CLASSIFICATION OF E-CUSTOMER SESSIONS BASED ON SUPPORT VECTOR MACHINE

Grażyna Suchacka,
Magdalena Skolimowska-Kulig
Institute of Mathematics and Informatics
Opole University
ul. Oleska 48, 45-052, Opole, Poland
E-mail: gsuchacka@math.uni.opole.pl,
mskolimow@math.uni.opole.pl

Aneta Potempa
GEA Technika Cieplna Sp. z o.o.
ul. Kobaltowa 2, 45-641, Opole
E-mail: aneta.potempa@gea.com

KEYWORDS

Data mining, web mining, web usage mining, Support Vector Machine, SVM, supervised classification, e-commerce, online store, Web store, Web users, Web traffic, log file analysis, R-project.

ABSTRACT

A key feature of high-traffic e-commerce sites is the ability to offer a predictive and personalized service to Web users. Visitors to online stores are potential buyers but in reality very few visits finally result in a product purchase. Thus, it would be especially valuable for online retailers to predict buyers against browsers based on some session features (e.g. session duration, the number of downloaded pages, the kind of realized Web interactions) and some HTTP-level information (the number of HTTP requests, the volume of data transfer in session). In this paper, we recast online purchase predictions as a classification problem. Every user session in a web store is represented as a 23-element vector in the session feature space. Based on historical data from an online bookstore an SVM classification model is proposed, dividing user sessions into two classes: browsing sessions and buying sessions. The best SVM classifier proved to be very effective, with a predictive accuracy of over 99% and the probability of predicting a buying session of almost 95%.

INTRODUCTION

Along with the ubiquity of the Internet, enhanced by the rapid development of mobile applications and secure electronic payment technologies, electronic commerce has been becoming the every-day reality for more and more people. E-customers are increasingly demanding and expect the highest quality service at any time. On the other hand, online retailers need to be able to know their customers and to predict their needs, which is largely made possible by online and offline web analytics software.

At the same time, due to its high practical potential, web usage mining (WUM) has become an active research area, proposing various data mining methods for extracting useful information on user behavior based on historical data. WUM methods have been applied to effectively manage relationships with e-customers in respect of customer identification, attraction, retention, and development (Kitayama et al. 2002), (Xu and Wang 2011). Patterns hidden in customer navigation paths on websites and in purchase transactions have also been intensively explored to build e-commerce information systems (Duan et al. 2012), to design effective product recommendation stategies (Cho et al. 2013), (Kuang and Li 2014), as well as to identify users with high purchasing intentions and to predict sales (Mohammadnezhad and Mahdavi 2012), (Poggi et al. 2007), (Suchacka and Chodak 2013).

In this paper, we consider the problem of predicting buying sessions in a web store. To this end, we propose applying one of popular data mining methods: Support Vector Machine (SVM). It is a powerful prediction model based on supervised classification, which makes it possible to recognize complex patterns in high-dimensional data sets (Vapnik and Chervonenkis 1974), (Cortes and Vapnik 1995). So far it has been successfully applied to many areas of the web. Examples of SVM applications in text and hypertext document classification include Web page classification (Li et al. 2004), (Li et al. 2001), keyword extraction (Zhang et al. 2006), document metadata extraction (Han et al. 2003), and website trust assessment (Soiraya et al. 2008). SVM was also applied in content-based multimedia classification and retrieval (Wang et al. 2011), (Mandel et al. 2006), (Guo and Li 2003), network intrusion detection (Zhang et al. 2011), (Du et al. 2009), and quality of web service management (Liu and Wang 2009), (Ali et al. 2012).

SVM classifiers have also been successfully applied in electronic commerce. In (Zhu and Zhang 2010) SVM was integrated with principal component analysis to manage credit in electronic commerce. In (Yu et al. 2011) and (Chen et al. 2012) customer churn forecasting frameworks based on SVM were proposed to accurately forecast and prevent e-customer churn. SVM methods proved effective in supporting customers in online shopping through improving product recommendation techniques (Cheung et al. 2000), (Xia et al. 2006), supporting decision-making based on mining e-customer reviews (Soliman et al. 2012), and designing intelligent software agents that perform product selection and associated functions on behalf of their clients (Cui 2003).

In (Hop 2013) an SVM classifier was applied to predict the probability of making an online purchase, like in

Proceedings 29th European Conference on Modelling and Simulation ©ECMS Valeri M. Mladenov, Petia Georgieva, Grisha Spasov, Galidiya Petrova (Editors)
ISBN: 978-0-9932440-0-1 / ISBN: 978-0-9932440-1-8 (CD)

our study. However, their research was based on high-level transaction data, including product prices, delivery availability of the products, and detailed e-customer data (lifetime of the customer account, the customer age, last order date, and so on). In contrast, we reconstruct customer visits only from low-level HTTP data recorderd in web server logs. Moreover, we compare SVM classification models with four various kernel functions (radial, linear, polynomial, and sigmoid), and show that the choice of a kernel is of key importance for the effectiveness of the SVM classifier.

SUPPORT VECTOR MACHINE

The goal of the supervised learning classification is to construct a classifier (a learner) based on the training data set which allows one to divide future observations into considered classes. The training data set consists of pairs of an input object (a vector of feature measurements) and an output value (a label of a class containing an observation): $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, N$, where $\mathbf{x}_i \in \mathbb{R}^p$ is a vector of observed features and y_i is a label (in case of two classes $y_i \in \{0, 1\}$).

The evaluation of the accuracy of the classifier and the value of possible classification error is based on a test data set (separated from the training set) consisting of objects with known labels.

One of the methods of supervised learning classification is the maximal margin classifier. The idea of the algorithm invented by (Vapnik and Chervonenkis 1974) is to find the hyperplane that creates the biggest margin between the training points for considered two classes, i.e. a good separation is achieved by the hyperplane that has the largest distance from the nearest training data point of any class. However, since in many cases no separating hyperplane exists or it could be worthwhile to misclassify a few observations to improve the classifying of the remaining ones, the task of margin maximalization needs modifications. The support vector machine suggested by (Cortes and Vapnik 1995) is such an extension of the maximal margin classifier. The SVM classifier finds the maximal margin separating hyperplane for observations from the training data set transformed by φ into the space of the higher dimension (sometimes infinite) with the constraint on the number of misclassifications. The optimization problem which yields to that classifying rule is to minimize the Lagrangian:

$$L(\alpha) = \frac{1}{2}\alpha'\mathbf{Q}\alpha - \mathbf{e}'\alpha$$

subject to $0 \leq \alpha_i \leq C$ and $\mathbf{y}'\alpha = 0$, where $\mathbf{e}$ is the unity vector, C is the upper bound of the number of misclassifications (so-called *cost*), $\mathbf{Q}$ is the matrix of $Q_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ and $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)'\varphi(\mathbf{x}_j)$ is the kernel function, $i, j = 1, 2, \ldots, N$. Due to the application of the kernel function, there is no need to obtain the exact form of the transformation φ. The solution of the optimization problem depends on a subset of vectors of features - the support vectors. The most common kernels are of the form:

- *linear*
$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i'\mathbf{x}_j;$$
- *polynomial of degree d*
$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\mathbf{x}_i'\mathbf{x}_j + a)^d;$$
- *radial basis*
$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2\right);$$
- *sigmoid*
$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\mathbf{x}_i'\mathbf{x}_j + a\right);$$

where a, d and γ are the kernel function's coefficients.

An important issue in constructing the SVM classifier is the selection of parameters influencing the effectiveness and the generalization capability of the classifier: the cost parameter C that determines the penalty for misclassifications, the kernel function, and kernel function's parameters, influencing the function shape or dimensionality.

RESEARCH METHODOLOGY

We consider an online store to be available as a website where a web user may browse the site's pages, search for products, add them to a virtual shopping cart, and confirm a purchase. To build and evaluate SVM classifiers we used real data from commercial Web server access logs, recorded in April 2014. The server hosted the website of an online bookstore and some pages with entertainment content, like movies, quizzes, simple games, etc.

Reconstruction of User Sessions From Log File Data

Based on HTTP requests read from logs, user sessions were reconstructed. A single user session corresponds to a single visit of a web user to the online store. The session involves realization of various web interactions, e.g. downloading pages with product information, adding products to the shopping carts, etc. A key interaction is that connected with the confirmation of a purchase transaction - we call it the *checkout success* interaction.

A web user may be a human interacting with a bookstore site via a web browser or a web robot. We identified web users based on two fields of HTTP requests: the IP address and the user agent string. We assumed that subsequent sessions of the same user are separated by a minimum 30-minute interval of user inactivity.

Selection of Variables Characterizing User Sessions

Each user session in a web store may be characterized by a number of features connected with the session characteristics (session duration, the number of pages visited in session, the kind of web interactions realized in session, information on the source of the session), as well as with some HTTP-level information (the number of HTTP requests sent in session, volume of data transfer). We propose using these features as variables in the supervised learning classification of user sessions in order to predict purchases in the store. The following 23 variables were used for classification:

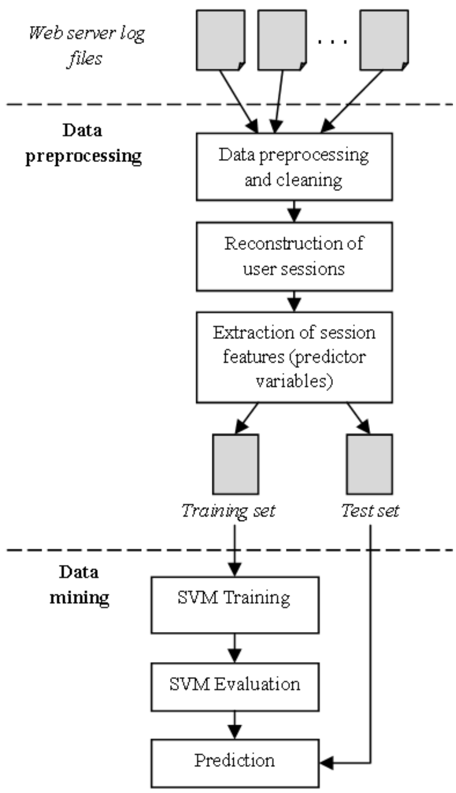
- *Checkout_success* - the variable indicating whether the session contained the *checkout success* interaction (1 if a purchase was realized successfully in session and 0 otherwise);
- *Checkout_try* - the number of Web interactions other than the *checkout success*, connected with the checkout process;
- *Home* - the number of visits to the home page of the web store;
- *Register_success* - the number of web interactions connected with a successful user registration (i.e. creating a user account) in the store;
- *Register_try* - the number of web interactions other than a successful user registration, connected with the registration process;
- *Login_success* - the number of web interactions connected with a successful user logging into the site;
- *Log_off* - the number of web interactions connected with a user logout;
- *Search* - the number of web searching interactions;
- *Browse* - the number of web browsing interactions;
- *Details* - the number of visited pages with detailed information about products;
- *Add* - the number of products added to the shopping cart;
- *Shipping_checkout* - the number of Web interactions corresponding to checking information on shipping terms and cost during the checkout process;
- *Shipping* - the number of web interactions corresponding to checking shipping information besides the checkout process;
- *Information* - the number of visited pages containing information about the store;
- *Entertainment* - the number of visited pages with the entertainment content;
- *Pages_no* - the number of all pages visited in session;
- *Requests_no* - the number of HTTP requests downloaded in session;
- *Transfer* - the volume of data (in kilobytes) downloaded in session;
- *Duration* - the session duration (in seconds);
- *Page_per_time* - the mean time per page (in seconds);
- *Source* - the source of the visit (a reference from paid search engine results, a natural search engine result, a reference from the e-mail newsletter, an entrance through social media sites, an internal reference from a page within the same website outside the web store, or other source);
- *Is_bot* - the variable indicating whether the session was performed by a web robot;
- *Is_admin* - the variable indicating whether the session was performed by the web site administrator (or the administrative software).

All of the above session features were considered in our SVM classification models.

Data Classification and Modeling

Classification of user sessions, aimed at predicting online purchases, involves distinguishing between two session classes, depending on the fact of whether the session containes the *checkout success* interaction or not (i.e. whether the variable *Checkout_success* is 1 or 0).

The data set of all observations (user sessions) was divided into two subsets: a training set and a test set. Then we applied the Support Vector Machine method to construct session classifiers. We used R-project, a free software environment for statistical computing (R-project) with e1071 package (e1071). This research was performed as part of the master's thesis (Potempa 2014). Firstly, for each SVM model a classifier was built and tuned to the highest possible accuracy based on the training set. Then, the performance of the classifier was estimated based on the test set. At the end, results of all SVM models were compared and the quality of the classifiers was assessed.



Figures 1: Analysis Process

DATA ANALYSIS RESULTS

The data set contained 39 000 observations. The training set was generated by drawing 26 000 observations (in this set the variable *Checkout_success* was 1 in 146 observations). The test set contained the remaining 13 000 observations (including 72 observations with the positive value of *Checkout_success*).

Model Tuning

The optimal values of SVM parameters, i.e. the cost parameter, the kernel function and the kernel function's parameters, can be determined by iteratively training SVM models in various configurations and selecting the best model. R-project allows one to look through a relatively large space of SVM parameters to select their optimal values and to find the optimal model for the given kernels.

Four kernel functions were applied in our SVM classification models: radial, linear, polynomial, and sigmoid. The optimal parameter values determined for each kernel function for the same training set are presented in Tab. 1. The SVM type was C-classification in all cases.

Table 1: Optimal Parameter Values and the Resultant Numbers of Support Vectors for the SVM Models with Different Kernel Functions

	SVM kernel			
	Radial	Linear	Polynomial	Sigmoid
Cost (C)	10	0.1	0.01	0.01
Gamma (γ)	0.5	-	1	0.5
Degree (d)	-	-	3	-
Coef.0 (a)	-	-	10	0
Number of support vectors	1186	91	42	293

Model Evaluation

We evaluate the performance of four SVM classification models using the notions of true and false positives and negatives, error rate, accuracy, and sensitivity.

Since we are interested in predicting online purchases, i.e. determining whether a session will eventually contain the *checkout success* interaction or not, a classification of *buying session* is considered to be positive and a classification of *browsing session* is considered to be negative. Depending on the accuracy of classification for the test set we can determine the following measures:

- *true positives* (*TP*) - the number of *buying sessions* which were correctly classified as *buying sessions*;
- *false positives* (*FP*) - the number of *browsing sessions* incorrectly classified as *buying sessions*;
- *true negatives* (*TN*) - the number of *browsing sessions* correctly classified as *browsing sessions*;
- *false negatives* (*FN*) - the number of *buying sessions* incorrectly classified as *browsing sessions*.

Classification results can be described with the confusion matrix of correct and incorrect classifications (Larose 2005). The columns represent the predicted classifications and the rows represent the actual classifications, for all sessions in the test set. The confusion matrices for optimal SVM models with different kernel functions are presented in Tab. 2-5.

Table 2: Confusion Matrix for the SVM Model with a Radial Kernel Function

	Sessions classified as *browsing sessions*	Sessions classified as *buying sessions*
Browsing sessions	12 928	0
Buying sessions	58	14

Table 3: Confusion Matrix for the SVM Model with a Linear Kernel Function

	Sessions classified as *browsing sessions*	Sessions classified as *buying sessions*
Browsing sessions	12 924	4
Buying sessions	4	68

Table 4: Confusion Matrix for the SVM Model with a Polynomial Kernel Function

	Sessions classified as *browsing sessions*	Sessions classified as *buying sessions*
Browsing sessions	12 923	5
Buying sessions	12	60

Table 5: Confusion Matrix for the SVM Model with a Sigmoid Kernel Function

	Sessions classified as *browsing sessions*	Sessions classified as *buying sessions*
Browsing sessions	12 928	0
Buying sessions	72	0

Let us consider the confusion matrix obtained for the SVM model with a radial kernel function (Tab. 2). This model gave 12 986 classifications (predictions) of a *browsing session*, i.e. 12 986 *negatives*, including 12 928 *true negatives* (*TN*) and 58 *false negatives* (*FN*). All of the 14 positive classifications were *true positives* (*TP*), in fact (there were no *false positives* - *FP* in this model).

Comparing the results obtained for all SVM models one can observe that the model with a linear kernel function was the most effective in predicting buyers: as many as 68 buying sessions were correctly classified and only four of such sessions were misclassified. The model with a polynomial kernel function was quite effective as well, resulting in 60 correct buyers' predictions. On the other hand, the models with radial and sigmoid kernels proved to be completely ineffective in this respect.

These results are confirmed by such evaluation measures as the overall error rate, accuracy, and sensitivity (tab. 6). The *error rate* can be determined as the sum of the *false positives* and *false negatives*, divided by the total number of classifications:

$$Error\ rate = \frac{FP{+}FN}{FP{+}FN{+}TP{+}TN}$$

The predictive accuracy of the model can be determined as the rate of all correct classifications:

$$Accuracy = \frac{TP{+}TN}{FP{+}FN{+}TP{+}TN}$$

In the e-commerce scenario, the ability to identify buyers against browsers would be especially desirable for an online retailer. Thus, the performance measure that is especially valuable from our point of view is the model sensitivity, which provides an estimate of the probability of predicting a buying session:

$$Sensisitivity = \frac{TP}{TP{+}FN}$$

The performance rates for optimal SVM models are presented in tab. 6. As one can observe, the quality of SVM classifiers is differentiated depending on the kernel function. The accuracy rates, reflecting the overall percentages of correct classifications, are very high for all SVM classifiers, ranging from 99.45% for a sigmoid kernel model to 99.94% for a linear kernel model. High accuracy rates correspond to low error rates, ranging from 0.06% to 0.55%. However, the models based on the radial and sigmoid kernel functions proved to be ineffective in predicting buying sessions, which is a key ability of a user session classifier in the e-commerce scenario.

Table 6: Assessment of the Quality of Classifiers for the SVM Models with Different Kernel Functions

Kernel function	Error rate [%]	Accuracy [%]	Sensitivity [%]
Radial	0.45	99.55	19.44
Linear	0.06	99.94	94.44
Polynomial	0.13	99.87	83.33
Sigmoid	0.55	99.45	0

To summarise, the best performing classification model was the SVM with a linear kernel function with the cost parameter of 0.1 which resulted in 91 support vectors.

CONCLUSIONS

Using data mining methods to predict online purchases is an important research area. We proposed applying Support Vector Machine to the classification of user sessions reconstructed from log file data for a Web store. We constructed and evaluated a few SVM classification models, dividing sessions into two classes: browsing sessions and buying sessions. Results show that the SVM classifier with a linear kernel was very effective both in respect of the overall predictive accuracy and the ability to predict buying sessions.

The SVM method is computationally very expensive, especially the time of training an SVM classifier is significant. However in the environment under consideration it can be done offline. Furhermore, it has to be noted that different web stores may reveal different user session characteristics and different navigational or purchasing patterns. Thus, the classifier obtained in our research is tuned for a given web store, and other web stores would require constructing and tuning their own classifiers (according to the same methodology as the one presented in this paper).

It would be interesting to integrate the proposed classifier into the process of managing active user sessions in the web store and to verify its efficiency in real time. Given an active user session, one could use the classifier to predict which class the active session belongs to and to offer a specially personalized service to the potential buyers, e.g. to provide them with a priority service during a server overload. We leave this issue to our future work. We also plan to compare results of the approach presented in this paper with other supervised classification methods, such as decision trees and a nearest neighbor method.

REFERENCES

Ali, W.; S.M. Shamsuddin; and A.S. Ismail. 2012. "Intelligent web proxy caching approaches based on machine learning techniques." *Decision Support Systems* 53, No.3 (Jun), 565-579.

Chen, Z.-Y.; Z.-P. Fan; and M. Sun. 2012. "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data." *European Journal of Operational Research* 223, No.2 (Dec), 461-472.

Cheung, W.K.; J.T. Kwok; M.H. Law; and K. Tsui. 2000. "Mining customer preference ratings for product recommendation using the support vector machine and the latent class model". In *Proceedings of the 2nd International Conference on Data Mining Methods and Databases for Engineering* (Cambridge, UK, Jul.). 601-610.

Cho, Y.S.; S.C. Moon; S.-p. Jeong; I.-B. Oh; and K.H. Ryu. 2013. "Clustering method using item preference based on RFM for recommendation system in u-commerce." *Lecture Notes in Electrical Engineering* 214, *Ubiquitous Information Technologies and Applications*, 353-362.

Cortes, C. and V. Vapnik. 1995. "Support-vector networks". *Machine Learning* 20, No.3 (Sep), 273-297.

Cui, D. 2003. "Product selection agents: a development framework and preliminary application." Doctoral Dissertation, Division of Research and Advanced Studies, University of Cincinnati.

Duan, Q; J. Li; and Y. Wang. 2012. "The application of fuzzy association rule mining in e-commerce information system mining." *Advanced Engineering Forum* 6-7 (Sep), 631-635.

Du, H.; S. Teng; and Q. Zhu. 2009. "Intrusion detection based on fuzzy support vector machines". In *Proceedings of the International Conference on Networks Security, Wireless Communications and Trusted Computing* (Wuhan, China, Apr.25-26), 639-642.

e1071: Misc Functions of the Department of Statistics (e1071), TU Wien, http://cran.r-project.org/web/packages/e1071/index.html

Guo, G. and S.Z. Li. 2003. "Content-based audio classification and retrieval by support vector machines." *IEEE Transactions on Neural Networks* 14, No.1 (Jan), 209-215.

Han, H.; C.L. Giles; E. Manavoglu; H. Zha; Z. Zhang; and E.A. Fox. 2003. "Automatic document metadata extraction using support vector machines". In *Proceedings of Joint Conference on Digital Libraries* (Houston, Texas, May 27-31), 37-48.

Hop, W. 2013. "Web-shop order prediction using machine learning." Masters Thesis, Erasmus University Rotterdam.

Kitayama, M.; R. Matsubara; and Y. Izui. 2002. "Application of data mining to customer profile analysis in the power electric industry". In *IEEE Power Engineering Society Winter Meeting* (Jan.27-31), Vol. 1, 632-634.

Kuang, G. and Y. Li. 2014. "Using fuzzy association rules to design e-commerce personalized recommendation system." *TELKOMNIKA Indonesian Journal of Electrical Engineering* 12, No.2 (Feb), 1519-1527.

Larose, D.T. 2005. *Discovering knowledge in data: An introduction to data mining*. Wiley.

Li, L.; W. Liu; Q. Xu; and Y. Zhu. 2004. "A topic-specific Chinese web page classifier based on support vector machine." *Journal of Computer Applications* 24, No.4, 58-61.

Li, X.L.; J.M. Liu; and Z.Z. Shi. 2001. "A Chinese web page classifier based on support vector machine and unsupervised clustering." *Chinese Journal of Computers* 24, No. 1, 62-68.

Liu, G. and H. Wang. 2009. "A method for real-time monitoring the service quality of website based on support vector machine." *Journal of Wuhan University* (*Natural Science Edition*), No.2, 239-243.

Mandel, M.I.; G.E. Poliner; and D.P.W. Ellis. 2006. "Support vector machine active learning for music retrieval." *Multimedia Systems* 12, No.1 (Aug), 3-13.

Mohammadnezhad, M. and M. Mahdavi. 2012. "Providing a model for predicting tour sale in mobile e-tourism recommender systems." *International Journal of Information Technology Convergence and Services* (IJITCS) 2, No.1 (Feb), 1-8.

Poggi, N.; T. Moreno; J.L. Berral; R. Gavald; and J. Torres. 2007. "Web customer modeling for automated session prioritization on high traffic sites". In *Proceedings of the 11th International Conference User Modeling* (Corfu, Greece, Jul.25-29), Lecture Notes in Computer Science 4511, 450-454.

Potempa, A. 2014. "Selected methods of supervised learning classification." (original title in Polish: "Wybrane metody klasyfikacji pod nadzorem"). Masters Thesis, Institute of Mathematics and Informatics, Opole University.

R Project, *The R project for statistical computing*, http://www.r-project.org.

Soliman, T.H.A.; M.A. Elmasry; A.R. Hedar; and M.M. Doss. 2012. "Utilizing support vector machines in mining online customer reviews". In *Proceedings of the 22nd International Conference on Computer Theory and Applications* ICCTA'12 (Alexandria, Egypt, Oct.13-15), Vol. 1, IEEE, 192-197.

Soiraya, B.; A. Mingkhwan; and C. Haruechaiyasak. 2008. "E-commerce web site trust asssessment based on text analysis." *International Journal of Business and Information* 3, No.1, pp. 86-114.

Suchacka, G. and G. Chodak. 2013. "Practical aspects of log file analysis for e-commerce". In *Proceedings of the 20th International Conference Computer Networks* (Lwwek lski, Poland, Jun.17-21), Communications in Computer and Information Science 370. Springer, Berlin Heidelberg, 562-572.

Vapnik, V.N. and A.Y. Chervonenkis. 1974. *Teoriya raspoznavaniya obrazov: statisticheskie problemy obucheniya.* (Russian)[Theory of pattern recognition: Statistical problems of learning.] Nauka. Moscow.

Wang, X.-Y.; J.-W. Chen; and H.-Y. Yang. 2011. "A new integrated SVM classifiers for relevance feedback content-based image retrieval using EM parameter estimation." *Applied Soft Computing* 11, No.2 (Mar), 2787-2804.

Xia, Z.; Y. Dong; and G. Xing. 2006. "Support vector machines for collaborative filtering". In *Proceedings of the 44th Annual Southeast Regional Conference* (Melbourne, FL, Mar.10-12. ACM New York, NY, 169-174.

Xu, H. and L. Wang. 2011. "Application of analysis CRM based on association rules mining in variable precision rough set". In *Proceedings of International Conference on Computer Science, Environment and Ecoinformatics* (Wuhan, China, Aug.21-22), Communications in Computer and Information Science 216. Springer, Berlin Heidelberg, 418-423.

Yu, X.; S. Guo; J. Guo; and X. Huang. 2011. "An extended support vector machine forecasting framework for customer churn in e-commerce." *Expert Systems with Applications* 38, No.3 (Mar), 1425-1430.

Zhang, K.; H. Xu; J. Tang; and J. Li. 2006. "Keyword extraction using support vector machine". In *Proceedings of the 7th International Conference Advances in Web-Age Information Management* (Hong Kong, China, Jun.17-19), Lecture Notes in Computer Science 4016, Part III. Springer, Berlin Heidelberg, 85-96.

Zhang, W.; X. Fang; S. Teng; and X. Fu. 2011. "A cooperative network intrusion detection based on SVMs". In *Proceedings of the 6th International Conference on Pervasive Computing and Applications* (Port Elizabeth, South Africa, Oct.26-28). IEEE, 129-138.

Zhu, Y. and Y. Zhang. 2010. "The support vector machine based on the principal component in the credit management of electronic commerce". In *Proceedings of the 2nd International Conference on e-Business and Information System Security* (Wuhan, China, May 22-23). IEEE, 1-3.

AUTHOR BIOGRAPHIES

GRAŻYNA SUCHACKA received MS degrees in Computer Science and in Management from Wroclaw University of Technology, Poland. She received her Ph.D. degree in Computer Science from Wroclaw University of Technology. Now she is an assistant professor at the Institute of Mathematics and Informatics at Opole University, Poland. Her research interests include web mining, web analytics, and quality of web service with special regard to electronic commerce. Her e-mail address is: gsuchacka@math.uni.opole.pl.

MAGDALENA SKOLIMOWSKA-KULIG received her Ph.D. degree in Mathematics from University of Wroclaw, Poland. Now she works as an assistant professor at the Institute of Mathematics and Informatics at Opole University, Poland. Her research interests include theory of probability and mathematical statistics. Her e-mail address is: mskolimow@math.uni.opole.pl.

ANETA POTEMPA received BS degrees in Mathematics from Opole University, Poland and in Economy, accounting and audit from The WSB Schools of Banking in Opole, Poland. She received an MS degree in Financial Mathematics from Opole University in 2014. She works in the monitoring department in GEA Technika Cieplna Sp. z o. o., the company specializing in the production of industrial heat exchangers. Her e-mail address is: aneta.potempa@gea.com.