ENSEMBLE CLUSTERING BASED ON FEATURE SELECTION APPROACH TO LEARNING RELATIONAL DATA

KUNG KE SHIN

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCES (SOFTWARE ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2015

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, excerpts, equations, summaries and references, which have been duly acknowledge.

This thesis may be made available within the university library and may be photocopied or loaned to other libraries for the purposes of consultation.

JUNE 22, 2015

_____

KUNG KE SHIN

BK11110131

CERTIFIED BY

.....

Associate Professor DR Rayner Alfred

SUPERVISOR

ACKNOWLEDGEMENT

This thesis is the main result of my bachelor degree studies, undertaken during the period from September 2014 to June 2015 in University Malaysia Sabah.

First and foremost, I would like to express my utmost gratitude to my supervisor, **Associate Professor DR Rayner Alfred** for his help, advise, support, guidance and constant encouragement. He has provided me with a powerful knowledge on my research studies, besides guiding me with ethical and life values as I walked through this journey of mine. I would like to take this chance to thank DR. Chin Kim On, DR. Lau Hui Keng and DR. Mohd Hanafi Ahmad Hijazi for their constant support and advice regarding the works. Also, I would like to express special thanks to staffs of Faulty of Computing and Informatics, UMS for their assistance in the in this projects.

Finally, I would like to express my gratitude to my family members for their undivided love and morale support. Without them, I will not be what I am today. I thank them for guiding and forming me on this journey of mine. Last but not least, I would like to thank all my friends and course mates, who have been here with me throughout my degree life, cheering me up during my tough times.

ABSTRACT

The employment of classification in learning big relational data is an important research field. Learning big relational data often involves large feature dimensionality and this can be very time consuming. Many approaches have been developed to learn relational data. One of the approaches used to learn relational data is DARA. The DARA algorithm is designed to summarize data with one-to-many relations. However, DARA suffers a major drawback when the cardinalities of attributes are very high because the size of the vector space representation depends on the number of unique values that exist for all attributes in the dataset. A feature selection process can be introduced to overcome this problem. However, different feature selection methods used will produce different sets of selected features and thus produces different classification results. The final results obtained based on these sets of features selected can be further optimized by computing the consensus result in order to achieve a good classification result. This can be achieved by introducing an ensemble technique into the framework. Ensembles are frequently used to improve the predictive accuracies of multiple classifiers by producing the final consensus result from multiple classifiers. In this project, a two-layered genetic algorithm-based feature selection is proposed to form the basic ensemble in order to improve the classification performance in learning relational datasets. Results from the experiments show that the proposed method is able to improve the accuracies of classification tasks and k-NN classifiers with Euclidean distance as similarity measurements outperformed other classifiers.

ABSTRAK

Kegunaan klasifikasi dalam pembelajaran data berhubung besar adalah satu bidang penyelidikan yang penting. Selalunya, pembelajaran data berhubung melibatkan kedimensian ciri yang besar dan ini amat mengambil masa. Banyak kaedah telah dicipta untuk mempelajari data berhubung. Salah satu kaedah yang telah digunakan untuk mempelajari data berhubung adalah DARA. Algorithma DARA direka untuk meringkaskan data yang mempunyai hubungan satu-ke-banyak. Namun, DARA mengadapi satu kelemahan utama apabila kardinaliti ciri amat tinggi kerana saiz perwakilan ruang vektor bergantung kepada nilai-nilai unik yang wujud untuk semua ciri-ciri dalam dataset. Proses pemilihan ciri boleh diperkenalkan untuk mengatasi masalah ini. Walau bagaimanapun, kaedah pemilihan ciri berbezaan yang digunakan akan menghasilkan set ciri-ciri yang berbeza dan dengan itu keputusan klasifikasi yang berbeza akan dihasilkan juga. Keputusan akhir yang diperolehi berdasarkan kepada set ciri-ciri terpilih boleh terus dioptimumkan dengan mengira hasil kesepakatan dalam usaha untuk mencapai keputusan klasifikasi yang baik. Ini boleh dicapai dengan memperkenalkan teknik ensemble ke dalam rangka kerja tersebut. Teknik ensemble sering digunakan untuk meningkatkan ketepatan ramalan pelbagai pengelas dengan menghasilkan hasil konsensus akhir daripada pelbagai pengelas. Dalam projek ini, pemilihan ciri dua lapis yang berdasarkan kepada algoritma genetik telah dicadangkan untuk membentuk ensemble asas untuk meningkatkan prestasi klasifikasi dalam pembelajaran dataset hubungan. Hasil daripada eksperimen menunjukkan bahawa kaedah yang dicadangkan mampu meningkatkan ketepatan tugas klasifikasi dan pengelas k-NN dengan jarak Euclidean sebagai ukuran persamaan mencapai keputusan yang lebih baik daripada pengelas lain.

List of Figures

List of Tables

Contents

CHAPTER 1

INTRODUCTION

1.1 Overview

The use of computer, electronic devices and the World Wide Web have become significantly important to the world as they bring conveniences to human's daily lives. Back in 1993, there was only less than one percent of the world's population who has access to the internet. Today, the internet is accessible to nearly 3 billion people or 40% of the world's population. The growth of the electronic devices and computer usage in every aspect of human's daily activities had made electronic data extremely valuable. Such growth is because the traditional way of storing data are based on physical basis for example information recorded on papers. With the volume of generated data increasing exponentially, the traditional approaches of storing the data become much less efficient in most domains and storing them electronically is always chosen as substitution. These electronic data can then be used in data mining to perform many tasks including classification. Up to this day, great amount of electronic data have been generated which end up making them big data.

In data mining, electronic data or information stored in electronic devices are being analyzed and summarized in data mining process to discover previously

unknown patterns, associations, changes, anomalies and significant structures for classification purpose. Extraction of these patterns usually involves analyzing a large quantity of data from database, data warehouses, or other information repositories. In classification tasks, these patterns are exploited to make predictions and the accuracies define the usefulness of the patterns discovered. Discovering interesting patterns often involves features selection where subsets of relevant features or variables are selected from the data for use in constructing data mining model. Traditional data mining algorithm look for patterns in a single table of a database by analyzing the features in the table. While most structured database store data in multiple tables and they are linked to each other to reduce redundancies, looking for patterns in a single table may leave out important features in other tables and thus lead to inefficiencies. In this project, a two-layered genetic algorithm-based ensemble classifier will be proposed in the attempt to improve the accuracy performance in learning big relational data.

Ensemble classifier is a robust technique frequently used by researchers to improve performance of classification in big relational data. Ensemble consist collection of different classifiers where each of the classifier will perform classification task to an entry and the final decision will be made by using voting system. Genetic algorithm represents an intelligent exploitation of random search which has been widely applied for problem solving in many real world applications including optimization problems and search problems. Designed to simulate the natural selection process, genetic algorithm handles randomization problem by utilizing the concept "survival of the fitness" which in normal case bad solutions will be omitted while exploring solutions in regions with better performance. With this, we can see that genetic algorithm is robust as it maintains randomness of a search and trying to

look for best solutions within the search space. Moreover, genetic algorithm outperformed many typical optimization techniques for searching in a large dimensional surface. Therefore, genetic algorithm has always been one of the top approaches being used for feature selection optimization in large search space problems.

1.2 Problem Background

Relational data mining is different from traditional data mining methods, in which all features obtained from multiple tables that exist in a structured relational database can be collected, selected and exploited during the data mining process. In other words, a relational data mining involves learning of target table that has a one-to-many relationship with records stored non-target tables in order to look for patterns across multiple tables and thus having the potential to outperform traditional data mining techniques in many cases. This might seem to be an ideal approach to discover useful and interesting patterns from relational databases. Unfortunately when it comes to learning big relational data with high degree of one-to-many association, joining features from multiple tables may cause information loss. Therefore, data transformation becomes a tedious trial-and-error work and the classification result is often not very promising especially when the number of tables and the degree of one-to-many association are large.

Data transformation process could be used to transform a relational data representation into a vector space model representation by combining records stored in the target table in a relational database that has a one-to-many relationship with records stored in other non-target table. Transformed data will be summarized and

append to the target table as new features. There is a drawback for this method, transforming these features may cause the vector space dimensionality to grow larger and causes the search dimension to be relatively large as well. In classification tasks, analyzing and processing each and every feature in the large search space makes the process very costly especially in term of time and computational resources. Therefore, various feature selection techniques are applied to help solving the aforementioned problem for classification task in relational data mining.

1.3 Problem Statement

Feature selection is the process to omit irrelevant features to the search goal and select a subset of relevant features for use before data mining process take place. In order words, feature selection can reduce the search space to achieve cost effectiveness [1]. This is important because the predictive accuracy of any classification tasks depends on the quality of the input data and good features selection technique is always crucial in classification task for big relational data to minimize costs and redundancies. However, selecting good features in big relational data has been a very challenging task that many researchers face due to the large number of different subsets of relevant features must be evaluated. In some cases, using random selection of features in big relational data shows improvement in accuracy performance of the classifiers. Despite that, using random subset of features has a major drawback that is it cannot guarantee the features selected are discriminant, causing poor classifiers to be generated [2]. In this case, genetic algorithm can be applied to perform the optimization task for the feature selection [3] [4] [5]. However, different classifiers (e.g., 1-Nearest Neighbor, 2-Nearest Neighbour, 3-Nearest Neighbor and etc.) may require different sets of features in order to

maximize the predictive accuracies. Thus, one of the main focuses in this research is to study whether classification tasks for big relational data can be improved by selecting common features that are derived from several sets of features obtained from different classifiers.

In an ensemble system, improvement is achieved when there are diversities among the classifiers which they do not make correlated error and the classifiers have certain accuracy in making predictions [6]. In other words if every classifier in the ensemble system has good performance in learning big relational data and they are using different features to perform the task as aforementioned, theoretically this ensemble system is able to improve the performance of classification task for big relational data. However, if all classifiers are used in the ensemble system, it can be very costly in term of computational resources because every classifier has to look into features in the big relational data and make predictions before final decision is made. This problem then makes the feature selection meaningless as it aims to optimize cost of the process and its objective is not achieved. Therefore, our second main focus in this research is to investigate whether using only a subset of effective classifier is able to generate an effective ensemble system to learn big relational data.

In order to investigate the two aforementioned problems, a two-layered genetic algorithm-based feature selection is proposed to form the basic ensemble in order to improve the classification performance in learning relational datasets. In the first layer, the proposed method involves the task of optimizing the process of selecting relevant subsets of features for several different classifiers. In the second layer, a genetic-based algorithm is applied to form the best ensemble classifier by selecting a subset of relevant classifiers obtained from the first layer described

previously. It is expected that the proposed method is able to improve the predictive accuracy in learning a big relational data. However, experiment must be done to prove such theory and therefore our last focus in this project is to investigate is there any significant improvement in term of predictive accuracy when the proposed framework of the evolutionary-based ensemble classifier is used in learning a big relational data.

1.4 Research Objectives

The purpose of this project is to propose an evolutionary-based ensemble system that can improve learning of big relational data. Focus of this project is as follow:

1) To propose and develop an evolutionary-based feature selection based ensemble method in order to learn big relational data with more effective and more efficient.
2) To investigate the effects of applying different similarity measurement method for k-Nearest Neighbor classifiers on the performance of the classification task.
3) To investigate the effects of varying the percentage of selected feature on the performance of the classification task.

1.5 Organization of the Report

This report consists of 5 chapters where the chapter one is the introduction to this project which includes problem background, problem statement, objectives of the research and project scope.

Chapter two consists of literature review of this research which includes general overview on data mining as a field and further explanation of feature

selections process, genetic algorithms and ensemble system include outlines of the steps involved in these processes. Other than that, chapter 2 will also further discuss the problem with a structured analysis of relevant researches done by other researchers, existing problems in relevant fields.

Chapter three consists of methodology which presents a two-layered genetic algorithm-based ensemble classifier that aims to improve the accuracy performance in learning big relational data. Explanation includes steps taken in attempt to optimize the process of selecting relevant features for different classifiers, approach in attempt to generate the most effective ensemble classifier, and the evaluation of the method to determine its performance.

Chapter four is preliminary experiment and experimental setting. The experimental design will be presented in this chapter to explain how experiments will be carried out. Details of the experiment including dataset, parameters, fitness and validation method will be discussed in this chapter to provide a clear picture of the experiment to users.

Chapter five is last part of this project which consist an overall analysis of the project. The result and performance of the experiments will be presented and analyzed to make a conclusion whether the research objectives are achieved and the hypothesis is accepted.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Nowadays, most scientific data are stored in relational databases. As the amount of data generated has been increasing exponentially, big relational data mining has become an interesting field of studies in recent years. Classification in big relational data mining is a very computational resources demanding task because of the data size and the attributes needed to process. Besides that in big relational data, selecting a subset of relevant features for classification purpose is difficult due to a large number of different subsets of relevant features must be evaluated and the search space is large. Ensemble has been applied in many ways to real world applications that involve classification of relational data because ensembles perform better than single classifier in many cases. Improvement of an ensemble based classifiers is achieved when there are certain diversity and accuracy among classifiers. Feature selection techniques are commonly applied to help reduce input dimensionality and produce good classifiers. However, there are possibilities that the features selected could not help to achieve diversities among classifiers in an ensemble system.

In the discussion above, we can briefly see that it is interesting to investigate whether selecting subset of features among the relevant features could generate good classifiers especially in learning of big relational data for the main purpose of dimensionality reduction. Beside this, it is also interesting to investigate the effectiveness of learning big relational data with ensemble generated by using subsets the classifiers generated using feature selection as discussed previously. Therefore, the literature review of this project will focus mainly on various techniques and related works on feature selection, ensemble and also feature selection-based ensemble.

2.2 Relational Data Mining

In relational data mining, the databases involved consist of a collection of data stored in a set of tables associated to each other through reference key from one table to another [7] [8]. Many approaches have been applied in learning relational data; some of these approaches are applied with ensemble by researchers to improve the performance of learning relational data. Probabilistic Relational Models (PRMs) [9] is a method extended from Bayesian networks and designed for relational learning that looks for good dependency structures that defines the relations between variables in tables from training databases in order to handle relational data. In [10], an ensemble of a set of PRM components is used in learning imbalanced relational data where each PRM model will have the samples with minority class and random subset of equal number of samples from the majority class to make each component of the ensemble balance.

A Relational Neighbor (RN) classifier [11] is a simple method that adopts the idea of "guilty by association" which makes predictions on relational data only based on class labels of related neighbors. It is shown that RN is able to perform competitively when compared to other relational classifiers including PRM. Random forest is applied by researchers [12] as classifiers within a hybrid relational learning framework which use both local attributes and flattened (aggregated) relational attributes. Their studies have shown that the prediction accuracy of the ensemble is usually better than individual classification tree.

Inductive Logic Programming (ILP) is another famous approach in learning relational data which was introduced by Muggleton in [13]. This method uses logic presentation. Predicates for logic representation are constructed based on the knowledge provided and the syntactic predicates helps to make hypothesis [14]. However, research shows that ILP-based methods are inefficient for databases with complex schemas. Other than that, it is not appropriate to for continuous values and missing values as well [15].

Graph-based approaches apply mathematical graph theory and make use of the graph based representation to search for graph patterns [15]. The main challenge faced by this method is the graphs are too big to fit into the main memory during processing [16].

Propositionalization based approaches capture and store relational representation in propositional form [17]. These propositional forms are known as new features and are usually stored as attributes in a vector form. Dynamic Aggregation for Relational Attributes (DARA) approach summarizes the entire

contents of non-target tables before the target table can be processed for knowledge discovery [18].

The motivation of this project is based on previous works by [19] where he has proposed a method called Dynamic Aggregation for Relational Attributes (DARA) to summarize data stored in relational databases that consist of data with one-to-many relations. In DARA algorithm, the entire contents of non-target tables are summarized to the target table. The relational data representation is transformed into a vector space representation after a data pre-processing stage. Then each feature extracted will go through model conversion and computation of component magnitude. Then data summarization will be performed which records stored in non-target table are clustered and will be given a label to indicate the group that the records belong to. Finally, the data can be appended to the target table as an additional column of as a new feature. The empirical results obtained show that DARA algorithm is able to improve the predictive accuracies of C4.5 classifier compared to other relational data mining methods. However, DARA has a major drawback which the vector space dimensionality will grow larger because it is affected by the increment of number of distinct values for each column in relational database.

In [20], the authors have proposed method called features transformation to overcome the aforementioned drawback in DARA algorithm. In their studies, they have implemented a further pre-processing step called features transformation in the data transformation process. They applied the TF-IDF (term frequency- inverse document frequency) as a statistical measure which expresses the importance of a feature. Features which have high frequency of present in a record will have a bigger

TF-IDF, but the TF-IDF value will be penalized if the same feature appears in other records too. Therefore, the dimensionality of record-pattern matrix can be reduced because all numerical values are required to be discretized before the feature selection process can be performed base of the feature scoring.

2.3 Related Works

2.3.1 Ensemble

An ensemble of classifiers is a collection of multiple classifiers which is a powerful technique commonly used to improve overall predictive accuracy by consolidating various diversities and accuracies between the classifiers. In other words, different classifiers will make different errors but remain a certain quality of performance. This concept is to make other classifiers disagree with the incorrect decisions made by other classifiers. Such condition decreases the error of ensembles monotonically with an increasing number of classifiers provided that a specific voting mechanism is used and the errors are independent and all the classifiers have the same probabilities of error, provided that probabilities are less than 50% [21]. [6] show that diversified classifiers that do not make the same errors are shown having the abilities to improve performance of ensembles. Sampling from the original datasets and training the classifiers with the datasets obtained from it is the most straightforward to have the classifiers made uncorrelated errors.

Breiman's bagging [22] and Freund and Schapire's boosting [23] are among the most popular ensembles methods that have shown impressive results for improving the predictive performance of ensembles. Both methods are proven to be effective in reducing generalization error [24].Bagging is a method applicable to any

base classifier that uses sampling with replacement or bootstrap in another word to produce new and different versions of training sets that have equal size as the original dataset. Each classifier is then trained on datasets and the outputs are combined using a simple voting to classify an entry. Boosting is a popular alternative to Bagging that uses adaptive sampling. The emphasis of this method is keeping a set of weight which every weight is for particular instance in the training sets. Every instance has the same initial weight and the weight is adjusted to let the classifier focus more on misclassified instances. This can be done by increasing the weights of misclassified objects. Ensemble has been applied in different classification techniques by many researches in attempt to solve various classification problems and to improve the performance of the classifiers. In the next section, we will discuss some examples of base classifiers.

An ensemble selection is a mechanism used by the ensemble to choose which base learners to include in the final ensemble [25] [26] [27]. Many approaches assign a weight to each base learner which specifies whether the individual is included in the ensemble [28] [29] [30]. The simplest way to determine the weight of an individual is by using the individual's fitness as its weight and this method suffers from a limitation that the weights do not reflect how well an individual cooperates with others in the ensemble. This limitation can be overcome by co-evolve the base learner and their weight values in parallel. A secondary training phase is used to optimize the individual weights, which can be done using genetic algorithm and bit-string representation where each bit specifies whether a member is included or not in the ensemble [26] [28].

Bibliography

[1] Y., Inza, I., & Larra~naga, P. Saeys, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 24, no. 4, pp. 2507-2517, 2007.

[2] S Bay, "Nearest neighbor classification from muletiple feature subsets," *Intelligent Data Analysis 3*, vol. (3), pp. 191-209, 1999.

[3] P., Ding, X., & Jiang, B Xia, "A GA-based feature selection and ensemble learning for high-dimensional datasets," vol. 1, pp. 7-12, 2009.

[4] L., & Jain, L. Kuncheva, "Designing classifier fusion systems by genetic algorithms," *Evolutionary Computation, IEEE Transactions*, vol. 4, no. 4, pp. 327-336, 2000.

[5] A., & Nascimento, D. Canuto, "A genetic-based approach to features selection for ensembles using a hybrid and adaptive fitness function," pp. 1-8, 2012.

[6] K., & Pazzani, M. Ali, "Error reduction through learning multiple descriptions.," *Machine Learning, 24(3)*, pp. 173-202, 1996.

[7] Saso Dzeroski, "Relational Data Mining," in *Data Mining and Knowledge Discovery Handbook, Oded Maimon and Lior Rokach.*: Springer US, 2010, pp. 887-911.

[8] Ping Ling and Xiangsheng Rong, "Double-Phase Locality Sensitive Hashing of neighborhood development for multi-relational data," *Computational Intelligence (UKCI), 2013 13th UK Workshop*, pp. 206-213, September 2013.

[9] L Getoor, "Multi-relational data mining using probabilistic relational models: research summary, In Proceedings of the First Workshops in Multi-relational data mining.," 2001.

[10] A.S., Venkatesh, S., West, G Ghanem, "Learning in imbalanced relational data," *Pattern Recogntion. ICPR 2008. 19th International Conference*, pp. 1-4, 2008.

[11] S. Macskassy F.Provost, "A simple relational classifier. In Proceedings of 2nd Workshop on Multi-Relational Data Mining (MRMM).," 2003.

[12] Chen. P, Bin Li Jian Xu, "Random forest for relational classification with application to terrorist profiling.," *Granular Computing, GRC'09. IEEE International Conference*, pp. 630-633, 2009.

[13] Stephen Muggleton, "Inductive Logic Programming," *New Generation Computing*, vol. 8, no. 4, pp. 295-318, 1991.

[14] Wei Zhang, "Multi-Relational Data Mining Based on Higher-Order Inductive Logic," *Intelligent Systems, 2009. GCIS'09. WRI Global Congress, Xiamen*, pp. 453-458, 2009.

[15] Jingfeng Guo, Lizhen Zheng, and Tieying Li, "An Efficient Graph-based Multi-relational Data Mining Algorithm," in *2007 International Conference on Computational Intelligence and Security*, Harbin, 2007, pp. 176-180.

[16] Lawrence B. Holder and Diance J.Cook, "Graph-Based Relational Mining: Current and Future Directions," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 90-93, July 2003.

[17] Dan Roth and Wen-tau Yih, "Propositionalization of Relational Learning: An Information Extraction Case Study," *Seventeenth International Joint Conference on Artificial Intelligence, Seattle*, 2001.

[18] M. Clergue, M. Collard, and L. Izquierdo W. Segretier, "An evolutionary data mining approach on hydrological data with classifier juries," *Evolutionary Computation (CEC), 2012 IEEE Congress, Brisbane*, pp. 1-8, June 2012.

[19] R Alfred, "Optimizing feature construction process for dynamic aggregation of relational attributes," *Journal of Computer Science*, vol. 5, no. 11, p. 864, 2009.

[20] C., Alfred, R., & Keng, L Kheau, "Dimensionality reduction in data summarization approach to elarning relational data," pp. 166-175, 2013.

[21] L., & Salamon, P. Hansen, "Neural network ensembles," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993-1001, 1990.

[22] L Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123-140, 1996.

[23] Y., & Schapire, R. Freund, "Experiments with a new boosting algorithm," 1996.

[24] J Quinlan, "Bagging, boosting, and C4.5," pp. 725-730, 1996.

[25] M. Sebag, M. Schoenauer, and M. Tomassini C. Cagne, "Ensemble learning for free with evolutionary algorithms?," *Proc. Genetic Evol. Comput. Conf.*, pp. 1782-1789, 2007.

[26] X. Yao and Y. Liu, "Making use of population information in evolutionary artificial neural networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions*, vol. 28, no. 3, pp. 417-425, Jun 1998.

[27] D.W. Optiz and J. W. Shavlik, "Generating accurate and diverse members of a neural-network ensemble," *Advance in Nueral Information Processing Systems. Cambridge, MA, USA: MIT Press*, pp. 535-541, 1996.

[28] N. Chawla and J.Sylvester, "Exploring diversity in ensembles: Improving the performance on unbalanced datasets," *Proc 7th International Conference. MCS*, pp. 397-406, 2007.

[29] P.Tino, and X. Yao H.Chen, "Predictive ensemble pruning by expectation propagation," *Knowledge and Data Engineering, IEEE Transactions*, vol. 21, no. 7, pp. 999-1013, March 2009.

[30] A.J Patel and J.S Patel, "Ensemble systems and incremental learning," *Intelligent Systems and Signal Processing (ISSP), 2013 international Conference*, pp. 365-368, March 2013.

[31] P.S Yashwant M.D Hanif, "Cybercrime detection techniques based on support vector machines," *Artificial Intelligence Research*, vol. 2, no. 1, 2013.

[32] H., Pang, S., Je, Hm, Kim, D., & Yang Bang, S Kim, "Constructing support vector machine ensemble," *Pattern Recognition*, vol. 36, no. 12, pp. 2757-2767, 2003.

[33] D., Tang, X., Li, X., & Wu, X Tao, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval.," *Pattern Analysis And Machine Intelligence, IEEE Transactions*, vol. 28, no. 7, pp. 1088-1099, 2006.

[34] L., Kong, F., & Shen, Z He, "Artificial Neural Network Ensemble for Land Cover Classification," vol. 2, pp. 10054-10057, 2006.

[35] Y., & Yu, C Wang, "Predicting project success using ANN-ensemble classificaiton models," pp. 47-51, 2011.

[36] Iwona Pozniak-Koszalka, Leszek Koszalka Mateusz Budnik, "The Usage of the k-Nearest Neighbour Classifier with Classifier Ensemble," *2012 12th International Conference on Computational Science and Its Applications*, pp. 170-173, 2012.

[37] Wenjia Wang, "Heterogeneous Bayesian ensembles for classifying spam emails," *Neural Networks (IJCNN), The 2010 International Joint Conference*, pp. 1-8, July 2010.

[38] Yaochu Jin Shenkai Gu, "Heterogeneous classifier ensembles for EEG-based motor imaginary detection," *Computational Intelligence (UKCI), 2012 12th UK Workshop*, pp. 1-8, sept 2012.

[39] S Khalid and S. Arshad, "Framework for Constructing Hybrid Classifier Using Weight Learning to Combine Heterogeneous Classifiers," *Computational Intelligence, Modelling and Simulation (CIMSim), 2012 Fifth International Conference*, pp. 163-168, Sept 2013.

[40] G., Katakis, I., & Vlahavas, I Tsoumakas, "Effective voting of heterogeneous classifiers.," pp. 465-476, 2004.

[41] L., Zhenjiang, L., Yafeng, Y., Zaixia, T., Junjun, G., & Bofeng, Z Yue, "Selective and heterogeneous svm ensemble for demand forecasting," pp. 1519-1524, 2010.

[42] I., & Elisseeff, A. Guyon, "An introduction to variable and feature selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1167-1182, 2003.

[43] J., & Chawla, N Sylvester, "Evolutionary ensemble creation and thinning," pp. 5148-5155, 2006.

[44] H., Khoshgoftaar, T., & Napolitano, A Wang, "An Empirical Study on Wrapper-Based Feature Selection for Software Engineering Data," vol. 2, pp. 84-89, 2013.

[45] C., & Whitley, D. Guerra-Salcedo, "Genetic approach to feature selection for ensemble creation," 1999.

[46] K., Li, B., Zhang, J., & Du, J. Liu, "Ensemble component selection for improving ICA based microarray data prediction models," *Pattern Recognition*, vol. 42, no. 7, pp. 1274-1283, 2009.

[47] Y., & Na, W. Ming-hai, "Research on the ensemble learning classification algorithm based on the novel feature selection method," pp. 263-267.

[48] W., WeiJuan, L., Rui, L., & Xuyang, W Yan, "Feature selection based on bagging ensemble learning algorithm," pp. 734-736, 2009.

[49] L., Li, D., & Xiao, J Xie, "Feature selection based transfer ensemble model for customer churn prediction," vol. 2, pp. 134-137, 2011.

[50] K., Neto, A., Canuto, A., & Dias, F Vale, "Static and Dynamic Weights in Ensemble Systems Built by Class-Based Feature Selection Methods," pp. 61-66, 2010.

[51] T.M. Mitchell, "Machine Learning," *New York: McGraw-Hill Companies Inc*, pp. 230-247, 1997.

[52] D Opitz, "Feature selection for ensembles," pp. 379-384, 1999.

[53] M., Punch, W., Goodman, E., Kuhn, L., & Jain, A Raymer, "Dimensionality reduction using genetic algorithms," *Evoluitionary Computation, IEEE Transactions*, vol. 4, no. 2, pp. 164-171.

[54] M., & Patnaik, L Srinivas, "Adaptive probabilities of crossover and mutation in genetic algorithms," *Systems, Man and Cyvernetcis, IEEE Transaction*, vol. 24, no. 4, pp. 656-667, 1994.

[55] Lawrence B. Holder and Diane J. Cook, "Graph-Based Relational Mining: Current and Future Directions," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 90-93, July 2003.

[56] Hongyu Guo and Herna L. Viktor, "Mining Relational Databases with Multi-view Learning," in *Proceedings of the 4th international workshop on Multi-relational mining*, New York, 2005, pp. 15-24.

[57] Dan Roth and Wen-tau Yih, "Propositionalization of Relational Learning: An Information Extraction Case Study," in *Seventeenth International Joint Conference on Artificial Intelligence*, Seattle, 2001.

[58] Chung Seng Kheau, Rayner Alfred, and Lau Hui Keng, "Dimensionality Reduction in Data Summarization Approach to Learning Relational Data," in *Intelligent Information and Database Systems Lecture Notes in Computer Science*. Berlin, Germany: Springer, 2013, pp. 166-175.