• Cover page with the project title and the names of the group members with their student IDs

anyone

• Abstract

Sun:

E.g.

k-means: In this paper we combine the largest minimum distance algorithm and the traditional K-Means algorithm to propose an improved K-Means clustering algorithm. This improved algorithm can make up the shortcomings for the traditional K-Means algorithm to determine the initial focal point. The improved K-Means algorithm effectively solved two disadvantages of the traditional algorithm, the first one is greater dependence to choice the initial focal point, and another one is easy to be trapped in local minimum

GMM-EM: A commonly used tool for estimating the parameters of a mixture model is the Expectation–Maximization (EM) algorithm, which is an iterative procedure that can serve as a maximum-likelihood estimator. The EM algorithm has well-documented drawbacks, such as the need for good initial values and the possibility of being trapped in local optima. Nevertheless, because of its appealing properties, EM plays an important role in estimating the parameters of mixture models. To overcome these initialization problems with EM, in this paper, we propose the Rough-Enhanced-Bayes mixture estimation (REBMIX) algorithm as a more effective initialization algorithm.

• Table of contents

Anyone:

I. Introduction A. Background information B. Statement of the problem C. Objectives of the study

II. Literature Review A. Previous studies on the topic B. Gaps in the literature

III. Methodology A. Research design B. Data collection methods C. Sampling procedure

IV. Results A. Presentation of findings B. Analysis of data V. Discussion A. Interpretation of results B. Implications of the findings

VI. Conclusion A. Summary of the study B. Recommendations for future research

• List of abbreviations in alphabetical order

Anyone:

Example 1: List of Abbreviations for a Medical Research Paper

AIDS - Acquired Immune Deficiency Syndrome BMI - Body Mass Index CVD - Cardiovascular Disease DM - Diabetes Mellitus HIV - Human Immunodeficien

• List of symbols

Anyone:

List of Symbols for a Physics Thesis

a - Acceleration c - Speed of Light E - Energy F - Force G - Gravitational Constant h - Planck's Constant m - Mass p - Momentum r - Radius v - Velocity

• List of Figures

anyone

a - Acceleration c - Speed of Light E - Energy F - Force G - Gravitational Constant h - Planck's Constant m - Mass p - Momentum r - Radius v - Velocity

• List of Tables

anyone

Table 1: Market Analysis by Product Type Table 2: Sales Forecasts by Region Table 3: Price Comparison of Competitors' Products Table 4: Operating Expenses by Quarter Table 5: Break-Even Analysis Table 6: Cash Flow Projections Table 7: Return on Investment Analysis Table 8: Key Performance Indicators

**• Introduction**

**Sun**

E.g. Cluster analysis is based on various kinds of objects’ differences and uses distance functions’regulations to make model classification [3]. Whether the classification is really make a difference or not is rest with the distribution form of pattern character vectors. If the contributions of dots of vectors is clustered and sample dots in the same group are concentrated and sample dots in different groups are distant, it will be easy to use distance functions to classify the dots, which will as far as possible make statistics in the same group be similar and statistics in different group be different. The eigenvector of the whole sample pattern congregation can be treated as dots which distribute in feature space. The distance function between dots may act as the measure of similarity of patterns. According to the proximity of dots’ distance, the measure can be used to classify patterns. In this paper we combine the largest minimum distance algorithm and the traditional K-Means algorithm to propose an improved K-Means clustering algorithm. This improved algorithm can make up the shortcomings for the traditional K-Means algorithm to determine the initial focal point. The improved. K-Means algorithm effectively solved the disadvantage that the traditional K-Means algorithm depends too much on the selection of initial focal points.

**• Scope and objectives of the project**

Zunao

**• Detailed methodology and implementation**

k-means methodology: Sun

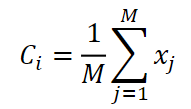
e.g. K-means is a clustering algorithm that aims to group data points based on similarity. It selects K centroids and assigns data points to the nearest centroid, iteratively updating the centroid's position as the mean of its assigned data points. The process repeats until convergence. K-means is commonly used for tasks such as customer and image segmentation, but its performance can be affected by the initial centroid choice, and it may not always produce the optimal solution.

It is a distance-based clustering algorithm that divides data into a number of clusters in numerical attributes.

1. Determine the number of clusters K and the number of maximum iterations.

2. Perform the initialization process K midpoint cluster, then the equation of centroid count

feature:



Equation 1 is done as much as p dimensions from i = 1 to i = p

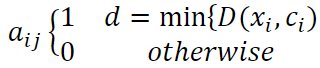
3. Connect any observation data to the nearest cluster. Euclidean distance spacing measurements

can be found using equation 2.



4. Reallocation of data to each group based on comparison of distance between data with each

group's centroid [9].



5. Recalculate the cluster midpoint position.

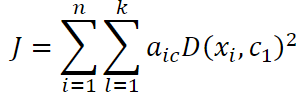
It is the value of the membership of point xi to the centres of the group c1, d is the shortest

distance from the data xi to the group K after being compared, and c1 is the centre of the group

to 1. The objective function used by this method is based on the distance and the value of the

data membership in the group. The objective function according to MacQueen (1967) can be

determined using equation.



n is the amount of data, k is the number of groups, ai1 is the membership value of the data

point xi to the c1 group followed a has a value of 0 or 1. If the data is an anngota of a group,

the value ai1 = 1. If not, the value ai1 = 0.

6. If there is a change in the cluster midpoint position or number of iterations <the maximum

number of iterations, return to step 3. If not, then return the clustering result.

---------------------------------------------------------------------------------------------------

k-means implementation:

Zunao

e.g.

GMM-EM methodology: Sun

1. g. the GMM-EM (Gaussian Mixture Model - Expectation-Maximization) methodology is its application in image segmentation, where the algorithm is used to group pixels with similar intensity values into distinct regions, allowing for the identification and extraction of objects or features of interest in the image.

GMM-EM implementation: Zunao

• Experimental results

Zunao

• Conclusion

Zunao

• References

Sun

Zunao

• Appendices if necessary

Zunao