Cluster Mapping with the help of New Extended MCF Algorithm and MCF Algorithm to Recommend an Ice Cream to the Diabetic Patient

Suhas M. Gaikwad
Symbiosis Institute of
Technology (SIT), Pune
Affiliated to Symbiosis
International University (SIU),
Pune

Rahul R. Joshi
Symbiosis Institute of
Technology (SIT), Pune
Affiliated to Symbiosis
International University (SIU),
Pune

Preeti Mulay, PhD
Symbiosis Institute of
Technology (SIT), Pune
Affiliated to Symbiosis
International University (SIU),
Pune

ABSTRACT

The research for suggesting an ice cream for a diabetic patient is carried out in data mining by using clustering and mapping between the data for ice cream and diabetic patients. Here, using MFCA, which is proposed and explained in this paper, does mapping of ice cream dataset with diabetic patient dataset. The results obtained from MCFA algorithm and the new extended MCF algorithm are explained and verified and it is observed that they are having the relevance.

General Terms

Pattern Recognition, Security, and Algorithms, clustering et. al.

Keywords

Modified Cluster Formation Algorithm, Extended modified cluster formation Algorithm, Ice cream attribute, Diabetic Patient, Matlab.

1. INTRODUCTION

The month of July is proclaimed as "National Ice Cream Month". Every year this month, in the United States is celebrated as Ice Cream Month. The study suggests, favorite ice cream flavors can be used to predict the personality of people [4, 8, 10] for e.g. people who loves vanilla ice cream are likely to be impulsive, easily suggestible and idealistic. Likewise, those who love chocolate ice cream are likely to be dramatic, lively and charming. The data mining is useful for identifying and predicting pattern from given data series and verification of the obtained pattern or results can be carried out with the help of new techniques [2, 3]. In this paper, the focus in on suggesting an ice cream to a diabetic patient by using data mining. Here, two different algorithms are considered and they are applied on two datasets viz., ice cream dataset and diabetic patient datasets. Here, closeness factor between two data series is found by considering clusters formation and mapping between the two datasets, which are ice cream and diabetic patient datasets, is observed [1, 6, 7].

2. METHODOLOGY INVOLVED IN MCFA ALGORITHM

For all the data sets, there are S1 and S2 be the two data series [17].

a. S1= S (i), S2=S (i+1)

b. Calculate the sum of each column
$$T \ni T(j) = S_1(j) + S_2(j)$$

c. Calculate the sum of each series, which can be used to calculate the probability ratio.
$$P = \frac{\sum_{j=1}^{n} S_1(j)}{\sum_{j=1}^{n} T(j)}$$

d. Calculate error for each series, $c(j) = \dfrac{p*T(j) - sl(j)}{\sqrt{T(j)*\,p*(1-p)}}$

e. Calculate weight of each series, $W(j) = \sqrt{T(j)}$

f. Calculate G (closeness factor) for these two series,
$$G(j) = \frac{\sum_{j=1}^{n} c(j)^2 \times w(j)}{\sum_{j=1}^{n} w(j)}$$

The stat crunch [16] for an ice cream datasets having 10000 units is taken from Pearson website. After solving the steps mentioned in the above algorithm from a to f, the range of closeness factor is obtained which is from 0.045500354 to 0.136190053. Similarly, for diabetic patient dataset of 10000 units, the range of closeness factor is from 0.01913278 to 0.434617605.

Now, cluster formation depending upon the closeness factor is considered. So, for ice cream dataset, there are only 2 clusters. These clusters are shown in table 1.

Table 1: Ice Cream Cluster's

Cluster 1	Cluster 2
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098
0.045500354	0.100169098

Also, for diabetic patient's dataset there are 5 clusters, which are shown in table 2 as given below.

Table 2: Clusters for Diabetic Patients

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
0.01913278	0.100454229	0.200681794	0.30044568	0.401195004
0.021573531	0.100454229	0.200681794	0.30044568	0.401195004
0.02283005	0.100454229	0.200681794	0.30044568	0.401195004
0.024110058	0.100454229	0.200681794	0.30044568	0.401195004
0.024110058	0.100454229	0.200681794	0.30044568	0.401195004
0.024758738	0.100454229	0.200681794	0.30044568	0.401195004
0.024758738	0.100454229	0.200681794	0.30044568	0.401195004
0.024758738	0.100454229	0.200681794	0.30044568	0.401195004

The cluster centriod is calculated from average distance between the formed clusters. The centriod for cluster 1 of ice cream datasets is 0.068670542, whereas centriod of cluster1 of diabetic patient dataset is 0.057610173. The details about centroid are mentioned in table 3.

Cluster	Centroid of Cluster for a ice cream dataset of 7020 unit	Centroid of Cluster for a diabetic patient dataset of 7020 unit
Cluster 1	0.068670542	0.057610173
Cluster 2	0.11750986	0.148106059

Thus, the mapping between ice cream and diabetic patient datasets is possible. The cluster 1 of ice cream datasets cannot map with the cluster cluster1 of diabetic patient dataset, but cluster2 of ice cream datasets can be mapped with the cluster 2 of diabetic patient dataset.

3. EXTENDED MCFA ALGORITHM

For all the data sets, there are S1 and S2 be the two data series [17]. However, extended MCFA algorithm can be used to give clusters in the ice cream and for diabetic patients. Here, sugar is an impactful attribute whose range for considered datasets is observed and plotting the on MATLAB

1 S1= S (i), S2=S (i+1)
2 Calculate the sum of each column

3 $T \ni T(j) = S_1(j) + S_2(j)$

4 Calculate the sum of each series, which can be used to calculate the probability ratio

$$P = \frac{\sum_{j=1}^{n} S_i(j)}{\sum_{j=1}^{n} T(j)}$$

4. Calculate the cluster distance

$$W(j) = \sqrt{T(j)}$$

5 expected value of cluster

$$e = p * \{\sum_{j=1}^{n} \frac{S_1(j)}{1} - (mean of \sum_{j=1}^{n} s1)\}^2$$

Extended MCFA algorithm consist of the two

The stat crunch [16] for an ice cream datasets having 10000 units is taken from Pearson website. After solving the steps mentioned in the above algorithm from 1 to 5, the range of closeness factor is obtained which is from table 4 is 0.001111111 to 107.5811689. Similarly, for diabetic patient dataset of 10000 units, the range of closeness factor is from 0.080994898 to 378.9432181.

Table 4: Ice Cream Cluster's

Cluster 1	Cluster 2
0.001111111	101.0727848
0.120689655	101.2525252
0.120967742	101.5467034
0.121121212	102.4390244
0.120422857	103.765432
0.120121612	104.2888889
0.120930212	105.1948053
0.12254902	106.8117976
0.127358491	107.5811689

Similarly table 5 for diabetic patient dataset of 10000 units, the range of closeness factor is from 0.080994898 to 378.9432181.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
0.080994898	101.0204081	156.0257353	217.0240274	320.6830773
0.091954023	102.9065861	157.1302606	227.0270272	322.7435063
0.112037037	103.0637254	159.3395316	229.9346591	330.9677415
0.12463977	104.8767968	162.7031249	230.9567009	334.9957631
0.152210884	106.1052631	170.9739265	233.488372	341.9481266
0.185840708	107.4793387	175.6958174	243.8095237	357.0056074
0.183206107	108.9192548	180.7361111	251.5706213	363.4652122
0.183206109	108.9192548	181.8206898	256.1026695	378.9432181

Table 5: Complexity table

Algorithm name	Time complexity	Clusters formed for 4020 unit data	Clusters formed for 7020 unit data	Attribute Visualization	Domain OF Data
MCFBA	2.522 sec	2 for ice cream data 4 for diabetic patient dataset	2 for ice cream data 5 for diabetic patient dataset	No Attribute Visualization	Data mining
Extended MCFBA Algorithm	0.025 sec	2 for ice cream data 4 for diabetic patient dataset	2 for ice cream data 4 for diabetic patient dataset	No Attribute Visualization	Data mining

4. CONCLUSION

The process of suggesting an ice cream to a diabetic is verified by two methods which are in turn two different algorithms are used for the same. Firstly closeness factor and then cluster formation and mapping of cluster for considered datasets is carried out. However, extended MCFA algorithm can be used to give clusters in the ice cream and for diabetic patients. Here, sugar is an impactful attribute whose range for considered datasets is observed and plotting the on MATLAB shows the same. So mapping between considered datasets that is an ice cream and a diabetic patient datasets can be possible and is achievable. This mapping is verified by the results obtained from of MCFA and extended MCFA algorithm and it is seen that they are analogous to each other.

5. RELATED WORKS AND ITS RELEVANCE WITH THE CONCEPT PROPOSED IN THIS PAPER

Analytical Hierarchy Process (AHP) can be used to verify cluster mapping in order to suggest an ice cream to the diabetic patient. Also, System dynamics modeling can be done in order to suggest an ice cream to the diabetic patient based on sugar content in the ice cream and sugar intake of the day by a diabetic patient. In this paper, by considering sugar as a main attribute for ice cream and diabetic patient dataset, it is proved mapping between these two datasets is possible. Also, by using the same, suitable ice cream can suggest to a diabetic patient.

6. REFERENCES

[1] S. Chauhan, M. Imdad, W. Sintunavarat, and Y. Shen, "Unified fixed point theorems for mappings in fuzzy metric spaces via implicit relations," *Journal of the Egyptian Mathematical Society,* 2014.

[2] T. Y. Chen and J. H. Huang, "Application of data mining in a global optimization algorithm," *Advances in Engineering Software,* vol. 66, pp. 24-33, 2013.

[3] M. J. Cracknell and A. M. Reading, "Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information," *Computers & Geosciences,* vol. 63, pp. 22-33, 2014.

[4] V. M. da Silva, V. P. R. Minim, M. A. M. Ferreira, P. H. d. P. Souza, L. E. d. S. Moraes, and L. A. Minim, "Study of the perception of consumers in relation to different ice cream concepts," *Food Quality and Preference,* vol. 36, pp. 161-168, 2014.

[5] J. P. Donate, P. Cortez, G. G. Sánchez, and A. S. de Miguel, "Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble," *Neurocomputing,* vol. 109, pp. 27-32, 2013.

[6] D. N. Flynn, "Building a Better Model: A Novel Approach for Mapping Organisational and Functional Structure," *Procedia Computer Science,* vol. 44, pp. 194-203, 2015.

[7] S. Ghosh and S. Mitra, "Clustering large data with uncertainty," *Applied Soft Computing,* vol. 13, pp. 1639-1645, 2013.

[8] M. Grossi, R. Lazzarini, M. Lanzoni, and B. Riccò, "A novel technique to control ice cream freezing by electrical characteristics analysis," *Journal of Food Engineering,* vol. 106, pp. 347-354, 2011.

[9] K. Jung, K. C. Morris, K. W. Lyons, S. Leong, and H. Cho, "Mapping Strategic Goals and Operational Performance Metrics for Smart Manufacturing Systems," *Procedia Computer Science,* vol. 44, pp. 184-193, 2015.

[10] T. Kanit, S. Forest, D. Jeulin, F. N'Guyen, and S. Singleton, "Virtual improvement of ice cream properties by computational homogenization of microstructures," *Mechanics Research Communications,* vol. 38, pp. 136-140, 2011.

[11] C. Lemke and B. Gabrys, "Meta-learning for time series forecasting and forecast combination," *Neurocomputing,* vol. 73, pp. 2006-2016, 2010.

[12] M. K. Pakhira, "Finding Number of Clusters before Finding Clusters," *Procedia Technology,* vol. 4, pp. 27-37, 2012.

[13] I. Railean, P. Lenca, S. Moga, and M. Borda, "Closeness Preference – A new interestingness measure for sequential rules mining," *Knowledge-Based Systems,* vol. 44, pp. 48-56, 2013.

[14] Preeti Mulay, Dr. Parag A. Kulkarni, "Knowledge augmentation via incremental clustering, new technology for effective knowledge management", ACM digital library, International journal of business information systems, vol. 12, issue 1, Dec 2013.

[15] Suhas Machhindra Gaikwad, Dr. Preeti Mulay, Rahul Raghvendra Joshi. "Analytical Hierarchy Process to Recommend an Ice Cream to a Diabetic Patient based on Sugar Content in it." procedia Elsevier (2015).

[16] StatCrunch – Data analysis on the Web. http://www.statcrunch.com.

[17] Prachi M. Joshi, Dr. Parag A. Kulkarni A Novel Approach for Clustering based on Pattern Analysis International Journal of Computer Applications (0975 – 8887) Volume 25– No.4, July 2011

[18] Gaikwad, Suhas Machhindra, Preeti Mulay, and Rahul Raghvendra Joshi. "Analytical Hierarchy Process to Recommend an Ice Cream to a Diabetic Patient Based on Sugar Content in it." Procedia Computer Science 50 (2015): 64-72.

[19] Gaikwad, Suhas Machhindra, Preeti Mulay, and Rahul Raghvendra Joshi. "Attribute Visualization and Cluster Mapping With The Help of New Proposed Algorithm and Modified Cluster Formation Algorithm To Recommend An Ice Cream To The Diabetic Patient Based on Sugar Contain In It." International Journal of Applied Engineering Research 10.8 (2015).