A STUDY ON MACHINE LEARNING: ALGORITHMS AND APPLICATIONS

Sunny Arora, Vishal Kumar

Guru Kashi University, Talwandi Sabo

Abstract

Machine Learning was the phenomenal outcomewhen Computer Science and Statistics joined forces. Computer Science focuses on building machines that solve particular problems, and tries to identify if problems are solvable at all. The main approach that Statistics fundamentally employs is data inference, modellinghypothesises and measuring reliability of the conclusions.The defining ideaof Machine Learning is a little different but partially dependent on both nonetheless. Whereas Computer Science concentrate on manually programming computers, MLaddressesthe problem of getting computers to re-program themselves whenever exposed to new data based on some initial learning strategies provided. On the other hand, Statistics focuses on data inference and probability, Machine Learning includes additional concerns about thefeasibility and effectiveness of architectures and algorithms to process those data, compounding several learning tasks into a compact one and performance measures.

Keywords: *accuracy, algorithm, data, Machine Learning, mathematics, statistics, training*

Introduction

The researchers proposed that how a machine could learn from experience most probably would not be significantly different than how an animal or a human mind learn with time and experience. However, the research concentrated on solving machine learning problems using learning methods of human brain did not yield much promising result so far than the researches concerned with statistical - computational approach. This might be due to the fact that human or animal psychology remains not fully understandable to date. Regardless of these difficulties, collaboration between human learning and machine learning is increasing for machine learning is being used to explain several learning techniques seeing in human or animals. For example, machine learning method of temporal difference was proposed to explain neural signals in animal learning. It is fairly expected that this collaboration is to grow considerably in comingyears.

Data Mining, Artificial Intelligence and MachineLearning

In practise, these three disciplines are so intertwined and overlapping that it's almost to draw a boundary or hierarchy among the three. To put it in other words, these three fields are symbiotically related and a combination of these approachesmay be used as a tactic to produce more efficient andsensitiveoutputs.

Roughly, Data mining is basically about interpreting any kind of data, but it lays the foundation for both artificial intelligence and machine learning. In practice, it not only sample information from various sources but it analyses and recognises pattern and correlations that exists in those information that would have been difficult to interpret manually. Hence, data mining is not a mere method to prove a hypothesis but method for drawing relevant hypotheses.That mined data and the corresponding patterns and hypotheses may be utilised the basis for both machine learning and artificial intelligence.

Artificial intelligence may be broadly defined asmachinesthose having the ability to solve a given problem on their own without any human intervention. The solutions are notprogrammed directly into the system but the necessary data and the AI interpreting that data produce a solution by itself. The interpretation that goes underneath is nothing but a data mining algorithm.

Machine learning takes promote the approach to an advanced level by providing the data essential for a machine to train and modify suitably when exposed to new data. This is known as "training". It focuses onextracting information from considerably largesets of data, and then detects and identifies underlying patterns using various statistical measures to improve its ability to interpret new data and produce more effective results. Evidently, some parameters should be "tuned" at the incipient levelfor betterproductivity.

Machine learning is thefoothold of artificial intelligence. It is improbable to design any machinehaving abilitiesassociated with intelligence, like language or vision, to get there at once. That task would have been almost impossible to solve. Moreover, a system cannot be considered completely intelligent if it lacked the ability to learn and improve from its previous exposures.

Using Unlabelled Data in SupervisedLearning

Supervised learning algorithms approximate the relation between features and labels by defining anestimator f : X→Y for a particulargroup of pre-labeled training data {□x_i, y_i□}. The main challengein thisapproach ispre-labeleddata is not always readily available. So before applying Supervised Classification, data need to be preprocessed, filtered and labeled using unsupervised learning, feature extraction, dimensionality reduction etc. there by adding to the total cost. This hike in cost can be reduced effectively if the Supervised algorithm can make use of unlabelled data (e.g., images) as well. Interestingly, in many special instances of learning problems with additional assumptions, unlabelled data can indeed be warranted to improve the expected accuracy of supervised learning. Like,consider classifying web pages or detecting spam emails. Currently active researchers are seriously taking into account new algorithms or new learning problems to exploit unlabelled data efficiently.

Linking Different MlAlgorithms

VariousML algorithms have been introduced and experimented on in a number of domains. One trail of research aims to discover thepossible correlations among the existing ML algorithms, and appropriate case or scenarios to use a particular algorithm. Consider, these two supervised classification algorithms, Naive Bayes and Logistic Regression. Both of them approach many data sets distinctly, but their equivalence can be demonstrated when implemented to specific types of training data (i.e., when the criteria of Naive Bayes classifier are fulfilled, and the number of examples in trying set tends to infinity). In general, the conceptualunderstanding of ML algorithms, theirconvergence features, and their respectiveeffectiveness and limitations to date remain a radical researchconcern.

Best Strategical Approach for Learners which collects their OwnData

A border research discipline focuses on learning systems that instead of mechanically using data collected by some other means, actively collects data for its own processing and learning. The research is devoted into finding the most effective strategy to completely hand over the control to the learning algorithm. For example consider a drug testing systemwhich try to learn the success of the drug while monitoring the exposed patients for possible unknown side effects and try to in turn minimising them.

CATEGORISATION OF MACHINE LEARNING ALGORITHMS:

An overwhelming number of Machine Learning Algorithms have been designed and introduced over past years. Not everyone of them are widely known. Some of them did not satisfy or solve the problem, so another was introduced in its place. Here the algorithms are broadly grouped into two category and those two groups are further sub-divided. This section try to name most popular ML algorithms and the next section compares three most widely used MLalgorithms.

A. GROUPEDBY LEARNINGSTYLE

1. Supervised learning — Input data or training data has a pre-determined label e.g. True/False, Positive/Negative, Spam/Not Spam etc. A function or a classifier is built and trained to predict the label of test data. The classifier is properly tuned (parameter values are adjusted)to achieve a suitable level ofaccuracy.
2. Unsupervised learning --- Input data or training data is not labelled. A classifier is designed by deducing existing patterns or cluster in the trainingdatasets.
3. Semi-supervised learning --- Training dataset contains both labeled and unlabelled data. The classifieris trained to learn the patterns to classify and label the data as well as topredict.
4. Reinforcement learning --- The algorithm is trained to map action to situation so that the reward or feedback signal is maximised. The classifier is not programmed directlyto choose the action, but instead trained tofindthemost rewarding actions by trial anderror.
5. Transduction --- Though it shares similar traits with supervise learning, but it does not develop a explicit classifier.Itattempts to predict the output based on training data, training label, andtestdata.
6. Learning to learn --- The classifier is trainedto learn fromthe bias it induced during previousstages.
7. It is necessary and efficient to organise the ML algorithms with respect to learning methods when one need to consider the significance of the training data and choose the classification rule that provide the greater level of accuracy.

B. *ALGORITHMS GROUPED BYSIMILARITY*

1. RegressionAlgorithms

Regression analysis is part of predictive analytics and exploits the co-relation between dependent (target) and independent variables. The notable regression models are:Linear Regression, Logistic Regression, Stepwise Regression, Ordinary Least Squares Regression (OLSR), Multivariate Adaptive Regression Splines (MARS) , Locally Estimated Scatterplot Smoothing (LOESS) etc.

2. Instance-basedAlgorithms

Instance-based or memory-based learning model stores instances of training data instead of developing an precise definition of target function. Whenever a new problem or example is encountered, it is examined in accordance with the stored instances in order to determine or predict the target function value.It can simply replace a stored instance by a new one if that is a better fit than the former. Due to this, they are also known as winner-take-all method.

Examples: K-Nearest Neighbour (KNN), Learning Vector Quantisation (LVQ), Self-Organising Map (SOM), Locally Weighted Learning (LWL)etc.

3. RegularisationAlgorithm

Regularisation is simply the process of counteracting overfitting or abate the outliers. Regularisation is just a simple yet powerful modification that is augmented with other existing ML models typically Regressive Models. It smoothes up the regression line by castigatingany bent of the curve that tries to match the outliers.

Examples:Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO) , Elastic Net, Least-Angle Regression (LARS) etc.

4. Decision Tree Algorithms

A decision tree constructs atree like structure involving of possible solutions to a problem based on certain constraints. It is so namedfor it begins with a single simple decision or root, which then forks off into a number of branches until a decision or prediction is made, forming a tree.They are favoured for its ability to formalise the problem in hand process that in turn helps identifying potential solutions faster and more accurately than others.

Examples: Classification and Regression Tree (CART), Iterative Dichotomiser 3 (ID3), C4.5 and C5.0, Chi-squared AutomaticInteraction Detection (CHAID) , Decision Stump, M5, Conditional Decision Trees etc.

5. BayesianAlgorithms

A group of ML algorithms employ Bayes' Theorem to solve classification and regression problems.

Examples:Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Averaged One-Dependence Estimators (AODE), Bayesian Belief Network (BBN), Bayesian Network (BN) etc.

6. Support Vector Machine(SVM)

SVM is so popular a ML technique that it can be a group of its own. Ituses a separating hyperplane or a decision plane todemarcate decision boundaries among a set of data pointsclassified with different labels. It is a strictly supervised classification algorithm. In other words, the algorithm develops an optimal hyperplane utilising input data or training data and this decision plane in turnscategories new examples. Based on the kernel in use, SVM can perform both linear and nonlinear classification.

7. ClusteringAlgorithms

Clustering is concerned with using ingrained pattern in datasets to classify and label the data accordingly.

Examples:K-Means, K-Medians, Affinity Propagation, Spectral Clustering, Ward hierarchical clustering, Agglomerative clustering. DBSCAN, Gaussian Mixtures, Birch, Mean Shift, Expectation Maximisation (EM) etc.

8. Association Rule LearningAlgorithms

Association rules help discovercorrelation between apparentlyunassociated data. They are widely used by e-commerce websites to predict customer behaviours and future needs to promote certain appealing products to him.

Examples:Apriori algorithm, Eclat algorithm etc.

9. Artificial Neural Network (ANN)Algorithms

A model based on the built and operations of actual neural networks of humans or animals.ANNs are regarded as non-linear modelsas it tries to discover complex associations between input and output data. But it draws sample from data rather than considering the entire set and thereby reducing cost and time.

Examples: Perceptron, Back- Propagation, Hop-field Network, Radial Basis Function Network (RBFN)etc.

10. Deep LearningAlgorithms

These are more modernised versions of ANNs that capitalise on the profuse supply of data today.They utilize larger neural networks to solve semi-supervised problems where major portion of an abound data is unlabelledor not classified.

Examples:Deep Boltzmann Machine (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN), Stacked Auto-Encodersetc.

11. Dimensionality ReductionAlgorithms

Dimensionality reduction is typically employed to reduce a larger data set to its most discriminative components to contain relevant information and describe itwith fewer features. This gives a proper visualisationfor data withnumerous features or of high dimensionality and helps in implementing supervised classification more efficiently.Examples: Principal Component Analysis (PCA), Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), Sammon Mapping, Multidimensional Scaling (MDS), Projection Pursuit, Linear Discriminant Analysis (LDA), Mixture Discriminant Analysis (MDA), Quadratic Discriminant Analysis (QDA), Flexible Discriminant Analysis (FDA)etc.

12. EnsembleAlgorithms

The main purpose of an ensemble method is to integrate the projections of several weaker estimators that are singly trained in order to boost up or enhance generalisability or robustness over a single estimator. The types of learners and the means to incorporate them is carefully chosen as to maximise the accuracy.

Examples: Boosting, Bootstrapped Aggregation (Bagging), AdaBoost, Stacked Generalisation (blending), Gradient Boosting Machines (GBM), Gradient Boosted Regression Trees (GBRT), Random Forest, Extremely Randomised Trees etc.

APPLICATIONS

One clear sign of advancement in ML is its important real-life applications, some of which are briefly described here.It is to be noted that until 1985 there was no signifiant commercial applications of ML algorithms.

A. SpeechRecognition

All current speech recognition systems available in the market use Machine Learning approaches to train the system for better accuracy. In practise, most of such systems implement learning in two distinct phases: pre-shipping speaker- independent training and post-shipping speaker-dependent training.

B. ComputerVision

Majority of recent vision systems, e.g., facial recognition softwares, systems capable of automatic classification microscopic images of cells, employ machine learning approaches for better accuracy. For example, the US Post Office uses a computer vision system with a handwriting analyser thus trained to sort letters with handwritten addresses automatically with an accuracy level as high as 85%.

C. Bio-Surveillance

Severalgovernment initiatives to track probable outbreaks of diseasesuses ML algorithms. Consider the RODS project in western Pennsylvania. This project collects admissions reports to emergency rooms in the hospitals there, and the an ML software system is trained using the profiles of admitted patientsin order to detect aberrant symptoms, their patterns and areal distribution. Research is ongoing to incorporatesome additional data in the system, like over-the- counter medicines' purchase history to provide more trainingdata. Complexity of this kind of complex and dynamic data sets can be handled efficiently using automated learning methodsonly.

D. Robot or AutomationControl

ML methods are largely used in robot and automated systems. For example, consider the use of ML to obtain control tactics for stable flight and aerobatics of helicopter. The self-driving cars developed by Google usesML to train from collected terrain data.

E. Empirical ScienceExperiments

A large group data-intensive science disciplines use ML methods in several of it researches. For example, ML is being implemented in genetics, to identify unusual celestial objects in astronomy, and in Neuroscience and psychological analysis.

The other small scale yet important application of ML involves spam filtering, fraud detection, topic identification and predictive analytics (e.g., weather forecast, stock market prediction, market survey etc.).

Conclusion

The foremosttarget of ML researchers is to design more efficient (in terms of both time and space)and practical general purpose learning methods that can perform better over a widespread domain. In the context of ML, the efficiency with which a method utilises data resources that is also an important performance paradigm along with time and space complexity. Higher accuracy of prediction and humanly interpretable prediction rules are also of high importance.Being completely data-driven and having the ability to examine a large amount of data in smaller intervals of time, ML algorithms have an edge over manual or direct programming. Also they are often more accurate and not prone to human bias. Customisation of software according to the environment it is deployed to. Consider, speech recognition software that has to be customised according to the needs of the customer. Like e-commerce sites that customise the products displayed according to customers or email reader that enables spam detection as per user preferences. Direct programming lacks the ability to adapt when exposed to different environment.

References

1. T. M. Mitchell, Machine Learning, McGraw-Hill International,1997.
2. T.M. Mitchel, The Discipline of Machine Learning, CMU-ML-06-108,2006
3. N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines. Cambridge University Press,2000.
4. E. Osuna, R. Freund, and F. Girosi. Support vector machines: training and applications. AI Memo 1602, MIT, May1997.
5. V. Vapnik. Statistical Learning Theory. John Wiley & Sons,1998.
6. C.J.C. Burges. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2):1-47,1998.
7. TaiwoOladipupoAyodele,TypesofMachineLearningAlgorithms,NewAdvancesinMachineLearning, YagangZhang(Ed.),InTech, 2010
8. T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios,A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov,M. Greaves, J. Welling, Never-Ending Learning, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence,2014
9. Pedregosa*et al.*,Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830,2011.
10. Wang, J. and Jebara, T. and Chang, S.-F. Semi-supervised learning using greedy max-cut.Journal of Machine Learning Research , Volume 14(1), 771-8002013
11. Chapelle, O. and Sindhwani, V. and Keerthi, S. S. Optimization Techniques for Semi-Supervised Support Vector Machines, Journal of Machine Learning Research , Volume 9, 203–233,2013
12. J. Baxter. A model of inductive bias learning. Journal of Artificial Intelligence Research, 12:149–198,2000.
13. S. Ben-David and R. Schuller. Exploiting task relatedness for multiple task learning. In Conference on Learning Theory,2003.

14. W. Dai, G. Xue, Q. Yang, and Y. Yu, Transferring Naive Bayes classifiers for text classification.AAAI Conference on Artificial Intelligence, 2007.
15. H. Hlynsson. Transfer learning using the minimum description length principle with a decision tree application. Master's thesis, University of Amsterdam,2007.
16. Z. Marx, M. Rosenstein, L. Kaelbling, and T. Dietterich. Transfer learning with an ensemble of background tasks. In NIPS Workshop on Transfer Learning,2005.
17. R Conway and D Strip, Selective partial access to a database, In Proceedings of ACM Annual Conference, 85 - 89,1976
18. P D Stachour and B M Thuraisingham Design of LDV A multilevel secure relational databasemanagement system, IEEE Trans. Knowledge and Data Eng., Volume 2, Issue 2, 190 - 209,1990
19. R Oppliger, Internet security: Firewalls and beyond, Comm. ACM, Volume 40, Issue 5, 92 -102,1997
20. RakeshAgrawal, RamakrishnanSrikant, Privacy Preserving Data Mining, SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data, Volume 29 Issue 2,Pages 439-450,2000
21. A. Carlson, J. Betteridge, B.Kisiel, B.Settles,E. R.HruschkaJr,and T. M. Mitchell, Toward an architecture for never-ending language learning, AAAI, volume 5, 3,2010
22. X. Chen, A. Shrivastava, and A. Gupta, Neil: Extracting visual knowledge from web data, In Proceedings of ICCV,2013.
23. P. Donmezand J. G. Carbonell, Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In Proceedings of the 17th ACM conference on In- formation and knowledge management, 619–628. ACM,2008
24. T. M.Mitchell, J. Allen, P. Chalasani, J. Cheng, O. Etzioni, M. N. RinguetteandJ. C. Schlimmer, Theo: A framework for self-improving systems, Arch. for Intelli- gence 323–356,1991
25. Gregory, P. A. and Gail, A. C. Self-supervised ARTMAP Neural Networks, Volume 23, 265-282,2010
26. Cour,T.andSapp,B.andTaskar,B.Learning from partiallabels,JournalofMachine LearningResearch,Volume12,1501-15362012
27. Adankon, M. and Cheriet, M. Genetic algorithm-based training for semi-supervised SVM, Neural Computing and Applications , Volume 19(8), 1197-1206,2010