I.H., Lanchares, J. (eds.) Parallel Architectures and Bioinspired Algorithms. SCI, vol. 415, pp. 243–266. Springer, Heidelberg (2012)
12. Herrera, L.J., Fernandes, C.M., Mora, A.M., Migotina, D., Largo, R., Guillén, A., Rosa, A.C.: Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification. International journal of neural systems **23**(3) (2013)
13. Liébana-Cabanillas, F.: El papel de los sistemas de pago en los nuevos entornos electrónicos, Doctoral Thesis, Marketing and Market Research Department, University of Granada (2012)
14. Long, X.X., Li, H.D., Fan, W., Xu, Q.S., Liang, Y.Z.: A model population analysis method for variable selection based on mutual information. Chemometrics and Intelligent Laboratory Systems **121**, 75–81 (2012)
15. Mallat, N., Rossi, M., Tuunainen, V., Rni, A.: The impact of use context on mobile services acceptance: the case of mobile ticketing. Information & Management **46**(3), 190–195 (2009)
16. May, R.J., Maier, H.R., Dandy, G.C., Fernando, T.M.K.: Non-linear variable selection for artificial neural networks using partial mutual information. Environmental Modelling & Software **23**(10), 1312–1326 (2008)
17. National Statistics Institute, Survey about Equipment and Use of Information and Communication Technologies in Households (2012). www.ine.es
18. OCass, A., Fenech, T.: Web retailing adopction: exploring the nature of internet users web retailing behavior. Journal of Retailing and Consumer services **10**, 81–94 (2003)
19. Oliver, R.L.: A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research **17**, 460–469 (1980)
20. Pavlou, P.A.: A theory of Planned Behavior Perspective to the Consumer Adoption of Electronic Commerce. MIS Quarterly **30**(1), 115–143 (2002)
21. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8), 1226–1238 (2005)
22. Premkumar, G., Rammurthy, K., Liu, H.: Internet Messaging: An Examination of the Impact of Attitudinal, Normative and Control Belief Systems. Information & Management **45**, 451–457 (2008)
23. Singh, J., Sirdeshmukh, D.: Agency and Trust Mechanisms in Consumer Satisfaction and Loyalty Judgments. Journal of the Academy of Marketing Science **28**(1), 150–167 (2000)
24. Smith, M., Brynjolfsson, E.: Consumer decision making at an internet shopbot: Brand still matters. The Journal of Industrial Economics **49**(4), 541–558 (2001)
25. Zhang, L.: Business model analysis for online social shopping companies. Case Company: Run To Shop Oy. Doctoral Thesis, Department of Business Technology/Logistics, Helsinki School of Economics (2009)
26. Herrera, L.J., Pomares, H., Rojas, I., Verleysen, M., Guilén, A.: Effective input variable selection for function approximation. In: Kollias, S.D., Stafylopatis, A., Duch, W., Oja, E. (eds.) ICANN 2006. LNCS, vol. 4131, pp. 41–50. Springer, Heidelberg (2006)