An Ensemble Method Based Aggregated Model by Analyzing Data of Existing Precipitation Prediction Models*

Ramyaa[1] and Kallol Das[2]

***Abstract*— Most of the existing precipitation prediction models are not predicting well enough. Most of the cases these models are over-predicting. Sometimes the rate of false positive is way high. Again, some models are predicting good in some places and worse in other places, i.e., some of them are good at mountain areas, some of them are good at desert areas etc. The goal of this research is to reduce the error rates of the existing prediction models. There are lots of existing researches going on implementing new models to predict precipitation. But, the false positive rate didn't reduced that much. We propose an ensemble approach to develop a New Aggregated Model to predict precipitation based on the dataset of some existing prediction models.**

***Index Terms*— Machine Learning, Precipitation Prediction, Ensemble Method**

I. INTRODUCTION

Predicting correct amount of precipitation for a particular day is always tough. Existing well established precipitation prediction models are not accurate enough. Sometimes the error rates are way high. Lots of research going on to improve the prediction accuracy, i.e., to decrease the error rates.

Basically, most of the research that has been done so far implemented new models to predict precipitation from some real features. Traditional statistical analysis techniques were mostly used previously for precipitation prediction. J. C. Thompson [2] proposed a numerical method to predict precipitation. This prediction model was based on a graphical integration technique by using a number of independent variables. Later machine learning started performing more accurately over traditional statistical analysis.

Wei-Chiang Hong [1] proposed a hybrid model of RNNs and SVMs (named as RSVR) to forecast the precipitation amounts. Chaotic Particle Swarm Optimization (CPSO) algorithm has been used to select the parameters of the SVR model. Selected parameters were used to predict precipitation amount. Theoretically that research was showing significantly small Normalized Mean Square Error rate, but, the predicting forecast for verification data and testing data had right shifted result in the time domain.

Emilcy Hernandez et al. [3] proposed a deep learning architecture for the next day precipitation prediction. In total, forty-seven features, including temperature, humidity, wind direction, pressure, previous rainfalls etc., have been used as input in this research to predict the amount of precipitation for the next day. According to the result of this research, new model is less accurate for days with light rainfall.

Beda Luitel et al. [4]

In summary, many research has been proposed new models from the real weather data by considering a good amount of features e.g., temperature, wind speed, humidity etc. Most of the cases, the error rates for the prediction data of those new models are high. [1] [3] Again, some models are more accurate for days with heavy rainfall, but less accurate for light rainfall. [3] Some models are good in mountain, desert cite cite

A very few works have been done to improve these models. In other word, many more new models are proposing instead of trying to improve the existing models. We have rainfall dataset of 39 Prediction Models and a Real Verification dataset for 39 different days in total and for each day we have data for 20 different times. **Our approach is to propose a new aggregated model from these existing prediction models' which can perform better than other existing prediction models.** In other word, our goal is to analyze the error rates of these 39 prediction models data and propose better model which has lower error rates than the existing prediction models.

a

a

a

a

a

a

a

a

a

a

a

a

a

II. PROBLEM STATEMENT

The Research Problem we are dealing with is that all the existing prediction models have a high error rates. Over-predicting rates of these models are quite high in lots of places. Different models are performing better in different places, but doing worse in other places. So, here comes our research question- can we implement a new model from these existing models data which has comparatively lower over-prediction and lower false positive rate? Since, different models give wrong prediction in different areas, we assume that an aggregated model would perform better.

*This work was supported by Oklahoma University

[1]Ramyaa is with Faculty of Computer Science, New Mexico Tech, 801 Leroy Pl, New Mexico, USA ramyaa.ramyaa at gmail.com

[2]Kallol Das is with the Department of Computer Science, New Mexico Tech, 801 Leroy Pl, New Mexico, USA kalloldash at gmail.com

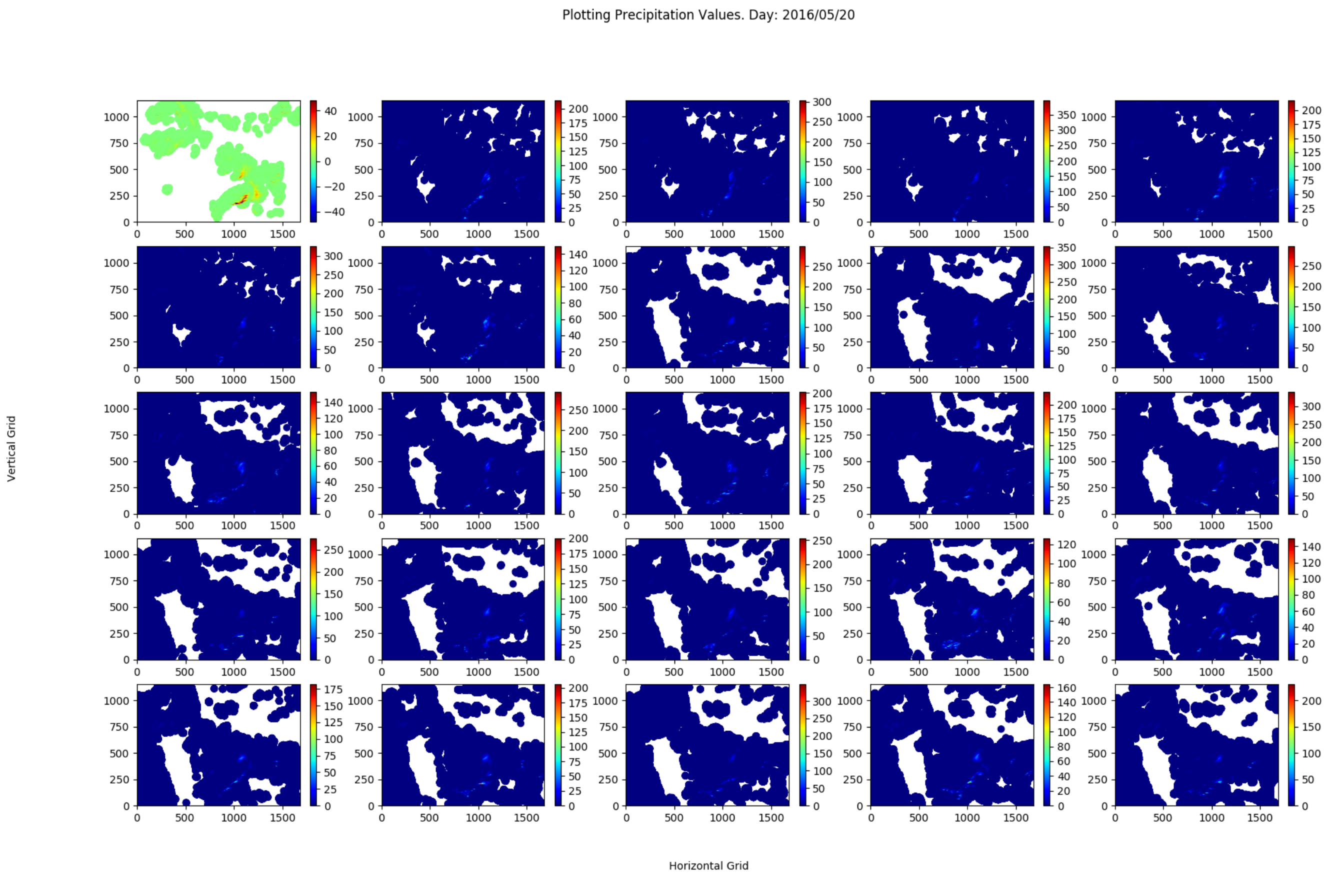

Fig. 1. Visualization of precipitation real and prediction data. The first image (image in the top left corner) is the real precipitation data and rest of the 24 images are visualizing precipitation prediction data from 24 prediction models.

Fig. 1. illustrates the comparison between all the prediction models data we have and the real verification data. All the prediction models data have some areas where real data (image in the top left corner) doesn't have any precipitation in those particular area.

III. Methodology

Some prediction models didn't have data for a large amount of time. Again, some days didn't have data for lots of prediction models. The data has been used to develop a new aggregated model has 24 prediction models for 20 days in total and 10 different times for each day.

The Mean Absolute Error and the Root Mean Square Error for the existing prediction models data have been calculated over all days and times. It has been observed from these

IV. Experimental Result

Most of the existing rainfall prediction models are not predicting well enough. Most of the cases these models are overpredicting. Sometimes the rate of false positive is way high. Again, some models are predicting good in some places and worse in other places. In other word, some of them are good at mountain areas; some of them are good at desert areas. There are lots of existing researches going on

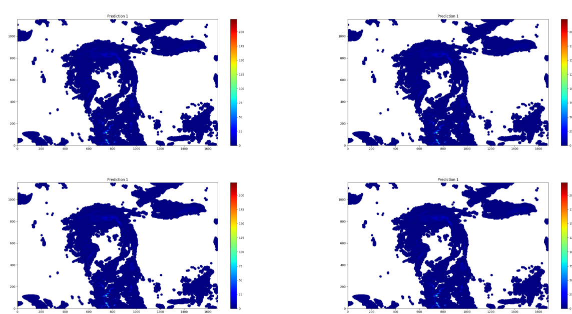

Fig. 2. Visualization of precipitation real and prediction data. The first image (image in the top left corner) is the real precipitation data and rest of the 24 images are visualizing precipitation prediction data from 24 prediction models.

implementing new models. But, the false positive rate didnt reduced that much.

V. Conclusions

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

APPENDIX

Appendixes should appear before the acknowledgment.

ACKNOWLEDGMENT

The authors would like to thank the professor of Computer Science Department, New Mexico Tech, Dr. Hamdy Soliman for providing the access in his powerful machine for this research work.

REFERENCES

[1] Wei-Chiang Hong, Rainfall forecasting by technological machine learning models, Applied Mathematics and Computation, Volume 200, Issue 1, 2008, Pages 41-57, ISSN 0096-3003

[2] THOMPSON, J.C., 1950: A NUMERICAL METHOD FOR FORECASTING RAINFALL IN THE LOS ANGELES AREA. Mon. Wea. Rev., 78, 113124

[3] Hernndez E., Sanchez-Anguix V., Julian V., Palanca J., Duque N. (2016) Rainfall Prediction: A Deep Learning Approach. In: Martnez-lvarez F., Troncoso A., Quintin H., Corchado E. (eds) Hybrid Artificial Intelligent Systems. HAIS 2016. Lecture Notes in Computer Science, vol 9648. Springer, Cham

[4] Beda Luitel, Gabriele Villarini, Gabriel A. Vecchi,Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones, Journal of Hydrology, Volume 556, 2018, Pages 1026-1037, ISSN 0022-1694

A. Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the US-letter paper size. Please do not use it for A4 paper since the margin requirements for A4 papers may be different from Letter paper size.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations

VI. MATH

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as 3.5-inch disk drive.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: Wb/m2 or webers per square meter, not webers/m2. Spell out units when they appear in text: . . . a few henries, not . . . a few H.
- Use a zero before decimal points: 0.25, not .25. Use cm3, not cc. (bullet list)

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled. Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus (/), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in

$$\alpha + \beta = \chi \tag{1}$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use (1), not Eq. (1) or equation (1), except at the beginning of a sentence: Equation (1) is . . .

D. Some Common Mistakes

- The word data is plural, not singular.
- The subscript for the permeability of vacuum ?0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter o.
- In American English, commas, semi-/colons, periods, question and exclamation marks are located within

quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)

- A graph within a graph is an inset, not an insert. The word alternatively is preferred to the word alternately (unless you really mean something that alternates).
- Do not use the word essentially to mean approximately or effectively.
- In your paper title, if the words that uses can accurately replace the word using, capitalize the u; if not, keep using lower-cased.
- Be aware of the different meanings of the homophones affect and effect, complement and compliment, discreet and discrete, principal and principle.
- Do not confuse imply and infer.
- The prefix non is not a word; it should be joined to the word it modifies, usually without a hyphen.
- There is no period after the et in the Latin abbreviation et al..
- The abbreviation i.e. means that is, and the abbreviation e.g. means for example.

VII. USING THE TEMPLATE

Use this sample document as your LaTeX source file to create your document. Save this file as **root.tex**. You have to make sure to use the cls file that came with this distribution. If you use a different style file, you cannot expect to get required margins. Note also that when you are creating your out PDF file, the source file is only part of the equation. *Your TEX $\rightarrow$ PDF filter determines the output file size. Even if you make all the specifications to output a letter file in the source - if you filter is set to produce A4, you will only get A4 output.*

It is impossible to account for all possible situation, one would encounter using TEX. If you are using multiple TEX files you must make sure that the “MAIN“ source file is called root.tex - this is particularly important if your conference is using PaperPlaza's built in TEX to PDF conversion tool.

A. Headings, etc

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named Heading 1, Heading 2, Heading 3, and Heading 4 are prescribed.

B. *Figures and Tables*

Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation Fig. 1, even at the beginning of a sentence.

TABLE I

AN EXAMPLE OF A TABLE

One	Two
Three	Four

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an document, this method is somewhat more stable than directly inserting a picture.

Fig. 3. Inductance of oscillation winding on amorphous magnetic core versus DC bias magnetic field

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity Magnetization, or Magnetization, M, not just M. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write Magnetization (A/m) or Magnetization A[m(1)], not just A/m. Do not label axes with a ratio of quantities and units. For example, write Temperature (K), not Temperature/K.