An Application of Rough Set Theory for Clustering Performance Expectancy of Indonesian e-Government Dataset

Deden Witarsyah Jacob[1], Mohd Farhan Md Fudzee[2], Mohamad Aizi Salamat[2], Rd Rohmat Saedudin[1], Iwan Tri Riyadi Yanto[3], and Tutut Herawan[4]

[1]Department of Industrial Engineering
Telkom University, Bandung, West Java, Indonesia
[2]Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
[3] Department of Information Systems, University of Malaya
Kampus III UAD, Jalan Prof Dr Soepomo, Yogyakarta, Indonesia
[4]Faculty of Computer Science and Information Technology, University of Malaya
50603 Pantai Valley, Kuala Lumpur, Malaysia

{dedenw,rdrohmat}@telkomuniversity.ac.id, farhan@uthm.edu.my, aizi@uthm.edu.my, yanto.itr@is.uad.ac.id, tutut@um.edu.my

Abstract: Performance expectancy has been studied as an important factor which influences e-government. Therefore, grouping of e-government users involving performance expectancy factor is still challenging. Computational model can be explored as an efficient clustering technique for grouping e-government users. This paper presents an application of rough set theory for clustering performance expectancy of e-government user. The propose technique base on the selection of the best clustering attribute where the maximum dependency of attribute in e-government data is used. The datasets are taken from a survey aimed to understand of the adoption issue in e-government service usage at Bandung city in Indonesia. At this stage of the research, we point how a soft set approach for data clustering can be used to select the best clustering attribute. The result presents useful information for decision maker in order to make policy concerning theirs people and may potentially give a recommendation how to design and develop e-government system in improving public service.

Keywords: Clustering; Rough set theory; Performance expectancy; e-Government.

1. Introduction

Growth in public familiarity with information and communication technologies (ICTs) in the world, the internet in particular, has opened up opportunities for the public sector to embrace the technologies and use them to better serve citizens. The implementation of e-government systems has been attracting increased research interest, and is believed to constitute one of the most important IT implementation and organizational change challenges of the future [30,31]. Electronic government is designed as a process of interaction between government and society. Carter and Belanger [3] and Pavlou [21] states that one important factor for the success of e-government services is the acceptance and willingness of people to use e-government services.

Venkatesh *et al.* defines "Performance Expectancy" as the degree to which one believes in using the system will help the person to gain performance on the job [29]. In this concept there is a combination of variables obtained from the model of previous studies of the model acceptance and use of technology. The variables are: 1. Perceived usefulness, 2. Extrinsic Motivation, 3. Job Fit, 4. Relative advantage, and the last, Outcome Expectations. In this concept there is a combination of variables obtained from the model of previous studies of the model acceptance and use of technology. The clear explanation about performance expectance could be seen in Table 1 below:

Table 1: Variables in performance expectancy

No.	Variable	Definition	Studies
1	Perceived usefulness	The extent to which a person believes that using a particular system would enhance his performance.	Venkatesh, *et al*. [29] and Davis, *et al*. [10].
2	Extrinsic motivation	The perception that the user wants to perform an activity because it is considered as a tool in achieving valuable results that differ from the activity itself.	Venkatesh, *et al*. [29] and Davis, *et al*. [10].
3	Job fit	How the capabilities of a system increases the performance of individual work.	Venkatesh, *et al*. [29] and Davis, *et al*. [10].
4	Relative advantage	The extent to which use innovation something perceived to be better than using its predecessor.	Venkatesh, *et al*. [29] and Barney [2].
5	Outcome expectations	Expectations are the result (outcome expectations) is associated with the consequences of his behavior.	Venkatesh, *et al*. [29], Compeau and Higgins [9].

Meanwhile, Davis, F.D. [10]; Adams, *et al*. [1] defined performance expectancy as a level where a person believes that the use of a particular subject will be able to improve the work performance of the person. Chin and Todd [8] adds the dimension of expediency TI, which makes the work easier, rewarding, increase productivity, enhance the effectiveness of, and improve job performance. It can be concluded that a person's trust and feel by using an information technology will be very useful and can enhance performance and job performance.

Huang states that the clustering operation is required in a number of data analysis tasks, such as unsupervised classification and data summation, as well as segmentation of large homogeneous data sets into smaller homogeneous subsets that can be easily managed, separately odeled and analyzed [32]. Meanwhile, a well-known approach for data clustering is using rough set theory [34,34,36]. For example, Mazlack, He, Zhu, and Coppock had developed a rough set approach in choosing partitioning attributes [33]. One of the successful pioneering rough clustering for categorical data techniques is Minimum–Minimum Roughness (MMR) proposed by Parmar, Wu, and Blackhurst [38].

However, pure rough set theory is not well suited for analyzing noisy information systems. A knowledge discovery system must be tolerant to the occurrence of noise. For example, in the previous work on constructing student models through mining students classification-test answer sheets by Wang and Hung [37], much noise was found in the classification tables, either the feature values or the class values, created by students. Their empirical results showed that attention should be paid to handle the noisy information in order to reach a satisfactory prediction accuracy [37].

Computational model such as rough set theory can be explored as an efficient clustering technique for grouping e-government users. This paper presents an application of rough set theory for clustering performance expectancy of e-government user. The propose technique is based on the selection of the best clustering attribute where the maximum dependency of attribute in e-government data is used. The data were taken from a survey aimed to identify of citizen behavior in using e-government. Descriptive statistics is used to find out the Mean (M) and Standard Deviation (SD) to identify the potential sources of study behavior. It is ran in SPSS version 22.0 and the results show that there are 5 potential sources of study performance expectancy.

The remainder of this paper is organized as follows. Section 2 describes proposed method. Section 3 describes the study's performance expectancy of e government data set. Section 4 describes experiment result. Finally, the conclusions of this work are reported in section 5.

2. Proposed Method

2.1. Rough Set Theory

Motivation for rough set theory is needed to represent a subset of a universe in terms of equivalence classes of a partition of the universe. In this section, the basic concept of rough set theory is presented. The notion of information system provides a convenient tool for the representation of objects in terms of their attribute values. An *information system* is a 4-tuple (quadruple) $S=(U,A,V,f)$, where $U=\{u_1,u_2,\cdots,u_{|U|}\}$ is a non-empty finite set of objects, $A=\{a_1,a_2,\cdots,a_{|A|}\}$ is a non-empty finite set of attributes, $V=\bigcup_{a\in A}V_a$, V_a is the domain (value set) of attribute a, $f:U\times A\rightarrow V$ is an information function such that $f(u,a)\in V_a$, for every $(u,a)\in U\times A$, called information (knowledge) function (Pawlak & Skowron, 2007). Two elements $x,y\in U$ in *S* is said to be *B-indiscernible* (indiscernible by the set of attribute $B\subseteq A$ in S) if and only if $f(x,a)=f(y,a)$, for every $a\in B$ (Pawlak & Skowron, 2007). An indiscernible relation induced by the set of attribute *B*, denoted by $IND(B)$, is an equivalence relation. It is well-known that an equivalence relation can induce a unique partition. The partition of *U* induced by $IND(B)$ in $S=(U,A,V,f)$ denoted by U/B and the equivalence class in the partition U/B contains $x\in U$ and denotes by $[x]_B$. Let *B* be any subset of *A* in *S* and let *X* be any subset of *U*, the *B-lower approximation* of *X,* denoted by $\underline{B}(X)$ and *B-upper approximation* of *X*, denoted by $\overline{B}(X)$ respectively, are defined by $\underline{B}(X)=\{x\in U \mid [x]_B\subseteq X\}$and $\overline{B}(X)=\{x\in U \mid [x]_B\cap X\neq\phi\}$. The *accuracy of approximation* of any subset $X\subseteq U$ with respect to $B\subseteq A$, denoted by $\alpha_B(X)$ is measured by $\alpha_B(X) = |\underline{B}(X)| / |\overline{B}(X)|$, where $|X|$ denotes the cardinality of *X*. For empty set ϕ, it is defined that $\alpha_B(\phi) = 1$ (Pawlak & Skowron, 2007). Obviously, $0 \le \alpha_B(X) \le 1$. If *X* is a union of some equivalence classes of *U*, then $\alpha_B(X) = 1$. Thus, the set *X* is *crisp* (precise) with respect to *B*. And, if *X* is not a union of some equivalence classes of *U*, then $\alpha_B(X) < 1$. Thus, the set *X* is *rough* (imprecise) with respect to *B*. This means that the higher the accuracy of approximation of any subset $X \subseteq U$, the more precise (the less imprecise) of itself (Pawlak & Skowron, 2007).

2.2. Maximum Dependency of Attribute Technique

Definitions 1 and 2 below describe the notions of dependency of attributes in general and from the point of view of rough set theory, respectively.

Definition 1. *Let $S = (U, A, V, f)$ be an information system and let D and C be any subsets of A. D is functionally depends on C, denoted $C \Rightarrow D$, if each value of D is associated exactly one value of C.*

Definition 2. *Let $S = (U, A, V, f)$ be an information system and let D and C be any subsets of A. Dependency attribute D on C in a degree k $(0 \le k \le 1)$, is denoted by $C \Rightarrow_k D$. The degree k is defined by*

$$k = \frac{\sum_{X \in U/D} |\underline{C}(X)|}{|U|}. \tag{1}$$

D is said to be fully depends (in a degree of k) on C if $k = 1$. Otherwise, D is partially depends on C.

Thus, *D* fully (partially) depends on *C*, if all (some) elements of the universe *U* can be uniquely classified to equivalence classes of the clustering U/D, employing *C*. Based on Definition 2, we can select the clustering attributes based on the maximum degree of *k*.

2.3. MDA Algorithm

In this sub-section, we will present the proposed technique, which we refer to as Maximum Dependency of Attributes (MDA). Figure 1 shows the pseudo-code of the MDA algorithm.

Algorithm: MDA
Input: Data set without clustering attribute
Output: Clustering attribute
Begin
Step 1. Compute the equivalence classes using the indiscernibility relation on each attribute.
Step 2. Determine the dependency degree of attribute a_i with respect to all a_j, where $i \neq j$.
Step 3. Select the maximum of dependency degree of each attribute.

Step 4. Select the clustering attribute based on the maximum degree of dependency of attributes.
End

Figure 1: The MDA algorithm

The maximum degree of dependency of attributes is the more accurate (higher of accuracy of approximation) for selecting clustering attribute. The justification that the higher of the degree of dependency of attributes implies the more accurate for selecting clustering attribute is stated in the proposition 1.

Proposition 1. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. If D depends totally on C, then*

$$\alpha_D(X) \le \alpha_C(X), \textit{ for every } X \subseteq U.$$

3. The study's performance expectancy of e government dataset

The data set was taken from a survey in Bandung. A total population were 200 people take part in this survey. The profile of the respondents is used to provide a description of the characteristics of the sample, so it is very useful in the discussion of the results of the study investigators. The majority of respondents were women, i.e. 105 people, and the respondents were male is as much as 95 peoples. To analysis the data, for distribution of study performance expectancy scores, it follows likert-scale, i.e., 1 very not agree; 2 not agree; and, 3 neutral; 4 Agree and 5 very agree. In this survey, the study performance expectancy questionnaire has been test for reliability with alpha score yielded 0.699 and accessing content validity. Table 2 describes each attribute of performance expectancy study include the mean, standard deviation, variance and range.

Table 2: Summary of the study's performance expectancy of e government dataset

		Perceived Usefulness	Extrinsic Motivation	Job-fit	Relative Advantage	Outcome Expectations
N	Valid	200	200	200	200	200
	Missing	0	0	0	0	0
Mean		4.010	3.680	3.675	3.685	3.170
Std. Deviation		.4701	.6078	.6720	.7673	.7708
Variance		.221	.369	.452	.589	.594
Range		3.0	3.0	3.0	4.0	3.0

3.1. Perceived usefulness

Perceived usefulness is a leading source with M=4.010 and SD=0.4701. Perceived usefulness refers to is the extent to which the person believes that using a particular system would enhance his job performance (Davis, 1989). Table 3 describes data distribution include frequency and percent.

Table 3: Summary of Perceived Usefulness data distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.0	1	.5	.5	.5
	3.0	18	9.0	9.0	9.5
	4.0	159	79.5	79.5	89.0
	5.0	22	11.0	11.0	100.0
	Total	200	100.0	100.0	

3.2. Extrinsic motivation

The second source is extrinsic motivation (refers to Table 4), it refers to the perception that users would and want to do an activity because it is considered a valuable role in achieving a different result from the activity itself, such as improved job performance, earnings, or promotions (Davis *et al*., 1992). This variable has M =3.680 and SD = 0.680.

Table 4: Summary of extrinsic motivation data distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.0	4	2.0	2.0	2.0
	3.0	67	33.5	33.5	35.5
	4.0	118	59.0	59.0	94.5
	5.0	11	5.5	5.5	100.0
	Total	200	100.0	100.0	

3.3. Job Fit

The third source is job fit with M=3.675 and SD=0.6720. This variable describes how to improve individual performance base on the system capabilities (Thompson *et al*. 1991). Table 5 is the result of data distribution of job fit.

Table 5: Summary of job-fit data distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.0	5	2.5	2.5	2.5
	3.0	73	36.5	36.5	39.0
	4.0	104	52.0	52.0	91.0
	5.0	18	9.0	9.0	100.0
	Total	200	100.0	100.0	

3.4. Relative advantage

The fourth source is relative advantage with M=3.685 and SD=0.7673. It refers to the extent to which use of an innovation is considered to be better than using its predecessor (Compeau and Higgins 1995b; Compeau *et al*. 1999). Table 6 portrays the result of distribution data.

Table 6: Summary of relative advantage data distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.0	3	1.5	1.5	1.5
	2.0	14	7.0	7.0	8.5
	3.0	40	20.0	20.0	28.5
	4.0	129	64.5	64.5	93.0
	5.0	14	7.0	7.0	100.0
	Total	200	100.0	100.0	

3.5. Outcome Expectations

The last source is outcome expectations (refers to Table 7), this variable refers to dealing with the consequences of behavior, based on empirical evidence, is separated into performance expectations (job-related) and personal expectations (individual goals) (Compeau and Higgins 1995b; Compeau *et al.* 1999). Table 6 representative of the data distribution include frequency and percent.

Table 7: Summary of outcome expectations data distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2.0	40	20.0	20.0	20.0
	3.0	91	45.5	45.5	65.5
	4.0	64	32.0	32.0	97.5
	5.0	5	2.5	2.5	100.0
	Total	200	100.0	100.0	

4. Experiment Results

In order to apply the proposed technique, a prototype implementation system is developed using MATLAB version 7.6.0.324 (R2008a). The algorithm is executed sequentially on a processor Intel Core 2 Duo CPUs. The total main memory is 1G and the operating system is Windows XP SP3.

There are five attributes of e government performance expectancy; Perceived Usefulness (PU), Extrinsic Motivation (EM), Job-Fit (JF), Relative Advantage (RA), Outcome Expectations (OE). The MDA result is shown in Table 8. The selected attribute is Job-fit with the value 0.075.

Table 8: MDA results of e government performance expectancy dataset

Attribute (w.r.t.)	Mean of Attributes Dependency				Max
	EM	JF	RA	OE	
PU	0.02	0	0.015	0.025	0.025
	PU	JF	RA	OE	
EM	0.005	0	0.015	0	0.015

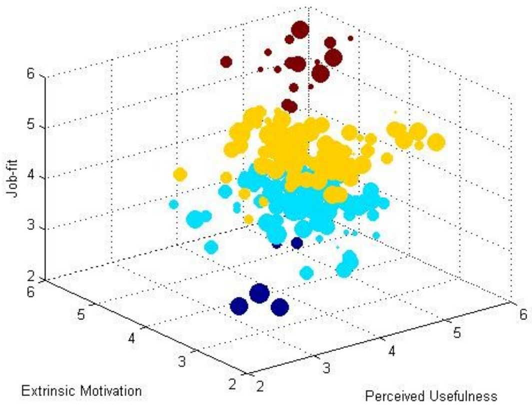
	PU	EM	RA	OE	
JF	0.005	0.075	0.015	0	**0.075**
	PU	EM	JF	OE	
RA	0.005	0.055	0.025	0.025	0.055
	PU	EM	JF	RA	
OE	0.005	0.02	0	0.015	0.02

From Table 8, we can see that attribute Job-fit has the highest dependency degree. Therefore we select it as a clustering attribute and consequently we have four clusters as described in Table 9.

Table 9: MDA results of e government performance expectancy dataset

Cluster Number	Number of Objects
1	5
2	73
3	104
4	18

The visualization of the clusters is captured in Figure 2 as follow.



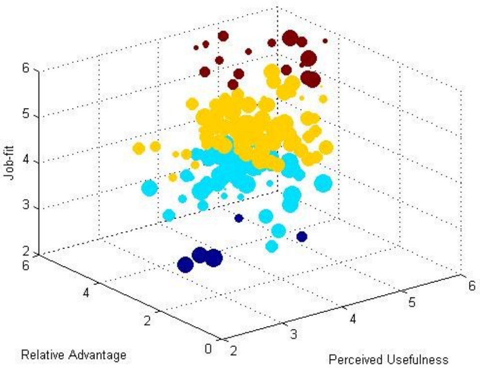


Figure 2: Cluster visualization

5. Conclusion

Computational model can be explored as an efficient clustering technique for grouping e-government users. This paper has presented an application of rough set theory for clustering performance expectancy of e-government user. The maximum dependency of attribute has been used as attribute selection to study performance expectancy. The technique is based on the highest dependency of attributes I information system. We elaborate the technique approach through performance expectancy of Indonesian e-government dataset which consist of five variable sources among people in Bandung, i.e., perceived usefulness, extrinsic motivation, job fit, relative advantage, and the last outcome expectations. The results show that variable precision rough set can be used to groups people in each study's performance expectancy.

Acknowledgments. This work was supported by Graduated Research Assistant under Geran Kontrak vot U559 Universiti Tun Hussein Onn Malaysia.

References

[1] Adams, D. A; Nelson, R. R.; Todd, P. A. (1992), "Perceived usefulness, ease of use, and usage of information technology: A replication", MIS Quarterly 16: 227–247, doi:10.2307/249577

[2] Barney, J., & Hansen, M. (1994). Trustworthiness as a source of competitive advantage", *Strategic Management Journal*, 15, 175–190.

[3] Carter, L., & Bélanger, F. (2005). The utilization of e-government services: Citizen trust, innovation and acceptance factors. *Information Systems Journal*, 15(1), 5–25.

[4] Carter, L. and F. Belanger, "Citizen adoption of electronic government initiatives", 37th Hawaii International Conference on System Sciences, Hawaii, 2004.

[5] Carter, L. and F. Belanger, "Diffusion of innovation & citizen adoption of e-government", The *Fifth International Conference on Electronic Commerce* (ICECR-5), Pittsburg, PA, 2003, hal.57-63.

[6] Carter, L. and F. Belanger, "Trust, and risk in e-government adoption", *The Journal of Strategic Information Systems,* vol. 17, no. 2, hal. 165- 176, 2008.