A Neuro-Fuzzy Integrated Clustering for Weather Knowledge Analysis

Sonakshi Dahiya
CSE & IT Dept.
ITM University
Gurgaon, India

Yogita Gigras
Astt. Professor,CSE & IT Dept.
ITM University
Gurgaon, India

ABSTRACT

Weather Information processing and knowledge extraction is one of the challenging applications of data mining. This process area requires authenticated and intelligent processing to obtain accurate information from the knowledge set. In this work, an intelligent clustering mechanism is defined to acquire such information. This neuro-fuzzy based model is applied on raw dataset defined with various weather characteristics including humidity, temperature, rainfall etc. The work is divided in three main stages. In first stage, the filtration over the dataset is performed to get more relevant information set. In second stage, the clustering is performed to divide the information set in knowledge groups. In final stage, the filtration over the knowledge set is performed to acquire the most effective knowledge. The results show the effective information analysis is obtained from the work.

Keywords

Neuro-Fuzzy; Weather Forecasting; Knowledgeset; Clustering.

1. INTRODUCTION

Clustering is one of most common data processing activity defined to generate the data patterns. These data patterns are considered as the measurement vectors defined in multidimensional space so that the relevant information can be collected in different data groups. The formation of these clusters is based on the similarity analysis. The clustering is effective to provide the data filtration as well as data reduction so that the information discovery can be done in an effective way. The cluster formation is based on the association analysis and structural analysis. This group generation and formation is based on the evidence analysis applied on the constraint specification. In this present work, a process variable based analysis approach is defined to divide the dataset in various data groups. The work is here defined based on multiple parameter or objects. In this work, the weather forecasting dataset is taken for more than 20 years. The information set considered here is defined with various associated features including the humidity, temperature and rainfall. All the dataset values are taken at different time instances obtained throughout the year. This dataset is collected for a specific city.

In this work, the cluster analysis approach is defined to analyze the behavior relation between these data values. This relationship establishment is also based on multiple vectors including the distance, variation, similarity measures. In this work, clustering effective classification model is defined to divide the dataset in certain groups. The pattern analysis is the key feature considered in this work for relevant class formation. This kind of information processing is the specialized data mining application defined under time domain. To acquire the information from such dataset, the time series segmentation is required.

1.1 Time Series Segmentation

This kind of segmentation process is considered as most intelligent and relevant information processing system defined for real environment. This kind of information process uses the neuron specific model to train the information set and adapt the required information from it. This segmentation process is defined in 2D space with specification of neurons to form the clusters. The learning model is here applied to generate the data patterns so that the neighbor information and distribution processing is performed. Such kind of data processing includes the cluster formation on larger dataset. The topological ordering is performed to identify the similarity between the data groups and this similarity analysis is handled by the neuron specific constraints. The intelligent neuron processing is defined to identify the cluster requirement at early stage and later on the data segmentation is applied. As the process continue, the cluster splition and merge operations can also be applied in an integrated way.

This kind of segmentation is also defined under specification of clustering model so that that the data division and the reduction is performed in effective way. The data patterns are generated based on the temporal specification so that the dataset formation and the relative information switching can be obtained from the work. This kind of information processing also defined the analysis as the integrated stage to the process. This processing model includes the analysis applied at the earlier stage so that the information division can be performed on larger dataset. In this work, the year specific time series segmentation is applied. The optimization is here been done using neuro fuzzy based rule specification.

1.2 Neuro-Fuzzy

In this work, neuro-fuzzy model is been applied to acquire the most effective information from the dataset. The layered model is here defined to process the time series data and to obtain the effective information results. The model used the filtered dataset as the input neuron set for the process. The constraints are here defined under the specification of forecasting features. The complex information format is defined with informational view to perform data modeling. Based on this, the data mapping is performed under constraint specification. Along with this, the predictive decision vectors are defined along with the specification of training time, validation rule etc. Based on this learning rule specification, the effective information validation is done. Finally the learning process is applied under fuzzy rules to acquire the information and to present the results.

In this paper, the weather forecasting dataset processing is done to obtain the knowledge extraction from the dataset. The knowledge is here represented to categorize the data values under different vectors so that the relative decisions can be performed. In this paper, two stage model is presented. In this section, the requirement of clustering process as the classification model is defined. The section also explored the requirement and features of time series segmentation. In section II, the work defined by earlier researchers is discussed. In section III, the proposed research methodology is explained. In section IV, the conclusion obtained from the work is presented.

2. EXISTING WORK

Different researcher presented the work on time series data to acquire the effective information and generate the relative results. The prediction and classification work are presented by different researchers. Some of such work is described in this section. Bragin[1] presented the work to process on real time remote sensing data. This data is obtained from underground sensor network to perform the soil information analysis. Author defined the data object specification for relevant information processing and information derivation. Takashi Kido[2] defined a work on pattern specification and mining under information disease identification with associated constraint specification. Bernard J. Vigier[3] provided a way to separate the damage identification on rot infection and provided the analysis on the low level information processing and generating the information narrow band so that the feature improved extraction will be obtained.

Anshuk.a Srivastava[4] processed on disease prediction based on plant information set. Author identify the crop under multiple features so that the detection of disease will be done at early stage of work. Ting Li[5] defined work on flea beetle prediction for feature generation and analysis. Author defined a work on learning rate analysis so that the data features can be improved. ZalizahAwang[6] defined work on pattern information extraction on real time dataset. Author defined the mining operations with association rules so that the range specification information set will be generated. Author defined the validation rule to process the data and to acquire the useful information from it.

Zhang Lei[7] defined the characterization of information data for gold mine area. Author collected the information study based sample set generation and processing under target analysis so that the information content analysis will be obtained. Author defined the soil information processing under spectrometric method so that the physical and analytical information will be obtained from the dataset. A. K. Tripathy[8] defined a data mining processing on network data so that the disease prediction processing will be obtained. Author defined the mining operation and sensor data processing so that the distributed network processing and the technological improvement will be done. E. L. Petersen[9] presented a work on even specific handling on crique dataset under rate level analysis. Author processed the sample set for the estimation of rates and generates the critical information with low level prediction and recommendation based on standard information analysis. Amir F. Atiya[10] defined the conceptual phenomenon based on the sparse prediction of data under model specification. This kind of information generation is maintained on network information and based on the network operations. Lean Yu[11] defined an agent based work on group data processing and information forecasting for price data. Author defined a model under intelligent

fuzzification approach so that the information regulation and low level computation over the dataset will be obtained. Author provided the study to extend the relative information system to derive the information set. Jibing Gong[12] defined a approach for price trend analysis so that the logistic prediction based regression model is presented by the author. Author generated the feature index to express the dataset features and represent it under trading information vector. Author provided the predictive study on significant information processing for time sensitive information. Author also provided the consideration to financial dataset.

3. RESEARCH METHODOLOGY

In this present work, the time series information processing is done on weather forecasting dataset to extract the relevant and effective information. The work is here defined as the improved clustering mechanism presented as the three stage model. Each stage of this model is defined with the specification of relative rules and the constraints. This model is shown in fig 1.
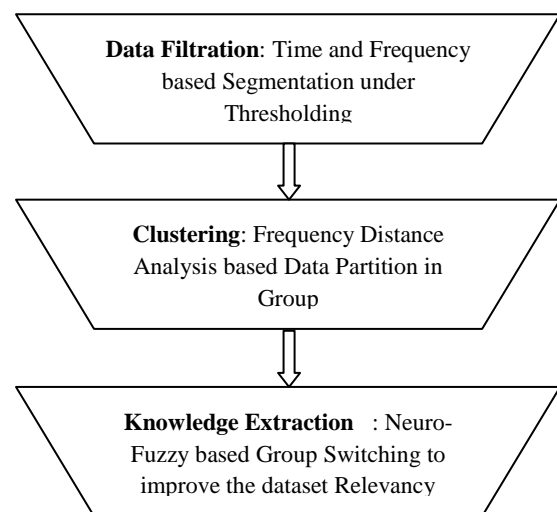
As shown in the figure, the work is divided in three main layers or the stages. In first stage of this work, the information processing is done based on the statistical analysis. The year based data division is here done to obtain the instance frequency. Once the statistics collected, the threshold value is specify to remove the year values having lesser number of instances. After this stage, the data years that cannot provide the effective information are eliminated. As the dataset is having more relevant information so that more accurate results are expected on this filtered dataset. In second stage, the attribute level analysis is applied on the dataset to generate the data groups. The grouping is here defined based on the data value and the specification of number of required groups. The distance analysis is performed to decide the group member. Based on this stage, the high level segmentation is done and the data values itself are adjusted in specific group. The grouping or clustering is here applied on all attributes independently. After formation of these groups, the attribute selective data classification is done. Now to improve the class relevancy, the neuro-fuzzy approach is applied. This approach accepts the dataset values as the neurons and assigns the weights under the fuzzy rules. All the parameters related to time series data are considered individually and collectively. The data membership is here analyzed with the specification of rules. If the membership is appropriate, no more data processing is done otherwise, the data switching among the

classes is done. The process is repeated till all the classes are not properly identified for all data members. The algorithm used to perform the data clustering is given here,

Algorithm(Dataset,N)
/*Here Dataset is the weather forecast data defined to store the weather details, N is the number of records present in the dataset*/
{

1. Set Number of Clusters initially called C
2. For j=1 to N
   [Process all weather transactions]
   {
3. For i=1 to C
   [Process for All Clusters]
   {
4. Rec=GetData(Dataset,j)
   [Obtain the dataset record from weather forecast table]
5. Dist(i)=GetDistance(Rec,Cluster(i))
   [Obtain the Disease Measure from the cluster]
   }
6. C1=Min(Dist)
   [Obtain the Minimum Distance cluster from the list]
7. Set Member(C1).Add(Dataset(j))
   [Transfer the value in particular Dataset]
8. For i=1 to C
   [process All Clusters]
   {
9. SDist=SDist+GetDist(Cluster(i),Dataset(j))
   [Obtain the Aggregative Distance of all elements from center]
10. ADist=SDist/C
    [Obtian the Average Distance
11. Update(Clusters, ADist)
    [Update ClusterPosition based on distance measures]
12. If (Updated Distace<>ActualDistance)
    [if switching of cluster is performed]
    {
13. Repeat Process from Step 2
14. Return Clusters
}

At the final stage of this work, the Neuro-Fuzzy system is applied to perform to control the prediction process and to generate the effective data segmentation on weather dataset. In this work, an estimation to the clustered values is been performed using neuro fuzzy based modeling. This model is here applied to perform the data classification based on self-analysis and will divide the dataset in set of classes so that effective data processing will be done. The stages for the work are given here under

- The clustered data is taken as the input to the neuro fuzzy process
- Analyze the dataset and perform the selection of most effective attribute set that can participate for result
- Analyze the weights for the data values based on the frequency analysis and the error data values
- Train the dataset with the specification self class
- Identify the class members based on intelligent data processing.
- The evaluation of estimated class is done under performance parameter

4. RESULTS

The presented work is here experimented on weather forecasting dataset. The data is collected from web source. The available dataset is defined with 7 attributes and 10000 reading of weather with features defined in Table 1

Table 1. Dataset Features

Attribute	Datatype
City	String
Temperature	Number
Humidity	Number
Rain	Number
Year	Number
Origin	Number

The dataset is here processed in three stage model to perform the data filtration and then data class generation using improved segmentation approach. After this model implementation, the analysis on the generated classes is done under different vectors. The results obtained from the work in terms of group analysis are given in fig 2.
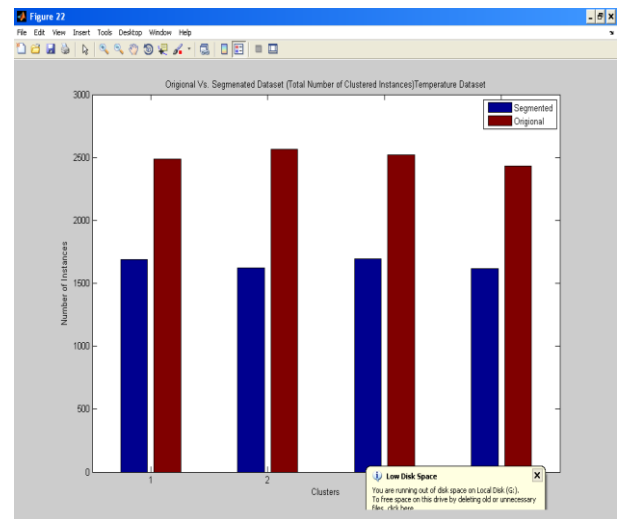

Fig 2: Class Instance Analysis (Temp.)

Here fig 2 is showing the number of istnace members obtained after model implementation. Here the analsyis is defined specific for temperature attribute.
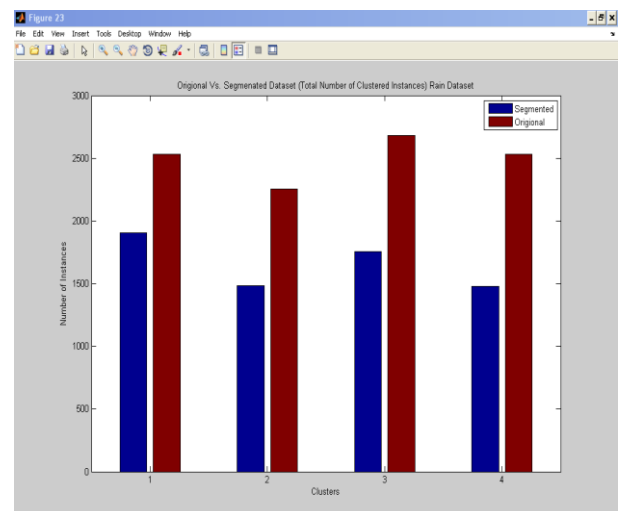

Fig 3: Class Instance Analysis (Rain)

Here fig 3 is showing the number of istnace members obtained after model implementation. Here the analsyis is defined specific for rain attribute.
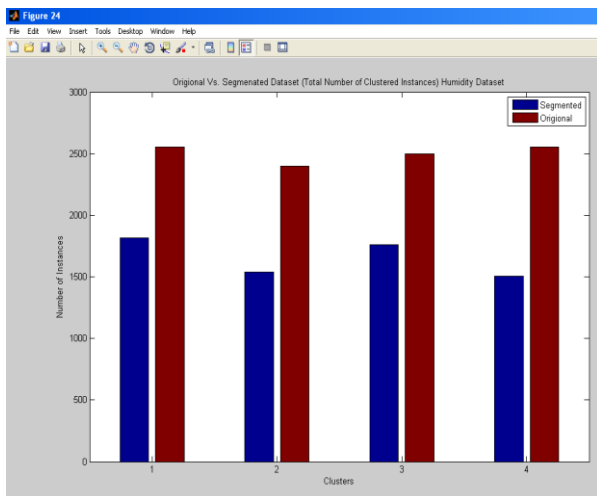

Fig 4: Class Instance Analysis(Humidity)

Here fig 4 is showing the number of istnace members obtained after model implementation. Here the analsyis is defined specific for humidity attribute.

The results show the clear class level division obtained over the dataset using neuron modeling.

5. CONCLUSION

In this present work, a three stage model is defined to perform class division on real time weather forecasting dataset. The work is here defined to improve the clustering process by including the pre and post stage. In pre stage, the time series segmentation is applied to filter the dataset. In post stage, the neuro-fuzzy is applied to optimize the class results. The results show the clear class division under instance distance.

6. REFERENCES

[1] Bragin, S. I., et al. "Applications of Passive Microwave Devices for Imaging System Designed for Detecting Objects in Soils." Microwave Conference, 31st European. IEEE, 2001.

[2] Daimon, Makoto, et al. "Association of the ABCA1 gene polymorphisms with type 2 DM in a Japanese population." Biochemical and biophysical research communications 329.1 (2005): 205-210.

[3] Vigier, B.J.; Pattey, E.; Strachan, I.B., "Narrowband vegetation indexes and detection of disease damage in soybeans," Geoscience and Remote Sensing Letters, IEEE , vol.1, no.4, pp.255,259, Oct. 2004

[4] Srivastava, A.; Shugen Ma; Inoue, K., "Development of a sensor for automatic detection of downey mildew disease," Intelligent Mechatronics and Automation, 2004. Proceedings. 2004 International Conference on , vol., no., pp.562,567, 26-31 Aug. 2004

[5] Ting Li; Jingfeng Yang; Xiaoqin Peng; Zhimin Chen; Chengyang Luo, "Prediction and Early Warning Method for Flea Beetle Based on Semi-supervised Learning Algorithm," Natural Computation, 2008. ICNC '08. Fourth International Conference on , vol.4, no., pp.217,221, 18-20 Oct. 2008

[6] Zalizah Awang Long," Parameter Setting Procedure via Quick Parameter Evaluation in Frequent Pattern Mining for Outbreak Detection", 2009 2nd Conference on Data Mining and Optimization 978-1-4244-4944-6/09 ©2009 IEEE

[7] Zhang Lei, "Characteristic of Lead Content in Soil and Crop in Zhaoyuan Gold Mine Area," Biomedical Engineering and Biotechnology (iCBEB), 2012 International Conference on , vol., no., pp.1251,1254, 28-30 May 2012

[8] Tripathy, A.K.; Adinarayana, J.; Merchant, S.N.; Desai, U.B.; Ninomiya, S.; Hirafuji, M.; Kiura, T., "Data mining and wireless sensor network for groundnut pest/disease precision protection," Parallel Computing Technologies (PARCOMPTECH), 2013 National Conference on , vol., no., pp.1,8, 21-23 Feb. 2013

[9] Petersen, E.L.; Pickel, J.C.; Adams, J.H.; Smith, E.C., "Rate prediction for single event effects-a critique," Nuclear Science, IEEE Transactions on , vol.39, no.6, pp.1577,1599, Dec 1992

[10] Amir F. Atiya," Packet Loss Rate Prediction Using the Sparse Basis Prediction Model", IEEE Transactions On Neural Networks 1045-9227© 2007 IEEE

[11] Lean Yu; Shouyang Wang; Kin Keung Lai, "A generalized Intelligent-agent-based fuzzy group forecasting model for oil price prediction," Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on , vol., no., pp.489,493, 12-15 Oct. 2008

[12] Jibing Gong, "A New Approach of Stock Price Trend Prediction Based on Logistic Regression Model", 2009 International Conference on New Trends in Information and Service Science 978-0-7695-3687-3/09© 2009 IEEE