Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549679

PREDICTION OF HEATWAVE USING ADVANCE SOFT COMPUTING TECHNIQUE

Ratnakar Das[1], Jibitesh Mishra [2], Sujogya Mishra [3], P. K. Pattnaik [4], S. Das[5], G. Pallai[6]
[1,2]Dept. of Computer Science & Application, Odisha University of Technology and Research, Bhubaneswar, Odisha 751029, India
[3,4]Dept. of Mathematics, Odisha University of Technology and Research, Bhubaneswar, Odisha 751029, India
[5]Department of Electrical Engineering, GITA Autonomous College, Bhubaneswar, Odisha, India
[6]Department of Electronics and Communication Engineering, GITA Autonomous College, Bhubaneswar, Odisha, India
Emails: rkdashpuri@gmail.com[1], jmishra@cet.edu.in[2], sujogya123@gmail.com[3], papun.pattnaik@gmail.com[4], sahasranshu.das7@gmail.com[5], gpalai28@gmail.com[6]

Abstract: At present, there is no suitable instrument available to simulate modeling the thermal performance of various areas of our states due to its complicated meteorological behavior. To accurately predict the heatwave, we studied the research gaps and the currently ongoing research on the prediction of heatwaves. For the prediction of heatwave accurately, we had considered two soft computing concepts, a. Rough Set Theory (RST) and b. Support Vector Machine. All the ongoing research on the prediction of heatwaves is based on the future prediction with an error margin. All the available techniques use a particular pattern of heatwave data, and these methods do not apply to vague data. This paper used an innovative RST and SVM technique, which can be applied to vague and imprecise data sets to produce the best outcomes. These outcomes are validated using a statistical method.

Keywords: Heat Wave, Rough Set Theory, Support Vector Machine, Soft Computing, Statistical method.

INTRODUCTION

The expected temperature of the last five-year period (0.75 0C-0.95 0C increase from the expected annual report for temperature in the last decade) was an all-time high, according to the report from the National Center for Environmental Information of the National Oceanic atmospheric administration. Heatwave globally has significantly changed the atmospheric condition in the recent past, cumulative the probability and gravity of meteorological and abnormal atmospheric conditions [1, 2]. Several researchers have researched social issues using the incomplete data set. Das et al. [3] discuss the managerial policy using a rough set. Mishra et al. [4] discuss the legal justice for government and Non-Government organizations by using Rough Set. Mishra et al. [5] discuss the future and sustainability of educational institutions run by private management. Mishra et al. [6] discuss various types of classification using a rough set to increase the sustainability of software industries. Nayak et al. [7] discuss the global medical challenges of malaria using a rough set. Much research has been conducted on various business establishments using rough sets applied to a vague pattern of business data set. Das et al. [8] applied the Rough Set technique to Software industries to get significant results about the future outcomes and its sustainability in the long run for cardiac diseases. Das et al. [9] discuss the future and remedies for the design of a mathematical model to predict rainfall. Mishra et al. [10] analyzed various population parameters of the different business sectors to find the exact importance parameter responsible for its development. Mishra et al. [11] discuss the development of business logic using a rough set. Park et al. [12] discuss the destruction and damage due to heatwaves, and they analyzed the time-series data set of the heatwave to predict the damage of heatwaves and methods to overcome these situations. Que et al. [13] used a hybrid technique using ANFIS, SVM, and ANN to estimate global solar radiation in a specific environment. The significance of solar energy and its utility in developing countries discussed by Liu et al. [14] and Besharat et al. [15] emphasized estimating solar energy by many empirical correlation techniques and using various meteorological and environmental factors. The principle of the satellite-derived technique is capable of approximating cosmological energy data taken from a huge geographical region, but it is comparatively novel and may agonize from a lack of available data. Weather forecasting using Stochastic weather generators is useful for producing daily approximations from significant regulars but not for validating the model if restrained data are not present. However, because the insolation data are not easily accessible in some locations, many models based on temperature have been discussed, and several assessments and changes have afterward been made by Chen et al. [16] and Olatomiwa et al. [17].

Various soft computing methods have been cast off in the current time for approximation of worldwide radiation due to sunlight, with ANFIS and ANN has been providing the most significant result. Mohammadi et al. [18] implemented a combination of ANFIS and SVM methods to predict worldwide solar radiation, including air temperatures in Bandar Abbas, situated in the south of Iran. Piri et al. [19] compared four sunshine periods using experiential analysis and the SVM technique to evaluate worldwide solar radiation in two different localities of Iran. Ramadan et al. [20] discussed an innovative SVM technique to forecast global solar radiation over Teheran, Iran. The SVM_rbf forecasting was compared with various soft computing methods like ANN and AIFN. The outcomes confirmed that the SVM_rbf provided better results than all hybrid soft computing models. Many research work [21-23] on soft computing have predicted different environmental and medical issues.

RESEARCH METHODOLOGY

This paper uses two concepts: Rough set and SVM; data sets are taken from various regions of our state Odisha. The seasonal data concerning heatwaves affecting our environment are described in the following table. I.The data was collected randomly from different parts of our country. The basic concept of data collection is based on the best available sources. From the above table, it is clear that environmental hazards due to heatwaves had a uniform trend throughout India. So, we focus our attention on environmental hazards due to heatwaves. In this paper, our principal objective is to significantly find the attribute affecting the environment. To achieve our result, we used the concept of a rough set [24] and combined it with the statistical method.

TABLE I.

Date	Number of Environmental hazards due to the heatwave in India
5.03.2019	1032
16.04.2019	1225
29.04.2019	1306
15.06.2019	1225
17.08.2019	1125
17.12.2019	925
18.02.2020	1035
15.03.2020	1115
18.10.2020	825
25.12.2020	778

Background for data analysis

The analytical part of this work is based on collecting data from various regions of our country. The data is present in the table given below:

TABLE II. Survey data related to Various Climate

States	Cold Wave	Heavy Rainfall	Heat Wave	Moderate Climate	Low Pressure	High Humidity	Total
Uttar Pradesh	15000	15000	15000	5000	5000	5000	60,000
Andhra Pradesh	5000	15000	15000	15000	15000	5000	70,000
West Bengal	15000	15000	10000	20000	5000	10,000	57,000
Coastal-Odisha	10,000	10000	15000	15000	5000	10,000	65,000
Kerala	10,000	15,000	15000	5000	5000	5000	55,000

Karnataka	5,000	5,000	15,000	15000	15000	15,000	70,000
Bihar	15,000	15,000	25,000	15000	15000	15,000	100,000
Delhi (U.T)	5,000	12,000	8,000	15000	5000	5,000	50,000

The conditional attribute in this case <Cold Wave, Heavy Rainfall, Heatwave, Moderate Climate, Low Pressure>.

Further Data analysis

We use the basic technique of rough set to find the significant attribute which affects the environment. To achieve the result, we designed the entire Table-II into a decision-making table similar to the rough set decision table, with detail of the rough set described in the preceding sections.

Fundamentals of Rough Set

The basics of a Rough set are of the form <U, C,{d}> where Universal set (U), conditional attribute (C), and decision attributes ({d}). This can be represented in a table called a decision table.

TABLE-III Rough Set (Initial Table)

VX	ax	bx	cx	dx
VX_1	1	4	7	10
VX_2	2	5	8	11
VX_3	3	6	9	12

In this case, records (V<V1,V2,V3>), conditional attributes (<asx, bsx, csx>) with values <1 to 9> and decision attributes (ds) with values <10,11,12>.

TABLE-IV Renaming Conditional Attributes

Serial Number	Conditional Attributes	Renaming conditional
1	Cold Wave	11
2	Heavy Rainfall	22
3	Heat Wave	33
4	Moderate Climate	44

5	Low Pressure	55

The decision attribute decides the most significant attribute responsible for affecting the environment among the five attributes described in the above table, renamed dx. The values of the conditional attributes are <high, moderate> renamed as <bsx, csx>, the value of the decision attributes 'dsx' significant & pointless renamed as <1, 2> respectively.

ANALYSIS USING ROUGH SET ALGORITHM

Algorithm to find the reduct:
Input:
QPR, Conditional Attribute Set, ETM, Decision attribute set,α is the equivalence classes concerning R, L is the initial equivalence class.
Output: R, sets with a unique feature

i. D:= { }, R: = { }
ii. repeat
iii. D: = R
iv. ∀ y ∈ (QPR – R)
v. if $\alpha_{RU\{y\}}$ (EMT) >α_R(L)
vi. D: = R U {y}
vii. R: = D
viii. until α_D(DSL) = α_{SRT}(EMT)
ix. return D

To start we had considered 100,000 records initially using statistical approach(correlation) technique, we get 6-records, with dissimilar features i.e., <VX1, VX2, VX3, VX4, VX5, VX6> for six conditional attributes <11 ,22 ,33, 44, 55> with their values described as High and Moderate renamed as bsx, csx as described in table-IV. The decision attribute ds to find the significant attributes it's valued <1,2>

TABLE VII. Preliminary Data Table

VX	11	22	33	44	55	dsx
VX_1	bsx	bsx	csx	csx	bsx	1
VX_2	bsx	bsx	bsx	csx	bsx	1
VX_3	csx	csx	bsx	bsx	bsx	2
VX_4	csx	csx	csx	bsx	bsx	2
VX_5	csx	csx	csx	bsx	csx	2
VX_6	bsx	csx	bsx	bsx	csx	2

II(11)={(rsk$_1$,rsk$_2$,rsk$_6$),(rsk$_3$,rsk$_4$,rsk$_6$)},II(22)={{rsk$_1$,rsk$_2$),(rsk$_3$,rsk$_4$,rsk$_5$,rsk$_6$)},II(33) = {(rsk$_3$,rsk$_4$,rsk$_5$),(rsk$_2$,rsk$_3$,rsk$_6$)},II(44) = {{rsk$_1$,rsk$_2$},{rsk$_3$,rsk$_4$,rsk$_5$,rsk$_6$}},II(55) = {{rsk$_1$,rsk$_2$,rsk$_3$,rsk$_4$},{rsk$_5$,rsk$_6$}},II(11,22)={{rsk$_1$,rsk$_2$},{rsk$_3$,rsk$_4$,rsk$_5$},rsk$_6$},II(11,33)={{rsk$_1$},{rsk$_2$,rsk$_6$},{rsk$_3$},{rsk$_4$,rsk$_5$}},II(11,44)={{rsk$_1$,rsk$_2$},{rsk$_2$,rsk$_6$},{rsk$_3$},{rsk$_4$,rsk$_5$}},

$II(11,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(22,33) = \{\{rsk_1\},\{rsk_2\},\{rsk_3,rsk_6\},\{rsk_4,rsk_5\}\}$, $II(22,44)=\{\{rsk_1,rsk_2\},\{rsk_3,rsk_4,rsk_5,rsk_6\}\}$, $II(22,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5,rsk_6\}\}$, $II(33,44) = \{\{rsk_1\},\{rsk_2\},\{rsk_3,rsk_6\},\{rsk_4,rsk_5\}\}$, $II(44,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5,rsk_6\}\}$, $II(11,22,33) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4,rsk_5\},\{rsk_6\}\}$, $II(11,33,44) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4,rsk_5\},\{rsk_6\}\}$, $II(11,44,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(11,22,44) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4,rsk_5\},\{rsk_6\}\}$, $II(11,22,55) = \{\{rsk_1,rsk_2\},\{rsk_3\},\{rsk_4,rsk_5\},\{rsk_6\}\}$, $II(11,33,55) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4,rsk_5\},\{rsk_6\}\}$, $II(11,44,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(22,33,44) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4,rsk_5\},\{rsk_6\}\}$, $II(22,33,55) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(22,44,55) = \{\{rsk_1,rsk_2\},\{rsk_3,rsk_4\},\{rsk_5,rsk_6\}\}$, $II(11,22,33,44) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(11,22,33,55) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(11,33,44,55) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4\},\{rsk_5\},\{rsk_6\}\}$, $II(22,33,44,55) = \{\{rsk_1\},\{rsk_2\},\{rsk_3\},\{rsk_4\},\{rsk_5\},\{rsk_6\}\}$.

Core={33}, as {33} =∩{Reduct<(1,2,3,4,5)>}

As we get heatwaves to affect the environment significantly in the future. Further, we classify the conditional attribute responsible for heatwave magnification.

TABLE VIII Reduct Table

Serial Number	Reduct
1	(22,33,55)
2	(11,22,33,44)
3	(11,22,33,55)
4	(11,33,44,55)
5	(22,33,44,55)

Further Analysis Using Rough Set

We consider the attributes such as <Industrialization, Air Pollution due to motor vehicles, Lack of greenery, Uncontrolled wind speed, and radiation, uses of equipment's like air conditioner & Air Coolers, we rename these conditional attributes as <1,2,3,4,5> and their values are renamed as significant as 'asx' and insignificant as 'bsx', and decision attribute dv and its values are noteworthy and pointless as 'csx' and 'dsx'. We applied the Rough set concept to the data collected from various parts of our country. Initially, we collected 100,000 data from various sources, as described in table-II column-3, using the correlation technique. We had divided the entire data set into six dissimilar records such as <AX_1, AX_2, AX_3, AX_4, AX_5, AX_6>.

TABLE IX. Data Table for Heatwave

AX	1	2	3	4	5	dsx
AX_1	asx	asx	bsx	bsx	asx	csx
AX_2	asx	asx	asx	bsx	asx	csx
AX_3	bsx	bsx	asx	asx	asx	dsx
AX_4	bsx	bsx	bsx	asx	asx	dsx
AX_5	bsx	bsx	bsx	asx	bsx	dsx
AX_6	asx	bsx	asx	asx	bsx	dsx

Finding the reduct similar to the analysis performed in section-3.2, we got the core as conditional attribute-3, i.e., Lack of greenery is a major issue in incrementing a heatwave. Although there are several natural ways to preserve the greenery, advanced technology must be implemented to preserve the greenery for a long period.

Statistical validation

We applied the chi-square distribution (Statistical test) to validate our claim. We prefer chi-square distribution over another statistical test as it is a non-parametric test that does not follow any particular distribution.

H0(Null Hypothesis)-: There is no significance of least amount greenery (increment of heat wave)

H1(Alternate Hypothesis)-: Lack of greenery has strong significance increment of heat wave

Observations Samples are 10,10,15,10,15,15,5,5,10,5

Expected Samples are: 10%, 10%, 20%, 25%, 10%, 10%, 15%, 15%, 10%, 15%.

Expected Values are: 10,10,20,25,10,10,15.15,10,15

$$\chi^2 = \sum \frac{(\text{Observed value-Expected values})^2}{\text{Expected values}} = 32.3$$

$\chi^2(0.05, 9) = 16.919$ tabular chi-square value as $\chi^2_{\text{calculated}}$ greater than χ^2_{table} so rejected the null hypothesis and accept the alternate hypothesis.

Concluding Remarks and Future Scope

In this our major thrust to find significant attribute responsible for environmental hazard, by using the rough set technique, we found that heatwave is the conditional attribute mainly responsible for environmental hazards, by further classification about heatwave we found that Lack of greenery will increase the heatwave in the future. Although the survey was conducted on sampling distribution, we expect this result to signify the population as we collected our sample in a heterogeneous environment. This concept can apply to various domains like wind speed predictions, rainfall prediction, and various meteorological phenomena.

References

- R. Djalante, Key assessments from the IPCC special report on global warming of 1.5 ◦C and the implications for the Sendai framework for disaster risk reduction. Prog. Disaster Sci., 1, 100001, 2019.
- P. Peduzzi, The disaster risk, global change, and sustainability nexus. Sustainability, 11, 957, 2019.

- R. Das, D. Mishra, S. Mishra, Rough Set Applications for the Classification of Managerial Policies Effect on Industries (IJCA), 80, 6, 2017.
- S. Mishra, S.P. Mohanty, S.K Pradhan,Reasons for employees need justice from legal bodies: A rough set approach, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), IEEE Conference.
- S. Mishra, S.P. Mohanty, S.K. Pradhan,Rough Set Approach for Development of College Industry a New Theory, IJCA, 116, 6, 2015.
- S. Mishra, A. Mohmaed, P.K. Pattnaik, K. Muduli, T.S.T. Ahmad, Soft Computing Techniques to Identify the Symptoms for COVID-19. Lecture Notes on Data Engineering and Communications Technologies, 86, 283–293, 2022.
- S.K. Nayak, S.K. Pradhan, S. Mishra, S. Pradhan, P.K Pattnaik Rough Set Technique to Predict Symptoms for Malaria, 15th INDIACOM IEEE- Conference, 2021.
- S. Das, S.K. Pradhan, S. Mishra, S. Pradhan, P.K. Pattnaik, 2021. Analysis of Heart Diseases Using Soft Computing Technique. Proceedings - 2021 19th OITS International Conference on Information Technology, OCIT 2021, 178–184, 2021.
- R. Das, J. Mishra, S. Mishra, P.K. Pattnaik, Design of mathematical model for the prediction of rainfall. Journal of Interdisciplinary Mathematics, 25(3), 587–613, 2022.
- S. Mishra, S.P. Mohanty, and S.K. Pradhan. Article: Rough Set Approach for the Classification of Advertisement in the Development of Business Establishment. International Journal of Computer Applications, 110 (5), 34-41, 2015.
- S. Mishra, S.P. Mohanty, and S.K. Pradhan,Development of a Business Logic using Simple Application of Rough Set Theory, IJCSI, 11 (5), 2014.
- M. Park, D. Jung, S. Lee, and S. Park, Heatwave Damage Prediction Using Random Forest Model in Korea, Appl. Sci., 10, 8237, 2020. doi:10.3390/app10228237.
- V.H. Queja, J. Almoroxa, Javier A. Arnaldob, L. Saitoc, ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment, Journal of Atmospheric and Solar-Terrestrial Physics, 2017.
- X. Liu, X. Mei, Y. Li, Y. Zhang, Q. Wang, J.R. Jensen, Calibration of the Ångström–Prescott coefficients (a, b) under different time scales and their impacts in estimating global solar radiation in the Yellow River basin. Agric. For Meteorol. 149, 697–710, 2009.
- F. Besharat, A.A. Dehghan, A.R. Faghih, Empirical models for estimating global solar radiation: a review and case study. Renew. Sustain. Energy Rev. 21, 798–821, 2013.
- J.-L. Chen, H.-B. Liu, W. Wu. D.-T. Xie, Estimation of monthly solar radiation from measured temperatures using support vector machines–a case study. Renew. Energy, 36, 413–420, 2011.
- L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, C. Sudheer, A support vector machine–firefly algorithm-based model for global solar radiation prediction. Sol. Energy 115, 632–644, 2015. http://dx.doi.org/10.1016/j.solener.2015.03.015
- H. Khorasanizadeh, K. Mohammadi, Prediction of daily global solar radiation by day of the year in four cities located in the sunny regions of Iran. Energy Convers. Manag. 76, 385–392, 2013.http://dx.doi.org/10.1016/j.enconman.2013.07.073.

- J. Piri, S. Shamshirband, D. Petković, C.W. Tong, M.H. ur Rehman, Prediction of the solar radiation on the Earth using support vector regression technique. Infrared Phys. Technol. 68, 179–185, 2015. http://dx.doi.org/10.1016/ j. infrared.2014.12.006.
- Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband, B. Khoshnevisan, Potential of radial basis function-based support vector regression for global solar radiation prediction. Renew. Sustain. Energy Rev. 39, 1005–1011, 2014. http:// dx.doi.org/10.1016/j.rser.2014.07.108
- R. Das, J. Mishra, S. Mishra, P.K. Pattnaik, S. Das, Mathematical Modeling using Rough Set and Random Forest Model to Predict Wind Speed.2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), 207- 213, 2022.
- S.K. Nayak, S.K. Pradhan, S. Mishra, S. Pradhan, P.K. Pattnaik, Prediction of Cardiac Arrest using Support Vector Machine and Rough Set, 9th International Conference on Computing for Sustainable Global Development (INDIACom), 164-172, 2022.
- S. Das, S.K. Pradhan, S. Mishra, S. Pradhan, P.K. Pattnaik, A Machine Learning-based Approach for Detection of Pneumonia by Analyzing Chest X-Ray Images, 9th International Conference on Computing for Sustainable Global Development (INDIACom), 177-183, 2022.
- Z. Pawlak, Rough sets International Journal of Computer & Information Sciences 11, 341–356, 1982.