Knowledge Assimilation of Machines Using Various Approaches

Ebin P M, Kavitha Nair R, Pradeeba V

Abstract*: Machine learning (ML), Artificial Intelligence (AI) and Data science are some of the top trending topics today. Machine learning can be seen as a branch of Artificial Intelligence and using machine learning; programs can scan and process huge databases. One of the core objectives of machine learning is to construct algorithms that can learn from the previous data and make predictions on new input data also called an automated learning. Knowledge assimilation of machines can be done through supervised learning, unsupervised learning, semi supervised learning and reinforcement learning. In this article, we present two most widely used supervised learning algorithms for knowledge assimilation. The machines learn things from data, usually known as training data, and apply the knowledge to different circumstances and this learning is a continuous process.

Index Terms*: Decision tree, Machine learning, Supervised learning, Support vector machine.

I. INTRODUCTION

In machine learning, once an algorithm learns what to do with data, it can do its work automatically. The input to a learning algorithm is normally called training data and it can be textual, numerical, video, audio or any multimedia. The output is any expertise. We are creating the best model using machine learning algorithms for attaining our aim. Training data and test data [6] are the main concepts in machine learning. Test data set is a set of observations used to evaluate the performance of the model that we created. No training sets are included in the test set because it leads to a difficulty to assess whether the algorithm has learned to generalize from the training set data or has simply memorized it. If a program that generalizes well, then can perform a task using new data efficiently and effectively. Different learning methods can be used in various applications like vision processing, Language processing, Pattern recognition, Robotics and so on. We know that learning is the process of converting experience in to expertise or knowledge.

Revised Manuscript Received on April 07, 2019.

Ebin P M, Bachelor's degree in Computer Science and Engineering from "The Institution of Engineers (INDIA)" of Kolkata in 2011. Institution of Engineers (INDIA)" of Kolkata in 2011 and master's degree in Computer Science and Engineering from Hindustan university, Chennai in 2013.

Kavitha Nair R, Assistant professor in KITS Engineering College and graduated B.Tech in Computer Science & Engineering in 2012.

Pradeeba V, Assistant professor in KITS Engineering College and graduated BE in Computer Science & Engineering in 2012, Computer Science & Engineering in 2012 and Post-graduation in Computer Science and Engineering from Cape Institute of Technology,Levengipuram.

Memorization and generalization are the two sides of these algorithms and a balancing between memorization and generalization is needed. In addition to the two set of observations(Training data set and Test data set), there is another observation set called validation set for validating the model which we created. We can allocate 50 percent or more of the data to the training set, 25 percent to the test set and the remaining to validation set.

Memorization and generalization are the two sides of these algorithms and a balancing between memorization and generalization is needed. In addition to the two set of observations(Training data set and Test data set), there is another observation set called validation set for validating the model which we created. We can allocate 50 percent or more of the data to the training set, 25 percent to the test set and the remaining to validation set.

Nowadays the training data will be scarce and it is a big challenge to get the real time training data. For resolving that problem we can use a method called cross validation [2] .Here the training data is partitioned. For example the training data is partitioned in to five equal parts labeled L, M, N, O and P. Initially the model is trained on partitions M through P, and tested on partition L. In the next iteration, the model is trained on partitions L, N, O, P and tested on partition M. This is repeated until the testing and training performs using all of the partitions.

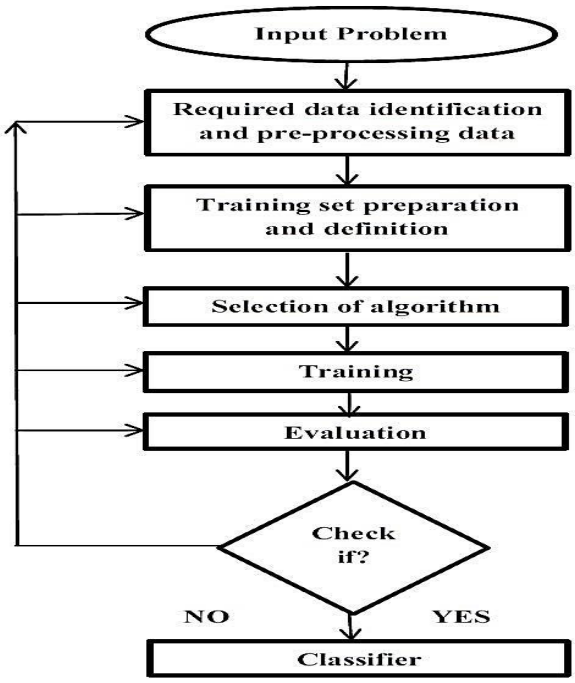

Fig 1: Workflow diagram of Machine Learning

Machine Learning field is basically divided in to three

subdomains: Supervised learning, unsupervised learning and Reinforcement learning. The supervised learning requires training with known labeled data [1] (The corresponding correct output), while in unsupervised learning the data or instances are unlabeled which forms clusters. Learning from feedback received through interactions with an external environment is called reinforcement learning [4]. Here the system adjusts to dynamic conditions, evaluate its performance and react accordingly.

II. LEARNING UNDER SUPERVISION

Association analysis can be achieved through supervised learning techniques. This type of learning or knowledge assimilation needs external assistance. We teach or train the machine using some set of data also called training data set which are well labeled (we know the output and we are trying to map from input to the target output using a general rule). This is a kind of learning which can be seen in real-world. A supervisor (teacher) teach the machine about how to predict the correct answer using some predefined well known training data set.

Learning under supervision learns a function from available training data, which can be used for mapping new and fresh examples. Classification [3] and regression are the two types of supervised learning methods. Classification outputs are discrete and it attempts to find the appropriate class labels (analyzing positive or negative sentiment, male or female persons etc.). Using supervised classification algorithms we are trying to develop a model which has a good generalization capacity. Regression method trains on and predicts a continuous-valued response like predicting real estate prices or predicting the salary for a new job based on its description etc.

For achieving better result, algorithm selection criteria are most important. Prediction accuracy should be considered for algorithm selection criteria. It can be measured by the formula [2]

Accuracy = Number of correct classifications/ Total number of test cases

To calculate classifier's accuracy some researchers use Cross validation method.

III. DECISION TREE ALGORITHMS

Decision tree [5] is an important data mining method and the humans generally follows this method while making a decision. Using decision tree we can create a training model which can predict class through learning some decision rules. In the tree representation the internal nodes corresponds to an attribute and the leaf node corresponds to a class label. We can merge multiple decision trees [2] and can build a random forest.

The challenging phase for making a decision tree is the identification of the correct root node, which should be the best attribute. This is normally called as attribute selection. A random approach for selecting the root node will create low accuracy and bad result. The two popular attribute selection methods are Information gain and Gini index. For calculating information gain there is another concept called entropy (a measure of uncertainty or randomness).

For a binary classification problem with only two classes if all examples are positive or all are negative the entropy will be zero (low), if half of the records are positive classes and half are of negative class then entropy is one (high). To calculate the information gain, first we calculate the entropy of target then calculate entropy for every attribute. Using information gain formula we will subtract this entropy from the entropy of target and the result is information gain. For example, suppose our class label contains 12 positive and 12 negative instances and then Entropy of target is

$$E(12,12) = -1\times((p(+ve)\times\log(p(+ve)) + (p(-ve)\times\log(p(-ve)))$$

$$= -1\times ((12/24)\times\log(12/24)) + (12/24)\times\log(12/24))$$

$$= 1$$

Table 1: A Sample Training Set

NO	A	B	C	D	CLASS
1	5.1	3.8	1.8	1.2	P
2	6	3.5	1.6	1.5	P
3	5	3	1.5	1.6	P
4	6.2	3.9	1.1	1.6	P
5	5.4	2.5	1.5	1.4	P
6	4.6	2.6	4.6	0.5	P
7	4.2	3.5	1.6	0.8	P
8	6.5	2.8	4.8	1.6	P
9	3.7	2	4.2	0.5	P
10	4.6	2.5	4.5	0.9	P
11	6.6	2.9	1.9	0.8	N
12	6.1	3.4	1.3	1.8	N
13	5.1	3.4	1.5	1.7	N
14	6	3.6	1.4	1.4	N
15	4.9	3.2	4.4	1.3	N
16	4.2	2	4.6	0.5	N
17	5.5	3	4.4	0.6	N
18	6.4	3.3	4.6	0.5	N
19	5	2.2	4.3	1.8	N
20	6.6	3.3	1.8	0.5	N

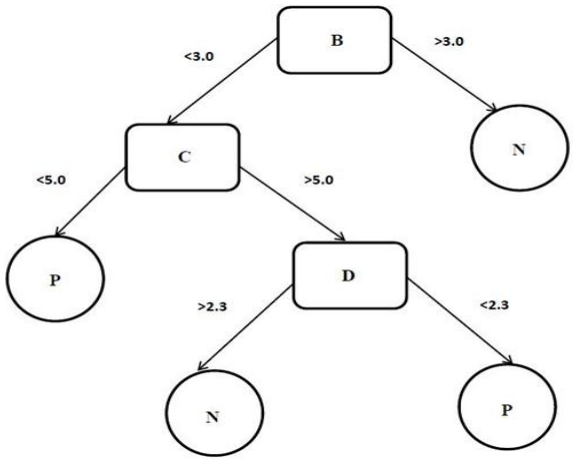

Fig 2: A Sample Decision Tree

Researchers have developed different decision tree algorithms [7] over a

period of time with enhancement. It has an ability to handle various types of data. CHAID (Chi-squared Automatic Interaction Detector) developed by Gordon V Kass in 1980 [8] is a decision tree learning algorithm.

CART (Classification and Regression Tree) [9] produces either classification or regression trees using Gini index as its impurity measure. ID3 (Iterative Dichotomiser 3) provides the entropy reduction in the level of maximum. To handle missing values C4.5 (extension of ID3) is another algorithm that uses information gain as splitting criteria.

C5.0/Sec 5(extension of C4.5) algorithm applied in big data. It split the sample data set based on the field that provides maximum information gain. Missing attributes and multi valued attributes from the data set can easily handle by C5.0. Hunt's algorithm uses divide and conquers approach or top-down approach to create a decision tree. It maintains an optimal split for every stage according to some threshold values. Some of the software's for decision trees are WEKA, GATree [10], Alice d'ISoft, OC1 etc.

IV. SUPPORT VECTOR MACHINE (SVM)

SVM is based on supervised learning methods used for classification and regression (SVR). SVM is based on linear classification. SVM is used for labeling the dataset. There are four types of SVM used in classification

A. The maximal region classifier

B. Kernelized version

C. Soft margin version

D. Combination of all the above

SVM used in regression is called Support Vector Regression (SVR) [11]. A special property of SVM is, it minimize the empirical classification error and maximize the geometric margin. Hence SVM is called Maximal Region Classifier. SVM maps input vector to high dimensional space and a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane which separates the data. This plane maximize the distance between these parallel hyper planes or margin is larger means better will be the classification error.

Assume data points of the form $\{(m_1,n_1),(m_2,n_2),(m_3,n_3),\ldots..,(m_x,n_y)\}$ where $n_x=1/-1$ is the constant belongs to m_x and m_x is the number of sample.

By means of dividing hyper plane the training data takes the form

$$\mathbf{L.\ m+n=0 \quad \rightarrow 0}$$

where n is scalar and L is p-dimensional vector**.** The offset parameter n is added to increase the margin .If the offset is absent then the hyper plane passes through the margin which restricts the solution. The parallel hyper planes can be in the following equation

$$\mathbf{L.m+b = 1}$$
$$\mathbf{L.m+b = -1}$$

According to geometry the distance between the hyper plane is 2/|l|.

$$\mathbf{L.m_i - n>=1 \text{ or } L.m_i - n<=-1}$$

The above can be written as

$$\mathbf{n_i\ (L.m_i - n)>=1,1<=i<=x;}$$

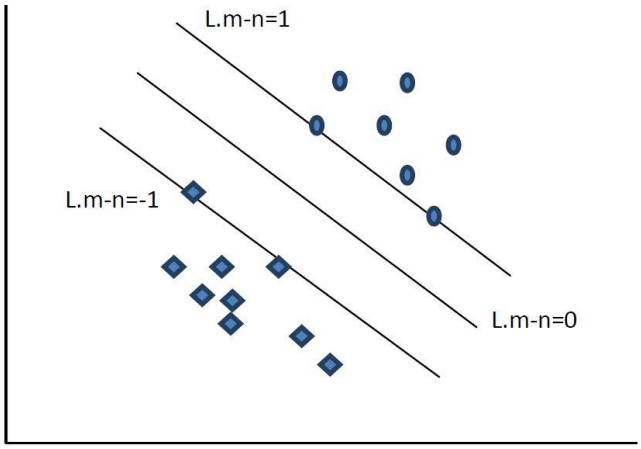

Fig 3: SVM trained samples from 2 classes for a maximal margin hyper plane

The samples along the hyper planes are called support vectors.

Kernelized version

In kernel version, training vectors m_i is mapped to higher dimensional space by the function Φ. Then a linear separating hyper plane is find out in the dimensional space p>0 is the penalty parameter of the error term. The kernel function is defined as

$$\mathbf{K(m_i,m_j)= \Phi(m_i)^T.\ \Phi(m_j)}$$

SVM have an intelligent way to prevent overfitting and can work with large number of features without too much complication. In SVM, a classifier is used that maximizes the separation between points and decision surface.

Soft margin version

Soft margin explains a way of developing classes that separates the classes. This constitutes Support Vector Classifier. This results in higher rate of robustness of individual observations and better classification of training observations.

Non-linearly separable data

There may be chances of mixing up of data which is really difficult to separate as given below

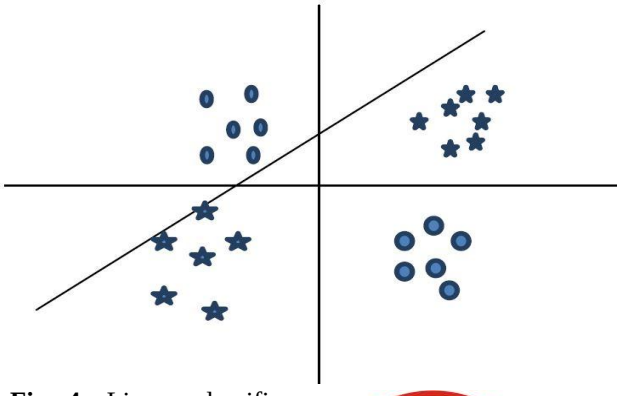

Fig 4: Linear classifier SVM

From the above figure we can predict only 75% of accuracy of data. In the above figure we can see some of the data passes very closest to the margin line. The best solution for better classification is to project the data to a space where it can be linearly separated and to find a hyper plane in this.
Here we can project the dataset into a 3 dimensional plane where the coordinates will be:

$$X_1 = x_1^2$$
$$X_2 = x_2^2$$
$$X_3 = \sqrt{2}x_1x_2$$

When the data project back to the original two-dimensional data the figure becomes as follows

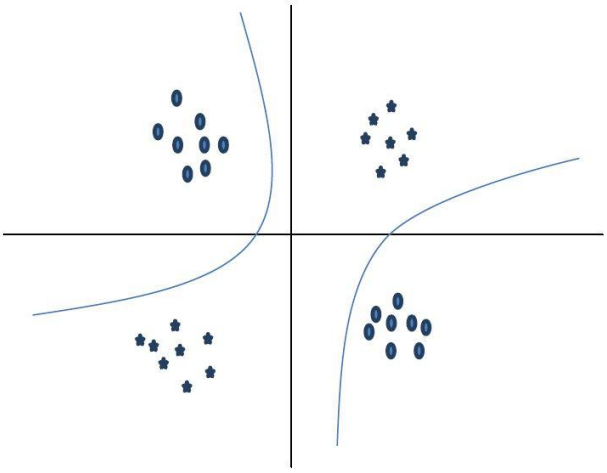

Fig 5: Non-linearly separable data

The above separation gives 100% accuracy as the margin does not pass too close to the data set. The shape of separation boundary depends on the projection and in a projected space it is a hyper plane always.

V. CONCLUSION

This review paper studied the two most important supervised learning algorithms namely Decision Tree and Support Vector Machine (SVM). It gives some fundamental ideas and comparisons about decision tree making and SVM classifications for students and researchers. The type of Machine Learning algorithms strongly depends on the type of problems that we want to solve.

ACKNOWLEDGMENT

I would like to express my gratitude and special thanks to my colleagues for their consistent assistance in completing this work

REFERENCES

1. Ayon Dey. Machine Learning Algorithms: A Review, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179.
2. Iqbal muhammad and Zhu yan. Supervised Machine Learning Approaches: A Survey, ICTACT Journal On Soft Computing, April 2015, Volume: 05(03),2229-6956
3. S. B. Kotsiantis. Supervised Machine Learning: A Review of Classification Techniques, Informatica, Vol. 31(3), 2007, 249-268.
4. Richard S. Sutton and Andrew G.Barto. Reinforcement Learning: An Introduction, Cambridge, MA: MIT Press, 1998.
5. Pierre Geurts, Alexandre Irrthum, Louis Wehenkel. Supervised learning with decision tree-based methods in computational and systems biology, Molecular BioSystems, Vol. 5(12), 2009, 1593-1605.
6. Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar. Foundations of Machine Learning, One Rogers Street Cambridge MA: The MIT Press, 2012.
7. Mr.Brijain R Patel, Mr.Kushik k Rana. A Survey on Decision Tree Algorithm For Classification, (IJEDR) International Journal of Engineering Development and Research, Volume2 (1), 2014, 2321-9939.
8. Gordan.V.Kass. An exploratory Technique for inverstigation large quantities of categorical dataApplied Statics, vol 29(2), 119-127.
9. Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, California.
10. A. Papagelis and D. Kalles (2000). GATree: Genetically evolved decision trees. Proc. 12th International Conference On Tools With Artificial Intelligence, 203-206.
11. Durgesh k. srivastava, Lekha bhambhu. Data Classification Using Support Vector Machine, Journal of Theoretical and Applied Information Technology(JATIT), Volume 12(1)

AUTHORS PROFILE

Er.EBIN P M CEng (I), M.Tech received bachelor's degree in Computer Science and Engineering from "The Institution of Engineers (INDIA)" of Kolkata in 2011 and master's degree in Computer Science and Engineering from Hindustan University, Chennai in 2013.He is working in KITS engineering college as an assistant professor and has published various papers in different national and international journals. Machine Learning and AI are his area of interest.

Er.KAVITHA NAIR R, is working as an assistant professor in KITS Engineering College and graduated B.Tech in Computer Science & Engineering in 2012 and Post-graduation in Software Engineering from Noorul Islam Centre for Higher Education. She has published various papers in national and international journals.

Er. PRADEEBA V is working as an assistant professor in KITS Engineering College and graduated BE in Computer Science & Engineering in 2012 and Post-graduation in Computer Science and Engineering from Cape Institute of Technology,Levengipuram. She has published various papers in national and international journals.