Improvement in Fuzzy C Means Clustering

According to the distance cost function proposed, we find the new values of the distance cost function. Based on the minimum distance cost function, we can get the optimal clustering value. The clustering value is calculated through the improved FCM. The algorithm is applied on the iris data set as well as a second data modeling dataset and so on.

The optimal cluster centers are taken using R Studio and

Analyzed using basic python script.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Data Set*** | ***No of Instances*** | ***No of attributes*** | ***No of classes*** |
| ***Iris[2]*** | ***150*** | ***4*** | ***3*** |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Data Set*** | ***No of Instances*** | ***No of Attributes*** | ***No of classes*** |
| ***User modeling Data set[2]*** | ***150*** | ***5*** | ***4*** |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Data Set[2]*** | ***No of Instances*** | ***No of Attributes*** | ***No of classes*** |
| ***Memographic.data set*** | ***250*** | ***6*** | ***2*** |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Data Set[2]*** | ***No of Instances*** | ***No of Attributes*** | ***No of classes*** |
| ***Tae.data*** | ***150*** | ***6*** | ***3*** |

***Result table for iris dataset***

| K-  C | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- |
| 2 | 2.442780675 | 2.9682716 | 3.2310170625 | 3.38866434 | 3.493762525 |
| 3 | 1.35858791667 | 1.64380671111 | 1.78641610833 | 1.87198174667 | 1.92902550556 |
| 4 | 1.5574710125 | 1.85442008333 | 2.00289461875 | 2.09197934 | 2.15136915417 |
| 5 | 2.91816187 | 2.82188372 | 2.773744645 | 2.7448612 | 2.72560557 |
| 6 | 1.93915484167 | 2.00851866667 | 2.04320057917 | 2.06400972667 | 2.07788249167 |

***Result table for user modeling dataset***

| K/c | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- |
| 2 | 1.650667275 | 1.6230761 | 1.6092805125 | 1.60100316 | 1.595484925 |
| 3 | 1.14887013333 | 1.13016873333 | 1.12081803333 | 1.11520761333 | 1.08831049444 |
| 4 | 0.7663925875 | 0.764360816667 | 0.76334493125 | 0.7627354 | 0.7485160375 |
| 5 | 0.56728735 | 0.578540133333 | 0.594166525 | 0.58754236 | 0.56956594 |
| 6 | 0.36370595 | 0.356695388889 | 0.353190108333 | 0.35108694 | 0.341666902778 |

***Result table for tae dataset***

| K  C | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- |
| 2 | 10.53084575 | 12.020463 | 12.765271625 | 13.2121568 | 13.51008025 |
| 3 | 8.0264858333 | 8.0800275556 | 8.1067984167 | 8.1228609333 | 8.3705348333 |
| 4 | 9.3998975 | 11.2403935 | 12.1606415 | 12.7127903 | 11.82817325 |
| 5 | 12.5610163 | 11.4310906667 | 10.86612785 | 10.52715016 | 10.2523877 |
| 6 | 18.0389349167 | 17.7997914444 | 17.6802197083 | 17.6084766667 | 16.6064035278 |

***Result table for mesmographic dataset***

| 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- |
| 13.43887 | 15.6876856667 | 15.8120935 | 15.4867382 | 15.9365013333 |
| 13.4758365 | 15.6229891111 | 16.6965654167 | 17.3407112 | 17.7661108333 |
| 12.5610163 | 11.4310906667 | 10.86612785 | 10.52715016 | 10.2523877 |
| 20.1223404 | 19.2908208 | 18.875061 | 18.62560512 | 18.4578647333 |
| 18.0389349167 | 17.7997914 | 17.6802197083 | 17.6084766667 | 16.606403527 |

There after we calculate the difference between two succsssive c values and find the min difference and then choose the lower C value as final optimal value

Final analysis of min distance from above tables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data set | Optimal K vale | Optimal C set | Min value of difference | Expected  cluster  value | Actual  cluster  value |
| iris | 2 | (3,4) | 0.198883096 | 3 | 3 |
| User modeling | 6 | (4,5) | 0.01797642 | 4 | 4 |
| Tae data set | 2 | (3,4) | 1.373411667 | 3 | 3 |
| Mesographic data set | 2 | (2,3) | 0.0369665 | 2 | 2 |