TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.2, February 2014, pp. 1487 ~ 1494
DOI: http://dx.doi.org/10.11591/telkomnika.v12i2.3635

Review on Studies and Advances of Machine Learning Approaches

Yongqing Wang*, Qingxiu Li
Department of Computer Science and Applications, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou, 450015, China
*Corresponding author, e-mail: wyq-yongqing@163.com

Abstract

Artificial intelligence is a frontier field of computer science, and achieved considerable progress in the past few decades. Being an important research branch of artificial intelligence, machine learning has been successfully applied to many fields in recent years, such as expert system, automatic reasoning, natural language processing, pattern recognition, computer vision, intelligent robots, and so on. This article comprehensively introduces the main strategies of machine learning, and summarizes the existing problems and challenges.

Keywords: *computer science, artificial intelligence, machine learning, learning mechanism*

1. Introduction

Learning is one of the key features that humans differ from other lower forms of life, and means an important intelligent behavior of humans. In the process of development and popularization of modern science and technology, all kinds of definitions of learning had been given. H. A. Simon believed that, learning is the adaptive changes made by a system, which could make it more effective in the next time when required to complete the same or similar task. R. S. Michalski argued that, learning is the expression of construction or modification to the experienced things. These two views both have its own focus. The first puts more emphasis on the effect of learning behavior, the second puts more emphasis on the expression of learning object. What we prefer to is the view pointed by Simon, "If a system can improve its performance by implementing some process, it is learning".

With the rapid development of modern technology, especially the rapid development of computer technology, people wondering that, can we make the computers has the same or similar learning ability as the humans. The problem got affirmative answer in the field of machine learning. In general, machine learning is to make machines constantly improve their performances, realize self-improvement, and acquire new knowledge and new skills by the use of existing knowledge provided by humans, based on the identification of object.

As an important research field of artificial intelligence, the mainly tasks of machine learning are learning mechanism, learning methods, and the oriented task [1]. With the development of artificial intelligence, machine learning had many successful research results and applications in all the fields of artificial intelligence, and became an energetic and challenging research subject.

How to divide the development stage of machine learning? Different scholars gave different answers. Vapnik, one of the founders of the statistical learning theory, divided its research and development history into four stages [2]. A) The emergence of learning machine. B) The establishment of the foundation of learning theory. C) The appearance of Neural Networks. D) The foundation of Statistical Learning Theory.

The development of machine learning is very rapid, and there are a lot of excellent learning strategies, we can classify them from different angles. According to the learning ability of system, they can be divided into supervised learning, semi-supervised learning, and unsupervised learning. According to the nature of learning object, they can be divided into concept learning and process learning. According to representations of knowledge, they can be divided into logical representations learning, productive representations learning, and frame

Received June 25, 2013; Revised August 27, 2013; Accepted September 29, 2013

representations learning. According to the reasoning way, they can be divided into deduction-based learning and induction-based learning. According to the comprehensive property, they can be divided into Inductive Learning, Analytical Learning, Connective Learning, Genetic Algorithm and Classifier System, and so on. In this article, according to the appearance time of classification criteria, we divide these strategies into two classes, traditional strategies and modern strategies.

2. Traditional Strategies of Machine Learning

According to the different complexity of reasoning strategies adopted in the system of learning process, traditional strategies can be divided into five types, Rote Learning, Learning by Being Told, Inductive Learning, Learning by Analogy, and Explanation-Based Learning.

2.1. Rote Learning

Rote Learning is the simplest, most primitive, and most basic learning strategy, which can achieve the purpose of learning through memorizing and evaluating the external environment provided information. What the learning system going to do is storing the evaluated knowledge to the knowledge base, from which it can retrieve the corresponding knowledge to directly solve the required problems. 1979, Hayes-Roth, Lenat and Klahr put forward an interesting idea about Rote Learning (Seen in Figure 1). They point out that we can take the Rote Learning as the first level in classification of data reduction, and data reduction has the same purpose with computer language to compile the original information into executable information. For Rote Learning, we need to pay attention to three important issues. A) Storage and organization of information. Using the proper storage style, and making the searching process as fast as possible. B) The stability of environment and applicability of stored information. Learning system must ensure that the stored information adapted to the changes of external environment. C) Tradeoff between storage and computing. Rote learning does not have any reasoning ability, all the information placed in computer will be added in some new, it is essentially using storage space for the processing time. Therefore, Rote Learning must fully weigh the relationship between storage and computing.

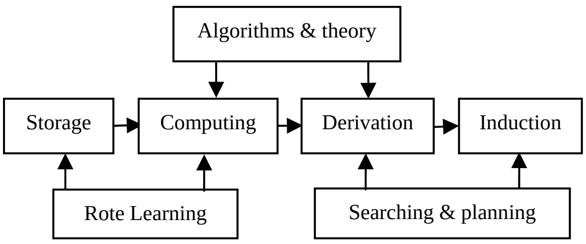

Figure 1. The Diagram of Levels in Reduced Data

2.2. Learning by Being Told

In Learning by Being Told, system can be provided with general instructions or proposals by external environment, which would be specifically converted to detailed knowledge and then sent into knowledge base (Seen in Figure 2).

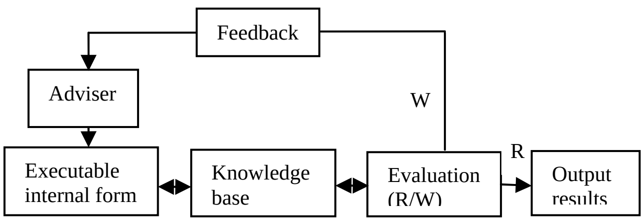

Figure 2. The Diagram of Learning by Being Told

The produced knowledge will be evaluated repeatedly in the learning process, which results in the knowledge base of system continuously improved. Learning by Being Told is a more practical learning method, which can be used to obtain the expert knowledge. It can avoid the difficulties brought by analysis and induction on its own system to produce new knowledge, and can do without domain experts' understanding about the system internal knowledge representation and organization, so obtains widely applications in extensive fields.

2.3. Inductive Learning

Inductive Learning is the application of inductive reasoning to learning (Seen in Figure 3). Inductive reasoning is a reasoning adopting inductive method, which means to conclude general knowledge from enough cases, and is a kind of reasoning from individual to general. The popular inductive methods include enumeration induction, association induction, analogy induction, inverse reasoning induction and eliminate induction, etc. For the reason that we cannot investigate all of related cases during induction in most cases, so the induced conclusion cannot guarantee its absolute accuracy, it only can be believed to be true to some degree, which is an important feature of inductive reasoning. Inductive Learning can be divided into case learning, observation and discovery learning according to whether or not there exists the guidance of teachers. Inductive Learning is the most basic and mature learning method, which has been widely applied in the research and application fields of artificial intelligence.

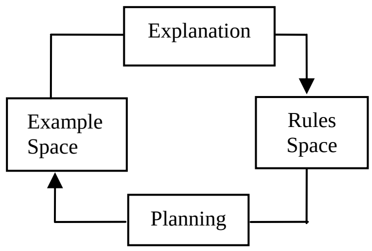

Figure 3. The Diagram of Inductive Learning

2.4. Learning by Analogy

Analogy is an important method for human to know the world, and an important measure to guide human to learn new things, conduct creative thinking. Learning by Analogy is a learning method by comparing similar things. Analogical reasoning is the foundation of Learning by Analogy, whose process can be divided into four steps, recall and association, selection, establishing corresponding relation and conversion. The methods of Learning by Analogy can be usually divided into attribute analogy learning and conversion analogy learning.

Learning by Analogy plays an important role in the progress history of human science and technology, by which many scientific discoveries had been obtained. But, the main difficulty in Learning by Analogy is association for analogy source, namely, given a target domain, and then find out from the countless intricate structures one or several candidate analogy sources. In the actual application, analogy sources are given by the user, which actually results in that the system can only repeat the analogies people known, but can't help people learn new knowledge.

2.5. Explanation-based Learning

In recent years, Explanation-based Learning (EBL) has rising as a learning method in the field of machine learning. On use of relevant domains' knowledge, analyzing the currently provided cases, it can construct interpretation, create corresponding knowledge, as well. In 1986, Mitchell et al. proposed a unified EBG algorithm for Explanation-based Learning, which established the explanation-based summarization process, and utilized the logic representation of knowledge and deductive reasoning for problem solving (Seen in Figure 4).

During the learning process of Explanation-based Learning, system uses relevant domains' knowledge to deduct gradually, and then constructs the explanation eventually to satisfy the target concept. Domain knowledge plays an important role in the formation of explanation, which requires domain knowledge is perfect, complete enough, can explain all the cases to be handled. But in the real world, most of the domain knowledge does not have this feature, knowledge is unavoidable incomplete. Accordingly, we need to study how to make EBL still effective in incomplete domain knowledge. Meanwhile, we also need to study how to modify the incomplete domain knowledge to have stronger explanation ability.

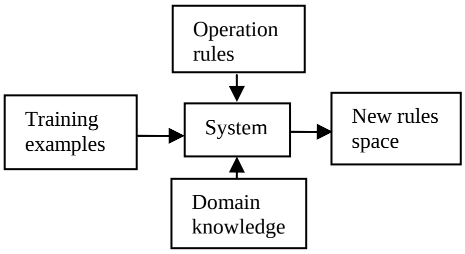

Figure 4. The Diagram of EBG Algorithm

3. Modern Strategies of Machine Learning

According to the classification criteria adopted in other scholars [3, 4], we can divide the modern strategies into the following five types.

3.1. Neural Networks-based Learning

Neural Networks (NNs) is composed of some simple neuron-like units, and the connection with weight between each of them (Seen in Figure 5). The state of every unit is decided by the inputs of other units, which is connected with it (Seen in Figure 6). NNs is different from Symbolic Learning, called Connective Learning. NNS trains network by using all kinds of examples, produces the internal representation of network, by which to identify the other inputted examples. Due to its highly parallel distributed processing ability, NNs has made great successes and developments in recent years. The well-known models or algorithms of NNs including Single-layer Perception, Hopfield Network, Boltzmann Machine and Back Propagation (BP) algorithm.

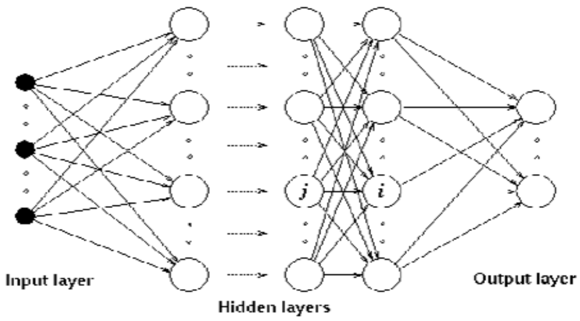

Figure 5. The Structure of Neural Networks

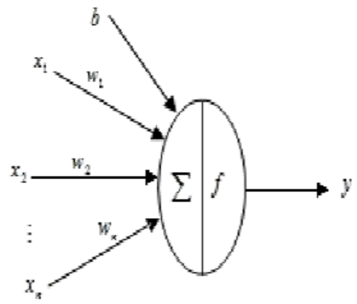

Figure 6. The Structure of a Single Neuron

NNs has made successful applications in many fields [5-7]. However, in practical applications, because of the lack of prior knowledge, users often need to choose the appropriate neural network models, algorithms and parameters through lots of time-consuming tests, which results in the application effect is completely depends on the user's experiences, and the lack of the rigorous theoretical system's guide. In addition, other defects of NNs, such as over-fitting and local minima problems, have recently limited its developments both in researches and in applications.

3.2. Statistical Learning Theory-based Learning

Compared with the traditional statistics, the Statistical Learning Theory (SLT), founded by Vapnik in 1960s to 1970s, is a kind of theory specially researches the laws of machine learning under limited samples. It not only considers the requirements of asymptotic performance, but also pursuits the optimal results with limited information [2].

In particular, in the period from 1992 to 1995, the SLT-Based Support Vector Machine (SVM) method [2], [8-10], proposed by Vapnik, fundamentally solves some of the defects of the traditional statistics theory. For example, SVM can effectively handle the problem of high dimensional data with limited samples, and has the merits of good generalization ability, convergence to the global optimum, and dimension insensitivity. The principle of SVM can be concisely interpreted in Figure 7 and Figure 8.

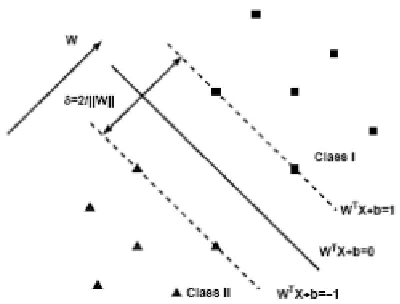

Figure 7. The Diagram of Optimal Separating Hyper-plane with Maximal Margin

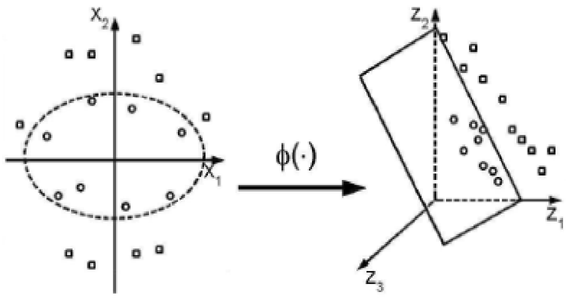

Figure 8. The Diagram of Kernel Function: Nonlinear problems are linear in kernel-induced space

After the developments in recent years, SVM has become a hot spot of machine learning, and has been successfully applied in many fields [11-15]. Such as, face detection, speech recognition, image recognition, signal processing, handwritten numeral recognition, automatic text classification, machine translation, etc. At present, the SVM method has become the youngest and the most practical content in SLT, the related theories and applications are quickly evolving.

3.3. Reinforcement Learning

Reinforcement Learning, also called Reward Learning, or Evaluation Learning, whose principle can be early traced back to the experiment of Pavlovian Conditioning. The theory of Reinforcement Learning formed in the 1980s, based on trial and error method, dynamic programming and temporal difference method [3].

Simon pointed out, in mechanism, this kind of machine learning theory is different from the learning in the sense of artificial intelligence. The main difference is, it emphasizes the adaptation to the change of environment, which means the need to establish a kind of feedback-based learning theory (Seen in Figure 9). Its basic principle can be briefly described as follows. If some behavior strategy of agent can leads to the positive rewards of environment, then the agent will strengthen the trend to generate the behavior strategy [16].

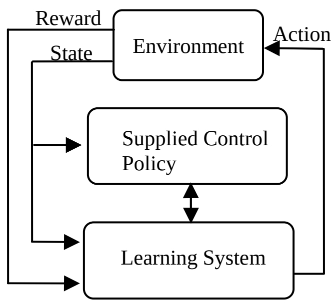

Figure 9. The Diagram of Reinforcement Learning

The main algorithms of Reinforcement Learning including Temporal Difference Method, O-Learning Algorithm, Adaptive Heuristic Critic Algorithm, and the Reinforcement Learning based on other learning machine [3].

At present, Reinforcement Learning has widely researches and applications in the fields of artificial intelligence, machine learning and automatic control. Especially, after the breakthrough progress achieved in the mathematics foundation, Reinforcement Learning has become one of focuses in the field of machine learning [17].

3.4. Ensemble Learning

The method of traditional machine learning is looking for a classifier, which is proximal to the actual classification function in the hypothesis space [18]. Ensemble Learning combines the classification results of several single classifiers to decide the final classification result on new samples [19, 20], by which better performances than single classifiers can be obtained.

The simplest Ensemble Learning is Bagging [21]. It generates several different training sets by resampling, trains classifier in each set, and then combines the results by Voting. This method is very effective, for the fact that it increases the bias slightly, and reduces the variance greatly [22].

In Boosting method, each training sample have an ever-changing weight, the training of new classifiers concentrates on the samples tend to be misclassified in the last time. AdaBoost is the fundamental algorithm in Boosting, from which derived most of the other algorithms of Boosting. The popular algorithms include Discrete AdaBoost, Real AdaBoost, LogitBoost and Gentle AdaBoost [23].

In addition to Bagging and Boosting, there are some other methods. Stacking [24] utilizes the outputs of every single classifier to form the inputs of higher classifiers. Meta Learning [25] uses all the outputs of basic Learning Machines to formulate a meta-classifier. Ensembles of other basic classifiers, such as ensemble of Naive Bayesian, ensemble of Decision Tree [26], ensemble of NNs [27-29], and ensemble of KNN [30]. Besides, there are some online ensemble learning methods [31, 32].

Although the Ensemble Learning has made many achievements in some fields recently, but there are still some hard topics to be further studied in future, such as the selective ensemble, the comprehensibility of ensemble [33].

3.5. Active Learning

At present, the reality situation machine learning faces is, the unlabelled samples are numerous and facile, the labeled samples are scarce and rare. In order to solve the above problem, the professor Angluin of Yale University put forward the method of Active Learning [34] in 1988. What it differs from the previous algorithms is, Active Learning simulates the process of human learning, improves the generalization capability of classifier iteratively by choosing some unlabelled samples to training set. Due to the wide applicability and efficient use of human experts, Active Learning attracts extensive attention and develops rapidly, and becomes one of the most important directions in the field of machine learning [35].

According to the sampling strategies in the choice of unlabelled samples, Active Learning can be divided into the following types, Membership Query Synthesis, Stream-Based Active Learning and Pool-Based Active Learning [36].

The defect of Membership Query Synthesis is leaving all the unlabelled samples to human experts for labeling, without considering the actual distribution of samples [37]. The defect of Stream-Based Active Learning is the lack of universality to the problems of different learning.

In order to solve the above-mentioned problems, Lewis proposes Pool-Based Active Learning [38]. The sampling strategy is, making up a ‘pool’ with unlabelled samples, in which to calculate and compare the information of unlabelled samples, choosing the samples with more information for labeling.

The sampling strategies of Pool-Based Active Learning inherit the advantages of the two formers, and overcome their shortcomings, and thus become the most widely-used sampling strategy. It has many successful applications in the fields of text classification, information extraction, image retrieval, video retrieval and cancer testing.

Although Active Learning has excellent performances in the actual applications, however, there are still some difficult problems to be solved in future, such as the sampling strategies under inaccessible conditions, online labeling, and the price-sensitive problem [35].

4. Conclusion

The research in Machine Learning aims at making computers obtain knowledge from the real world the way like human, reveal the mechanism of human learning and the mystery of the human brain. Along with the deepening of research, it is not difficult to find that, no matter what kind of learning methods, all have some limitations or one-sidedness, which results in the failure to fully reveal the mechanism of brain thinking and human learning.

Machine Learning is a very active and vigorous research field, also a field full of difficulties and controversies, there are still a lot of problems and challenges not completely solved. They can be summarized as follows. The problems in Machine Learning include dimension disaster problem, PU learning problem, data generalization problem, and Over-fitting problem [39-41]. There are five issues about the challenges for Machine Learning to be further studied [33]. (a) Could the generalization ability be more accurate? (b) Could the training or testing speed be faster? (c) Could the learning mechanism be easier to understand? (d) Could all the data be used efficiently? (e) Could the costs be handled discriminately? Any breakthrough achieved in the problems and challenges mentioned above can make great contributions to the progress of Machine Learning.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 41001235), the Aeronautical Science Foundation of China (Grant No. 2011ZC55005), the Project of Henan Provincial Audit Office (Grant No. 2012-0922), the Scientific & Technological Research Key Project of Henan Provincial Education Office (Grant No. 13A520404), and the Project of Henan Provincial Scientific & Technological Office (Grant No. 132102210468).

References

[1] TM Mitchell. *Machine Learning*. New York: Mc Graw-Hill, 1997.
[2] V Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag. 1995.
[3] Yan Youbiao, *A Survey on Machine Learning and Its Main Strategy*. *Application Research Of Computers*. 2004; 21(7): 4-13.
[4] TG Dietterich, Machine-Learning Research Four Current Directions. *AI Magazine*. 1997; 18(4):97-136.
[5] JAK Suykens, J Vandewalle, BD Moor. Optimal control by least squares support vector machines. *Neural Networks*. 2001; 14(1): 23-35.
[6] T Joachims. *Optimizing search engines using clickthrough data*. Proc. 8th ACM SIGKDD. 2002; 133-142.
[7] S Tronci, M Giona, R Baratti. *Reconstruction of chaotic time series by neural models: a case study*, Neurocomputing. 2003; 55: 581-591.
[8] C Cortes, V Vapnik. Support vector networks. *Machine Learning*. 1995; 20: 273-297.
[9] V Vapnik. Three Fundamental Concepts of the Capacity of Learning Machines. *Physica A*. 1993; 200: 537-544.
[10] V Vapnik. *Statistical Learning Theory*. New York: J. Wiley. 1998.
[11] K Kim. Financial time series forecasting using support vector machines. *Neurocomputing*. 2003; 55: 307-319.
[12] Y Wang, IV Tetko. Gene selection from microarray data for cancer classification-a machine learning approach. *Computational Biology and Chemistry*. 2005; 29(1): 37-46.
[13] G Dror, R Sorek, S Shamir. Accurate identification of alternatively spliced exons using Support Vector Machine. *Bioinformatics*. 2005; 21(7): 897-901.
[14] IW Tsang, JT Kwok, PM Cheung. Core Vector Machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research*. 2005; 6: 363-392.
[15] S Asharaf, MN Murty, SK Shevade. *Multiclass Core Vector Machine*. Proc. of the 24th International Conference on Machine Learning. 2007; 41-48.
[16] Wen Wang. *Simple Analysis on the Research and Application of Machine Learning*. Computer & Information Technology. 2010; 22: 7-9.
[17] Zhongzhi Shi. *Knowledge Discovery*, Beijing: Tsinghua University Press. 2011.
[18] Tom M Mitchell. *Machine Learning*: McGraw Hill. 1997.

[19] TG Dietterich. *Ensemble Methods in Machine Learning*, In Multiple Classier Systems, Italy: Cagliari. 2000.
[20] G Valentini, F Masulli. *Ensembles of learning machines*, In Neural Nets WIRN Vietri02. Series Lecture Notes in Computer Sciences. 2002.
[21] L Breiman. Bagging Predictors. *Machine Learning*. 1996; 24(2): 123-140.
[22] Pedro Domingos. *A Few Useful Things to Know About Machine Learning*. In Communications of the ACM. 2012; 10.
[23] Chen Kai. *A Summary of Machine Learning and Related Algorithms*. Statistics and Information Forum. 2007; 22(5): 105-112.
[24] DH Wolpert. *Stacked Generalization*. Neural Networks. 1992; 5: 241-259.
[25] V Ricardo, D Youssef. A perspective view and survey of meta-learning. *Artificial Intelligence Review*. 2002; 18(2): 77-95.
[26] M Kearns, Y Mansour. *On the boosting ability of top-down decision tree learning algorithms*. Proc. of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing. 1996.
[27] ZH Zhou, J Wu, W Tang. Ensembling Neural Networks: Many Could Be Better than All. *Artificial Intelligence*. 2002; 137(2): 239-263.
[28] ZH Zhou, J Wu, W Tang, ZQ Chen. *Selectively Ensembling Neural Classifiers*, In International Joint Conference on Neural Networks. 2002; 2.
[29] LK Hansen, P Salamon. Neural Network Ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1990; 12(10): 993-1001.
[30] ShiXin Yu. *Feature Selection and Classifier Ensembles: A Study on Hyper-spectral Remote Sensing Data.* 2003.
[31] Alan Fern, Robert Givan. *Online Ensemble Learning: An Empirical Study*. Proc. of the 17th International Conference on Machine Learning. 2000.
[32] Tsymbal A, Puuronen S. *Bagging and Boosting with Dynamic Integration of Classifiers, Principles of Data Mining and Knowledge Discovery*. Proc. PKDD. 2000.
[33] Zhihua Zhou. *Machine learning and its challenges*. Nanjing. 2003.
[34] D Angluin. Queries and concept learning. *Machine Learning*. 1988; 2(4): 319-342.
[35] Wu Weining, Liu Yang. Advances in Active Learning Algorithms Based on Sampling Strategy. *Journal of Computer Research and Development*. 2012; 49(6): 1162-1173.
[36] B Settles. *Active learning literature survey*, Madison, Wisconsin: Computer Sciences. University of Wisconsin-Madison. 2009.
[37] EB Baum, K Lang. *Query learning can work poorly when a human oracle is used*, in Proc. of IEEE IJCNN. 1992; 335-340.
[38] D Lewis, J Catlett. *Heterogeneous uncertainty sampling for supervised learning*. Proc. of ICML. 1994; 148-156.
[39] Liu Xin. Learning Techniques for Automatic Test Pattern Generation using Boolean Satisfiability. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 4077-4085.
[40] Jue Wang. *Statistical Machine Learning*. Proc. of the CCML. Dalian. 2008.
[41] Mu Zhang. Neural Network Based Color Recognition for Bobbin Sorting Machine. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 3728-3735.