A Fuzzy Interpolation Approach of Prediction using Market Basket Analysis Algorithm in Weka Tool

Reshma Badhala
M.Tech IT
RCEW JAIPUR

Subhash Chandra Jat
Assistant Professor
RCEW JAIPUR

ABSTRACT
The advance deployment of I.T. has bring about a great innovation .technologies have been innovated at exponential rate and subsequently the bulk amount of data and information have been processed at high rate. There is a great need to manage those bulky data in more efficient manner in order to reach at some conclusion or making the data meaningful. In field of business, technologies and science many methodologies have been proposed and many numerical methods conventionally developed for implementing association rules however they were very useful for client in context of prediction and classifications.

Our proposed work altered a conventional calculation for adding supermarket information utilizing fuzzy information with association of fuzzy association rules. The contemplated approach will handle and manipulate the data in more efficient manner in order to reduce complexity and improve the dynamic approach of data dredging activities.

The proposed work focuses on time series information which is a bulk and dynamic information with some pattern and connoted information associated with that .It might be super market information's, climate predication information and other fuzzy information. so basically our work is primarily based on prediction how bulk of interrelated information and provide a proposed solution for précising super market information manipulations. In order to implement the proposed work we have made a data base having supermarket store information and with the assistance of conventional apriori algorithm approach with the applications of fuzzy methods. We have extended the conventional approach of association rules in order to depict the data more accurately comparatively. The proposed work basically proposed for dynamic and bulky data with some assumptions in order to reach at explicit results.

Keywords
WEKA, Fuzzy, fuzzy association rule, Information Technology.

1. INTRODUCTION
As Information technology (IT) is increasing at a quick rate, its ability to store, manage and manipulate knowledge in databases is changing into vital. Though I.T. development created processing easier, generating implicit data to support deciding has become a rigid and tedious task. data processing, initial projected by Agrawal et al., has become a core field of analysis and study within the info and soft computing domain. It develops relationships among things in such the simplest way that the presence of a selected item in an exceedingly dealings affects the presence of different item. within the past, Agrawal projected several different mining algorithms for causing association rules from transactions supported the abstraction of enormous itemsets. They additionally projected a technique for causing association rules from knowledge sets with the help of quantitative and categorical attributes. Their projected technique initial generates the quantity of partitions for every quantitative attribute, then maps all attainable values of every attribute onto a collection of consecutive integers. Fuzzy ideas have a good impact on knowledge dredging methodology. Currently, the fuzzy pure mathematics has wide been employed in intelligent systems because of its simplicity and similarity to human reasoning. many fuzzy learning algorithms for generating rules from a given set of knowledge are developed and employed in numerous domains.

1.1 Association Rule
For a given dealings info T, an association rule is an expression of the for $X \rightarrow Y$, Where X and Y are subsets of A and $X \rightarrow Y$, holds confidently c, if computer programme of transactions in T that support X additionally support Y, The rule $X \rightarrow Y$ has support s, if s% of transaction in T that support $X \cup Y$.

2. LITERATURE SURVEY
As we all know fuzzy pure mathematics was 1st introduced by Zadeh in 1965.Fuzzy pure mathematics is primarily connected with quantifying and reasoning victimization linguistic communication during which every word has ambiguous meanings. It are often delineated because the extension of ancient crisp sets, during which components should either be in or out the set. It handles the unfinished data by employing a systematic calculus that deals lingually and performs numerical computation victimisation membership functions. Fuzzy logical thinking system plays an important role in fuzzy system. It selects many numbers of fuzzy if then rules from multiple rules that ar accustomed effectively model the intelligent system in an exceedingly specific manner. the sole downside of fuzzy system is that it can't adapt itself in ever-changing external surroundings.

2.1 Crisp set
Crisp set (subset) C of reference set R reflects clearly that either part of crisp set belongs to reference set or it doesn't belongs to reference set. Suppose we've the set . For the crisp set we have a tendency to might write this in terms of a operate C that takes one to one, 2 to 1, 3 to 0, 4 to 1,s and five to zero, or we will write .

2.2 Fuzzy set
For a down like set F of a reference set X the elements of F might have an area with F to a degree within the middle of zero and one (and to boot might have an area with F to degree zero or 1). we will compose this by allotting a capability M that takes each individual from X to variety within the interim of real numbers from zero to one, [0, 1] to talk to its level of enrollment. Here "bigger" numbers speak to a a lot of

noteworthy level of enrollment within the down like set F. to Illustrate, for the reference set we have a tendency to might have a capability M that takes one to .4, 2 to 1, 3 to .6, 4 to .2, and 5 to 0, or , with three having a a lot of distinguished level of enrollment in F than four will, since .6>.2

2.3 Membership Function

For a given crisp set B, this function assigns a value μB(x) to every x € X such that

μB (x) = {1 iff x € B

μB (x) = {0 iffs x does not € B

This sort of capacity can be spoken to such that the qualities appointed to the components of the all inclusive set exist in indicated ranges, communicated as the enrollment evaluations of these components in the set. Bigger esteems speak to higher degrees of set enrollment. Such capacities are called as the participation work μB(x), by which a fluffy set B is typically characterized. This capacity is spoken to by

μB: x-[0, 1]

Where [0, 1] denotes the interval of real numbers from 0 to 1.

B is usually expressed as: - B= μ1 /x1 + μ2 /x2 +…+ μn/xn.

- **Classification**

As we know some objects share similar properties, so based on their properties, it is checked with the existing objects property. If it matches it is added into its group, otherwise neglected. This process is called classification. Data classification consists of two-steps, one is learning step and another is classification step. The first step consists of a classifier in which existing object property is stored. In next step, the algorithm scrutinizes the object with the existing object properties and classifies its category. A tuple Y is depicted by a x-dimensional attribute vector, Y= (y1, y2 ,… , yn.). As we know every tuple Y belongs to a pre-specified class. Since we are already know its class table in advance, so it is also called as supervised.

Now the computation is done for forecasting the correctness of a classifier. The test set do not depend on the training tuples, it means that they do not play any role in constructing the classifier. An instance of a classification task is classifying customers based on their monthly salary. When the customers in a bank apply for a loan they are asked several questions for example salary, duration of employment, present address. These queries answers help the loan manager to decide whether the applicant should be funded or not. This acts like a classification method. Data mining techniques which help in classifications are decision trees and nearest neighbor techniques.

- **Estimation**

A A shopkeeper keeps the record of a sell of a particular item. He feels that the particular item is sold 2 pieces per day. It means that he will keep in stock at least 60 pieces in a month. The prediction of total value for a product for a specific period of time is called estimation. Regression and neural networks are the techniques which can be applied for estimation in data mining.

- **Association Extraction and Frequent pattern**

Association extraction is a task which generates the items which occur together depending upon the occurrence of an item. The market basket analysis is one of the best instances of association extraction between the item set. For instance, suppose a customer wants to a bucket at the general store, there could be possibility to purchase a mug with it? If the shopkeeper knows what items a customer wants to purchase together can help general stores with attractive advertising, storage in the shop, pricing, promotions and inventory management. Frequent pattern is a method which is related to association extraction between the items but the objective is to identify the most frequent items together over a list of transaction. An antecedent is an event which occurs always before the consequent .The pattern can be symbolized as:

Consequent ⇐ Antecedent

is up to the analyzer that how many antecedent they want to use in a sequence. The most popular and demanding data mining technique for extraction between items and frequent pattern is the CARMA algorithm.

- **Clustering**

Clustering is a method of splitting a group of data objects into subsets. Every subset builds a cluster, in such an order that cluster objects are similar to one another. As clustering breaks large item set into smaller groups, it is also called as data segmentation. It can also be applied for outlier detection in handwritten character recognition systems for image recognition. For example the two scientists evaluated the star's temperature and its brightness. They found the three different groups of stars. Each group has totally different feature from the others. These helped them to cluster the stars having similar attributes.

1. Problem statement: "Interpolation of Stock Market Data with Fuzzy Conception Using Weka Tool" 2.3.2 Objectives:

1. The proposed algorithm calculates variance and standard deviation which enables us to decide that by how much the stock prices units would fluctuate about the mean stock price over a period of time

2. It allows the investors to decide the selling and buying of products by watching the fluctuation of stock prices over a period of time

3. Prediction of risk associated with products.

3. PROBLEM FORMULATION AND PROPOSED SOLUTION

Input: A time alternation TS with n abstracts points, a account of m associates functions for abstracts points, a predefined minimum abutment beginning α, a predefined minimum aplomb beginning λ, and a sliding window admeasurement ws .

STEP 1: Calculate the mean of given time series TS with m data points.

Mean ($\bar{x}$) = 1/n$\sum_{i=1}^{m} x_i$

STEP 2: Determine the variance of m data points of time series TS

Variance (σ^2) = $\sqrt{\sum_{i=1}^{n} 1/m(x_i - \bar{x})^2}$

STEP 3: Calculate the standard deviation of L data points

Standard deviation
(σ) =$1/m\ \sum_{i=1}^{n}(x_i - \bar{x})^2$

STEP 4: Convert the time series TSD into a list of subsequences W (TSD) according to the sliding-window size ws. That is, W (TSD) = {sb | sb = (db, db+1,……, db+ws-1), b = 1 to (m−ws+1),where db is the value of the b-th data point in TS.

STEP 5: Change the k-th (k = 1 to ws) quantitative admire vbk in every arrangement sb (b = 1 to m−ws + 1) into a creamy set fbk announced to as (fbk1/Rk1 + fbk2/Rk2 +. . . + fbkn/Rkn) utilizing the accustomed acceptance capacities, breadth Rkl is the l-th creamy commune of the k-th advice point in every subsequence, m is the abundance of creamy participations, and fbkl is vbk's creamy acceptance admire in breadth Rkl. Each Rkl is accepted as a creamy thing.

STEP 6: Compute the scalar cardinality of each fuzzy item Rkl as

$$\mathrm{Count}_{\mathrm{kl}} = \sum_{b=1}^{\mathrm{m-ws}+1} f_{bk1}$$

STEP 7: Group the above obtained fuzzy items to form the candidate 1-itemsets C1.

STEP 8: Check whether the support value (=count_k1/m−ws + 1) of each Rkl in C1 is greater than or equal to the predefined minimum support threshold α. If Rkl fulfil the above condition, collect it in the set of large 1-itemsets (L1). That is:

L1 = {Rkl |count_{k1} ≥α,1≤k≤b+ws−1 and 1≤l≤m}.

STEP 9: IF L1 is not null, then perform the next step; otherwise, terminate the algorithm.

STEP 10: Set t = 1, where t is used to represent the number of fuzzy items in the current item sets to be processed.

STEP 11: Join the large t-itemsets Lt to obtain the candidate (t+ 1)-itemsets Ct+1 in the same way as in the Apriori algorithm provided that two items obtained from the same order of data points in subsequences cannot exist in an itemset in Ct+1 at the same instant.

STEP 12: Continue the accompanying substeps for each recently framed (t + 1) - itemset I with fluffy things (I1, I2, . . ., It+1) in Ct+1:

(a) Compute the fluffy estimation of I in every subsequence sb as f_I^sb=f_I1^sb^f_I2^sb^f_I3^sb^… ..^f_(It+1)^sb

is the participation estimation of fluffy thing Ik in Sb. On the off chance that the base administrator is utilized for the convergence, at that point:

$$f_I^{sb} = \mathrm{Min}\ {}_{k=1}^{t+1} f_I^{sb}$$

If the support (=count_I/n−ws + 1) of I is greater than or equal to the predefined minimum support threshold α, put it in Lt+1.

STEP 13: If Lt+1 are null, then do the next step; otherwise, set t= t + 1 and repeat STEPs 11–

STEP 14: Generate the association rules for each large h-itemset I with items (I1, I2, . . .,Ih),

h≥2, using the following sub steps:

a) Form each possible association rule as follows:

I_1^…^I_{n-1}^I_{n+1} ^…^I_h →I_n, n = 1 to h.

(b) Calculate the confidence values of all association rules by the following formula:

$$\sum_{b=1}^{n-ws+1} f_I^{sb} \,/\, \sum_{b}^{\mathrm{n-ws}+1} (f_I^{sb} \wedge .... \wedge f_I^{sp})$$

STEP 15: (a) Output the fluffy affiliation rules with certainty esteems more prominent than or equivalent to the predefined certainty limit λ. from time arrangement information focuses TS.

(b) Measurement of appropriation of information focuses along their mean gave the mean is picked as the middle point.

4. COMPARISON RESULTS

The proposed work has indicated correlation with the base reference work [1].

CASEI-We have plotted information pixel as per the dataset of securities exchange .A greater dataset has been expanded the proficiency of the proposed calculation. The fig below demonstrated the diagram of scattered information focuses for least certainty versus enrollment work, not at all like the past work the order is accomplished all the more precisely and powerfully nearly.

CASE II-The following diagram has demonstrated diverse substance completely in which there is less equivocalness than the past work as the Graph has indicated three distinctive classes appearing with changed shading obviously. So this powerfully dispensed class is all the more appropriately working with clear representation.

CASE III-This case concentrates on haphazardly assigned securities exchange informational index which are conflicting in nature and one can't without much of a stretch perform arrangement and predication. In the past work the dataset was too little to investigate the dynamic measurement of any operation of Algorithm.

CASE IV –The resultant of Analysis ought to be sufficiently productive to remain by some particular outcome .On a dynamic dataset of securities exchange with variable sliding window, our proposed work has connected proficiently lastly appeared in the figure. The plot has indicated precisely n district with group result on a dynamic set not at all like the vague consequence of past calculation. The quality of this proposed procedure is not so much equivocalness but rather more exactness.

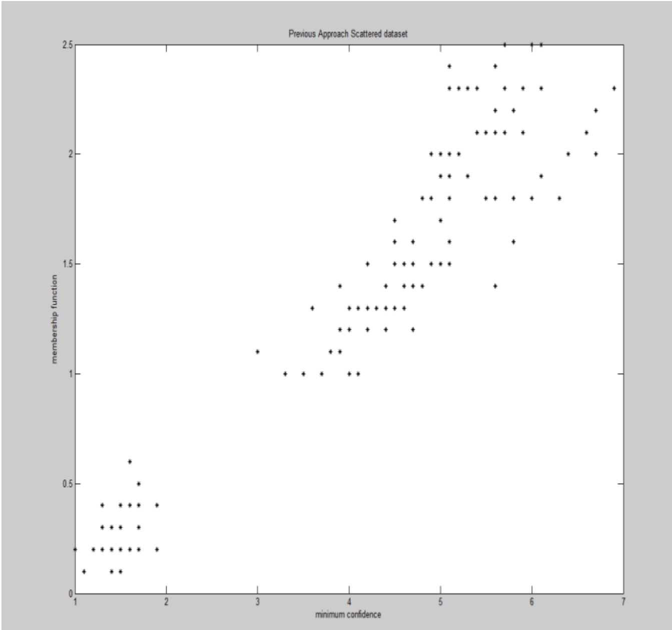

Fig 1: Scattered Data Points for Minimum Confidence vs. Membership Function

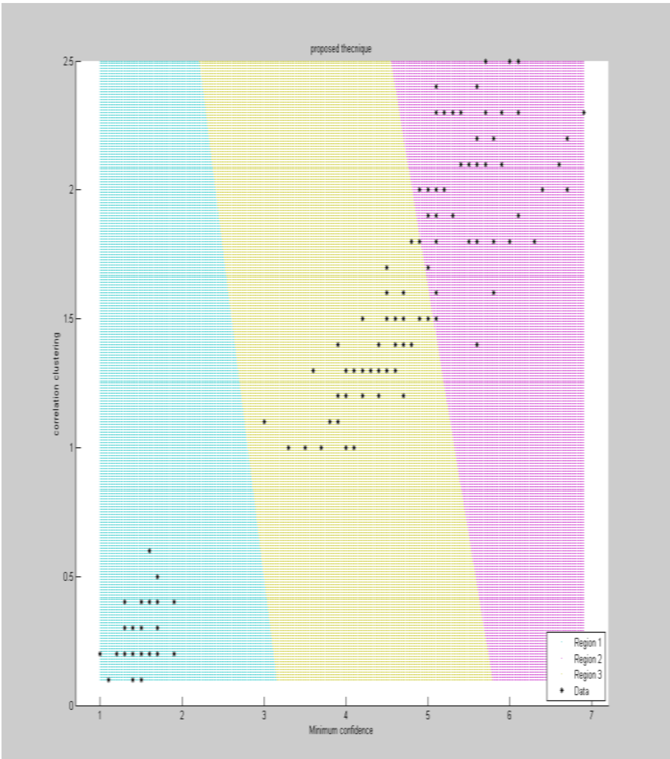

Fig 2: Graph has shown three different classes

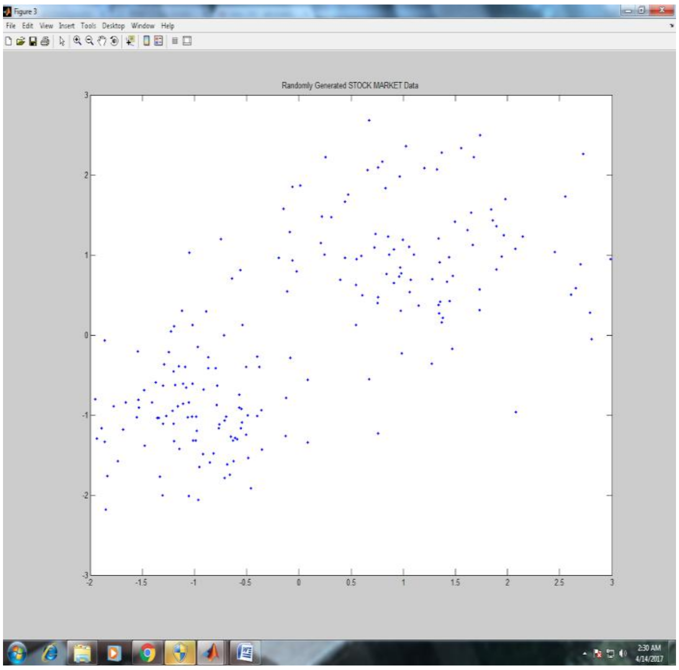

Fig 3: Randomly Allocated Stock Market Data Set

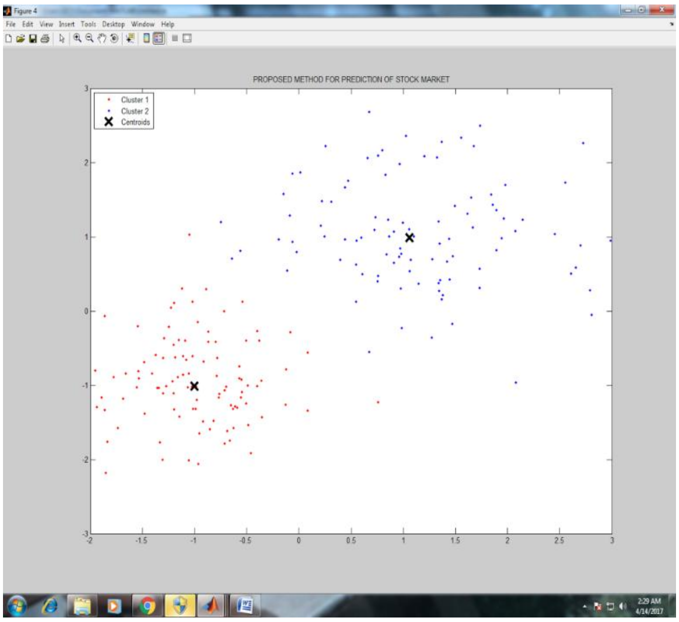

Fig 4: Dynamic Dataset of Stock Market with Variable Sliding Window

5. RESULTS AND CONCLUSION

All through the paper, it is endeavored to incite the fuzzy Association controls and diminish the superfluous fuzzy guidelines. Aside from this, the proposed calculation demonstrated the stock value scattering from the mean market cost over a timeframe which would help the financial specialist to comprehend the market variance. It would likewise clarify when of time the offering and purchasing movement should be possible. The proposed approach can expel loads of excess standards through legitimate sifting process, with the goal that clients can viably get to the principles. Future work proposes that the participation capacity can be set powerfully. In this paper, participation capacities are known ahead of time. More intricate operations could be made in not so distant future. Our work additionally measures the information scattering in time arrangement information i.e. General store information utilized here. It demonstrates the deviation of the costs from the mean of Super value information focuses assumed control over a timeframe which helps the financial specialists to choose whether to purchase or offer their offers or items. Chance related with the specific offer can likewise be anticipated by comprehension the acquired bend in the trial. We have actualized the considered work in WEKA apparatus utilizing Market Basket Analysis calculation especially keeping in mind the end goal to get more exact and proficient outcome alongside perception.

6. REFERENCES

[1]. Jiawei Han, Jian Pei, ―Data Mining Concepts and Techniques‖, ISBN-978-93-80931- 91 3, Elsevier 2013

[2]. Abdullah Al Mueen, ―Exact Primitives for Time Series Data Mining‖ University of California riverside 2014

[3]. B.Liu, W.Hsu, YMa, ―Mining Association Rules with Multiple Minimum Supports‖, International Conference on Knowledge Discovery and Data Mining, pp-337-341 1999

[4]. R.Agrawal, R.Srikant,‖Fast Algorithm for Mining Association Rules‖, International Conference on Very large Databases, pp -487-499, 1994

[5]. R.Agrawal, R.Srikant, ―Mining Sequential Patterns‖, The 11th International Conference On Data Engineering, pp-3-14, 1995

[6]. J-S.R.Jang, C-T.Sun, ―Neuro-Fuzzy and Soft Computing‖, ISBN-978-81-203-2243-1, PHI, 2011

[7]. T P Hong, K Y Lin and S L Wang, ―Fuzzy Data Mining for Interesting Generalized Association Rules‖, Fuzzy Sets & Symbols, Elsevier pp-255-269 2002

[8]. Cai, ―Mining Association Rules with Weighted Items‖ International Database Engineering and Applications Symposium 1998

[9]. T P Hong, K Y Lin, ―Induction of Fuzzy Rules and membership Functions from Training Examples‖, Fuzzy Sets and Systems, Elsevier pp-33-47, 1996

[10]. Richard J.Povinelli, ―Time Series Data Mining: Identifying Temporal patterns for Characterization and prediction of Time series and Events‖ Doctoral Thesis, Marquette University 1999

[11]. Christopher J. Neely, ―Technical Analysis in the Foreign Exchange Market: A Layman's Guide‖ A Review, 1997

[12]. R.Srikant, R.Agrawal, ―Mining Quantitative Association Rules in Large Relational Tables‖, International Conference on Management of Data, ACM pp-1-12, 199 47

[13]. Au, Chan, ―Mining Fuzzy Rules for Time Series Classification‖ International conference on Fuzzy system 2004 G. Das, K Lin, ―Rule Discovery from Time Series,‖ in: Proceedings of the 4th International Conference on knowledge discovery and data mining pp-16-221998