Predicting Electricity Consumption Using Hidden Parameters

Lata Chakor[1], Lata Musne[2], Priyanka Kumbhar[3], S. M. Shedole[4]

[1,2,3]*Department of Computer Engineering, SIT'S, Lonavala-410401, India*
[4]*Assistant Professor, Department of Computer Engineering, SIT'S, Lonavala-410401, India*

Abstract: **-data mining technique to forecast power demand of a biological region based on the metrological conditions. The value forecast analytical data mining technique is implement with the Hidden Marko Model. The morals of the factor such as heat, clamminess and municipal celebration on which influence operation depends and the everyday utilization morals compose the data. Data mining operation are perform on this chronological data to form a forecast model which is able of predict every day utilization provide the meteorological parameter. The steps of information detection of data process are implemented. The data is preprocessed and fed to HMM for guidance it. The educated HMM network is used to predict the electricity demand for the given meteorological conditions.**

Keywords: **Data Mining, Predictive modeling, Artificial Neural Networks, KDD, etc.**

I. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyze data from dissimilar perspective and abbreviation it into helpful in sequence - in sequence that can be old to add to profits, cut expenses, or both. Data mining software is one of a digit of systematic apparatus for analyze data. Some of the data mining technique is analytical model, cluster, and association psychoanalysis and difference discovery. Analytical model is a technique to forecast from monitor. Analytical model is a process that uses data mining and outlook to expect outcome. Each model is made up of a number of predictors, which are variables that are probable to power viewpoint results. Data preprocessing is a data mining technique that involve transform underdone data into an logical format. Real-world data is frequently partial, conflicting, and/or absent in confident behaviors or trend, and is likely to contain many errors. Data preprocessing is a established process of resolve such issues. Data preprocessing process is divided into different categories Data cleaning, Data transformation, Data reduction. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Fuzzy c-mean clustering algorithm. This algorithm works by transitory on add-on to each data point consequent to each cluster middle on the basis of a space between the cluster middle and the data position. More the data is close to to the cluster middle more is its relationship towards the exacting cluster middle. obviously, the outline of relationship of each data point should be alike to one. Later than each iteration relationship and cluster center are efficient. Hidden Markov model (HMM) is a statistical Markov model in which the system being model is unspecified to be a Markov process with unseen (hidden) state. An HMM can be obtainable as the simplest energetic Bayesian network.Markov models (like a Markov chain), the location is straight visible to the viewer, and so the place change accidental are the one control. In a hidden Markov model, the location is not traditional obvious, but the production, needy on the state, is noticeable. Every formal has a prospect dissemination done the thinkable production symbols. Thus, the arrangement of symbols created by an HMM stretches certain material about the arrangement of positions. a hidden Markov process can be pictured as a generalization of the tricky with additional (where each item from the urn is repaid to the unique urn before the next step).Redirect this example: in a room that is not noticeable to an viewer near is a sprite. The room holds urns X1, X2, X3, …each of which holds a acknowledged mixture of balls, each ball characterized y1, y2, y3, … The sprite selects an urn in that room and casually attractions a ball after that urn. It then sets the ball on a conveyor tie, where the viewer can note the arrangement of the balls but not the classification of urns after which they stood strained. The apparition has some process to select urns; the special of the urn for the *n*-th ball be contingent only upon accidental amount and the optimal of the urn for the $(n-1)$-th ball. The optimal of urn does not straight be contingent on the urns select before this lone prior urn; then, this is called a Markov process. It can be labelled by the greater part of Figure. The Markov process themself cannot be practical; only the arrangement of considered balls, thus this plan is called a "hidden Markov process". This is showed by the lesser slice of the drawing shown in Figure 1, where one can get that balls y1, y2, y3, y4 can be pinched in each formal.

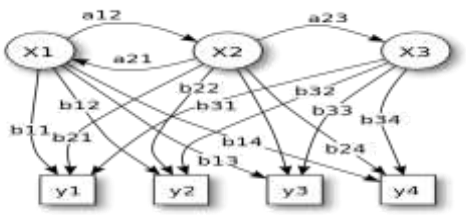


Figure1. Probabilistic factors of a hidden Markov model

X —conditions
y —probable observations
a — state conversion probabilities
b — output prospects

Uniform if the viewer identifies the structure of the urns and has just perceived a arrangement of three balls, *e.g.* y1, y2 and y3 on the express or belt, the viewer still cannot be definite which urn (*i.e.*, at which state) the spirit has strained the third ball from. Though, the viewers can effort out additional substantial, such as the lookout that the third ball come from each of the urns. Fuzzy classification is the process of collection origins into a fuzzy set whose associative purpose is distinct by the fact value of a fuzzy propositional purpose.

II. LITERATURE SURVEY

There are many steps complicated to prediction energy consumption like Fuzzy c-means clustering, Hidden Marko Model (HMM), Fuzzy classification. The major phase presented to forecast electricity feasting is Fuzzy c-means clustering offers the best result for overlain data set and relatively better than k-means algorithm.

Let X = {x1, x2, x3 ..., xn} be the set of data points and V = {v1, v2, v3 ..., vc} be the set of midpoints.

1) Randomly select 'c' cluster centers.

2) calculate the fuzzy membership 'µij' using:

3) Compute the fuzzy centers 'vj' using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

1) Repeat step 2) until the minimum 'J' value is achieved or ||U(k+1) - U(k)|| < β.

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, ...c$$

Where,

'K' is the iteration step.

'β' is the termination criterion between [0, 1].

'U = (µij) n*c' is the fuzzy membership matrix.

'J' is the objective function.

Different k-means wherever data point must completely fit to one cluster midpoint here data point is allocated association to each cluster midpoint as a result of which data point may belong to extra than one cluster midpoint.

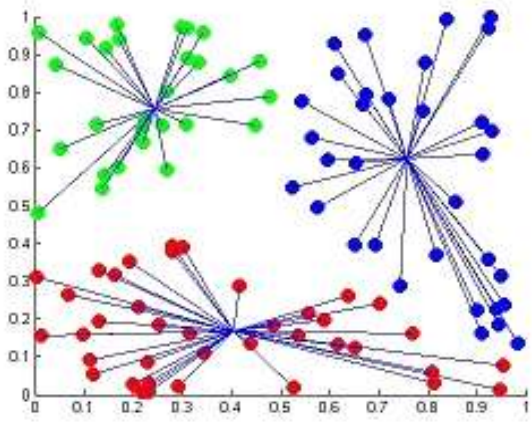


Fig I: Result of Fuzzy c-means clustering

Hidden Markov Model (HMM) is clarified as a arithmetic Markov model in which the system being demonstrated is expected to be a Markov process with overlooked (hidden) states. Poisson hidden Markov models (PHMM) are superior belongings of hidden Markov models anywhere a Poisson procedure has a amount which differs in suggestion with variations among the dissimilar conditions of a Markov model. PHMMs are not essentially Markova processes themselves because the original Markov chain or Markov process cannot be experiential and only the Poisson signal is experiential. Hidden Markov model is used in cryptanalysis, speech recognition, speech synthesis etc.

Fuzzy classification labeled as a is the process of group essentials into a fuzzy set whose connection function is distinct by the truth value of a fuzzy propositional function. Later the opening of the history of Fuzzy Sets Theory, Classification Models and Control Theory has been two essential grounds for their academic and applied progresses. In fact, countless difficulties indoors both arenas are logically formal by presenting fuzzy concepts. In some cases, a fuzzy method looks to offer a useful generalization of a too composite certainty. In a fuzzy classification process, one of the main aims is to combine the information we have obtained about each object. If such information is given by means of degrees of association, we can find in the literature that improver rules and some t-conforms are useful for starting conjunctive instructions. Material is in this way combined into one single index. Later we frequently do not know in advance the number of pieces of information we should collective, traditional methods undertake the actuality of a basic binary operative existence associative, in such a way that a consecutive application of such an operative will give us the combined information, no difficulty the measurement of information. But we know that not each operative is associative.

III. PROPOSED METHODOLOGY

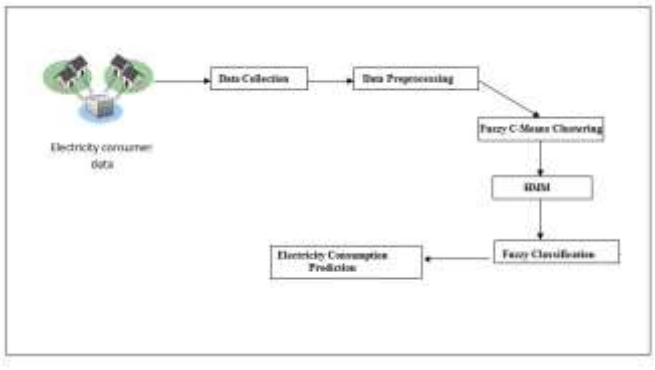


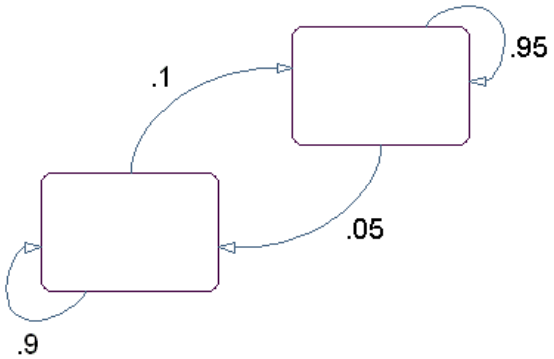
Figure 1: System Overview

Proposed System Implements Five Phase Algorithmic Procedure to forecast Larceny in Energy Practice founded on Design of electricity usage.

Phase 1: Data Reading and Pre-processing: Firstly whole data is been recite from dataset. Parallel process run to on Working out dataset and input Data Set. Only applicable factors are been nominated after input dataset.

Phase 2: Fuzzy C-Means Clustering: This Stage Fuzzy C means Clustering is been useful on Both Working out dataset and input data set producing clusters for coming treating. In K-means Centroid are been Prepared for clustering and based on Euclidean distance portion Data points has been clustered. Charted Data is too remained clustered.

Phase 3: HMM: A hidden Markov model (HMM) is one in which you see a arrangement of productions, but do not know the order of conditions the typical departed determined to produce the radioactivity. Estimates of hidden Markov models attempt to invention to progress the instruction of conditions since the experiential data.



HMM Working

HMM assistance in classifying Hidden States in data . In electricity consumption overlooked data of customer no is been recognized created on HMM preparation.

Phase 4: Fuzzy Classification: classification approaches agree you to decrease the dimensionality of a compound data set by group the data into a set number of classes. With outdated (crisp) classification methods, each model/position is located into one class or another. In crusty classification, class association is second, a model is a member of a class or not. Crusty class association ideals can be either "1" when that class is the best fit, or "0" (for all other classes). In fuzzy classification, a model can have association in countless dissimilar programs to dissimilar grades. Typically, the association values are inhibited so that all of the membership values for a specific model amount to 1.

Phase 5: Consumption Analysis: previous stage in Consumption analysis which recognizes consumption units and recognizes consumption values of user with higher values.

1. Forward Algorithm

// Input : Data Set **D**,

Detected Conditions $\mathbf{O_s}$ = { O_{s1} , O_{s2},O_{s3}}

Step 0: **Start**

Step 1: Recognize the Detected state Characteristic $\mathbf{O_{si}}$

Step 2: **FOR** i=0 to size of **D**

Step 3: Recognize Characteristic $\mathbf{O_{si}}$ and put in discrete Lists,

Step 4: FINISH FOR

Step 5: FOR i=0 to size of $\mathbf{O_{SL}}$

Step 6: recognize α using equation 1

Step 7: **FINISH FOR**

Step 8: **Break**

Equation1:

$$\alpha = O\ (t+1) \sum_{j=1}^{N} \text{Osi(O1)}$$

Where α = Forward Probability

O(t+1) = For every next element of Observed state

N= Size of the Observed state List

Osi(O1) = For Every observed state

2.Backward Algorithm

// Input : Data Set **D**

Detected Conditions $\mathbf{O_s}$ = { O_{s1} , O_{s2},O_{s3}}

Step 0: **Start**

Step 1: Recognize the Detected state Characteristic $\mathbf{O_{si}}$

Step 2: **FOR** i=0 to size of **D**

Step 3: Recognize Characteristic $\mathbf{O_{si}}$ and put in discrete list'$_s$

Step 4: FINISH FOR

Step 5: FOR i=0 to size of $\mathbf{O_{SL}}$

Step 6: recognize β using equation2

Step 7: **FINISH FOR**

Step 8: **Stop**

Equation 2 :

$$\beta = \lambda\, O\,(t+1) \sum_{j=1}^{N} \mathrm{Osi}(\mathrm{O1}+1)$$

Where α = Forward Probability

O(t+1) = For every next element of Observed state

N= Size of the Observed state List

Osi(O1) = For Every observed state

λ = Frequency of the occurred state

---

4 Fuzzy c-means clustering Algorithm

Let X = {x_1, x_2, x_3 ..., x_n} be the set of data points and V = {v_1, v_2, v_3 ..., v_c} be the set of centers.

1) Randomly select *'c'* cluster centers.

2) calculate the fuzzy membership $'\mu_{ij}'$ using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)} \quad \textbf{3)}$$

compute the fuzzy centers $'v_j'$ using:

4) Repeat step 2) and 3) until the minimum 'J' value is achieved or $\|U^{(k+1)} - U^{(k)}\| < \beta$.

Where,

'k' is the iteration step.
'β' is the termination criterion between [0,1]
'U = $(\mu_{ij})_{n*c}$' is the fuzzy membership matrix.
'J' is the objective function.

IV. RESULTS AND DISCUSSIONS

Proposed system of Theft Detection System is organized as a individual system using Net beans as development IDE for java technology which is reinforced by java media file structure. Presentation is evaluated constructed on the accuracy and recollection constraints. Accuracy is defined as the ratio of number of related Consumption to the entire amount of related and unrelated related Consumption identified. Relation efficiency of the system is well articulated by using precision parameters. While the recollection can be definite as the ratio of number of related Consumption identification the total figure of related Consumption are noticed not perceived. Complete exactness of the system is fine recounted by using recollection constraints.

System can be estimated using accuracy and recollection constraints, and they can be more openly explained as follows

• X = The numbers of related Consumption been are noticed,

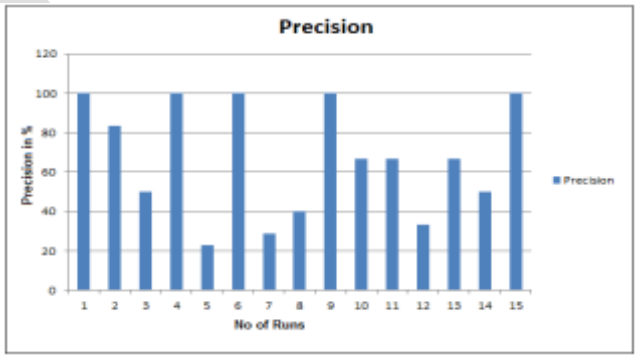
• Y = the number of related Consumption not noticed, and

• Z = the number of unrelated Consumption that are been noticed.

So, Accuracy = (X/ (X+ Z))*100

And Recollection = (X/ (X+ Y))*100

Fig.2. Average precision for Consumption Detection System



In Fig. 2, by detecting it is clear that the average precision gained for Consumption Detection System using HMM is almost 68%.

Figure3. Shows that the system gives 89.33% recollection for the Consumption method using HMM Fuzzy logic Dissemination. By relating these two graphs we can complete that the system with proposed methodology has higher performance.

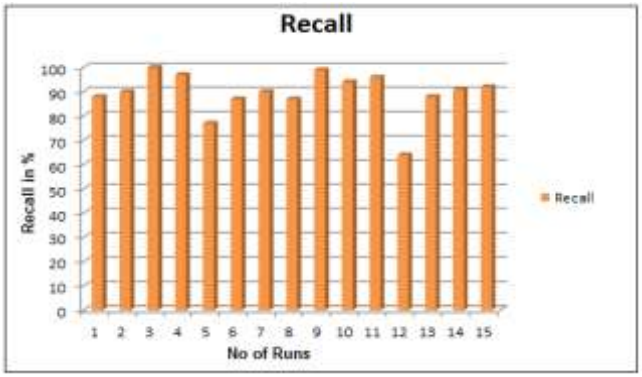


Fig.3. Average Recall for Consumption Detection System

V. CONCLUSION AND FUTURE SCOPE

Presented Artifact is Implementation Work on electricity Consumption Analysis Artifact presents invention methods in data mining domain. Presents analytical System which would contribution electricity boards to possible perceive Higher value consumption meters. This Exploration work is decently predictive and hence involves sureness and provision scrutiny .upcoming system can be improved and taken to higher level as product. Computational scrutiny and everyday values have been initiate to be accurate analytical and hence system has large scope of exploration.

REFERENCES

[1]. Jiacong Li, Ying Wang, Jinke Yu, Shaoyong Guo," An HMM-based Performance Diagnosis Approach for Hadoop Clusters" The 18th Asia-Pacific Network Operations and Management Symposium (APNOMS) 2016.

[2]. Md. Khaliluzzaman and Lamia Iqbal Dolon,"Analyzing MRI Segmentation Based on Wavelet and BEMD using Fuzzy C-Means Clustering" ©2016 IEEE.

[3]. Daniele Soria, Jonathan M. Garibaldi," Validation of a quantifier-based fuzzy classification system for breast cancer patients on external independent cohorts", © 2016 IEEE DOI 10.1109/ICMLA.2016.176.

[4]. DimitriosKoutsoukos*, GeorgiosAlexandridis t, GeorgiosSiolas+, Andreas Stafylopatis," A new approach to session identification by applying fuzzy c-means clustering on web logs", ©2016 IEEE.