Mohamed M. Mostafa[1]

CITIZENS AS CONSUMERS: PROFILING E-GOVERNMENT SERVICES' USERS IN EGYPT VIA DATA MINING TECHNIQUES

Abstract

This study uses data mining techniques to examine the effect of various demographic, cognitive and psychographic factors on Egyptian citizens' use of e-government services. Multi-layer perceptron neural network (MLP), probabilistic neural network (PNN), classification and regression trees (CART), and multivariate adaptive regression splines (MARS) are compared to a standard statistical method (linear discriminant analysis (LDA). The variable sets considered are sex, age, educational level, e-government services perceived usefulness, ease of use, compatibility, subjective norms, trust, civic mindedness, and attitudes. The study shows how it is possible to identify various dimensions of e-government services usage behavior by uncovering complex patterns in the dataset, and also shows the classification abilities of data mining techniques.

Keywords

Consumer profiling, Data mining, E-government services, Egypt, Neural networks

1. Introduction

One of the most intractable problems for anyone dealing with government is the sheer complexity of its organizational structure. For example, it has been estimated that the average government has between 50 and 70 different departments, agencies and regulatory bodies (Silcock, 2001). A number of government's different agencies may be involved in simple matters such as registering the birth of a child. In several countries there has been a growing pressure for governments to move online. In the Arab world, Dubai pioneered e-voting in elections for half the members of the United Arab Emirates' consultative assembly (The Economist, 2008). In Bahrain the e-government authority of Bahrain (E-GA) has recently launched the Enterprise Architecture Project initiative (EAP), which is considered to be the first of its kind in the Arab world. The initiative aims at streamlining government procedures by unifying the standards and procedures among all government entities in all matters related to information communication technology (Bahrain Tribune, 2009). Finally, in Egypt e-government currently provides 85 services to citizens including government forms, public policy information and tax filing (Hamed, 2008). Two main reasons are behind governments' decision to move online. First, a more enlightened view has begun in the ranks of government to treat the citizen like a consumer where transaction satisfaction is important. Second, pressures for governments to do more with less will force governments to provide services in a more efficient way. In fact, e-government offers substantial performance gains over the traditional model of government. For example, based on the analysis of 49

[1] Gulf University for Science and Technology, West Mishref, Kuwait, e-mail: moustafa.m@gust.edu.kw

empirical studies, Danziger and Andersen (2002) concluded that there were positive e-government impacts on data access and efficiency and productivity of government performance in both internal operations and external service functions. In fact it has been argued that a significant portion of the benefits created by e-government services are obtained by the government itself in terms of efficiency gains (Tung and Rieck, 2005). For example, the U.S. government generates around US$ 3 billion on its Web site (Clark, 2003).

Profiling e-government services users is very important because the first step in planning the target marketing strategy is to segment the market and develop profiles of the resulting market segments. In fact, the usefulness of market segmentation hinges upon accurate profiling.

2. Literature review and hypotheses development

2.1. Perceived usefulness

The perceived benefit factor is closely related to perceived usefulness in the TAM theoretical model. Raman and Leckenby (1998) used the concept of utilitarianism to explain online behavior. They found a positive link between utilitarianism and duration of visit of web ads. This construct, too, seems to be closely related to perceived usefulness identified in TAM. Rogers (1995), in his diffusion of innovation paradigm, also posits that the perceived benefit or relative advantage of innovation positively influences adoption rate. In a meta-analysis in the innovation research literature, Tornatzky and Klein (1982) concluded that relative advantage was positively related to adoption. In a similar vein, King and He (2006), in a meta-analysis of the TAM, found a strong positive link between perceived usefulness and behavioral intention ($\beta = 0.505$). It follows that

H1: perceived usefulness of e-government services positively influences users' intention to use these services.

2.2. Perceived ease of use

Perceived ease of adoption can affect adoption behavior since an innovation that is easy to use can considerably reduce the time and effort required by the user and, thus, increase the likelihood of adopting the technology. Most studies on technology acceptance showed that perceived ease of use directly influenced attitude towards use (e.g., Ahn et al., 2004; Bruner and Kumar, 2005; Chen et al., 2002). King and He (2006), in a meta-analysis of the TAM, found a strong positive link between perceived ease of use and behavioral intention ($\beta = 0.186$). In a study of technology adoption in government agencies, Gilbert, Balestrini and Littleboy (2004) found a statistically significant association between perceived ease of use and attitude, indicating the important role of the ease of use in the formation of users' attitudes. It follows that

H2: perceived ease of use of e-government services positively influences users' intention to use these services.

2.3. Compatibility

Prior studies indicated that compatibility had strong direct impact on behavioral intention in areas such as using group support systems (Van Slyke et al., 2002), adopting new methodology for software development (Hardgrave et al., 2003) and using university smart card systems (Lee and Cheng, 2003). In a recent study of e-payment adoption in China, He et al. (2006) found that only compatibility has a significant effect on respondents' intention to adopt the system. Compatibility may also influence behavioral intention through performance expectancy and effort expectancy (Schaper and Pervan, 2007). For example, Lee and Cheng (2003) showed that compatibility of telemedicine technology exerted a significant effect on perceived usefulness. It follows that

H3: perceived compatibility of e-government services positively influences users' intention to use these services.

2.4. Subjective norms

Subjective norm (also called social norm) refers to users' perception of whether other important people perceive they should engage in the behavior (Schepers and Wetzels, 2007). While TAM does not include subjective norm, the theory of reasoned action (TRA) identifies attitudes and subjective norms as the sole determinants of behavioral intention (Fishbein and Ajzen, 1975). The theory of planned behavior (TPB), an update of TRA, also included subjective norms. Venkatesh and Davis (2000) acknowledged this and updated the TAM (TAM2) by integrating subjective norms. Several studies found a positive relationship between subjective norms and behavioral intention (e.g. Yi et al., 2006; Lu et al., 2009). In a study examining culture-specific enablers and impediments to the adoption and use of the Internet in the Arab world, Loch et al. (2003) found that both social norms and the degree of technological culturation can impact the individual and organizational acceptance and use of the Internet. It follows that

H4: subjective norms positively influence users' intention to use e-government services.

2.5. Trust

Prior empirical research incorporated trust into TAM in several ways. For example Shih (2003) extended TAM by adding the perceived Web security construct and found that high perceived Web security directly increases consumer attitudes towards e-shopping. Results also support trust as an antecedent of usefulness (Pavlou, 2003), ease of use (Pavlou, 2003), attitude (Chen and Tan, 2004), and behavioral intention (Pavlou, 2003). Few studies explored the role of trust in e-government adoption. For example, Kim et al. (2008), found that trust in the organization using the technology and trust in government as responsible for the introduction of electronic services are important determinants of e-government services adoption. It follows that

H5: trust in e-government systems positively influence users' intention to use e-government services.

2.6. Civic mindedness

The concept of civic mindedness is central to any analysis of e-government services adoption (Dermody and Hanmer-Lloyd, 2004). Civic mindedness encompasses three aspects: social contact, prior interest in government, and media use for public affairs (Dimitrova and Chen, 2006). As cyberdemocracy represents an extension of democracy into the realm of information technology and electronic communication, it is expected that the use of electronic means by citizens to interact with government to be an extension of their civic and political involvement via traditional channels (Katchanovski and La Porte, 2005). Prior research on e-government suggests that e-government users are similar to those who use government traditional services and are more engaged in civic affairs (Dimitrova and Chen, 2006). It follows that

H6: civic mindedness positively influences users' intention to use e-government services.

2.7. Attitudes

The social psychology literature on behavioral research has established attitudes as important predictors of behavior, behavioral intention, and explanatory factors of variants in individual behavior (Kotchen and Reiling, 2000). Attitude is defined as an individual's overall evaluation of performing a behavior (Lu et al., 2009). Prior e-services research has established a positive link between attitudes and behavioral intention (e.g., Aggelidis and Chatzoglou, 2009; Agarwal et al., 2000). It follows that

H7: attitude towards e-government services positively influences users' intention to use e-government services.

3. Method

3.1. Sample

The empirical study involved the administration of self-completion questionnaire to citizens in three Egyptian cities. Data were collected using the drop-off, pick-up method (Craig and Douglas, 1999). The effective sample size, thus, was 776 with a response rate of 52%.

3.2. Measures

All questionnaire items, originally published in English, were translated into Arabic using the back translation technique (Brislin, 1986).

4. Results

To compare e-government services users and non-users the traditional LDA was used using the SPSS 16.0 package. In order to assess the overall fit of the discriminant function classification results were examined. In combination, the discriminant function achieved 94.3% classification accuracy.

Given its usefulness in data mining (Smith and Gupta, 2000), MLP is a logical choice for the problem studied here. Following Lim and Kirikoshi (2005), a quasi-Newton algorithm with weight updates occurring after each epoch was used for MLP training. The learning rate was set at 0.1. After 100 iterations the correct classification rate (CCR) reached 99.8% as seen in Figure 1. As can be observed, the MLP classifier predicted training sample with 99.8% accuracy and validation sample with 98.3% accuracy.

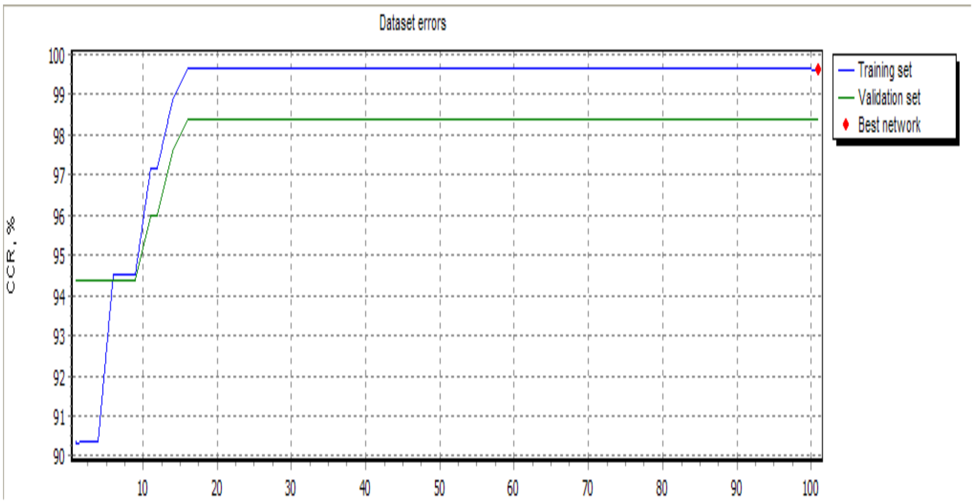

Figure 1: *Correct classification rate (CCR) for the MLP neural network*

The PNN classifier predicted both training and testing samples with 100% accuracy. CART is a nonparametric technique developed by Breiman et al. (1984) to classify group observations based on a set of characteristics into distinct groups, using the decision tree methodology. The technique was introduced to overcome the inherent limitations in the automatic interaction detector (AID) and the chi-square automatic interaction detector (CHAID) techniques. Unlike AID or CHAID, CART can work in classification tree mode withy categorical predictor variables, or in regression tree mode with interval or ratio scaled predictors. CART recursively splits a dataset into non-overlapping subgroups based on the independent variables until splitting is no longer possible (Baker and Song, 2008).

Following D'Alisa et al. (2006), the 10-fold validation approach with re-substitution was adopted. This consists of simulating 10 different samples by subtracting randomly each time 10% of the subjects and duplicating randomly another 10%. After each run, the original sample is restored. The final tree represents the best trade-off between variance explanation and variance stability across 10 "different" samples. Overall correct classification rate obtained from CART was 99.48%. Figure 2 depicts the final obtained pruned CART tree. From this figure we see that trust in e-government systems plays the most important role in rule induction.

MARS is a relatively novel data mining technique developed by Friedman (1991). This technique combines classical linear regression, mathematical construction of splines and binary recursive partitioning to produce a local model where the relationship between response and predictors are either linear or nonlinear through approximating the underlying function through a set of adaptive piecewise linear regression termed basis functions (BF) (Jesus and Angel, 2004). The power of MARS for building prediction and classification models has been demonstrated in many applications such as information technology productivity studies (Ko and Osei-Bryson, 2006), and genetics (York and Eaves, 2001). In this study we used MARS 2.0 package (Steinberg et al., 1999) to conduct the analysis. Overall correct classification rate obtained from MARS was 99.10% (sensitivity = 0.931 and specificity = 0.997).

5. Hypotheses testing

Group means for user and non-user groups on each of the independent variables used in the LDA were first calculated. Given the unequal group sizes of user versus non-user groups, group specific covariance matrices were used. Review of the significance levels of the individual attributes revealed that all the attributes except demographic variables displayed significant differences between the group means. Our results confirm previous studies which found e-government usage behavior to be correlated with perceived usefulness, ease of use, compatibility, trust, subjective norms and favorable attitudes to use the system (e.g. Helbig et al., 2009). For example Gilbert et al. (2004) found that trust is one of the strongest predictors of willingness to use e-government services.

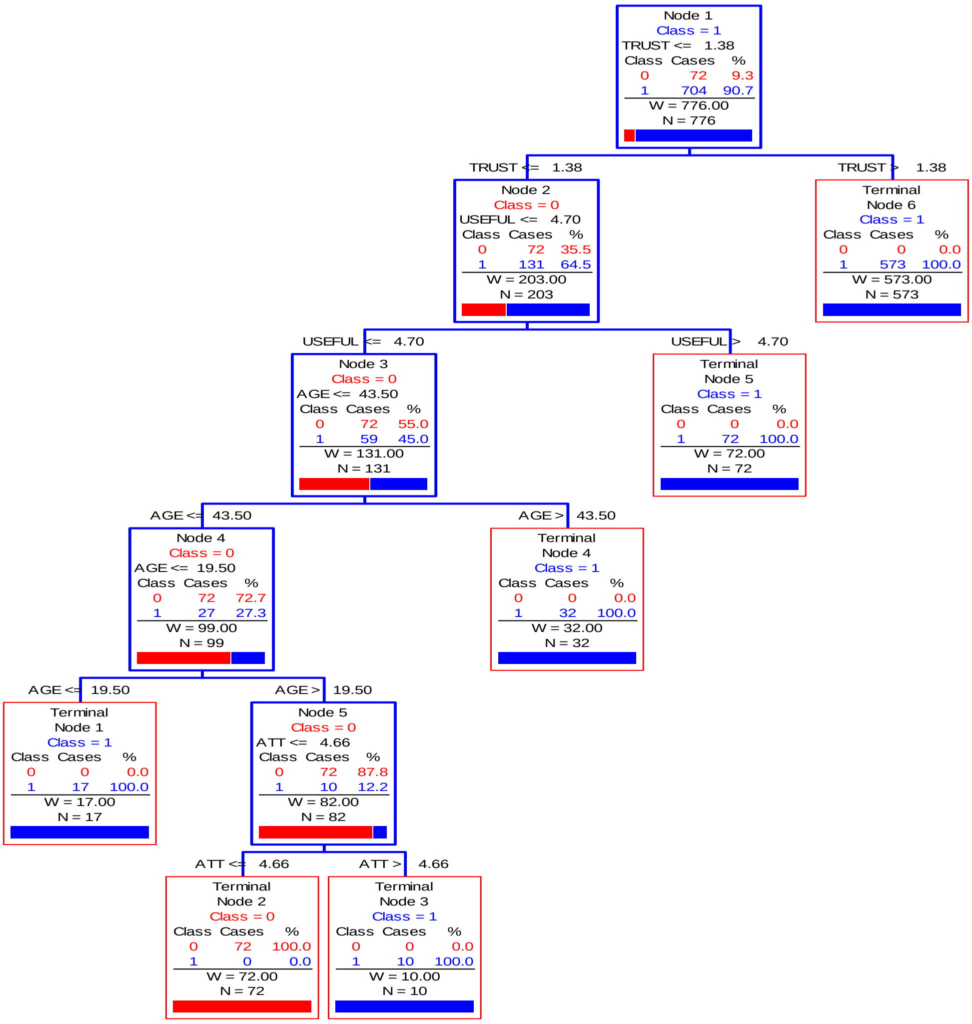

Figure 2: *CART decision tree of e-government services' users*

The finding that subjective norm is significant in predicting e-government services usage is consistent with Hofstede's (1991) cultural dimensions. In a high power distance collectivist culture such as Egypt one would expect significant others' opinions to have more impact on the individual because of face saving and group conformity, also a higher power distance would invoke a more influential role of peers.

It should be noted that we tested the research hypotheses using LDA as data mining techniques are of limited ability to statistically test and interpret hypotheses concerning the roles of specific variables that are included in the models as predictors. However, this limitation "should not be a serious drawback if one simply desires classification from the model" (Swicegood and Clark, 2001, p. 176).

6. Implications

The results of this study have several important implications for both theory and practice. From a theoretical perspective, the superior performance of data mining techniques found in this study confirms the theoretical work by Hecht-Nielson (1989) who has shown that machine learning techniques, such as neural network models, can learn input-output relationships to the point of making perfect forecasts with the data on which the model is trained. However, perfect forecasts with the training data do not guarantee optimal forecasts with the testing data due to differences in the two data sets. Our results also corroborate the findings of other researchers who have investigated the performance of machine learning techniques compared to other traditional statistical techniques, such as regression analysis, discriminant analysis, and logistic regression. For example, in a study of clinical diagnosis of cancers, Shan et al. (2002) found a hit ratio of 85% for the PNN model compared to 80% for the LDA model. In a study of credit-scoring models used in commercial and consumer lending decisions. Trust was found to be one of the most important factors in determining e-government services usage. Because governmental agencies may be required by law to share information with other agencies, the need for trust in the maintenance of accurate citizen information will increase. Thus, a strategic aim could be to develop a trustworthy relationship with the public, giving assurances that their data will be secure, and that the information contained on the Web would be both current and accurate. This can be done through tools and techniques that Web developers can use to increase and promote the security of e-government Web sites, such as firewalls and encryption technology. Therefore, e-government services need to be user-friendly, and citizens need to have confidence in the system. In this process, the government need to be careful to protect its brand and credibility.

7. Bibliography

1. Agarwal, J., Sambamurthy, V. and Stair, R. (2000). The evolving relationship between general and specific computer self-efficacy: an empirical assessment. Information Systems Research, 11(4), 418–430.
2. Aggelidis, V. and Chatzoglou, P. (2009). Using a modified technology acceptance model in hospitals. International Journal of Medical Informatics, 78(2), 115–126.
3. Ahn, T., Ryu, S. and Han, I. (2004). The impact of online and offline features on the user acceptance of Internet shopping malls. Electronic Commerce Research and Applications, 3(4), 405–420.
4. Bahrain Tribune (18.01.2009). E-GA launches new initiative, 3.
5. Baker, J. and Song, J. (2008). Exploring decision rules for sellers in business-to-consumer (B2C) Internet auctions. International Journal of E-Business Research, 4(1), 1–21.
6. Brieman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and regression trees. Belmont (CA): Wadsworth International Group.
7. Brislin, R. (1986). The wording and translation of research instruments. In W. Lonner and J. Berry (eds.), Field methods in cross-cultural research. Beverly Hills (CA): Sage.
8. Bruner, G. and Kumar, A. (2005). Explaining consumer acceptance of handheld Internet devices. Journal of Business Research, 58(5), 553–558
9. Chadwick, S. (2001). Communicating trust in e-commerce transactions. Management Communication Quarterly, 14(4), 653–658.
10. Chen, L., Gillensen, M. and Sherrell, D. (2002). Enticing online consumers: an extended technology acceptance perspective. Information & Management, 39(8), 705–719.
11. Clark, E. (2003). Manging the transformation to e-government: an Australian perspective. Thunderbird International Business Review, 45(4), 377–397.
12. Craig, C, and Douglas, S. (1999). International Marketing Research (2nd ed.). Indianapolis: John Wiley & Sons.
13. D'Alisa, S. et al. (2006). Depression is the main determinant of quality of life in multiple scelerosis: a classification-egression (CART) study. Disability and Rehabilitation, 28(5), 307–314.
14. Danziger, J. and Andersen, K. (2002). The impacts of information technology on public administration: an analysis of empirical research from the "golden age" of transformation. International Journal of Public Administration, 25(5), 591–627.

15. Dermody, J. and Hanmer-Lloyd, S. (2004). Segmenting youth voting behavior through trusting-distrusting relationships: a conceptual approach. International Journal of Nonprofit and Voluntary Sector Marketing, 9(3), 202–217.
16. Dimitrova, D. and Chen, Y. (2006). Profiling the adopters of e-government information services. Social Science Computer Review, 24(2), 172–188.
17. Fishbein, M. and Ajzen, I. (1975). Belief, attitude, intention and behaviour: an introduction to theory and research. Reading (MA): Addison-Wesley.
18. Friedman, J. (1991). Multivariate adaptive regression splines. Annals of Statistics, 19(1), 1–141.
19. Gilbert, D., Balestrini, P. and Littleboy, D. (2004). Barriers and benefits in the adoption of e-government. International Journal of Public Sector Management, 17(4), 286–301.
20. Hamed, T. (30.12.2008). E-government provides 85 services to citizens. El-Gomhouria, 9.
21. Hardgeave, B., Davis, F. and Riemenschneider, C. (2003). Investigating determinants of software developers' intention to follow methodologies. Journal of Management Information Systems, 20(1), 123–152.
22. He, Q., Duan, Y., Fu, Z. and Li, D. (2006). An innovation adoption study of online e-payment in Chinese companies. Journal of Electronic Commerce in Organizations, 4(1), 48–69.
23. Hecht-Nielson, R. (1989). Theory of the back-propagation neural network. Washington (DC): International Joint Conference on Neural Networks.
24. Hofstede, G. (1991). Cultures and organizations: software of the mind. London: McGraw Hill.
25. Jesus, M. and Angel, F. (2004). Comparison of statistical methods commonly used in predictive modelling. Journal of Vegetation Science, 15(2), 285–292.
26. Katchanovski, I. and La Porte, T. (2005). Cyberdemocracy or potemkin e-villages? Electronic governments in OECD and post-communist countries. International Journal of Public Administration, 28(7/8), 665–681.
27. Kim, D., Ferrin, D. and Rao, H. (2008). A trust-based decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decision Support Systems, 44(2), 544–564.
28. King, W. and He, J. (2006). A meta-analysis of the technology acceptance model. Information & Management, 43(6), 740–755.
29. Ko, M. and Osei-Bryson, K. (2006). Analyzing the impact of information technology investments using regression and data mining techniques. Journal of Enterprise Information Management, 19(4), 403–417.
30. Kotchen, M. and Reiling, S. (2000). Environmental attitudes, motivations, and contingent valuation of nonuse values: a case study involving endangered species. Ecological Economics, 32(1), 93–107.
31. Lee, C. and Cheng, Y. (2003). Comparing smart card adoption in Singapore and Australian universities. International Journal of Human-Computer Studies, 58(3), 307–325.
32. Lim, C. and Kirikoshi, T. (2005). Predicting the effects of physician-directed promotion on prescription yield and sales uptake using neural networks. Journal of Targeting, Measurement and Analysis for Marketing, 13, 156-167
33. Loch, K., Straub, D. and Kamel, S. (2003). Diffusing the Internet in the Arab world: the role of social norms and technological culturation. IEEE Transactions on Engineering Management, 50(1), 45–63.
34. Lu, Y., Zhou, T. and Wang, B. (2009). Exploring Chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. Computers in Human Behavior, 25(1), 29–39.
35. Pavlou, P. (2003). Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. International Journal of Electronic Commerce, 7(3), 101–134.
36. Raman, N. and Leckenby, J. (1998). Factors affecting consumers' "Webad" visits. European Journal of Marketing, 32(7/8), 737–748.
37. Rogers, E. (1995). Diffusion of innovation (4th ed.). New York: Free Press.
38. Schepers, J. and Wetzels, M. (2007). A meta-analysis of the technology acceptance model: investigating subjective norm and moderation effects. Information & Management, 44, 90–103.
39. Shan, Y., Zhao, R., Xu, G., Liebich, H. M. and Zhang, Y. (2002). Application of probabilistic neural network in the clinical diagnosis of cancers based on clinical chemistry data. Analytica Chimica Acta, 471(1), 77–86.
40. Shih, H. (2003). An empirical study on predicting user acceptance of e-shopping on the Web. Information & Management, 41(3), 351–368.
41. Silcock, R. (2001). What is e-government?. Parliamentary Affairs, 54, 88–101.
42. Smith, K. and Gupta, J. (2000). Neural networks in business: techniques and applications for the operations researcher. Computers & Operations Research, 27, 1023–1044.
43. Swicegood, P. and Clark, J. (2001). Off-site monitoring systems for prediction bank underperformance: a comparison of neural networks, discriminant analysis, and professional human judgment. International Journal of Intelligent Systems in Accounting, Finance & Management, 10(3), 169–186.

44. The electronic bureaucrat: a special report on technology and government. (2008). The Economist, 386(8567), 1–18.
45. Tornatzky, L. and Klein, K. (1982). Innovation characteristics and innovation adoption-implementation: a meta-analysis of findings. IEEE Transactions on Engineering Management, 29(1), 28–43.
46. Tung, F., Chang, S. and Chou, C. (2008). An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. International Journal of Medical Informatics, 77(5), 324–335.
47. Van Slyke, C., Lou, H. and Day, J. (2002). The impact of perceived innovation characteristics on intention to use groupware. Information Resources Management Journal, 15(1), 5–12.
48. Venkatesh, V. and Davis, F. (2000). A theoretical extension to the technology acceptance model: four longitudinal field studies. Management Science, 46(2), 186–204.
49. Yi, M., Jackson, J., Park, J. and Probst, J. (2006). Understanding information technology acceptance by individual professionals: toward an integrative view. Information & Management, 43(3), 350–363.