Solid State Technology
Volume: 63 Issue: 6
Publication Year: 2020

A Survey on Selection and Tuning of Machine Learning Models Dynamically for Cloud Network Analysis

Tanniru Annapurna 1,Dr.S.Govinda Rao 2

[1] CSE Department, GRIET, Hyderabad, Telangana, India.
[2] Professor, CSE Department, GRIET, Hyderabad, Telangana, India.
E-mail: [1] tanniruannapurna27@gmail.com , [2] govind.griet@gmail.com

***Abstract*—** In every possible area, machine learning has recently been used to harness its incredible strength. Networking and distributed computing systems have long been the main infrastructure for providing machine learning with powerful computational resources. This promising technology can also benefit from networking itself. This paper Focus on the introduction of MLN, which will not only help to address old, intractable network problems, but would also support new network applications. In thispaper, we summarise the basic workflow to disclose how to apply AI innovation to the systems administration space. AI (ML) models tuned to the informational index would effortlessly get deficient.

The model might be unimaginably exact at one point in time, however it might lose its precision sometime in the future because of changes in input information due to its functionality. Dynamic model learning choice is therenfore regularly required. In this post, we suggest a new approach for cloud automated selection and tuning of ML models that automates the design and selection of models and competes with current methods. To further explore data space before, we use unsupervised learning automated development of targeted supervised learning models.

***Keywords*-** autoselection, autotuning,cloud analytics, machine learning, ensemble learning.

.

I. Introduction

With the effective growth of the Internet, research in both academia and industry has taken place gained a lot of interest in the last few decades. Analysts and organization administrators can encounter different sorts of organizations (e.g. wired or wireless) and software (e.g. network security and live streaming). Each organization application likewise has its own highlights and execution specifications, with time and space can change dynamically. Owing to the Network diversity and complexity, various algorithms are also designed centered on the characteristics of the network and user requirements for different network scenarios. It's a daunting job to build successful algorithms and mechanisms to solve dynamic issues in a number of network scenarios. Recently, in a number of fields of use, bioinformatics, discourse acknowledgment and PC vision, AI (ML) techniques have made breakthroughs. Machine learning is an effort to build algorithms and models that, without following pre-defined rules, learn how to take decisions directly from the results. In general, the latest AI calculations fall into three classes: managed (SL), non-directed (USL) and improved (RL) learning. SL algorithms are learning how to classify or regress more specifically, although USL algorithms concentrate on classifying sample sets in various classes (i.e. clusters) with unmarked data values, the assignments originate from marked information. Operators figure out how to locate the best in RL calculations sequence of actions by interacting with the environment to maximise the cumulative reward (i.e., objective function). Deep learning (DL), learning transfer and adversarial generative learning are the new breakthroughs in an unprecedented way, networks (GAN) often have possible research and application directions. One of the most significant benefits of machine learning is coping with difficult issues. Machine learning can perform similar to or even better than human beings for certain activities including classification, regression and decision-making. Facial recognition and artificial intelligence in games are some examples. Since the field of networks frequently sees uncertainty, it is exciting to integrate deep learning algorithms for the network domain make the most of the powerful ML capabilities in the higher network domain efficiency problems that need efficient solutions. The integration machine learning in design the management of the network also provides an impetus for the creation of new network applications. In fact, the ML techniques were used for a long time in the field of networks.

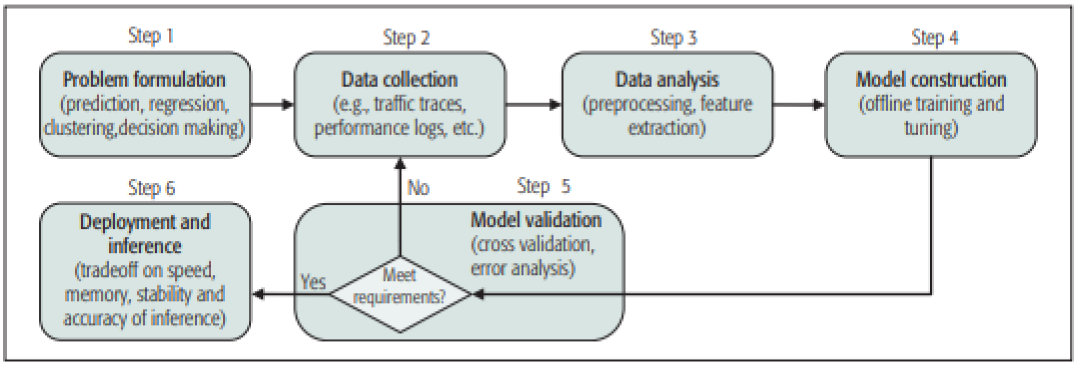


Fig.1: The typical workflow of machine learning for networking

As a consequence, in a single data space, certain prediction models may be extremely accurate, but not so on a relatively different data space. In the event that the training and testing data differ considerably after some time because of changes in outstanding burdens, arrangements, network geographies, and so forth. even robust models can easily become erroneous. Consequently, a single, static model of machine learning sometimes unsuccessful for a long time to come to achieve specific results. One potential solution to this problem is divided learning by dynamic model selection and time-dependent model parameters tuning. More precisely, a complex there is a need for a strategy for self-choice and self-tuning of AI models, which shows extensive fluctuation after some time.

II.RELATED WORK

A survey on feature selection methods :

Because of the accessibility of information for many factors, very high-dimensional data, a range of choices for selecting features are available in the literature. Feature selection techniques provide us with a way to minimise processing time, boost prediction efficiency, and better understand the data in applications for machine learning or identification of patterns. In this article, we include a review any of the approaches present in the literature. The purpose is to include a general prologue to variable end that can be applied to a wide scope of sorts of AI issues. We focus on filtering, wrapper and embedding techniques. To illustrate the applicability of feature selection techniques, we also use some of the techniques for selecting features on standard datasets.

We also attempted to include an introduction to feature selection strategies in this article. The literature on the techniques of feature selection covers applications for machine learning and pattern recognition very widely. Only a single dataset can be used to compare feature selection algorithms, as for different data, each underlying algorithm will work differently. Selection of features techniques suggest that more data in computers is not necessarily healthy Applications for learning. For the data at hand, we can implement various algorithms and we can choose a last element determination calculation with straightforward yield rating esteems. The determination calculation for the current application can be chosen based on the accompanying contemplations: straightforwardness, soundness, diminished number of capacities, exactness of order, stockpiling requirements and specifications for computing. Overall, the collection of features that may also have advantages, such as providing insights into the data, improving the classification model, and improving the consistency of the data generalisation and identifying irrelevant variables. We use the accuracy of classifier and reduced number of features of the Classifier results in this paper to compare the techniques of selection of features. Also, we have successfully used the Indicator Performance and Fault Prediction Data Analysis of Fault Mode Set.

Energy-Efficient Virtual Machines Consolidation in Cloud Data Centers using Reinforcement Learning :

In cloud data centres, complex consolidation strategies maximise resource usage and reduce energy consumption. In order to minimise energy consumption, workload variability should be considered to assess when idle or when idle underutilised hosts turn to sleep mode. We suggest a Strengthening approach to learning-based powerful solidification (RL-DC) in this paper to reduce the number of active hosts needed of resources. Of the RL-DC Uses an agent by using a common reinforcement learning method learn the best policy for choosing the power mode for the host. Deciding when the host should be moved to sleep or active mode, the agent learns from previous experience and strengthens as the workload shifts. RL-DC thus requires no previous workload knowledge what's more, it adjusts dynamic to the world to accomplish online energy and execution the board. The test results of more than 1,000 PlanetLab virtual machines on real workload tracks show that RL-DC minimises and maintains energy consumption the necessary energy consumption output levels.

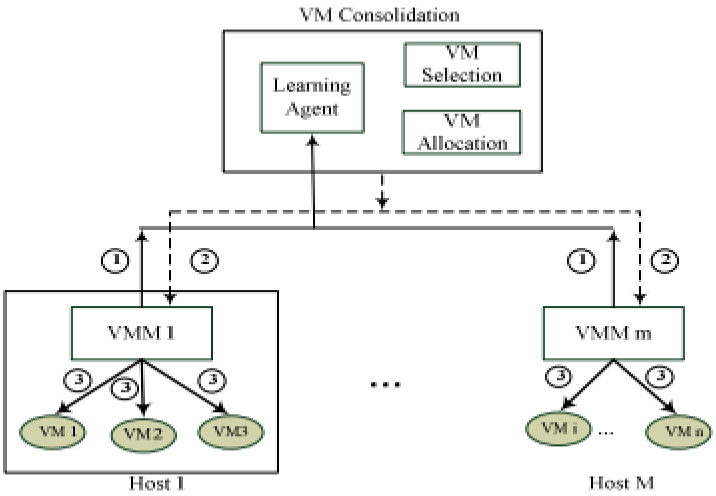


Fig.2: Millions Instructions Per Second (MIPS) model

In this paper, we presented in the cloud data centres dynamic consolidation approach to reducing and breaching energy consumption of SLA. Without previous awareness of the environment and workload, the reinforcement learning technique is used learn how to detect host power mode policy. The method should then adjust the number of active hosts to the number of active hosts currently needed for resources. In CloudSim, compared to existing dynamic consolidation approaches, the proposed reinforcement of complex learning-based consolidation, simulation, will effectively minimise energy costs and the rate of infringement SLA.

III.METHODOLOGY

All This autoscaling function has been updated in our work for the evaluation and selection of machine learning algorithm models and parameters in the containers. In order to transfer messages between containers, a messaging client was used. Within containers , machine learning algorithms are run. The DevOps Cloud auto Model Selection and Tuning Architecture is shown in Figure 3. His detailed block diagram is shown in Figure 2. This configuration simulates the atmosphere of the cloud, where each of the docker containers may be placed, but on the various physical devices, at separate network / geographical locations or within the same network. Cloud micro service instantiation and deletion are the basis for model selection. Containers are brought online that introduce fresh models, while the stale and outdated ones are being taken off-line. For the transmission of messages between containers, we use Kafka. Usage of the Docker — Use of composing allows individual containers to be carried online or taken offline in DevOps fashion (and thus the machine learning workers in them). A different machine learning (ML) algorithm is modelled within each employee container. These machine learning containers will be able to allow some off-line learning in online learning as long as their precision increases, some will deliberately forecast. The sender compartment goes about as an information streaming generator that streams the information continuously.

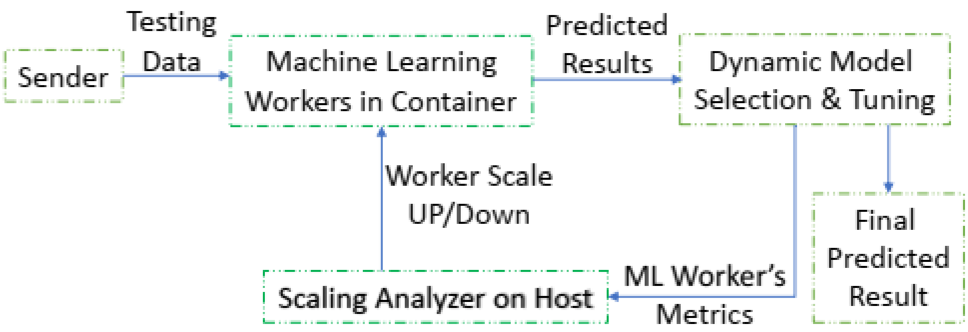


Fig.3: Cloud DevOps Architecture

Automated machine learning techniques enhanced machine learning models, including automatic ensemble generation, have been widely studied, hyper parameter tuning, and feature selection. The comparison between various models of AI and the determination of the best have additionally been recorded, giving an agent perspective on the best in class just as of continuous exploration in this field. While these methodologies advance learning models, the dynamic approach continues to be of practical interest to choose between these tuned models and automatic tuning and choice in the Cloud DevOps climate. Albeit past work has been fruitful in finding ideal calculations for AI, the motivation behind this paper is to give a perplexing technique and setting for model determination and tuning in conditions where signal quality and information content regularly change and where models need to be selected be tuned changed in DevOps system.

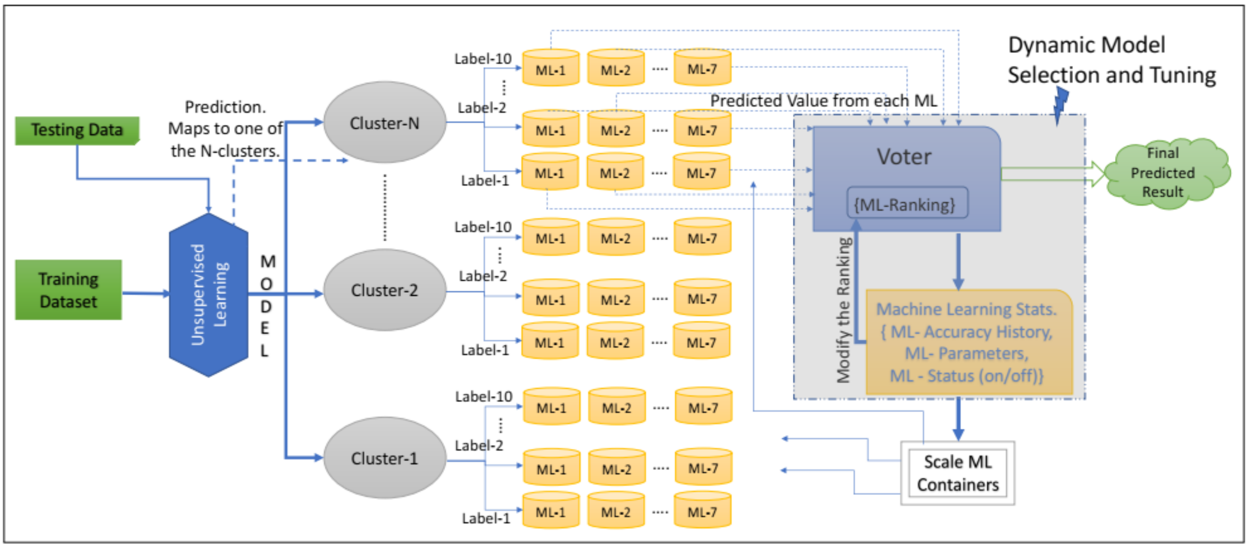


Fig.4: Reflexive DevMLOps Architecture for Autoselection and Autotuning

This paper proposes 3 algorithms:

1.Model Auto Selection Algorithm: In this algorithm we will choose model from algorithm who is correctly predicting class label of given test data or providing better accuracy.

2.Model Selection with Unsupervised Learning + Supervised Learning Algorithm: In this algorithm we will cluster data and choose only that cluster to generate training model who is giving better accuracy.

3.Parameter Tuning with Autoselection Algorithm: Accuracy of one algorithm will be compare with other algorithm to auto select that algorithm who has best accuracy.

We implement 5 algorithms such as Random Forest, Decision Tree, KNN, Stochastic Gradient and Naïve Bayes along with proposed algorithms we can take the help of these 5 algorithms for implementing the proposed concept.

Random Forest: Arbitrary Forests or Random Decision Forests are a gathering learning strategy for grouping, relapse and other purposes purposes activities work through the construction at the training time of a large number of dynamic trees and class results, which is the mode of the person. Tree class (classification) or mean forecast (regression). Random decision the forest corrects the trend of over-fitting their decision tree preparation.

Decision Tree: The choice tree is a stream outline like structure in which each inner hub is located is located represents a test function (e.g. if heads or tails come up with a coin flip), each leaf node represents the class mark (decision made after all features are computed) and branches are combinations of features leading to those class labels.

K-Nearest Neighbors Algorithm**:** The k-nest approach used for classification and regression is the non-parametric approach neighbours (k-NN) algorithm. The input consists of the nearest examples of training in the feature space in both cases. whether K-NN is used for classification or regression, depending on results.

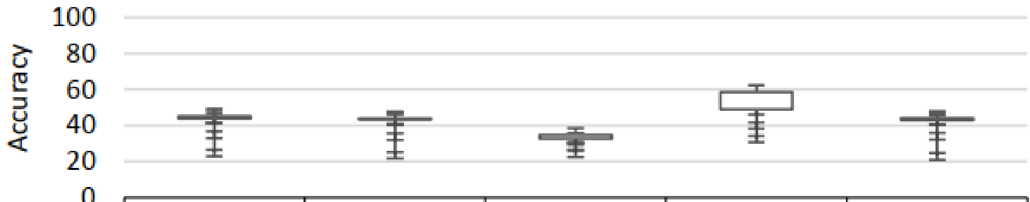
Stochastic Gradient**:** Iterative approach for optimising objective function with sufficient The properties of smoothness (e.g. differentiable or sub-differentiable) are stochastic gradients of descent. (often abbreviated as SGD). It can be viewed as a stochastic gradient descent optimization approximation, since it substitutes for the actual gradient (calculated from the data set as a whole) with the real gradient optimization estimation (calculated from a randomly selected data subset). In particular, in Big Data, this decreases the computational burden of implementation, leading to faster return iterations at a slightly lower convergence rate.

Naïve Bayes**:** Credulous Bayes classifiers in AI are a group of basic probabilistic classifiers dependent on the usage of the Bayes theorem with strong assumptions of (naïve) independence between characteristics. They are one of the most simple variants of the Bayesian network. But they should be combined with a kernel density calculation in order to achieve higher amounts of accuracy.

IV.EVALUATION OF THE CLASSIFICATION EXPERIMENT

Comparison of the precision of the classification of the multinomial Naïve Bayes, Random Forest , Decision Tree, Help Vector Machinery stochastic gradient optimisation algorithm linear kerneland (Gupta et al., 2014), and Broyden-Fletcher-Goldfarb-Shanno optimization algorithm Restricted Memory Logistic Regression (Mokhtari et al., 2015) classification methods for classification methods.

The results show that the multi-class logistic regression classification approach with the provided product-review data is the highest classification accuracy (minimum 32.43 per cent, max. 58.50 per cent) compared to the analysed classifiers. Classification of a multi-class method of logistic regression is less robust, as the average classification accuracy values are narrowly distributed relative to other approaches.



	Naïve Bayes	Random Forest	Decision Tree	Logistic Regression	Support Vector Machine
unigram	44,01	43,53	32,74	48,86	42,99
bigram	34,82	33,75	28,4	39,93	34,06
trigram	24,66	23,47	24,1	32,43	22,84
uni-/bigram	44,78	43,58	32,48	53,3	43,7
uni-/ bi-/ trigram	45,22	43,93	34,58	58,5	44,06

Fig.5: Average classification accuracy

Fig. Fig. 5 shows that Naïve Bayes, Random Forest, and Vector Machine Support's average classification accuracy values are identical (min in trigram: 33-34 per cent, max in uni / bi / tri-gram: 43-45 per cent) and Naïve Bayes achieved 1-2 per cent higher average classification accuracy compared to Random Forest and Support Vector Machine, but the gap was between the two is as follows as follows not statistical Production of analysed classification, except logistic regression, methods have greater stability and less distributed values of the average accuracy of the classification.

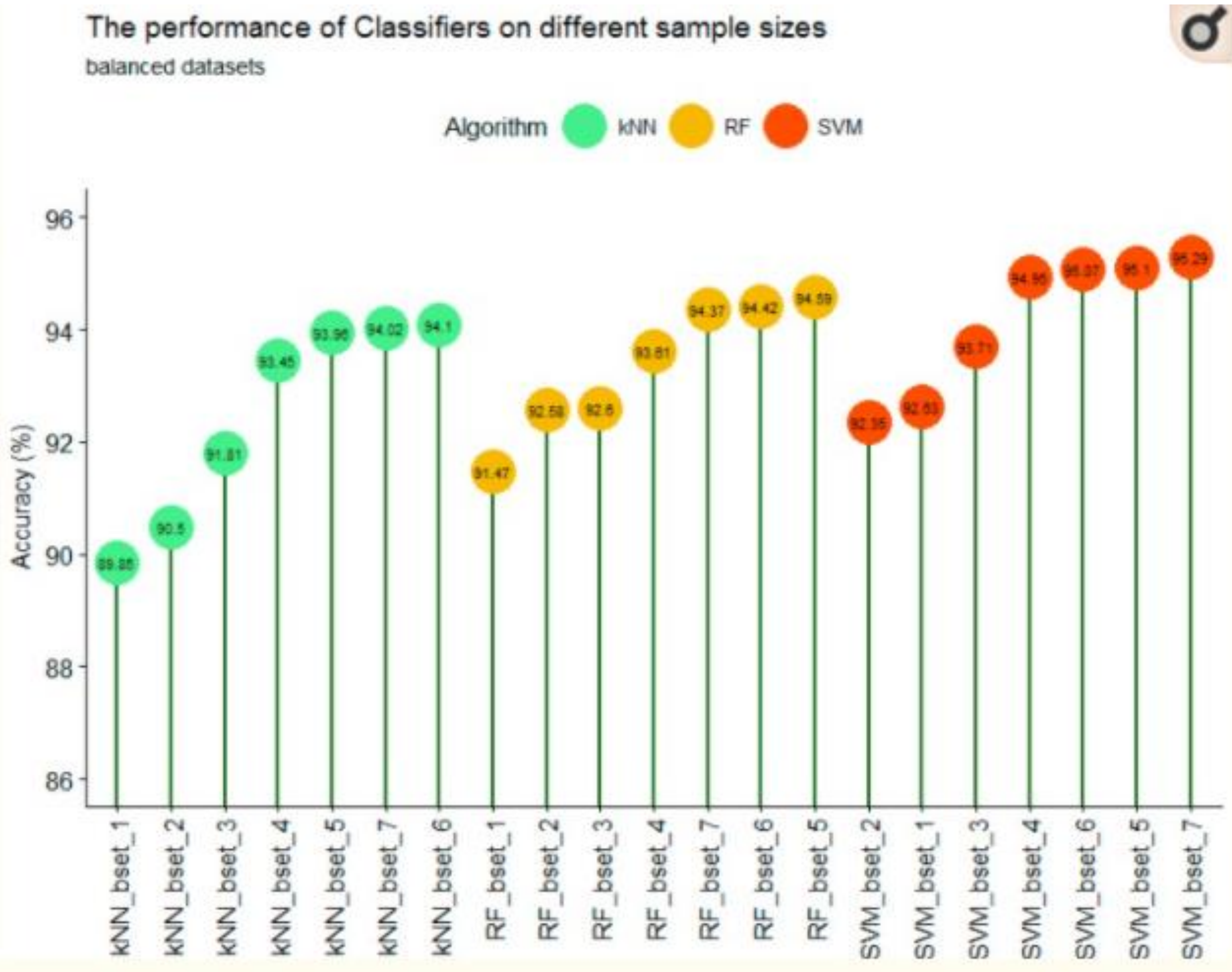


Fig.6: The performance of kNN, SVM, and RF classifiers on different balanced training sample sizes

The SVM classifier still provided the highest accuracy at 95.29 percent for balanced datasets (bset 1 to bset 7), followed by RF at 94.59 percent and kNN at 94.10 percent. The output of each classifier on different training sample sizes, however, differed only slightly (Figure 9). The training sample size had a training sample size for the kNN classifier with a limited training sample size (bset1 to bset4) strong effect on classification accuracy. The overall trend showed that the greater the size of the training sample, the higher the precision.

V.DISCUSSION

To accurately scale resources, a conventional autoscaler requires a thorough comprehension of application domain, cloud technology and operating load dynamics. This knowledge is not always accessible to the system administrator, and using an online output prediction neural network and a regression predictor is a potential solution estimate the efficiency and device metrics of post-scaling, respectively. Compared to the ideal static policy, the findings showed a resource cost reduction of 50 percent. A survey of cloud-based applications and services for self-aware, adaptive, runtime autoscaling is presented.

The collection of features is a process for automatically selecting the dataset features that most contribute to the prediction mark. The timed-dependent evolution of the data set records can be revealed by selected features, while unselected ones have a negative effect on predictive precision for ML models. A decrease in model over-fitting and decrease in training time are other benefits of feature selection. An automatic adjusting the parameters of the uphold vector machine (SVM) design acknowledgment classifier has been included. By limiting appraisals of SVM speculation blunder utilizing an inclination plummet calculation over a lot of boundaries, the parameters are tuned. This technique has many benefits, including a substantial reduction in runtime and stop holding back the validation data subsets. In comparison to cross-validation methods, the technique in this way utilizes the preparation set to enhance the boundaries.

VI.CONCLUSION

Model selection based on dynamic data evolution, a Reflexive DevMLOps method is presented to you. New models are trained offline and brought online when needed while older models are tuned offline as needed sufficient prediction accuracy can no longer be achieved. In order to adjust prediction of the accuracy of the different data sets, the online models are further tuned automatically. In the context of distributed learning, the criteria of dynamic feature selection are clarified. A contrast is made with the progress of ensemble learning to illustrate the dominance dynamic Model Selection of Static Ensemble Learning.

REFERENCES

[1] Machine learning for automated algorithm design. http://www. ml4aad.org/automl/.

[2] UNSW-NB15 Dataset dataset features and size description. https://www.unsw.adfa.edu.au/australian-centre-for-cybersecurity/cybersecurity/ADFA-NB15-Datasets/. Accessed: 2017-08-16.

[3] Using Machine Learning to Explore Neural Network Architecture google research blog. https://research.googleblog.com/2017/05/ using-machine-learning-to-explore.html. Accessed: 2017-09-25.

[4] ACETOZI, J. Docker. In Pro Java Clustering and Scalability. Springer, 2017, pp. 3–11.

[5] BAUER, E. Improving operational efficiency of applications via cloud computing. IEEE Cloud Computing 5, 1 (2018), 12–19.

[6] BOTTA, A., DE DONATO, W., PERSICO, V., AND PESCAPE´, A. Integration of cloud computing and internet of things: a survey. Future Generation Computer Systems 56 (2016), 684–700.

[7] CHANDRASHEKAR, G., AND SAHIN, F. A survey on feature selection methods. Computers & Electrical Engineering 40, 1 (2014), 16–28.

[8] CHAPELLE, O., VAPNIK, V., BOUSQUET, O., AND MUKHERJEE, S. Choosing multiple parameters for support vector machines. Machine learning 46, 1 (2002), 131–159.

[9] CHEN, T., AND BAHSOON, R. Survey and taxonomy of self-aware and self-adaptive autoscaling systems in the cloud. arXiv preprint arXiv:1609.03590 (2016).

[10] FARAHNAKIAN, F., LILJEBERG, P., AND PLOSILA, J. Energyefficient virtual machines consolidation in cloud data centers using reinforcement learning. In Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference on (2014), IEEE, pp. 500–507.