ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 5, May 2018

Analyzing the Food Quality Using K-Means Clustering

R. Sophia Porchelvi[1], B. Snekaa[2*] and M.Slochana[3]
[1]Associate Professor, [2*]Research Scholar, [3]Assistant Professor, Department of Mathematics, A.D.M College for Women (Autonomous), Nagapattinam-611001, Tamil Nadu, India.

Abstract: Clustering plays an important role in a life problem, since we live in a world full of data where we encounter a large amount of information. One of the vital means in dealing with these data is to classify or group them into a set of categories or clusters. In this paper we apply the k-Means clustering algorithm in analyzing the food quality.

Keywords: Clustering, Fuzzy clustering, k-Means clustering.

I. INTRODUCTION

Good nutrition is an important part of leading a healthy lifestyle. Healthy food also increases a person's chances of living longer. It improves one's mood and enhances his mental status. When a person's body is in stress, protein is often broken down into amino acids which aid the body to deal with stress. Protein-rich foods can help to enhance the protein level in the body hence it ensures that one is in good moods. Fuzzy clustering plays an important role in various streams and even in our day to day life like data mining, text mining, grouping the products, surveys, graph theory and etc.

K-Means clustering is used to analyze the food quality. The term "k-means" was first used by James MacQueen in 1967.Cluster analysis was originated in anthropology by Driver and Kroeber in 1932 and introduced to psychology by Zubin in 1938 and Robert Tryon in 1939.This paper is organized as follows: section II contains the concept of clustering, fuzzy clustering and k-Means clustering, section III deals with application of k-Means clustering, section IV has conclusion.

II. The Concept Of Clustering, Fuzzy Clustering And K-Means Clustering

A. Clustering

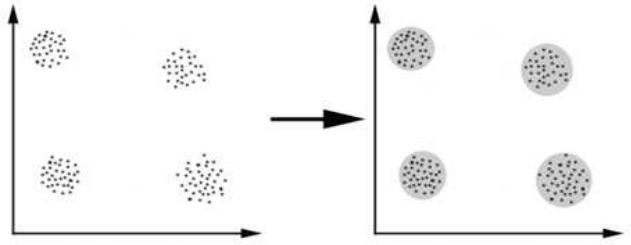
Clustering is the classification of objects into various groups, or more precisely, the partitioning of data set into clusters, so that in each cluster shares some common features, according to some defined distance measure. Clustering finds application in many fields. It is used to discover relevance knowledge in data.

B. Fuzzy Clustering

Fuzzy clustering is a form of partitioning in which every data point can belong to more than one cluster.

It is a classification of objects into various groups or more precisely, the partitioning of data sets into clusters, so that the data in each cluster share some common features, often proximity according to some defined measures.

A cluster is usually represented as grouping of similar data points around a center known as ***centroid*** or it may be defined as prototype data instance nearest to the centroid. A cluster can be represented with or without a well-defined boundary such that those clusters with well-defined boundaries are called ***crisp clusters*** whereas those without a well-defined boundary are called ***fuzzy clusters.***



Example of clustering in diagrammatical representation

C. k-Means Clustering

k-Means is one of the simple algorithm that solve the well-known clustering problem. Its procedure follows a very simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 5, May 2018

III. Application of K-Means Clustering

A. Food products grouping

Let us apply the k-Means clustering algorithm in analyzing the food quality using k-Means clustering and obtain four clusters. Here I have chosen seven food products which are not same but with some similarities to each other in the protein and fat content [15].

S.No	Nuts & Seeds	Protein content, P	Fat content, F
1	Almonds	**13**	**38**
2	Sesame seeds, Dry	**9**	**24**
3	Coconut, Shredded, Sweetened	**1**	**20**
4	Brazil nuts	**10**	**47**
5	Cashews, Unsalted	**12**	**32**
6	Pecans, Raw, Halves	**5**	**35**
7	Sunflower seeds	**11**	**26**

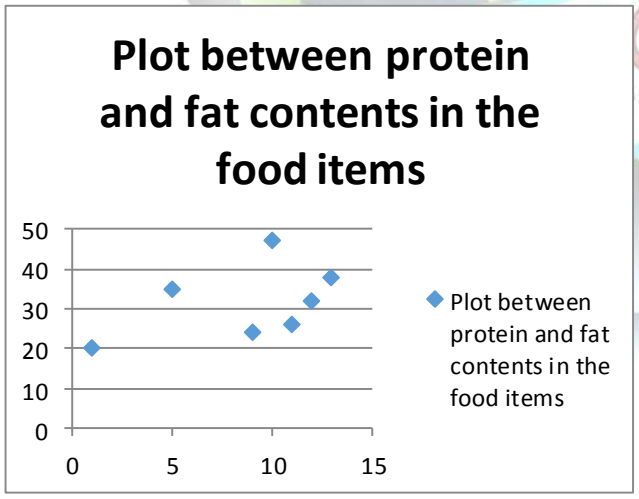


Table 1

Let us plot these points so that we can have better understanding of the problem. Also, we can select the three points which are farthest apart.

We observe that point 1 is close to point 5. So, both can be taken as one cluster. The resulting cluster is called C15 cluster. The value of P for C15 centroid is (13 + 12)/2 = 12.5 and F for C15 centroid is (38 + 32)/2 = 35.

Upon closer observation, the point 2 can be merged with the C7 cluster. The resulting cluster is called C27 cluster.

The values of P for C27 centroid is (9 + 11)/2 = 10 and F for C27 centroid is

(24 + 26)/2 =25.

The point 3 is close to point 6. They can be merged into C36 cluster.

The values of P for C36 centroid is (1 + 5)/2 = 3 and F for C36 centroid is

(20 + 35)/2 = 27.5.

The point 4 is not close to any other point. So, it is assigned to C4 with the value of P for C4 centroid as 10 and F for C4 centroid is 47.

Finally, four clusters with three centroids have been obtained.

Cluster number	Protein content, P	Fat content, F
C15	12.5	35
C27	10	25
C36	3	27.5
C4	10	47

Table 3

In the above example it was quite easy to estimate the distance between the points. In cases in which it is more difficult to estimate the distance, one has to use euclidean metric to measure the distance between two points to assign a point to a cluster.

IV. Conclusion

Here we have applied k-Means clustering technique in grouping the food products which are available in the market according to their similarities and dissimilarities based on protein and fat content in those food products, so that it helps us to decide which is suitable for our health and what not. And this method can be even applied on grouping the students who're using social networking sites and how the impact of these networking sites on the student's academic performances.

Conflict of interests: The author declared no conflict of interests.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 5, Issue 5, May 2018

REFERENCES

[1]. Bailey, Ken, "Numerical Taxonomy and Cluster Analysis". Typologies and Taxonomies. p. 34. ISBN 9780803952591,1994.

[2]. Tryon, Robert C, "Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality", Edwards Brothers, 1939.

[3]. HartiganJ. A, Wong M. A, "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C (Applied Statistics),28 (1): 100–108, 1979.

[4]. Hamerly G,Elkan C, "Alternatives to the k-means algorithm that find better clusterings" (PDF). Proceedings of the eleventh international conference on Information and knowledge management (CIKM), 2002.

[5]. Celebi M. E, Kingravi H. A, and Vela P. A, "A comparative study of efficient initialization methods for the k-means clustering algorithm", Expert Systems with Applications. 40 (1): 200–210, 2013.

[6]. P. S. Bradley and U. M. Fayyad,"Refining initial points for K-Means clustering", In Proc. 15th International Conf. on Machine Learning, pages 91–99. Morgan Kaufmann, San Francisco, CA, 1998.

[7]. T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH, "a new data clustering algorithm and its applications", volume 1, pages 141–182, 1997.

[8]. AmandeepKaur Mann, NavneetKaur Mann, "Review Paper on Clustering Techniques",Global Journal Of Computer Science And Technology Software & Data Engineering, VOL. 13, 2013.

[9]. Junatao Wang, XiaolongSu, "An Improved K-means Clustering Algorithm", Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on 27 may,2011 (pp. 4446).

[10]. Al-Daoud M. B, Venkateswarlu N. B, and Roberts S. A, "Fast K-means clustering algorithms", Report 95.18, School of Computer Studies, University of Leeds, June 1995.

[11]. Kanungo T., Mount D. M., Netanyahu N., Piatko C., Silverman R., and Wu A, "The efficient K-means clustering algorithm: analysis and implementation", IEEE Trans. Pattern Analysis Mach. Intell., 2002, 24(7), 881–892.

[12]. Bottou L, and Bengio Y, "Convergence properties of the K-means algorithm", Adv. Neural Infn Processing Systems, 1995, 7, 585–592.

[13]. Tibshirani R., Walther G., and Hastie T, "Estimating the number of clusters in a dataset via the gap statistic", Technical Report 208, Department of Statistics, Stanford University, California, 2000.

[14]. Ishioka T, "Extended K-means with an efficient estimation of the number of clusters", In Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2000), Hong Kong, PR China, December 2000, pp. 17–22.

[15]. Table of food nutrition's – Wikipedia.

BIOGRAPHY

Dr. R. SOPHIA PORCHELVI working as Associate Professor in Mathematics and Controller of Examinations for A.D.M College for Women (Autonomous), Nagapattinam, Tamil Nadu, India; and pursuing Ph.D. in Mathematics from Manonmaniam Sundaranar University (MSU), Tirunelveli, Tamil Nadu, India. Her research interest component Stochastic Processes (Queuing Models), Optimization Techniques, Mathematical Modelling, Fuzzy Mathematics, Image Processing (Stochastic models), and Mathematical Probability & Statistical Technique.

SNEKAA B received B.Sc (Mathematics) in A.D.M College for Women (Autonomous), Nagapattinam and M.Sc (Mathematics) in St.Joseph College of Arts and Science, Cuddalore, Tamil Nadu, India; she is currently doing Ph.D (Full Time) in Mathematics, A.D.M College for Women (Autonomous), Nagapattinam, Tamil Nadu, India. Her research interests in Fuzzy Mathematics (Multi - Criteria Decision Making Problems [MCDM] and Multi – Person Decision Making Problems [MPDM]) and its real life applications.

SLOCHANA M working as Assistant Professor of Mathematics, A.D.M.College for women (Autonomous), Nagapattinam, Tamil nadu,India. She is currently doing Ph.D (Part -time) A.D.M.College for women (Autonomous), Nagapattinam, Tamil nadu, India. Her research interests in Fuzzy mathematics and its applications.