International Journal of Computer Engineering in Research Trends
Multidisciplinary, Open Access, Peer-Reviewed and fully refereed
Research Paper Volume-5, Issue-12 ,2018 Regular Edition E-ISSN: 2349-7084

Data Science: Prediction and Analysis of Data using Multiple Classifier System

Veena Hosamani[1], Dr. H S Vimala[2]

[1]*Dept. Of CSE, University Visvesvaraya College Of Engineering, Bengaluru, India*
[2] *Dept. Of CSE, University Visvesvaraya College Of Engineering, Bengaluru, India*

e-mail: veena.hosamani5@gmail.com, vimalahs2009@gmail.com

Received: 17/Dec/2018, Revised: 19/Dec/2018, Accepted: 23/Dec/2018, Published: 02/Jan/2019

Abstract:-In modern times, with the trending technology, for classification of Big Data it is ubiquitous that Deep Neural network algorithms are used. The experiment was carried out considering relatively smaller data. In this paper, we propose, a model Multiple Classifier System, in which the different classifiers are ensembled. We have ensembled different classifiers like LR, LDA, KNN, CART, NB, and SVM. To check the performance of the Multiple Classifier System, we have used Iris flower dataset. When the neural networks and the Multiple Classifier System was compared with the performance, the MCS has shown graduation increase in the results.

Keywords: Multiple classifier systems, Ensemble, Data confidence, Machine learning.

1. Introduction

In machine learning, it is very popular to use multiple classifiers instead of one single classifier. Multiple classifiers give more generalised results when compared with a single classifier.

Every model has its advantage and disadvantages of combining multiple classifiers. As an advantage, instead of single classifier algorithms power, using multiple classifiers is more accurate. By combining various classifiers, it is more reliable. For classification of the data, to train the dataset and test the trained dataset for the generalisation is carried.

The combination of multiple classifiers for instance classification this task is the following three types:

1. Ensemble 2. Voting 3. Stacking

During the training session, [1] the results are accurate, but during the testing, the possibility of error rate is high due to over-fitting of data. The classification and prediction of the data, to get more generalized and accurate results are not up to the acceptation, due to an insufficient and inefficient dataset which are currently used [2]. To improve, the performance of the trained data to be accurate while testing, the research on the techniques to enhance the generalization and precise results of the application are being conducted [3], [4].

A Bayesian network [3], was used to diagnose the liver disorder model. It had a drawback, the assumption of possibility allotment as well as the setting of the respective parameter. The parameter of the data cannot be learned with a comparatively smaller dataset, the results are mainly depending on the possibility of the distribution which is an assumption, which is close to the real-time delivery of data.
In a case study [5], for the bankruptcy predictions problem, the support vector machine performs better than the back propagation neural networks, in getting the results with accuracy and generalization, as the data for the training is smaller in size. A neural network dominates in most of the

 https://doi.org/10.22362/ijcert/2018/v5/i12/v5i1201

applications but doesn't always gives the best results. To improve the performance, a suitable learning algorithm can be considered especially where the data is smaller in size.

Author Chopra [6], used a discriminative loss function, for which the results were to learn the similarity metric from data. This experiment was conducted to recognize the face, where the data was smaller in size. An energy-based model, which used the discriminative, loss function and tried to reduce the loss, to get the similarity metric, which resulted in matching latest features.

A case [2], the study was carried out; the dataset used was smaller in size. When the size of the labelled dataset is small, the size of the labelled dataset is small, the suggestions for the unsupervised learning in neural networks learning application are considered.

To improve the generalization, the unsupervised pre-training techniques are used. The generalization and accuracy can be improved, by the proposed multiple classifier systems, where the data is given as input to all classifier.

There are many different methods of machine learning techniques, due to which sometimes it is possible, that the information of the data is ignored at every single learning algorithms.

The advantage of combine multiple classifiers, the dissimilar information that is correlating can be extracted from the results of each classifier. Resulting which gives a more accurate and generalized classification of the smallest dataset too.

However, there are some disadvantages:

1. Overfitting problem, many research papers have stated that if combine multiple classifiers to classify instances from the dataset that time, it is prone to get overfitting problem.
2. The classification algorithm has their drawbacks so, when we are combining more than one classification algorithms in the test at that time, it will risk having the multiple classifier algorithms problems in one go.

2. Related Work

A crossing point strategy and an association technique are proposed for the class set decrease. Three techniques in light of the most elevated rank, tally, and calculated relapse be projected for class set re-ranking [7]. Outfit strategies are learning calculations to develop a set of classifiers as well as after that order fresh data focuses in winning a (weighted) vote of their predictions [8]. A few new systems, for the classification of hyperspectral remote detecting pictures because of multiple classifier systems (MCS) [9]. An ebb and flow focal point of excellent research in example arrangement is the mix of a few classifier frameworks, which can be manufactured after either the equivalent or diverse models or potentially datasets building approaches [10]. A comprehensive analysis of joining classifiers is accounted for and talked about. It incorporates, together, the mix of various classifiers on the equivalent include set as well as a combination of classifiers on different capabilities [11]. Recognising flaw classes, particularly for compound shortcomings, is a testing errand in mechanical blame conclusion [12].

A multi-master engineering, uniquely suited for check frameworks, which endeavours to offer the execution points of interest of a serial methodology while holding the consistent quality of a parallel blend plot [13]. The blended HMM-KNN word acknowledgement module of a bank check preparing framework created at CENPARMI [14].

The principle blend techniques, to produce it in favour of various levels of classifier yields - dynamic level, positioned rundown of classes, and estimations [15]. The perception assumes a critical job in self-awareness. In numerous situations, for example, the Observe-Decide-Act (ODA) circles, mindfulness is heaps of the framework. Perception creates the comprehension of the framework from the status and conduct of its self and its condition [16]. A design to break down and upgrade information unwavering quality and consistency. Specifically, to present a various levelled operator based information certainty assessment framework to distinguish incorrect or unimportant fundamental flag estimations [17]. Remote monitoring what's a more, symptomatic framework that gives an all-encompassing point of view of patients and their health conditions. To talk about how the idea of self-awareness can be utilized in different parts of the framework such as data accumulation through wearable sensors, certainty evaluation of the tangible information, the learning base of the patient's health situation, and robotization of

thinking about the health situation [18]. Self-awareness is the base in favour of a large number of methods these days. for example, self-optimization also, self-adaption [19]. the instructional exercise which gives an outline of the fundamental thoughts basic Support Vector (SV) machines for work estimation[20].

3. Existing System

With the different classifiers available in the machine learning, to find the accurate generalisation and prediction, the data set has to be checked with all the relevant classifiers individual. In the traditional method, the different classifiers like naive Bayesian, support vector machine, decision tree and neural networks were individually trained and tested to generalise and predict.

The disadvantage with individual training and testing of classifiers consume a lot of time, human resources and money. And sometimes many not get accurate expected results too.

4. Proposed System

A Self-Aware Multiple Classifier System, generally a classifier groups, which are examined more than two decades. Join an arrangement of individual classifiers with the end goal to enhance the execution and dependability of the agreement in an assortment of uses, for example, advanced flag preparing and design acknowledgement. Each classifier, for example, Neural Networks, support vector machine, as well as decisions treè, accomplishes exact outcome inside a particular field as well as in a few imperatives. In this way, self-aware multiple classifier systems acquires the coveted highlights of every classifier and can manage very perplexing issues, vulnerability, high-dimensionality of information, improvement issues.

The objective of a self-aware multiple classifier systems has made predominant classifiers that beats the component. A way which can be achieved highest execution otherwise precision shows up decent variety classifier. In this way, a selections suitable classifier can be considered primary concerns. Few emphasis information combined, few consolidate which yields also, and few had all the info as well as the return to fusions.

The three significant classifications which self-aware multiple classifier system approaches is to incorporate preparing test manipulating, parallel in combination, as well as concatenations combination. Bagging as well as Boosting calculations is well known self-aware multiple classifier system techniques dependent in managing the preparing tests that make multiplier preparation informational indexes produce multiplier speculations. Every data set utilized is prepared for learning calculation. Parallel combinations autonomously train that are picked classifier as well as afterwards uses a few procedures such as the lion's share casting a ballot to join the yields. Concatenations fusion consist, chain of classifiers that they yield in single numbered is sustained in following single info.

A few classifiers that gives the positioning scores to the class given issue even estimation is likelihood even confidence. Concatenations engineering, positioning which classifiers might utilize in refining the quantity in class as well as along with these lines the more certain class which will be sent for the following classifier. Other case, positioning scores utilizes the models such a regression demonstrated in vectors positioning score as well as that yields to classifier enter models in last decisions. It is not in observable attentions positioning approaches likely because of powerlessness in many classifiers which will deliver positioning scores. Besides, the methodology for the most part favoured is space example recognitions.

A Self-aware Multiples classifiers system, generally sacrifice quick computation higher precision which runs multiples classifier. In this way, results are highly computationally expensive. The proposed methodology has advantages in self-aware multiple classifier systems, the high execution as well as exactness and besides low computational expense. This methodology points particularly at managing those classification applications which experience the ill effects of the little size of information sets.

Thus, the proposed calculation, by misusing a few classifiers, enhances the exactness and generalization of the classification for the little informational collection. At the same time, it diminishes the computational expense of Self-aware multiple classifier systems, which keeps quantity in the classifier that races in the base as well as keeping away from the run to additionally in classifier while they are redundant.

5. MULTIPLE SYSTEM CLASSIFIER SYSTEM ALGORITHMS

The Mindfulness is the capacity of the framework to screen its state, conduct, and execution to refresh at least one of its parts to accomplish its objectives [16]. In any case, observing has gotten little consideration up until now. As of late, TaheriNejad et al. [17] distributed an investigation on different components of perception and their potential job in mindfulness. This prompted additionally look into which demonstrated the advantages of these components, for example, information consistent quality, consideration various applications. Among the components, perception will 'certainty' that provides food framework proportion consistent quality after effects to a calculation. That help in the mindful framework which has been better choices given the consistent quality of its subsystems, and utilizes its assets in a more productive way and light of the current circumstance. It has exploited this idea in the proposed calculation and shows how it enhances the execution and the consistent quality of the framework, mainly on account of little information, which keeps quantity in classifier it hurries the base.

Algorithm:

Multiple Classifier System

Input: Iris Dataset

Output: Generalisation and Accuracy

1. DataSet = Iris DataSet
2. Read DataSet
3. Initialization
4. Train data
5. Multiple classifier system
6. Test data
7. For model selection
8. Split = 10
9. End for
10. Output generalization
11. Calculate accuracy of the classifier
12. End

Fig 1: Multiple Classifier Algorithms

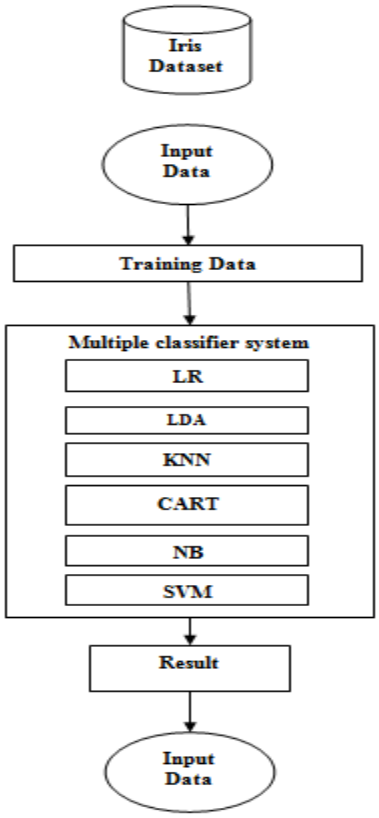

Fig : 2 Multiple Classifier system Flowchart

Figure 2 gives the detailed insight into the algorithm used, in this experiment. The flowchart explains the dataset is taken as the input. The input dataset is trained, and then it is given input to the multiple classifier systems, where each classifier gives its generalization and accuracy rate of the classifier. The results obtained with the model proposed are compared with the traditional methods.

6. Experimenters and Results

The experiment was conducted, using the IRIS flower dataset. The Iris dataset is very famous dataset used for many years. The dataset contains 3 classes of 50 instance each, where each class refers to an iris plant. One type is linearly separable from the other 2, and the latter are not linearly separable from other.

To conduct this experiment, machine learning packages are used such as pandas, sklearn, and matplotlib using the python3.6. The dataset was initially trained and test rigorously.

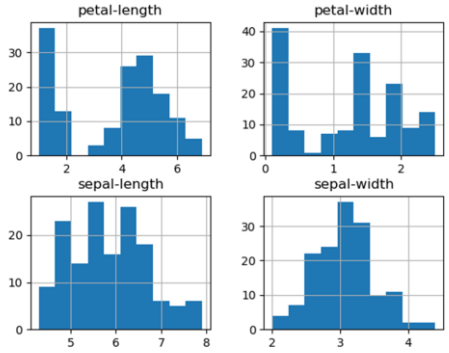

Fig. 3 Bar chart showing the classification based on the petal length, petal width, sepal length, sepal width respectively.

In fig 3, the results of the classification based on the petal lenght, petal width, sepal length, sepal width is generalised.

Table 1: Mean and std deviation of different classifier

SL.No	Classifier Name	Mean	Std Deviation
1.	LR	0.966667	0.040825
2.	LDA	0.975	0.038188
3.	KNN	0.983333	0.033333
4.	CART	0.975	0.038188
5.	NB	0.975	0.053359
6.	SVM	0.991667	0.025

The results of the multiple classifier system results concluding with the accuracy of 0.991667 which was given by the support vector machine classifier among different classifier, the table 1, presents the effects of different classifiers used in the multiple classifier systems.

Table 1 gives the results with the different classifiers obtained. In multiple classifier systems, we have combined six different classifiers to get the generalisation and the accuracy to the maximum level. For every classifier, the mean and the standard deviation has been calculated. Figure 4 shows the line graph of the different classifiers ensembled in the multiple classifier systems, the mean and the standard deviation is calculated for the graph shown.

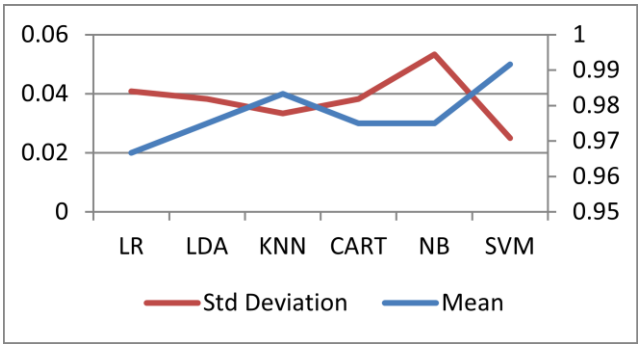

Fig. 4 line graph showing the mean and the st deviation of the multiple classifiers.

7. Conclusions

The proposed multiple classifier systems which results with the generalization and classification of the dataset with the confidence metrics. A confidence metrics is defined with the possibility of accurate classifications, which were calculated on behalf of each classifier algorithm while using the testing phase. The classifier selection outperforms, with each classifier by using the confidence metrics. The cost of training several single classifiers will be reduced, by using the multiple classifier system. For this case study, we have used logistic regression, linear discriminate analysis, K nearest neighbor, classification and Regression trees, Naïve Bayes and support vector machine classifier. Out of the entire classifiers support vector machine out performs with results of 0.991667 accuracy on the Iris dataset.

Future enhancement, the same self-aware multiple classifier system can be tried on huge amount of data i.e, Big Data.

8. Acknowledgment

I would extend my gratitude towards my parents Dr Mallappa Hosamani and Ratna for their support throughout the journey. My siblings Soumya and Sachin, for the support.

9. References

[1] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". *Nature*, 521(7553):436–444, 2015.

[2] Yoshua Bengio, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives". *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[3] Agnieszka Oni´sko, Marek J Druzdzel, and Hanna Wasyluk. "Learning bayesian network parameters from small data sets: Application of noisyor gates". *International Journal of Approximate Reasoning*, 27(2):165–182, 2001.

[4] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". *Journal of machine learning research*, 15(1):1929–1958, 2014.

[5] Kyung-Shik Shin, Taik Soo Lee, and Hyun-jung Kim. "An application of support vector machines in bankruptcy prediction model". *Expert Systems with Applications*, 28(1):127–135, 2005.

[6] Sumit Chopra, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification". In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 539–546. IEEE, 2005.

[7] Tin Kam Ho, Jonathan J. Hull, and Sargur N. Srihari. "Decision combination in multiple classifier systems". *IEEE transactions on pattern analysis and machine intelligence*, 16(1):66–75, 1994.

[8] Thomas G Dietterich et al. "Ensemble methods in machine learning". *Multiple classifier systems*, 1857:1–15, 2000.

[9] Peijun Du, Junshi Xia, Wei Zhang, Kun Tan, Yi Liu, and Sicong Liu. "Multiple classifier system for remote sensing image classification: A review. Sensors", 12(4):4764–4792, 2012.

[10] Michał Wo´zniak, Manuel Gra˜na, and Emilio Corchado. "A survey of multiple classifier systems as hybrid systems". *Information Fusion*, 16:3– 17, 2014.

[11] Robert PW Duin and David MJ Tax. "Experiments with classifier combining rules". In *International Workshop on Multiple Classifier Systems*, pages 16–29. Springer, 2000.

[12] Jinxiu Qu, Zhousuo Zhang, and Teng Gong. "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion". *Neurocomputing*, 171:837–853, 2016.

[13] Luigi P Cordella, Pasquale Foggia, Carlo Sansone, Francesco Tortorella, and Mario Vento. "A cascaded multiple expert system for verification". In *International Workshop on Multiple Classifier Systems*, pages 330–339. Springer, 2000.

[14] Didier Guillevic and Ching Y Suen. "Hmm-knn word recognition engine for bank cheque processing". In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 2, pages 1526–1529. IEEE, 1998.

[15] Ching Y Suen and Louisa Lam. "Multiple classifier combination methodologies for different output levels". In *International workshop on multiple classifier systems*, pages 52–66. Springer, 2000.

[16] Nikil Dutt, Axel Jantsch, and Santanu Sarma. "Toward smart embedded systems: A self-aware system-on-chip (soc) perspective". *ACM Trans. Embed. Comput. Syst.*, 15(2):22:1–22:27, February 2016.

[17] Nima TaheriNejad, Axel Jantsch, and David Pollreisz. "Comprehensive observation and its role in self-awareness: an emotion recognition system example". In *Proceedings of the Federated Conference on Computer Science and Information Systems*, Gdansk, Poland, 2016.

[18] Maximilian G¨otzinger, Nima Taherinejad, Amir M. Rahmani, Pasi Liljeberg, Axel Jantsch, and Hannu Tenhunen. "Enhancing the Early Warning Score System Using Data Confidence", pages 91–99. *Springer International Publishing*, Cham, 2017.

[19] Arman Anzanpour, Iman Azimi, Maximilian G¨otzinger, Amir M. Rahmani, Nima TaheriNejad, Pasi Liljeberg, Axel Jantsch, and Nikil Dutt. "Self-awareness in remote health monitoring systems using wearable electronics". In *Proceedings of Design and Test Europe Conference (DATE)*, Lausanne, Switzerland, March 2017.

[20] N. TaheriNejad, M. A. Shami, and S. M. P. D. "Self-aware sensing and attention-based data collection in multi-processor system-on-chips". In 2017 15th *IEEE International New Circuits and Systems Conference (NEWCAS)*.

Appendix

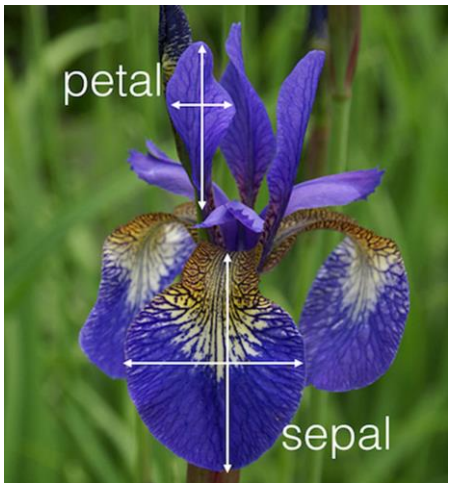

Fig. 5 Iris Flower sample picture

The Iris flower using the conducting the experiment which the features: petal length, petal width, sepal length, sepal width. Figure 5, shows a sample picture the Iris flower.The table 2, gives the sample data collected about the features exctracted from the flower in tabluar form. The use of this data set in cluster analysis however is not common, since the data set only contains two clusters with rather obvious separation. One of the clusters contains Iris setosa,while the other cluster contains both Iris virginica and Iris versicolor and is not separable without the species information Fisher used.

Table 2: Iris dataset

Petal length	Petal width	Sepal length	Sepal width	species
1.4	0.2	5.1	3.5	Setosa
1.4	0.2	4.9	3	Setosa
6.4	3.2	4.5	1.5	Versicolour
6.9	3.1	4.9	1.5	Versicolour
7.1	3	5.9	2.1	Verginica
4.9	2.5	4.5	1.7	Verginica
7.3	2.9	5.6	1.8	Verginica

Authors Profile

1. **Veena Hosamani**, pursuing ME at University Visvesvaraya College Of Engineering, K R circle, Bangalore in the department of computer science and engineering. Received BE from Tontadarya College Of Engineering, Gadag. From 2013 to 2016, was lecturer at Vasavi Jnana peetha polytechnic college, Vijaynagar, Bangalore.

2. **Dr. H. S Vimala** Associate professor, department of computer Science, University Vivesvaraya College of engineering, BUB, Bangalore. Bachelor of engineering in computer science and engineering from Mysore University. Master degree in computer science from bits, Pilani. Phd in Computer Science from Himalayan University, Itanagar.Thirty years of teaching experience in the University Of Visvesvaraya College Of Engineering, Bangalore.

