EMPIRICAL ANALYSIS OF CLASSIFIERS AND FEATURE SELECTION TECHNIQUES ON MOBILE PHONE DATA ACTIVITIES

Fandi Husen Harmaini and M. Mahmuddin
School of Computing, Universiti Utara Malaysia, Sintok Kedah, Malaysia
E-Mail: fandi.husen@yahoo.com, ady@uum.edu.my

ABSTRACT

Mobile phones nowadays become ubiquitous device and not only a device to facilitate communication, with some addition feature of hardware and software. There are many activities can be captured using mobile phone with many of features. However, not all of these features could benefit to the in processing and analyzer. The large number of features, in some cases, gives less accuracy influence the result. In the same time, a large feature takes requires longer time to build model. This paper aims to analyze accuracy impact of selected feature selection techniques and classifiers that taken on mobile phone activity data and evaluate the method. Furthermore, with use feature selection and discussed emphasis on accuracy impact on classified data of respective classifier, usage of features can be determined. To find the suitable combination between the classifier and the feature selection sometime is crucial. A series of tests conducted in Weka on the accuracy on feature selection shows a consistency on the results although with different order of features. The result found that combination of K* algorithm and correlation feature selection is the best combination with high accuracy rate and in the same time produce less feature subset.

Keywords: feature selection, classification, mobile cellphone activities, machine learning.

INTRODUCTION

Accuracy is very closely associated with systematic and random errors; it is a combination of both trueness and precision. In feature selection, accuracy impact is the percentage of correctly classified instance and time taken to build model. Sensor networks use small, inexpensive sensor with several special characteristics, with radio range and processing power, sensor permits very low energy consumption, perform limited and specific monitoring and sensing functions. Sensor networks also possible for short term storage and providing processed data as information [1]. Typical functions in a sensor network are sensing, collecting, processing, and transmitting sensed data.

Sensor data can facilitate automated or human-induced tactical/strategic decisions, if analyze efficiently and transform it into usable information. Sensor technology that is embedded in mobile phone has ability to receive and send data ubiquitously. This data not only encompasses of calls log, but also information on other motion activities including play game, walking, and many others. Handling data from mobile phone are challenging in terms of resource constraint, fast and huge data arrival and data transformation. The amount of excessive incoming data will make the sensor nodes exhausted and not effective. Data that comes from different sources is probably aggregated [2]. This case will effected important data. Furthermore, with large set of data in the stream, most likely data lost or contaminated will happen and may lead to density, redundancy and latency.

Mobile phone is a device that facilitates people to communicate each other in long distances. Nowadays, mobile phone is not been used only to make a call, but become a smart phone. This smart phone can help the user to perform activity, for example, short messages, chatting, play game, listening music, take a picture, watching movie and other that make people work easy. This type of mobile phone has central processing unit and random access memory like computer, this aims to make user feel like using computer but in ubiquitous and small size. Enhancing availability of sensor network in consumer products, with many potential applications makes mobile phone activities have recently gained attention as a research topic [3]. This sensor includes audio, GPS, image, light, temperature, direction, and accelerator sensors. Because of the small size, their substantial computing power, ability to receive and send data and their nearly ubiquitous use in society is feature of these "smart" mobile phones.

Several techniques of feature selection that consume many features may lead to the longest time to build model but gives better accuracy, and several techniques may lead to the fastest time process to build model but give worse accuracy with consume less features. This problems need to be analyzed to be references what techniques and approaches can be used to big data analysis.

FEATURE SELECTIONS FOR MOBILE PHONE ACTIVITIES

Feature selection process

The product of data pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. knowledge discovery during the training phase is more complicated when there some event of distorted and redundancy of information present noisy and unreliable data. Preparation and filtering data phase can considerate total time of processing. Data that have been analyses but not scanned carefully for such problem can produce distorted results. Final training set is the product of pre-processing data as suggested in [4] an algorithm for each of data pre-processing step. Furthermore, before

www.arpnjournals.com

running an analysis done, the representation and quality of data is first and foremost.

Feature selection is process of select a subset. There are two approaches of feature selection, a) forward selection, that is start with no variables and add them one by one, at each step adding the one that decreases error the most, until any further addition does not significantly decrease the error, or ii) second one is backward selection, that starts with all the variables and remove them one by one, at each step removing the one that decreases error the most until any further removal increases the error significantly. Feature selection basically has four benefits: i) to reduce feature space dimensionality that can reduce the need of storage and increase the speed algorithm, ii) to eliminate data redundancy, irrelevant feature or noise iii) to increase algorithm learning time and iv) to increase and improve quality of performance data. Only the most contribute subset sustains and discard the remaining unimportant dimension. The best subset contains the least number of dimensions that most contribute to accuracy after that, discard the remaining unimportant dimension.

Data might also contains noise features is when added to the document representation, an error on new data occur. To satisfy mining process, if data contain many redundant or irrelevant features that is when use feature selection technique. Irrelevant or redundant features are happen when selected features is no useful information to provide in any context or no more provide information than currently selected features. Excess in classification to eliminating noisy or irrelevant features is one of advantages feature selection [5]. Furthermore, to discover new knowledge, solid and quick models by developing them use a small subset of the original set of features, also able to focus on a subset of relevant features [6].

Feature selection techniques

In machine learning and statistics, dimension reduction is the process of reducing the number of random feature (also known as attribute or variables). In some cases, data analysis such as regression or classification can be done in the reduced space more accurately. The data transformation may be linear but many nonlinear dimensionality reduction techniques also exist [7]. The main feature selection techniques as follow that available in Weka [8] are:

Principle Component Analysis (PCA) works to reduce the dimensionality of dataset that composed of great number of interrelated variables. This goal can be reach by transforming it into a new set of variable, not correlated principal components and been ordered then the first few retain most of variation present in all of original variables [9]. In the transformation, initial principal component has most possibility of variation and the following component shift has the greatest variance possibility under limitation that is orthogonal to the earlier components. This analysis is sensitive to the scaling of the variables, whenever distinction variables have different units; this analysis is a slightly arbitrary method of analysis.

Information Gain (IG) is an attribute selection measure and is based on information gain entropy of Kullback-Leibler divergence in information theory and machine learning [10]. A notable problem occurs when IG is applied to attributes that can take on a large number of distinct values. For example, suppose that one is building a decision tree for some data describing the customers of a business. IG is often used to decide which of the attributes are the most relevant, so they can be tested near the root of the tree.

Chi Squared (ChS) based on a statistical applied test to sets categorical data to assess how often it is that any observed distinction between the sets appears by chance is called Pearson's χ^2 [11]. A common case for this is where the events each cover an outcome of a categorical variable. A simple example is the hypothesis that an ordinary six-sided die is "fair". Pearson's chi-squared test is used to assess two types of comparison: tests of goodness of fit and tests of independence. A test of goodness of fit establishes whether or not an observed frequency distribution differs from a theoretical distribution. A test of independence assesses whether paired observations on two variables, expressed in a contingency table, are independent of each other.

Gain Ratio (GR) magnified IG as it normalizes distribution of all attributes to final classification decision [12]. Problem with using gain ratio, in some situations the gain ratio modification overcompensates and can lead to preferring an attribute just because its intrinsic information is much lower than that for the other attributes. A standard fix is to choose the attribute that maximizes the GR, provided that the information gain or that attribute is at least as great as the average information gain for all the attributes examined.

Filtered Attribute (FA) filter methods use a custodian measure on the contrary of the error rate to imprint a feature subset. This technique runs a despotic subset evaluator on data that has been passed through an arbitrary filter. Filters that modify the order or number of attributes are not allowed.

OneR Attribute (ORA) builds rules based on a single feature for each feature in a dataset. OneR develops rules based on a single feature for each feature in a dataset. By splitting the dataset into training and test sets it is probably to calculate a classification accuracy score for each feature. Work in [13] has selected the highest scoring features and indicated that for most of the datasets the rule associated with this single feature displays comparably with state-of-the-art techniques in machine learning.

Relief F Attribute (RFA) used in binary classification (generalizable to polynomial classification by decomposition into a number of binary problems) proposed by [14]. Its strengths are that it is not dependent on heuristics, requires only linear time in the number of given features and training instances, and is noise-tolerant and robust to feature interactions, as well as being applicable for binary or continuous data. However, it does not discriminate between redundant features, and low numbers of training instances fool the algorithm.

Symmetrical Uncertainty (SU) evaluates the worth of a set attributes with respect to another set of attributes (an entropy based filter method). The symmetrical uncertainty between features and the target concept can be used to evaluate the goodness of features for classification.

Correlation Feature Selection (CFS) doubles evaluation formula with an appropriate correlation measure and a heuristic search strategy.

Consistency Subset (CS) evaluates the decent of a subset of attributes by the rank of consistency in the class prestige while the training instances are calculated onto the subset of attributes. Consistent sampling has important applications in resemblance calculation, and calculation of the number of distinct items in a data stream [15].

Filtered Subset (FS) search process time, each generated subset need to be evaluated by an evaluation criterion. If the new subset turns out to be better, it substitutes the previous subset [16]. Subset evaluation can be divided into two types, filter and wrapper based on their dependency on a data mining algorithm. The filter model evaluates the subset of attributes by examining the intrinsic.

Classification techniques

Machine Learning is the engine which powers the modern data-driven and can discover the optimal decision-making, estimate the output interest automatically, allowing to react in real time, configurable, and infinite scalability. Machine learning has algorithms that almost exclusively are iterative also has non-standard fault tolerance that can deal with unavailable partitions and aggregation function over huge objects [17].

Support Vector Machine (SVM). SVM is relatively new learning machine on statistical learning to analyze data, and recognizes patterns. By making it non-probabilistic binary linear classifier, SVM develops a model that assigns new sample into one category or the other. SVM has advantages the sensitive to noises and outliers but inconsistence conditional features [18]. The sensitivity against noises is useful for analyze accuracy of data sensor network that located in physical environment.

Artificial Neural Network (ANN). ANN algorithm can be adopted in sensor network easily and achieve simple parallel-distributed computation, data robustness, auto classification of sensor readings. On the other hand, neural networks algorithms not represent big burden to memory because of the simple computation. One of the important models of ANN is multilayer perceptron (MLP) contains multiple layers of nodes utilize supervised learning of backpropagation of training the network.

Radial Basis Function (RBF). RBF network have gained much popularity in recent times due to their ability to approximate complex nonlinear mappings directly from the input–output data with a simple topological structure. In the field of mathematical modelling, a radial basis function network is an ANN that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Selection of a learning algorithm for a particular application is critically dependent on its accuracy and speed [19].

Naïve Bayes (NB). NB classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. NB is based on Bayes' theorem also highly scalable, requiring a number of parameters linear in the number of features in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression [20] which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of features in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression [20], which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

Random Forests (RF). RF are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The algorithm for inducing a random forest was developed [21, 22] and "Random Forests" is their trademark was first proposed by [23]. Increasing the correlation increases the forest error rate and the strength of each individual tree in the forest. A tree with a low error rate is a strong classifier.

J48. J48 is the Java based decision tree of C4.5 [24] that predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. It creates decision trees of any depth. The internal nodes of a decision tree denote the different attributes; the branches between the nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes tell us the final value (classification) of the dependent variable [25].

Decision Tables (DT). DT is a precise yet compact way to model complex rule sets and their corresponding actions. DT is like flowcharts and if-then-else and switch-case statements, associate conditions with actions to perform, but in many cases do so in a more elegant way. Each decision corresponds to a variable, relation or predicate whose possible values are listed among the condition alternatives. Some DT use simple true/false values to represent the alternatives to a condition (if-then-else), other tables may use numbered alternatives (switch-case), and some tables even use fuzzy logic or probabilistic representations for condition alternatives. In a similar way, [26] stated that action entries can simply represent whether an action is to be performed, or in more advanced decision tables, the sequencing of actions to perform.

K-star (K*). K* is an instance-based classifier that is the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function [27]. It differs from other instance-based learners in that it uses an entropy-based

distance function that use generalization beyond the training data is delayed until a query is made to the system. The main advantage gained in employing a lazy learning method is that the target function will be approximated locally.

IMPLEMENTATION

Data collection

The procedure is conducted in this work as dispicted in Figure-1. It starts with collecting data to be analysed and feature selection technique is used to determine total selected features. This selected features, later will be verified with chosen classifiers to abotain the accuracy and taken time of each techniques.
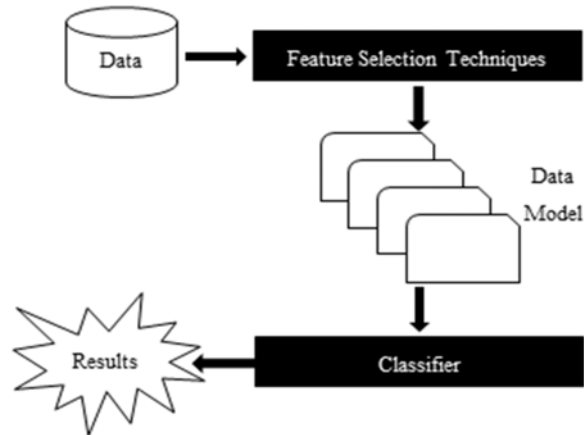

Figure-1. Simplified overall undertaken method.

Raw dataset is a set of data that been collected from a source. This raw data has not been subjected to processing or any other manipulation and are also referred to as primary data. This work is going to use mobile phone activity data as data collection. The data is consist of activities of human including walking (38.4%), jogging (30.0%), sitting (5.7%), standing (4.6%), upstairs (11.7%), and downstairs (9.8%). This data contain five thousand four hundred eighteen (5418) instances with forty four (44) attributes including class feature with no missing value. Details of this are summarized in Table-1. The data itself is collected from [28]. The reason this paper uses this data, because the transformed data is clear and comprehensive with detailed explanation in the studies. This data is collected from twenty-nine users as they performed daily activities then aggregated this time series data into examples that summarize the user activity over ten second intervals. Raw data that have been collected then transform into ARFF (Attribute-Relation File Format) to be examine in Weka [8].

Table-1. Decription of used dataset.

Parameter	Label	Description	Data Type
1-30	**X_1-X_{10}, Y_1-Y_{10}, Z_1-Z_{10}**	**Average acceleration values are the fraction of accelerometer samples**	**Numeric**
31-33	X_{AVG}, Y_{AVG}, Z_{AVG}	Average x, y, and z values over the 200 records	Numeric
34-36	X_{PEAK}, Y_{PEAK}, Z_{PEAK}	Approximations of the dominant frequency	Numeric
37-39	$X_{ABSOLDEV}$, $Y_{ABSOLDEV}$, $Z_{ABSOLDEV}$	Average absolute deviations from the mean value for each axis	Numeric
40-42	$X_{STANDDEV}$, $Y_{STANDDEV}$, $Z_{STANDDEV}$	Standard deviations for each axis	Numeric
43	RESULTANT	Average root of the sum of the values of each axis squared, $\sqrt{x_i^2 + y_i^2 + z_i^2}$	Numeric
44		Type of activities was performing.	Nominal

www.arpnjournals.com

Selected feature selection and classifier

From seventeen feature selection algorithm in WEKA, this work only can use eleven of them namely PCA, IG, ChS, GR, FA, ORA, RFA, SU, CFS, CS and FS, while the other six cannot be used due to technical errors including Latent Semantic Analysis, SVM Attribute, Cost Sensitive Attribute, Cost Sensitive Subset, Classifier Subset, and Wrapper Subset. Feature selection of this work use full training set as attribute selection mode and search method this work using Ranker or Greedy Stepwise based on feature selection characteristic. Ranker search method is a process to ranks features by their individual evaluation; meanwhile, greedy stepwise is a process that performs a greedy forward or backward search through the space of feature subsets. As might noteiced that all classification algorithm are based on supervised approach where we know that it produced better accuracy in classifying. This test was included time taken to build model as complement analyze. Eight classifiers that are used in this work and based on mostly used for data mining in development and research area, there are SVM, MLP, RBF, NB, RF, J48, DT and K*.

Classification is fall to supervised learning that the problem of identifying to which of the set of categories new observation belong, on the basis of training set of data containing observations whose category membership is known. Classifier that this work used is SVM, MLP, RBF Network, Naïve Bayes, Random Forest, J48, decision table, and K*. In this phase, test option of this work is use training set. These classifier are implemented by choose the sets of machine learning technique, based on what the techniques is classified.

RESULT AND DISCUSSIONS

There are a few metrices have been considered to access the performance of all possible techniques. Average time taken to generate the model and accuracy of each algorithm are the main consideration in this work. The datasets has been tested with different feature selection approach and majority of them took about 80 seconds to be generated. The result also shows that MLP requires the longest time to generate the data model. This situation confirms that MLP requires long training time. And in the same time, parameters that been chosen in MLP is a standard configuration setting. While the NB and K* produce the fastest model less than 1 second. Figure-2 depicts the summary of each selected technique of the average time taken. FS is the fastest feature selection while GR need the longest time to process.
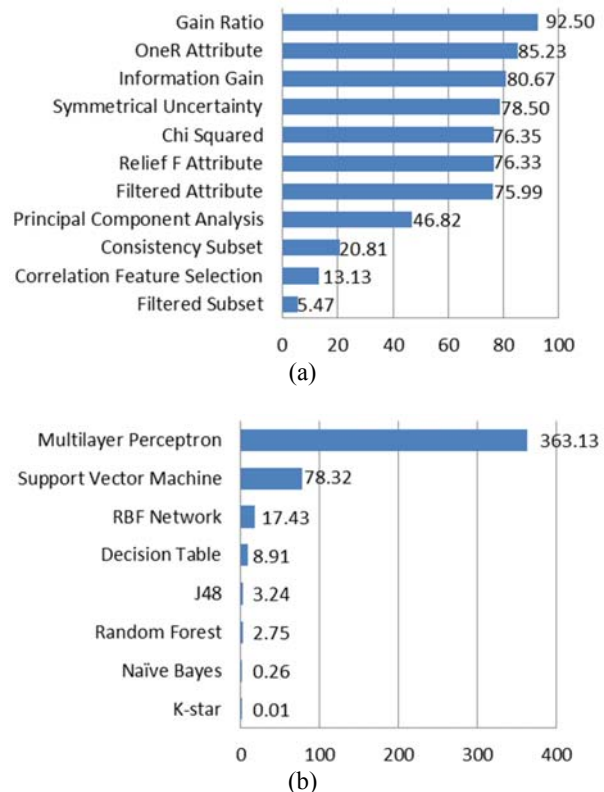

Figure-2. Average time (in sec.) for (a) feature selection techniques and (b) classification algorithms.

The accuracy performance of the selected algorithms is shown in Table-2. It's clearly displayed that majority of the algorithm perform accurately well with more than 80%. Table-2 also shows that feature selection algorithm such as IG, CHS, GR, FA, ORA, and RFA are basically produce the same correctness. K* is generally out-performed to other classifier which produces very high accurate result.

Each of the feature selection technique proposes total number of feature subset to be considered. In this matter, the aim is to identify the as small number of subset as possible and retain high accuracy of this feature subset. A fine-tuning process of both is needed to find balance and better final result. For that, this paper propose K* and CFS should be considered to be employed for the best result. Although FS has generated the least number of feature subset, yet in the same time the accuracy is not favorable compare to other algorithm. Based on all result, the average of fastest time taken to build the model on feature selection is FS with 5.47 s and K* in 0.01 s, while the average most accurate are IG, ChS, GR, FA, ORA, RFA and SU with 90.50% and on RF with 99.34 respectively. In contrast, the average longest time for feature selection is GR with 92.5 s and MLP with 363.13 s, while the average most inaccurate is PCA with 70.6% and on NB with 73.6%. CFS suggests only 6 features to be considered Z_{AVG}, Z_{PEAK}, $Y_{ABSOLDEV}$, $Z_{ABSOLDEV}$, $X_{STANDDEV}$ and $Y_{STANDDEV}$.

www.arpnjournals.com

Table-2. Result of the obtained accuracy.

	PCA	IG	ChS	GR	FA	ORA	RFA	SU	CFS	CS	FS
SVM	**38.5**	**99.8**	**99.8**	**99.8**	**99.8**	**99.8**	**99.8**	**99.8**	**91.1**	**90.4**	**76.6**
MLP	63.1	89.3	89.3	89.3	89.3	89.3	89.3	89.3	81.3	85.8	73.4
RBF	56.8	78.5	78.5	78.5	78.5	78.5	78.5	78.5	77.6	78.4	74.5
NB	56.4	75.5	75.5	75.5	75.5	75.5	75.5	75.5	75.8	77.0	71.7
RF	98.5	99.7	99.7	99.7	99.7	99.7	99.7	99.7	99.3	99.3	98.0
J48	92.4	97.4	97.4	97.4	97.4	97.4	97.4	97.4	95.2	96.9	84.1
DT	58.9	83.9	83.9	83.9	83.9	83.9	83.9	83.9	83.9	83.9	77.3
K*	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	76.4
#FS	31	43	43	43	43	43	43	43	6	13	3

#FS is total number of selected features

CONCLUSIONS

Several tested has been conducted to analyze which feature selection has the best accuracy impact. The average time is also been taken to build model. Based on the tests that have been done, there are many classifiers that produce the fastest time and the most accurate, and among all classifier, K* has shown the best accuracy impact with time taken to build model in less than a second and 100% accuracy. The future work of this study can be applied in the real world problem that related to big data analysis which requires high accuracy impact feature selection and fastest time to build model. The result of this project test can be reference on which feature selection and classifier that is good for next analysis. However, seventeen feature selection need to be compared and analysis to get complete result. Lastly, future study to analyze data with other transformed data is suggested to get comparison result more comprehensive.

It also can be concluded that every classifier and feature selection techniques will produce differently. This work can be extended further that related to big data analysis which requires high accuracy impact feature selection and fastest time to build model. Lastly, future study to analyse data with other transformed data is suggested to get comparison result more comprehensive.

REFERENCES

[1] D. Westhoff, *et al*. 2006. "Security Solutions for Wireless Sensor Networks," NEC Journal of Advanced Technology, vol. 59, pp. 2-6.

[2] C. Intanagonwiwat, *et al*. 2002. "Impact of network density on data aggregation in wireless sensor networks," in Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on, 2002, pp. 457-458.

[3] J. R. Kwapisz, *et al*. 2011. "Activity recognition using cell phone accelerometers," ACM SigKDD Explorations Newsletter, vol. 12, pp. 74-82.

[4] S. B. Kotsiantis, *et al*. 2006. "Data preprocessing for supervised learning," International Journal of Computer Science, vol. 1, pp. 111-117.

[5] B. Krishnapuram, *et al*. 2004. "Gene expression analysis: joint feature selection and classifier design," in Kernel Methods in Computational Biology, B. Schölkopf, *et al*., Eds., ed Cambridge, MA, : MIT Press, pp. 299-317.

[6] I. Guyon and A. Elisseeff. 2003. "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182.

[7] S. T. Roweis and L. K. Saul. 2000. "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science vol. 290, pp. 2323-2326.

[8] I. H. Witten and E. Frank. 2005. Data Mining: Practical machine learning tools and techniques: Morgan Kaufmann.

[9] I. Jolliffe. 2005. Principal Component Analysis: John Wiley and Sons, Ltd.

[10] S. Kullback and R. A. Leibler. 1951. "On information and sufficiency," Annals of Mathematical Statistics, vol. 22, pp. 79-86.

[11] K. Pearson, "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random

sampling," Philosophical Magazine Series 5, vol. 50, pp. 157-175, 1900/07/01 1900.

[12] A. Sharma and S. Dey. 2012. "A comparative study of feature selection and machine learning techniques for sentiment analysis," presented at the 2012 ACM Research in Applied Computation Symposium.

[13] R. C. Holte. 1993. "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets," Machine Learning, vol. 11, pp. 63-90.

[14] K. Kira and L. A. Rendell. 1992. "A practical approach to feature selection," presented at the 9th International Workshop on Machine Learning.

[15] K. Kutzkov and R. Pagh. 2014. "Consistent Subset Sampling," in Algorithm Theory – SWAT 2014. vol. 8503, R. Ravi and I. Gørtz, Eds., ed: Springer International Publishing, pp. 294-305.

[16] K. Gao, et al.. 2009. "An empirical investigation of filter attribute selection techniques for software quality classification," presented at the IEEE International Conference on Information Reuse and Integration, 2009 (IRI'09).

[17] T. Condie, et al. 2013. "Machine learning for big data," presented at the 2013 International conference on Management of data, ACM.

[18] H. Han, et al. 2012. "Comparative study of two uncertain support vector machines," presented at the IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI), 2012.

[19] G. B. Huang, et al. 2005. "A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation," IEEE Transactions on Neural Networks, vol. 16, pp. 57-67.

[20] S. Russell and P. Norvig. 2003. Artificial Intelligence: A Modern Approach, 2 ed.: Prentice Hall.

[21] L. Breiman. 2001. "Random Forests," Machine Learning, vol. 45, pp. 5-32.

[22] A. Liaw. 2012. "Documentation for R package randomForest," ed.

[23] T. K. Ho, "Random Decision Forest," presented at the 3rd International Conference on Document Analysis and Recognition, Montreal, QC.

[24] J. R. Quinlan. 1995. C4.5: Programs for Machine Learning: Morgan Kaufmann Publishers, 1993.

[25] W. J. Dixon and F. J. Massey. 1969. Introduction to statistical analysis vol. 344. New York: McGraw-Hill.

[26] G. Wets, et al. 1996. "Locational choice mod-elling using fuzzy decision tables," presented at the Biennial Conference of the North American Fuzzy In-formation Processing Society.

[27] J. G. Cleary and L. E. Trigg. 1995. "K*: An Instance-based Learner Using an Entropic Distance Measure," in 12th International Conference on Machine Learning Tahoe City, California, pp. 108-114.

[28] J. R. Kwapisz, et al. 2011. "Activity recognition using cell phone accelerometers," SIGKDD Explor. Newsl., vol. 12, pp. 74-82.