International Journal of Latest Trends in Engineering and Technology
Vol.(20)Issue(2), pp.023-026
DOI: http://dx.doi.org/10.21172/1.202.04
e-ISSN:2278-621X

AN APPROACH FOR IDENTIFICATION OF 'K'-MEANS CLUSTERING METHODS

Dr. Kanaka Durga Returi[1], Dr. Vaka Murali Mohan[2]

Abstract- **An Approach for Identification of 'k'-means clustering methods have been presented in this paper. One of the simplest clustering methods is k-means, in which the number of clusters k is chosen in advance, after which the goal is to partition the inputs into sets S1, ..., Sk in a way that minimizes the total sum of squared distances from each point to the mean of its assigned cluster. There are a lot of ways to assign n points to k clusters, which means that finding an optimal clustering is a very hard problem. We'll settle for an iterative algorithm that usually finds a good clustering.**

Keywords – Clustering, "k Means', Coloring & Decoloring.

I. INTRODUCTION

Whenever you look at some source of data, it's likely that the data will somehow form*clusters*. A data set showing where millionaires live probably has clusters in places likeBeverly Hills and Manhattan. A data set showing how many hours people work each week probably has a cluster around 40 (and if it's taken from a state with laws mandating special benefits for people who work at least 20 hours a week, it probably has another cluster right around 19). A data set of demographics of registered voters likely forms a variety of clusters (e.g., "soccer moms," "bored retirees," "unemployedmillennials") that pollsters and political consultants likely consider relevant.

Unlike some of the problems we've looked at, there is generally no "correct" clustering. An alternative clustering scheme might group some of the "unemployed millenials" with "grad students," others with "parents' basement dwellers." Neither scheme isnecessarily more correct instead, each is likely more optimal with respect to its own"how good are the clusters?" metric. Furthermore, the clusters won't label themselves. You'll have to do that by looking atthe data underlying each one.

Numerous methods must exist since various years in clustering methods some of them are Sanchoa et al [1] presented k-means clustering algorithm was applied to a wide range of physicochemical properties to identify groups of crudes oils with high affinity that possibly have similar behavior later on, in downstream operations. Geng Niu et al [2] Reported the cohesive hierarchical clustering based K-means clustering method is proposed and used for clustering analysis of typical scenarios of island power supply systems. Vaka Murali Mohan., Kanaka Durga R [3] explained the Digital image pixel processing and 2d-convolution methods. Vaka Murali Mohan et al [4] presented the Image processing representation using binary image; grayscale, color image, and histogram. Hakan Cevikalp et al [5] introduced a strong technique for semi-supervised research on neural networks designed for multi-label image grouping. Rizwan Qureshi et al [6] reported the performance for pattern appreciation to predictable 3 channel images in RGB. Ji Zhang et al [7] suggested visual semantic tree successfully establish significant image

Professor & Head, Department of Computer Science and Engineering, Malla Reddy College of Engineering For Women, Maisammaguda, Medchal Hyderabad, Telangana, India
Professor & Head, Department of Computer Science and Engineering, Malla Reddy College of Engineering For Women, Maisammaguda, Medchal Hyderabad, Telangana, India

classifications and accuracy rates. Jianqiang Song et al [8] recommended a great design 'multi-layer discriminative dictionary learning (MDDL)" through section limitation in classification of an image. Douglas C.Yoon et al [9] presented the digital Radiographic Image Processing and Investigation. Lei Wang et al [10] introduced an extraordinary structure of photography in 3D motion. Hamilton, P, W [11] described "process of digital images in histopathology, cytopathology and pathology-centric research".

II. METHODOLOGY

For us, each input will be a vector in d-dimensional space (which, as usual, we will represent as a list of numbers). Our goal will be to identify clusters of similar inputs and (sometimes) to find a representative value for each cluster. For example, each input could be (a numeric vector that somehow represents) the title of a blog post, in which case the goal might be to find clusters of similar posts, perhaps in order to understand what our users are blogging about. Or imagine that we have a picture containing thousands of (red, green, blue) colors and that we need to screen-print a 10-color version of it. Clustering can help us choose 10 colors that will minimize the total "color error."

One of the simplest clustering methods is *k-means*, in which the number of clusters k is chosen in advance, after which the goal is to partition the inputs into sets S_1, ..., S_k in a way that minimizes the total sum of squared distances from each point to the mean of its assigned cluster. There are a lot of ways to assign n points to k clusters, which means that finding an optimal clustering is a very hard problem. We'll settle for an iterative algorithm that usually finds a good clustering:

1. Start with a set of *k-means*, which are points in *d*-dimensional space.
2. Assign each point to the mean to which it is closest.
3. If no point's assignment has changed, stop and keep the clusters.
4. If some point's assignment has changed, recompute the means and return tostep 2.

Using the vector mean function, it's pretty simple to create a class that does this:

To celebrate DataSciencester's growth, your VP of User Rewards wants to organize several in-person meetups for your hometown users, complete with beer, pizza, and DataSciencester t-shirts. You know the locations of all your local users (Figure 1),and she'd like you to choose meetup locations that make it convenient for everyone to attend. Depending on how you look at it, you probably see two or three clusters. (It's easy todo visually because the data is only two-dimensional. With more dimensions, itwould be a lot harder to eyeball.)

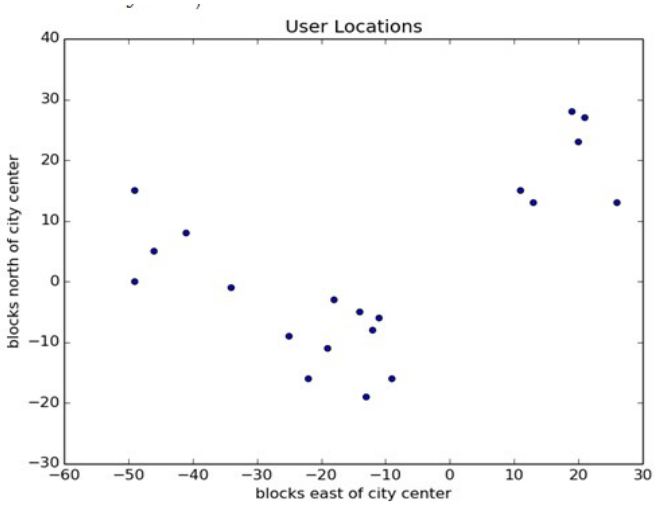

Figure 1: The locations of your hometown users

There are three clusters centered at [-44,5], [-16,-10], and [18, 20], and you look for meetup venues near those locations (Figure 2).

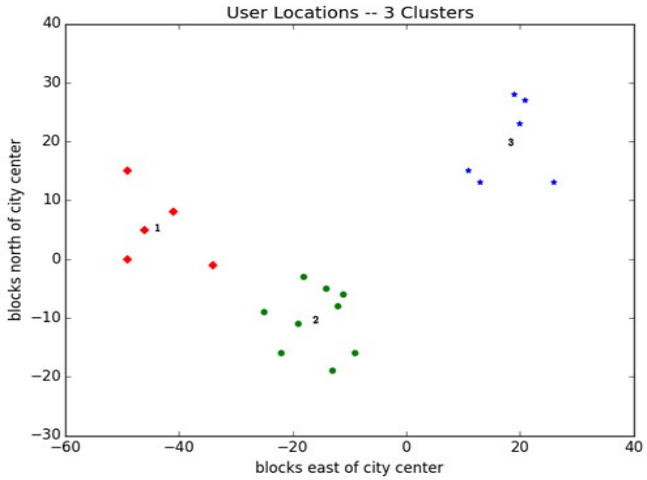

Figure 2: User locations grouped into three clusters

III. CHOOSING K

In the previous example, the choice of *k* was driven by factors outside of our control. In general, this won't be the case. There is a wide variety of ways to choose a *k*. One that's reasonably easy to understand involves plotting the sum of squared errors (between each point and the mean of its cluster) as a function of *k* and looking at where the graph "bends":

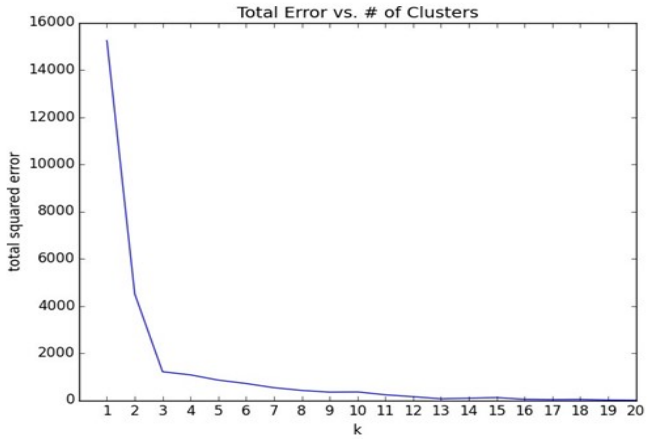

Figure 3: Choosing a k

Looking at Figure 3, this method agrees with our original eyeballing that 3 is the "right" number of clusters.

IV. CLUSTERING COLORS

The VP of Swag has designed attractive DataSciencester stickers that he'd like you to hand out at meetups. Unfortunately, your sticker printer can print at most five colors per sticker. And since the VP of Art is on sabbatical, the VP of Swag asks if there's some way you can take his design and modify it so that it only contains five colors. Computer images can be represented as two-dimensional array of pixels, where each pixel is itself a

three-dimensional vector (red, green, blue) indicating its color. Creating a five-color version of the image then entails:

1. Choosing five colors
2. Assigning one of those colors to each pixel

It turns out this is a great task for k-means clustering, which can partition the pixels into five clusters in red-green-blue space. If we then recolor the pixels in each clusterto the mean color, we're done. To start with, we'll need a way to load an image into Python. It is difficult to show color results in a black-and-white book, but Figure 4 shows grayscale versions of a full-color picture and the output of using this process toreduce it to five colors:

Figure 4: Original picture and its 5-means decoloring

V. CONCLUSION

An Approach for Identification of 'k'-means clustering methods have been presented in this paper. One of the simplest clustering methods is *k-means*, in which the number of clusters k is chosen in advance, after which the goal is to partition the inputs into sets S_1, ..., S_k in a way that minimizes. Start with a set of k-means, which are points in d-dimensional space. Assign each point to the mean to which it is closest. If no point's assignment has changed, stop and keep the clusters. If some point's assignment has changed, recompute the means and return to step 2.

REFERENCES

[1] Sanchoa., J.C.Ribeiro., M.S.Reis., F.G.Martins "Cluster analysis of crude oils with k-means based on their physicochemical properties" Computers & Chemical Engineering, Volume 157, January 2022, 107633.

[2] Geng Niu., Yu Ji., Zhihui Zhang., Wenbo Wang., Jikai Chen., Peng Yu "Clustering analysis of typical scenarios of island power supply system by using cohesive hierarchical clustering based K-Means clustering method" Energy Reports, Volume 7, Supplement 6, November 2021, Pages 250-256.

[3] Vaka Murali Mohan., Kanaka Durga R "Digital image pixel processing and 2d-convolution" Gedrag en Organisatie, ISSN: 0921 5077, Volume 33, Issue 1, January-March 2020, Pages 123-130.

[4] Vaka Murali Mohan et al "Image processing representation using binary image; grayscale, color image, and histogram (353–361), 978-81-322-2516-4, Advances in Intelligent Systems and Computing, 2016, 381.

[5] Hakan Cevikalp., Burak Benligiray., Omer Nezih Gerek "Semi-supervised robust deep neural networks for multi-label image classification" Pattern Recognition, Volume 100, April 2020, pp 107-164.

[6] Rizwan Qureshia., Muhammad Uzairc., Khurram Khurshid., Hong Yan "Hyperspectral document image processing: Applications, challenges and future prospects" Pattern Recognition, Volume 90, June 2019, Pages 12-22

[7] Ji Zhang., Kuizhi Mei., Yu Zheng., Jianping Fan "Learning multi-layer coarse-to-fine representations for large-scale image classification" Pattern Recognition, Volume 91, July 2019, Pages 175-189

[8] Jianqiang Song., Xuemei Xie., Guangming Shia., Weisheng Dong "Multi-layer discriminative dictionary learning with locality constraint for image classification" Pattern Recognition, Volume 91, 2019, Pages 135-146

[9] Douglas C.Yoon "Digital Radiographic Image Processing and Analysis" Dental Clinics of North America, Volume 62, Issue 3, 2018, Pages 341-359

[10] Lei Wang., Min Xu., Bo Liu., Shiwu Zhang "A Three-Dimensional Kinematics Analysis of a Koi Carp Pectoral Fin by Digital Image Processing" Journal of Bionic Engineering, Vol. 10, Issue 2, 2013, pp 210-221

[11] Peter W.Hamilton "How to take and process digital images for publication" Diagnostic Histopathology, Volume 16, Issue 10, October 2010, Pages 476-483.