Advances in Computer Science and Information Technology (ACSIT)
Print ISSN: 2393-9907; Online ISSN: 2393-9915; Volume 2, Number 1; January-March, 2015 pp. 53-55

http://www.krishisanskriti.org/acsit.html

Towards a Hybrid System with Cellular Automata and Data Mining for Forecasting Severe Weather Patterns

Pokkuluri Kiran Sree[1] and SSSN Usha Devi N[2]

[1]Sree Vishnu Engineering College for Women Bhimavaram
[2]University College of Engineering, JNTUK Kakinada
E-mail: [1]pkiransree@gmail.com, [2]usha.jntuk@yahoo.com

Abstract—*Early detection of possible occurrences of severe convective events would be useful in order to avoid, or at least mitigate, the environmental and socio-economic damages caused by such events. In this project, we investigate the use of data mining techniques strengthened with cellular automata in forecasting maximum temperature, rainfall, evaporation and wind speed arriving at a hybrid system. The proposed classifier considers six neighborhood Cellular Automata (CA), which was implemented efficiently by a modified CLONAL classifier for accurate weather prediction. 6CAMCC classifier together with the standard data mining techniques will certainly improve the performance of the hybrid system. The performance of the proposed algorithms has shown 6.4% improvement when compared using standard performance metrics and the algorithms for predicting the temperature, rainfall, evaporation and wind individually.*

1. INTRODUCTION

Climate Forecasting involves foreseeing how the current situation with the air will change. Present climate conditions are acquired by ground perceptions, perception from satellites, ships, flying machine, floats, blow ups and climate stations covering the whole planet. This incorporates data from over the seas, from the surface (ships and floats), from high in the climate (satellites) and underneath the seas (a system of extraordinary buoys called Argo).creating conjectures is a complex procedure which is continually being overhauled. Climate gauges made for 12 and 24 hours are regularly very precise. Gauges made for two and three days are normally great. In any case past around five days, estimate precision tumbles off quickly. The rate of information era and capacity far surpasses the rate of information investigations. This speaks to lost open doors regarding exploratory bits of knowledge not picked up and effects or adjustment methods not satisfactorily educated. While there is a developed writing in atmosphere insights and scattered applications of information mining, orderly endeavors in atmosphere information mining need. Numerous specialists have attempted to utilize information mining innovations as a part of regions identified with meteorology and climate forecast.

2. REVIEW OF WEATHER FORECASTING

Decision trees: Decision trees models are generally utilized as a part of information mining to look at the information and to incite the tree and its decides that will be utilized to make expectations. Various diverse calculations may be utilized for building choice trees including CHAID (Chi-squared Automatic Interaction Detection), CART (Classification And Regression Trees), Quest, and C5.0. A choice tree is a tree in which each one limb hub speaks to a decision between various choices, and each one leaf hub speaks to a choice. Contingent upon the calculation, every hub may have two or more extensions. For instance, CART creates trees with just two limbs at every hub. Such a tree is known as a double tree. At the point when more than two limbs are permitted this is known as a multiway tree . Fluffy Logic: Fuzzy Logic is a basic yet effective critical thinking strategy with far reaching materialness. It is presently utilized as a part of the fields of business, frameworks control, hardware and activity designing.

The Rule Base: The standard base is a situated of principles of the If-Then structure. The If segment of a principle alludes to the level of enrollment in one of the fluffy sets. The Then partition alludes to the result, or the related framework yield fluffy set. For instance, one standard could be expressed: If (dry & unsaturated & drying & excessively light & overcast) Then (low likelihood of mist) the following step is to determine a framework yield, a likelihood of haze development, from the relevant standards. Note that few sets of guidelines above have the same result, or framework yield. The estimation of the yield is allotted the estimation of the "most genuine", or strongest, tenet.

3. CELLULAR AUTOMATA & DESIGN

A cell robot (CA) comprises of a standard framework (grid) of cells (automata), each of which can be in one of a limited

number of states. At discrete time steps, all cells all the while upgrade their states relying upon their current state and those of their prompt neighbors (i.e., contingent upon the nearby neighborhood setup of each one cell). For this upgrade step, all cells utilize the same deterministic redesign standard, which records the new cell states for every conceivable neighborhood setup. This overhaul procedure is then rehashed ("iterated") for a specific number of time steps.

Cell Automata (CA) are scientific models of decentralized spatially expanded frameworks. They comprise of countless basic individual units, or "cells", which are joined just by regional standards, without the presence of a focal control in the framework. Each one cell is a straightforward limited robot that over and again upgrades its own particular state, where the new cell state relies on upon the cell's present state and those of its prompt (neighborhood) neighbors. Be that as it may, notwithstanding the constrained usefulness of every individual cell, and the cooperations being confined to neighborhood neighbors just, the framework overall is fit for delivering many-sided examples, and even of performing confounded processings. In that sense, they structure an option model of reckoning, one in which data handling is carried out in an appropriated and exceedingly parallel way. Due to these properties, CA have been utilized widely to study complex frameworks in nature, for example, liquid stream in physical science or example development in science, additionally to study data transforming (reckoning) in decentralized spatially amplified frameworks (common or manufactured). Here, we will give a concise outline of the distinctive courses in which calculations could be possible with cell automata.

The data sets are collects from NCDC (National Climatic Data Center). The inputs are processed with set of 6CAMCC. Six neighborhood is taken into consideration.

This section explains the steps involved in modified CLONAL algorithm in brief.

1. Generate initial antibody population (AIS-MACA rules) randomly and call it as Ab. It consists of two subsets of memory population Ab_m and reservoir population Ab_r.
2. Construct a set of Antigens population call it as Ag (DNA Sequence with Class/ Input).
3. Select an antigen Ag_j , from Ag the antigen population.
4. Expose every member of the antibody population to the selected antigen Ag_j , Check whether it is predicting the correct class or not and calculate affinity of the rule with the antigen via fitness equations .
5. Select m highest affinity antibodies (AIS-MACA rules) from Ab and place them in Pm.
6. Generate clones for each antibody, which will be proportional to the affinity as per the equation . Place the clones in the new population P_i.
7. Apply mutation to the newly formed population P_i where the degree is inversely proportional to their affinity as per equation . This produces a more mature population P_i^*.
8. Recalculate the affinity of the rule with the corresponding antigen as in step 4. Order the antibodies in descending order. (high fitness antibody will be on the top)
9. Compare the antibodies from P_i^* with the antibodies population from Ab_m. Select the better fitness rules and remove them from Pi^* and place them in Ab_m.
10. Randomly generate antibodies for introducing diversity. Compare the antibodies in Ab_r , the left out antibodies in P_i^* and randomly generate antibodies. Select the better fitness rules among three antibody sets and place them in Ab_r.
11. For every generation compare the antibodies in Ab_m and Ab_r and place the best in Ab_m.
12. The output of the classifier is set of rules in Ab_m (solution set).

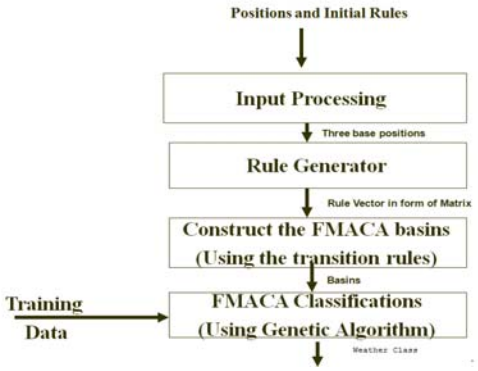

Fig. 1: Design of the Weather Forecast System

4. CAMCC RESULTS & DISCUSSION

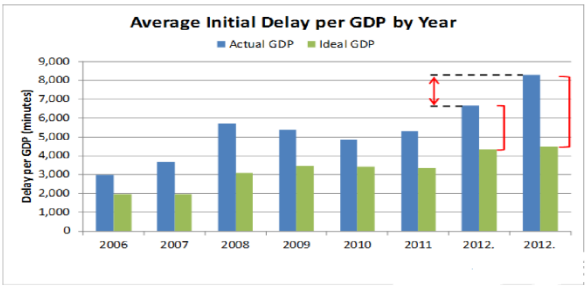

Fig. 2: GDP per year

The gross domestic product is taken from the statistics of USA. We have done examinations to assess both the MCS following strategy and the information mining methodology proposed in the paper. Firstly, we contrasted our MCS following methodology and the zone covering following technique proposed by Arnaud The physically "master eyescanning" strategy is additionally embraced as an execution benchmark of test results. Figure3 represents the exploratory results and correlations of the above routines. As demonstrated in Fig. 2,3 MCS no. measures the quantity of effectively followed Mcss; Error Rate measures the rate of slip in following, which is processed by utilizing the MCS no. as

numerator and the followed MCS no. of "master eye-filtering" system as denominator. This has demonstrated that the cloud following precision of our system makes an normal of 17% improvement over the following strategy proposed by Arnaud et al., and is close to the following precision of meteorologists. To take after on, the MCS information mining methodology is then tried on MCS following results. There are, altogether, 320 qualified Mcss that have been followed and described for information mining, among which 50 Mcss have moved out of the Tibetan Plateau (1051e): among them, 37 Mcss to "E", 9 Mcss to ""NE"" and 4 Mcss to "SE". A sum of 70% of the recognized MCS structures, that is, 224 Mcss, was utilized as train examples and the staying 30% was kept for testing. A gathering of deduction standards are produced and a situated of natural, physical model charts are plotted. records the ensuing choice guidelines of the C4.5 choice tree calculation used to arrange the evolvement patterns and moving trajectories of the Mcss moving out of the Tibetan Plateau at 500 hpa level. After the pruning process, the quantity of misclassifications on the experiments is 5 out of 96 Mcss and the mistake rate is 5.2%.

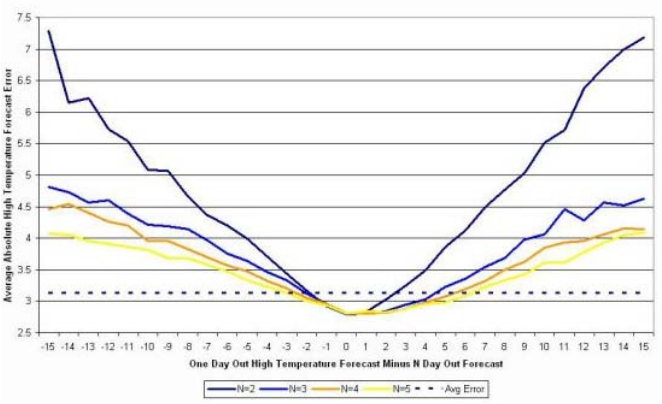

Fig. 3: Forecast on the Out High Temperature

5. CONCLUSION

We have successfully developed a preliminary system for predicting severe weather patterns. The performance of the proposed algorithms has shown 6.4% improvement when compared using standard performance metrics and the algorithms for predicting the temperature, rainfall, evaporation and wind individually. The accuracy of the classifier is due to good training algorithm with 6neighbour CAMCC.

REFERENCES

[1] Pokkuluri Kiran Sree et al, Investigating an Artificial Immune System to Strengthen the Protein Structure Prediction and Protein Coding Region Identification using Cellular Automata Classifier**.** International Journal of Bioinformatics Research and Applications ,Vol 5,Number 6,pp 647-662, ISSN : 1744-5493. (2009) (

[2] Pokkuluri Kiran Sree et al, Identification of Promoter Region in Genomic DNA Using Cellular Automata Based Text Clustering. The International Arab Journal of Information Technology (IAJIT),Volume 7,No 1,2010,pp 75-78.

[3] Pokkuluri Kiran Sree et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier for Coding Region Prediction in Human Genome, Journal of Bioinformatics and Intelligent Control, Vol. 3, 2014, pp 1-6.

[4] Pokkuluri Kiran Sree et al, A Fast Multiple Attractor Cellular Automata with Modified Clonal Classifier Promoter Region Prediction in Eukaryotes.Journal of Bioinformatics and Intelligent Control, Vol. 3, 1–6, 2014**.**

[5] Pokkuluri Kiran Sree et al, 5.MACA-MCC-DA: A Fast MACA with Modified Clonal Classifier Promoter Region Prediction in Drosophila and Arabidopsis. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 22-26,

[6] Pokkuluri Kiran Sree et al, Cellular Automata in Splice Site Prediction. European Journal of Biotechnology and Bioscience, 1 (6), 2014, pp 36-39

[7] Pokkuluri Kiran Sree et al, AIX-MACA-Y Multiple Attractor Cellular Automata Based Clonal Classifier for Promoter and Protein Coding Region Prediction. Journal of Bioinformatics and Intelligent Control 3, no. 1 (2014): 23-30.

[8] Pokkuluri Kiran Sree et al, PSMACA: An Automated Protein Structure Prediction Using MACA (Multiple Attractor Cellular Automata). Journal of Bioinformatics and Intelligent Control 2, no. 3 (2013): 211-215**.**

[9] Pokkuluri Kiran Sree et al, An extensive report on Cellular Automata based Artificial Immune System for strengthening Automated Protein Prediction. Advances in Biomedical Engineering Research (ABER) Volume 1 Issue 3, September 2013, pp 45-51.

[10] Pokkuluri Kiran Sree et al, A Novel Protein Coding Region Identifying Tool using Cellular Automata Classifier with Trust-Region Method and Parallel Scan Algorithm (NPCRITCACA). International Journal of Biotechnology & Biochemistry (IJBB) Volume 4, 177-189 Number 2 (December 2008).(Eight Years Old Journal) Listed in Indian Science Abstracts, ISSN: 0019-6339,Volume 45, Number 22, November 2009.

[11] Pokkuluri Kiran Sree et al, HMACA: Towards proposing Cellular Automata based tool for protein coding, promoter region identification and protein structure prediction. International Journal of Research in Computer Applications & Information Technology, Volume 1 Number 1, pp 26-31,2013.

[12] Pokkuluri Kiran Sree et al, PRMACA: A Promoter Region identification using Multiple Attractor Cellular Automata (MACA) in the proceedings CT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India- Vol I Advances in Intelligent Systems and Computing Volume 248, 2014, pp 393-399.

[13] Pokkuluri Kiran Sree et al, Towards Proposing an Artificial Immune System for strengthening PSMACA: An Automated Protein Structure Prediction using Multiple Attractor Cellular Automata proceedings of International Conference on Advances in electrical, electronics, mechanical and Computer Science(ICAEEMCS)-2013, ISBN: 978-93-81693-66-04 on September 2nd2013, Hyderabad.

[14] Pokkuluri Kiran Sree et al, Multiple Attractor Cellular Automata (MACA) for Addressing Major Problems in Bioinformatics in Review of Bioinformatics and Biometrics (RBB) Volume 2 Issue 3, September 2013, pp70-76.

[15] Pokkuluri Kiran Sree et al, Protein coding region Identification ,in proceedings of 2nd International Conference on Proteomics Bioinformatics, July 2-4, 2012 Embassy Suites Las Vegas, USA “,(Special Issue of Journal of Proteomics & Bioinformatics. (USA), Volume 5 Issue 6–123, ISSN:0974-276X,H

[16] Pokkuluri Kiran Sree et al, Hybrid Attractor Cellular Automata for Addressing Major Problems in Bioinformatics in Research and Reviews: Journal of Engineering and Technology, Volume 2 Issue 4, October-2013,pp 42-48.