International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7, Issue-6, March 2019

Iterative Gradient Ascent Expected Maximization Clustering for Weather Forecasting

Pooja S. B, R.V Siva Balan

Abstract : Weather forecasting is a significant process to be solved as it discovers future atmosphere for a given location. Few clustering techniques were intended in order to group similar weather data for predicting weather conditions. However, the clustering accuracy of the existing technique was not effectual when taking big dataset as input. In order to solve this limitation, a Iterative Gradient Ascent Expected Maximization Clustering (IGAEM) Model is proposed. The IGAEM Model predicts the future weather conditions with higher clustering accuracy and minimal time. In IGAEM Model; after selecting the relevant features, IGAEM Model applied Iterative Gradient Ascent Expected Maximization Clustering (IGAEMC) to accurately group the weather data into diverse clusters with lower amount of time utilization. Thus, IGAEM Model significantly increases the performance of weather forecasting as compared to existing works. The IGAEM Model conducts experimental evaluation using factors such as clustering accuracy clustering time and false positive rate with respect to a number of features and weather data from Atlantic hurricane database. The experimental results depict that the IGAEM Technique is able to enhance the clustering accuracy and reduce the clustering time of weather prediction when compared to state-of-the-art-works.

Keywords: Clustering, Features, Gradient Ascent Method, Iterative Gradient Ascent Expected Maximization Clustering, Weather Data

I. INTRODUCTION

Weather forecasting attains greater significance in day-to-day applications. As weather forecasting process discover conditions of atmosphere for particular location and time. Many clustering techniques are introduced in existing works for enhancing the prediction performance of weather data. But, the clustering process of the existing technique was not adequate when taking big dataset as input. Therefore, IGAEM Model is designed. The main objective of IGAEM Model is to predict the cyclone through the clustering weather data with high accuracy and minimal time.

A conjunct space cluster-based adaptive neuro-fuzzy inference system (CF-ANFIS) was introduced in [1] for the seasonal forecasting of tropical cyclones with higher accuracy. But, the false positive rate of clustering was not solved effectively. A K-means algorithm was applied in [2] to predict the Atlantic Tropical Cyclones. However, clustering time was very higher.

A novel technique was designed in [3] to enhance the clustering performance for weather predictions. But, the clustering accuracy was not improved. A new technique was introduced in [4] for examining temporal growth of uncertainty in the ensemble of weather forecasts from perturbed conditions.

The empirical formula of threshold distance was evaluated in [5] to classify the cyclone–cyclone interactions with or without mid-tropical depressions (TD). Cloud Computing was employed in [6] to solve the Big Data demands to enable transformation. Big geospatial data failed to include challenges during the lifecycle of data storage, access, manage, analysis, mining, and modeling.

The dynamic self-organized neural network inspired by the immune algorithm was applied in [7] for the prediction of weather data signals. However, the computational complexity of this algorithm was higher. Gaussian process regression (GPR) model was designed in [8] to obtain higher clustering accuracy for weather prediction. A novel method was presented in [9] that determine similarities among data for the seasonal forecast with the application of hierarchical clustering.

Fuzzy C-means clustering was developed in [10] for clustering numerical weather forecasts and enhancing statistical prediction performance. A survey of various techniques designed for weather prediction was analyzed in [11].

II. II. RELATED WORKS

Incremental K-Means Clustering was intended in [12] for forecasts weather events with minimal time. The true positive rate of weather data clustering is lower. Medium-range Weather Forecasting (ECMWF) model was presented in [13] to get higher weather prediction performance for the forecast of a cloudburst. The computational complexity involved during clustering was very higher.

Self-organizing map (SOM) based cluster analysis technique was introduced in [14] for accurate typhoon rainfall forecasts. The false positive rate of clustering was not solved. A survey of diverse data mining techniques designed for weather forecast analysis in [15].

Revised Manuscript Received on 30 March 2019.
***** Correspondence Author
Pooja S.B*, Department of computer science, Noorul Islam Centre for Higher Education, Kumaracoil, India.
R.V Siva Balan, , Department of MCA, Noorul Islam Centre for Higher Education, Kumaracoil, India.

A K-Means and Fuzzy C-Means (FCM) was designed in [16] for accomplishing weather forecasting with minimal false positive rate. The clustering performance was not sufficient for effectual future weather condition prediction. A Backpropagation neural network (BPNN) was presented in [17] to predict weather changes. A data with meteorological forecasting problems were addressed by designing set of informed default encoding choices in [18]. A non-parametric Spearman correlation-based method was introduced in [19] to rank and select relevant features for exact prediction. The feature selection performance was poor. Subspace clustering algorithm was used in [20] to choose imperative climatic predictor variables for obtaining improved prediction performance. The accuracy of feature selection was lower. The Sliding Window Algorithm was designed in [21] to predict a day's weather conditions. The weather forecasting performance was not adequate where it does not consider feature selection process. In order to resolve the above said existing issues, IGAEM Model is proposed. The main contribution of IGAEM Model is as follows,

- To increase the performance of weather forecasting with higher accuracy and minimal time as compared to state-of-the-art works, IGAEM Model is introduced. The IGAEM Model is designed with the application of Iterative Gradient Ascent Expected Maximization Clustering (IGAEMC).
- To improve the performance of feature selection with higher accuracy during weather prediction as compared to existing works, IGAEM Model is developed.
- To achieve enhanced clustering accuracy to execute weather forecasting with minimal time complexity as compared to traditional works, IGAEMC is designed in IGAEM Model. On the contrary to existing works, gradient ascent is applied in IGAEMC to maximize the likelihood function and thereby effectively clustering weather data for accurate future weather prediction.

The rest of the paper is planned as follows; Section 2 portrays the related works. In Section 3, the proposed IGAEM Model is explained with the aid of the architecture diagram. In Section 4, Experimental settings are presented and the analysis of results is discussed in Section 5. Section 6 depicts the conclusion of the paper.

III. III. ITERATIVE GRADIENT ASCENT EXPECTED MAXIMIZATION CLUSTERING MODEL

Weather forecasting is process of predicting weather condition in the given future time. A weather forecast presents significant information about future weather. There are diverse techniques designed in existing work using clustering for performing weather forecasting. However, the clustering process of the conventional technique was not sufficient. In order to enhance the weather prediction performance through clustering, Iterative Gradient Ascent Expected Maximization Clustering (IGAEM) Model is developed. Clustering is a process of grouping similar data together. An Iterative Gradient Ascent Expected Maximization Clustering (IGAEMC) is designed in IGAEM Model in order to increase the performances of clustering with minimal time complexity for effectively performing weather forecasting processes.

The IGAEMC is proposed by combining the expectation maximization clustering algorithm and Gradient Ascent method. On the contrary to existing works, Gradient Ascent method is exploited in IGAEMC with aim of increasing the clustering accuracy of data for future weather prediction. The Gradient Ascent method applied in IGAEMC discovers mean and variance between cluster center and weather data to effectively determine the maximum log-likelihood on the contrary to existing expectation maximization clustering algorithm. This helps for IGAEMC to accurately cluster the weather data into different clusters with a minimal amount of time for future event prediction.

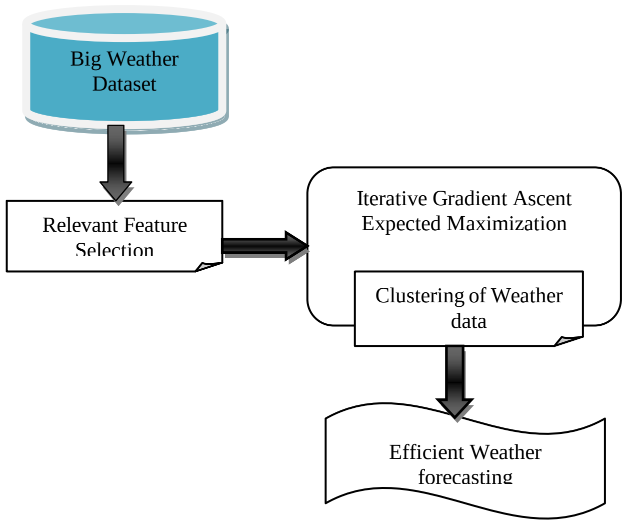

Figure 1 Architecture Diagram of IGAEM Model for Weather Forecasting

The IGAEMC computes probabilities of cluster memberships depending on one or more probability distributions. The goal of the IGAEMC is to maximize the overall probability or likelihood of the data with help of Gradient Ascent method and thereby obtains final clusters. An IGAEMC is an iterative method which alternates among two steps, expectation '(E)' and maximization '(M)'. The IGAEMC designed in IGAEM Model group the collection of weather information in the input dataset into a diverse number of clusters for effective prediction of future weather conditions. The clustering is performed by determining maximization expected likelihood probability between weather data. In statistics, IGAEMC is an iterative technique which measures likelihood probability between weather data with the selected features. After that, IGAEMC applied gradient ascent that maximizes expected log likelihood between data and cluster center to attain higher clustering accuracy for weather forecasting as compared to existing works. The process of IGAEMC is repetitive until all weather data are grouped. The process involved in IGAEMC for weather data clustering is depicted in below Figure 2.

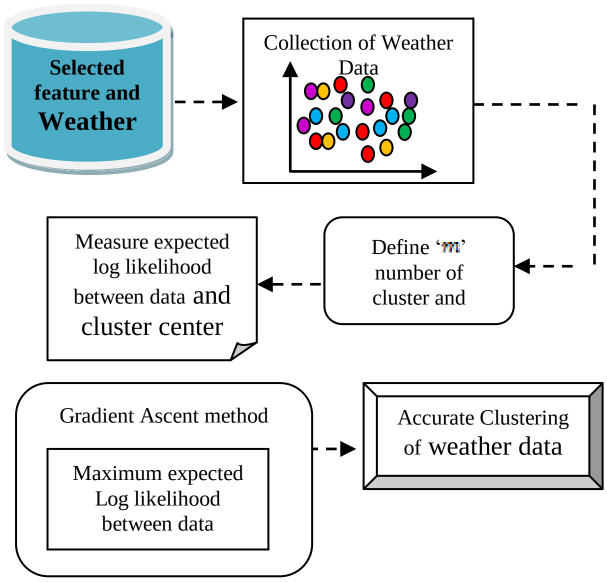

Figure 2 Processes of IGAEMC for Weather Data Clustering

Figure 2 presents the flow process of IGAEMC to enhance weather prediction performance through clustering. As depicted in figure 2, IGAEMC takes big weather dataset (i.e. Atlantic hurricane database) and selected features as input. Then, IGAEMC algorithm randomly initializes 'm' number of clusters and their centers. Next, IGAEMC performs expectation and maximization process with aim of grouping the weather data with lower time complexity. During Expectation process, IGAEMC measures likelihood between the cluster center and weather data. Here, Likelihood represents that probability of weather data belongs to a particular cluster. During the maximization process, IGAEMC computes maximum expected log likelihood between the cluster center and weather data with application of gradient ascent method. This supports for IGAEMC to effectively find the higher probability of weather data belongs to a particular cluster. Thus, IGAEMC significantly groups the similar weather data into a particular cluster with higher accuracy and minimal time.

Let us consider the number of weather data in the input dataset represented as '$WD_1, WD_2, \ldots, WD_n$'. The likelihood function of weather data points is estimated which represents the data belonging to a particular cluster or not. In expectation step, IGAEMC estimates log likelihood probability for each weather data in big dataset using below mathematical formulation,

$$E\,[\log(P\,(C_m\,|WD_i))] = Log\left(\prod_{i=1}^{n} \frac{e^{-(WD_i-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\right) \quad (1)$$

From equation (7), '$E\,[\log(P\,(C_m\,|WD_i))]$' refers to the expected likelihood probability of weather data 'WD_i'. Here, 'C_m' represents a cluster center and 'μ' denotes a mean value of cluster center whereas 'σ' indicates variance among cluster center and weather data. With measured expected likelihood probability, IGAEMC identifies the weather data is belongs to a cluster member. After that, the maximization process is carried out in IGAEMC to identify maximum expected log likelihood value by using Gradient Ascent method. Gradient ascent method is a process that discovers maximum expected log likelihood probability to improve clustering performance for effective weather forecasting. With determined maximum expected log likelihood probability, IGAEMC discovers the higher probability for each weather data belongs to a particular cluster. The Gradient ascent method in IGAEMC employed below expression to evaluate maximum log likelihood probability,

$$\gamma = arg\ max\ \ E\,[\log(P\,(C_m\,|WD_i))] \quad (2)$$

$$\gamma = arg\ max\ \left\{Log\left(\prod_{i=1}^{n} \frac{e^{-(WD_i-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\right)\right\} \quad (3)$$

From equation (8) and (9), 'γ' represents maximum log-likelihood estimation of weather data (WD_i) which maximize the log likelihood value measured from the expectation step. The '$arg\ max$' function assists for IGAEMC to determine a maximum likelihood probability for each weather data in input big dataset. This supports for IGAEMC to accurately group more similar weather data to a particular cluster with lower time. This process is cyclic until the entire data are clustered. The algorithmic processes of IGAEMC are presented in below.

[1] // Iterative Gradient Ascent Expected Maximization Clustering Algorithm

[2] Input : big dataset '(DS)' with number of weather data '$WD_i, WD_2, \ldots, WD_n$'

[3] **Output:** Improved clustering accuracy for weather prediction

[4] Step 1: Begin

[5] **Step 2:** Define 'm' cluster and center 'C_m'

[6] **Step 3:** **For** each data '$WD_i \in DS$'

[7] **Step 4:** **For** each cluster center 'C_m' **do**

[8] **Step 5:** Compute '$E\,[\log(P\,(C_m\,|WD_i))]$' using (7)

[9] **Step 6:** Apply gradient ascent to find maximum likelihood probability using (9)

[10] **Step 7:** Go to step 3 until all weather data are grouped

[11] **Step 8:** **End for**

[12] **Step 9:** **End for**

[13] Step 10: End

Algorithm 1 Iterative Gradient Ascent Expected Maximization Clustering Algorithm

Algorithm 2 explains the step by step processes of IGAEMC to obtain enhanced clustering performance for weather forecasting. As demonstrated in the above algorithm, IGAEMC initially assigns a number of clusters and centers randomly. Subsequently, IGAEMC estimates an expected log likelihood probability for each weather data. After that, IGAEMC utilized gradient ascent that computes maximum expected log likelihood probability to effectively group the more similar weather data in similar cluster with minimal amount of time. This process of IGAEMC continual until the entire weather data is clustered. Thus, IGAEM Model significantly enhances the data clustering accuracy for effective weather prediction when compared to state-of-the-art works.

IV. EXPERIMENTAL SETTINGS

In order to measure the proposed performance, IGAEM Model is implemented in Java Language using big weather dataset (i.e. Atlantic hurricane database) [22]. The Atlantic hurricane database contains a tropical cyclone historical database. Further, this dataset includes of Atlantic six-hourly information about the location, maximum winds, central pressure, and size of every known tropical cyclones and subtropical cyclones. In order to enhance the clustering accuracy of future weather prediction with lower time using this Atlantic hurricane database as compared to existing works, IGAEM Model is designed.

The Atlantic hurricane database comprises 75242 instances with 22 attributes such as ID, Name, Date, Time, Event, Status, Latitude, Longitude, Maximum Wind, Minimum Pressure, Low Wind NE, Low Wind SE, Low Wind SW, Low Wind NW, Moderate Wind NE, Moderate Wind SE, Moderate Wind SW, Moderate Wind NW, High Wind NE, High Wind SE, High Wind SW, High Wind NW. For conducting experimental work, IGAEM Model considers 1000-10000 weather data from Atlantic hurricane database. The performance of IGAEM Model is estimated in terms of feature selection accuracy, clustering accuracy, clustering time and false positive rate. The experimental result of IGAEM Model is compared with existing two methods namely, CF-ANFIS [1] and K-means clustering [2].

V. RESULT AND DISCUSSIONS

In this section, the performance result analysis of IGAEM Model is discussed. The performance of IGAEM Model is compared against with CF-ANFIS [1] and K-means clustering [2] respectively. The effectiveness of IGAEM Model is evaluated along with the following metrics with the help of tables and graphs.

A. Performance Result of Clustering Accuracy

The Clustering Accuracy (CS) is determined as the ratio of number of weather data that are correctly clustered to the total number of data taken as input. The clustering accuracy is estimated in terms of percentage (%) and mathematically calculated as,

$$CA = \frac{No.of\ data\ correclty\ clustered}{Total\ No.of\ data} * 100 \quad (4)$$

(11)

From equation (11), the clustering accuracy of weather forecasting is estimated with respect to a different number of weather data. When the clustering accuracy is higher, the model is said to be more efficient.

Sample calculation:

- **CF-ANFIS:** number of data correctly clustered is 826 and the total number of data is 1000. Then clustering accuracy is obtained as follows,

$$CA = \frac{826}{1000} * 100 = 83\ \% \quad (5)$$

- **K-means Clustering:** number of data accurately grouped is 815 and the total number of data is 1000. Then clustering accuracy is estimated as follows,

$$CA = \frac{815}{1000} * 100 = 82\ \% \quad (6)$$

- **IGAEM:** number of data exactly grouped is 888 and the total number of data is 1000. Then clustering accuracy is determined as follows,

$$FSA = \frac{888}{1000} * 100 = 89\ \% \quad (7)$$

Table 1 Tabulation Result of Clustering Accuracy

Number of Data	Clustering Accuracy (%)		
	CF-ANFIS	K-means Clustering	IGAEM
1000	83	82	89
2000	82	79	90
3000	86	85	91
4000	85	84	90
5000	89	84	94
6000	87	85	93
7000	86	83	89
8000	85	82	88
9000	84	83	92
10000	82	80	90

For measuring clustering accuracy during processes of weather prediction, IGAEM Model is executed in Java Language with dissimilar number of data in the range of 1000-10000. The clustering accuracy results of -IGAEM Model are compared with traditional two methods namely CF-ANFIS [1] and K-means clustering [2] to analyze the performance of proposed technique. When conducting an experimental evaluation using 5000 numbers of data from Atlantic hurricane database, IGAEM Model obtains 94 % clustering accuracy whereas state-of-the-art works CF-ANFIS [1] and K-means clustering [2] acquires 89 % and 84 % respectively. From that, it is significant that the clustering accuracy using proposed IGAEM Model is higher for future cyclone prediction than the existing methods [1], [2]. The performance result analysis of clustering accuracy is presented in below. Figure 3 demonstrates the impacts of clustering accuracy using three methods namely CF-ANFIS [1] and K-means clustering [2] and IGAEM Model. As exposed in Figure 3, the proposed PBH-SRKEC Technique attains higher clustering accuracy to discover future weather conditions effectively when compared to existing CF-ANFIS [1] and K-means clustering [2].

This is because of application of IGAEMC on the contrary to conventional works where it employed gradient ascent method to find maximum expected log likelihood probability.

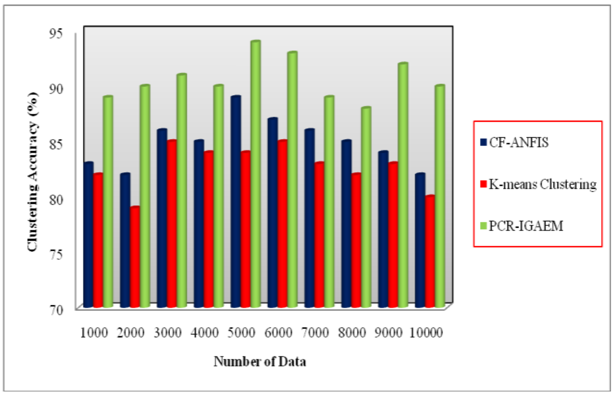

Figure 3 Measurements of Clustering Accuracy Vs Different Number of Data

With help of determined maximum expected log likelihood probability result, IGAEM Model accurately clusters the more similar weather data together in a particular cluster. This assists for IGAEM Model to increase ratio of number of weather data that are correctly clustered as compared to existing works. Therefore, IGAEM Model enhances the clustering accuracy by 7 % and 10 % as compared to CF-ANFIS [1] and K-means clustering [2] respectively.

B. Performance Result of Clustering Time

Clustering Time '(CT)' determines the amount of time taken for grouping weather data in a big dataset. The clustering time is estimated in terms of milliseconds (ms) which obtained using below mathematical representation.

$$CT = n * T\ (CSWD) \tag{8}$$

From equation (12), the amount of time needed for weather data clustering is estimated. Here, 'n' point outs the number of data in big weather dataset whereas '$T\ (CSWD)$' represents the time employed for single weather data clustering. When the clustering time is lower, the model is said to be more effective.

Sample calculation:

- **CF-ANFIS:** time taken for clustering single weather data is 0.031 and the total number of data is 1000. Then clustering time is estimated as follows,
  $$CT = 1000 * 0.031 = 31\ ms \tag{9}$$
- **K-means Clustering:** time required for clustering single weather data is 0.034 and the total number of data is 1000. Then clustering time is calculated as follows,
  $$CT = 1000 * 0.034 = 34\ ms \tag{10}$$
- **IGAEM:** time used for clustering single weather data is 0.027 and the total number of data is 1000. Then clustering time is obtained as follows,
  $$CT = 1000 * 0.027 = 27\ ms \tag{11}$$

Table 2 Tabulation Result of Clustering Time

Number of Data	Clustering Time (ms)		
	CF-ANFIS	K-means Clustering	IGAEM
1000	31	34	27
2000	42	48	34
3000	54	63	42
4000	68	80	52
5000	75	95	55
6000	72	90	49
7000	77	91	50
8000	80	96	52
9000	86	99	59
10000	90	100	75

To compute clustering time involved during process of weather data clustering, IGAEM Model is implemented in Java Language with a diverse number of data in the range of 1000-10000. The clustering time using IGAEM Model is compared with existing two methods namely CF-ANFIS [1] and K-means clustering [2]. When carried outing an experimental process using 6000 numbers of data from Atlantic hurricane database, IGAEM Model takes 49 ms clustering time whereas state-of-the-art works CF-ANFIS [1] and K-means clustering [2] utilizes 72 ms and 90 ms respectively. Thus, it is expressive that the clustering time using proposed IGAEM Model is lower for finding future cyclone effectively when compared to existing methods [1], [2]. The experimental result analysis of clustering time is depicted in below.

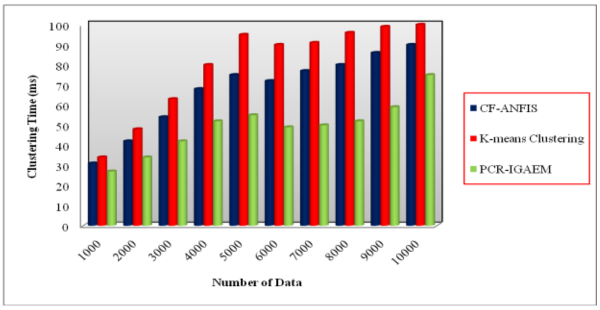

Figure 4 Measurements of Clustering Time Vs Different Number of Data

Figure 4 portrays the impacts of clustering time using three methods namely CF-ANFIS [1] and K-means clustering [2] and IGAEM Model. As revealed in Figure 4, the proposed IGAEM Model takes lower amount of clustering time for improving weather prediction performance when compared to existing CF-ANFIS [1] and K-means clustering [2]. This is owing to application of IGAEMC on the contrary to state-of-the-art works. During clustering processes of weather data, IGAEMC applied gradient ascent method to discover maximum expected log likelihood probability. This helps for IGAEM Model precisely groups the similar types of weather data in a particular cluster with minimal time. This supports for IGAEM Model take lower amount of time for clustering big weather data in input data set when compared to traditional works. Hence, IGAEM Model minimizes the clustering time by 25 % and 36 % as compared to CF-ANFIS [1] and K-means clustering [2] respectively.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

C. Experimental Result of False Positive Rate

The False Positive Rate '(FPR)' is evaluated as the ratio of number of weather data that are incorrectly clustered to the total number of data taken as input. The false positive rate is determined in terms of percentage (%) and expressed as,

$$FPR = \frac{No.of\ data\ incorrecl ty\ clustered}{Total\ No.of\ data} * 100 \quad (12)$$

From equation (12), the false positive rate of weather data clustering is calculated with respect to diverse number of weather data. When the false positive rate is minimal, the model is said to be more effective.

Sample calculation:

- **CF-ANFIS:** number of data wrongly clustered is 174 and the total number of data is 1000. Then false positive rate is formulated as follows,

$$CA = \frac{174}{1000} * 100 = 17\ \% \quad (13)$$

- **K-means Clustering:** number of data inaccurately clustered is 185 and the total number of data is 1000. Then false positive rate is obtained as follows,

$$CA = \frac{185}{1000} * 100 = 19\ \% \quad (14)$$

- **IGAEM:** number of data inexactly grouped is 112 and the total number of data is 1000. Then false positive rate is estimated as follows,

$$FSA = \frac{112}{1000} * 100 = 11\ \% \quad (14)$$

Table 3 Tabulation Result of False Positive Rate

Number of Data	False Positive Rate (%)		
	CF-ANFIS	K-means Clustering	IGAEM
1000	17	19	11
2000	18	21	10
3000	15	15	9
4000	15	16	10
5000	11	16	6
6000	13	15	7
7000	14	17	11
8000	15	18	12
9000	16	18	8
10000	18	20	10

In order to measure false positive rate of data clustering for effective weather prediction, IGAEM Model is implemented in Java Language with a different number of data in the range of 1000-10000. The false positive rate results using IGAEM Model is compared with existing two methods namely CF-ANFIS [1] and K-means clustering [2]. When accomplishing an experimental process using 7000 numbers of data from Atlantic hurricane database, IGAEM Model gets 11 % false positive rate whereas state-of-the-art works CF-ANFIS [1] and K-means clustering [2] obtains 14 % and 17 % respectively. From that, it is descriptive that the false positive rate using proposed IGAEM Model is minimal for accurately carried outing cyclone forecasting when compared to existing methods [1], [2]. The result analysis of false positive rate is demonstrated in below.

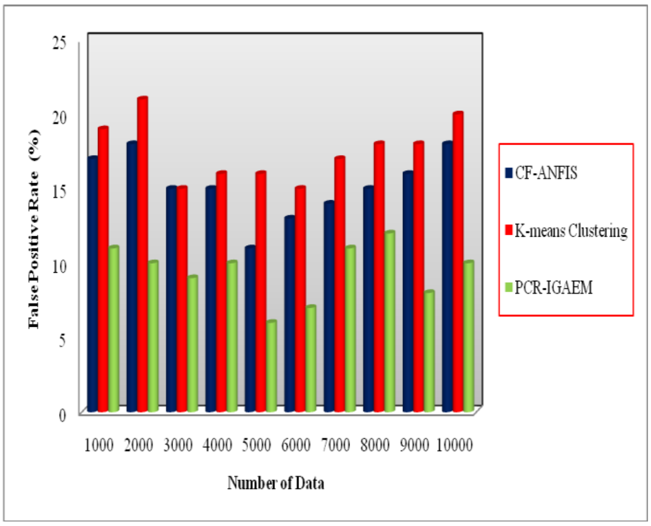

Figure 5 Measurements of False Positive Rate Vs Different Number of Data

Figure 5 describes the impacts of false positive rate using three methods namely CF-ANFIS [1] and K-means clustering [2] and IGAEM Model. As shown in Figure 5, the proposed -IGAEM Model obtains minimal false positive rate for weather prediction when compared to existing CF-ANFIS [1] and K-means clustering [2]. This is due to usage of IGAEMC in proposed IGAEM Model. The IGAEMC improves clustering performance with aid of gradient ascent method. Thus, IGAEM Model efficiently clusters the similar weather data together without any error. This support for IGAEM Model to minimizes ratio of number of weather data that are incorrectly clustered as compared to existing works. Therefore, IGAEM Model decreases the false positive rate by 38 % and 46 % as compared to CF-ANFIS [1] and K-means clustering [2] respectively.

VI. CONCLUSION

An effective IGAEM Model is intended with goal of acquiring higher clustering accuracy and minimal time for weather prediction. The goal of IGAEM Model is attained with aid of Iterative Gradient Ascent Expected Maximization Clustering (IGAEMC). The algorithmic processes of IGAEMC, IGAEM Model achieves higher accuracy for weather data clustering when compared to traditional works. Thus, IGAEM Model effectively performs weather prediction process when compared to state-of-the-art works. The performance of IGAEM Model is tested with the metrics such as, clustering accuracy and clustering time and false positive rate using Atlantic hurricane database and compared with existing works. With the experimental conducted for IGAEM Model, it is expressive that clustering accuracy is higher for grouping weather data when compared to conventional works.

The experimental results illustrate that IGAEM Model provides better performance with the reduction of clustering time for effectual weather forecasting as compared to existing works.

REFERENCES

1. Trong Hai Duong, Phi Hung Do, Sy Dzung Nguyen, Minh Hien Hoang, "ENSO-based tropical cyclone forecasting using CF-ANFIS", Vietnam Journal of Computer Science, Springer, Volume 3, Issue 2, Pages 81–91, May 2016
2. Irenea L. Corporal-Lodangcoa, Michael B. Richmana, Lance M. Lesliea, Peter J. Lamb, "Cluster Analysis of North Atlantic Tropical Cyclones", Procedia Computer Science, Elsevier, Volume 36, Pages 293 – 300, 2014
3. Alexander Kumpf, Bianca Tost, Marlene Baumgart, Michael Riemer, R¨udiger Westermann, and Marc Rautenhaus, "Visualizing Confidence in Cluster-based Ensemble Weather Forecast Analyses", IEEE Transactions on Visualization and Computer Graphics, Volume 24, Issue 1, Pages 109-119, January 2018
4. Florian Ferstl, Mathias Kanzler, Marc Rautenhaus, and Rudiger Westermann, "Time-hierarchical Clustering and Visualization of Weather Forecast Ensembles", IEEE Transactions on Visualization and Computer Graphics, Volume 23, Issue 1, Pages 831 – 840, January 2017
5. Yuei-An Liou, Ji-Chyun Liu, Meng-Xi Wu, Yueh-Jyun Lee, Chi-Han Cheng, Ching-Ping Kuei and Rong-Moo Hong, "Generalized Empirical Formulas of Threshold Distance to Characterize Cyclone–Cyclone Interactions", IEEE Transactions on Geoscience and Remote Sensing, Volume 54, Issue 6, Pages 3502 – 3512, June 2016
6. Chaowei Yang, Manzhu Yu, Fei Hu, Yongyao Jiang and Yun Li, "Utilizing Cloud Computing to address big geospatial data challenges", Computers, Environment and Urban Systems, Elsevier, Volume 61, 120–128, 2017
7. Abir Jaafar Hussain, Panos Liatsis, Mohammed Khalaf, Hissam Tawfik, Haya Al-Asker, "A dynamic neural network architecture with immunology inspired optimization for weather data forecasting", Big Data Research, Elsevier, Pages 1-38, May 2018
8. Gunasekaran Manogaran, Daphne Lopez, "A Gaussian process based big data processing framework in cluster computing environment", Cluster Computing, Springer, Pages 1–16, June 2017
9. Xing Yuan and Eric F. Wood, "On the clustering of climate models in ensemble seasonal forecasting", Geophysical Research Letters, Volume 39, Pages 1- 7, 2012
10. Ashkan Zarnani, Petr Musileka, and Jana Heckenbergerova, "Clustering numerical weather forecasts to obtain statistical prediction intervals", Meteorological Applications, Volume 21, Pages 605–618, 2014
11. Sara khan, Mohd Muqeem, Nashra Javed, "A Critical Review of Data Mining Techniques in Weather Forecasting", International Journal of Advanced Research in Computer and Communication Engineering, Volume 5, Issue 4, Pages 1091-1094, April 2016
12. Sanjay Chakraborty, N. K. Nagwani, Lopamudra Dey, "Weather Forecasting using Incremental K-means Clustering", CiiT International Journal of Biometrics and Bioinformatics, Volume 4, Issue 5, Pages 1-6, June 2012
13. Kavita Pabreja, "Clustering technique to interpret Numerical Weather Prediction output products for forecast of Cloudburst", International Journal of Computer Science and Information Technologies, Volume 3, Issue 1, Pages 2996 – 2999, 2012
14. Ming-Chang Wu, Jing-Shan Hong, Ling-Feng Hsiao, Li-Huan Hsu and Chieh-Ju Wang, "Effective Use of Ensemble Numerical Weather Predictions in Taiwan by Means of a SOM-Based Cluster Analysis Technique", Water, Volume 9, Pages 1-17, 2017
15. Nikita Gupta, Rashmi Narayanan, Anagha Chaudhari, "Implementation and Analysis of Data Mining Techniques for Weather Prediction", International Journal of Innovations & Advancement in Computer Science, Volume 6, Issue 11, Pages 101-104, November 2017
16. Candra Dewi, "Performance of Clustering on ANFIS for Weather Forecasting", Communication & Information Technology Journal, Volume 12, Issue 1, Pages 43–49, 2018
17. Prasanta Rao Jillella S.S, P Bhanu Sai Kiran, P. Nithin Chowdary, B. Rohit Kumar Reddy, Vishnu Murthy, "Weather Forecasting Using Artificial Neural Networks and Data Mining Techniques", International Journal Of Innovative Technology And Research, Volume 3, Issue.6, Pages 2534 – 2539, 2015
18. P. Samuel Quinan and Miriah Meyer, "Visually Comparing Weather Features in Forecasts", IEEE Transactions on Visualization and Computer Graphics, Volume 22, Issue 1, Pages 389-398, January 2016
19. Eduardo Soares, Pyramo Costa Jr, Bruno Costa and Daniel Leite, "Ensemble of evolving data clouds and fuzzy models for weather time series prediction", Applied Soft Computing, Volume 64, Pages 445-453, March 2018
20. Moumita Saha, Arun Chakraborty, and Pabitra Mitra, "Predictor-Year Subspace Clustering Based Ensemble Prediction of Indian Summer Monsoon", Hindawi Publishing Corporation, Advances in Meteorology, Volume 2016, Article ID 9031625, Pages 1-12, 2016
21. Piyush Kapoor and Sarabjeet Singh Bedi, "Weather Forecasting Using Sliding Window Algorithm", Hindawi Publishing Corporation, ISRN Signal Processing, Volume 2013, Article ID 156540, Pages 1-5, 2013
22. Atlantic hurricane database: https://www.kaggle.com/noaa/hurricane-database

AUTHORS PROFILE

Pooja S.B After completing her post graduation degree she joined as a research scholar in Noorul Islam centre for higher education. Her area of interest is cloud computing, big data, data mining, and remote sensing.,

Dr .R.V Siva Bala he is working as a associate professor in MCA department. After completing his doctorate degree. He guided more than 13 students in different fields. He has nearly 6 paper publications in SCI and Scopus indexed journals. His interested fields are image processing, software engineering,, cloud computing. Data mining, big data, network security and, remote sensing,