Chapter 1

Assessment of FCM, PCM, and UFPC Algorithms Through Internal and Fuzzy Cluster Validity Indices on Multidisciplinary Benchmark Datasets[1]

Berna Özbaşaran[2]

Gözde Ulutagay[3]

Abstract

In the contemporary context, due to the high volume of unlabelled data in fields such as medicine, agriculture, chemistry, and many more; unsupervised machine learning models have been topics of interest and of investment. One of the computationally inexpensive and fast models investigated in this paper will be the fuzzy form of K-Means known as Fuzzy C-Means (FCM). Since FCM like K-Means requires the cluster number beforehand it is also vital that the cluster validity indices be fuzzy. In this paper, the evolutionary steps of FCM will be compared by evaluating the models suggested to overcome the pitfalls of the FCM algorithm. As there are many other algorithms created for this purpose, the algorithms analyzed in this article will be Possibilistic C-Means (PCM) and Unsupervised Fuzzy Possibilistic C-Means (UFPC). The comparison of these models is crucial since the new parameters introduced affect the cluster number chosen as seen in the internal validity indices. For applying the algorithms 4 benchmark datasets will be studied in R that belong to fields from biology, chemistry, and demography. The researcher expects that the UFPC algorithm will surpass the others since, the algorithm uses parameters from both FCM and PCM, however, as real-life datasets are rather complex, it is significant that the analysis be compared to benchmark datasets as proposed in this article. The performance will be evaluated on 12 fuzzy clustering validity indices and 3 internal validity indices that being silhouette, gap, and WSS. Custom R libraries will be used to ease the process of applying the algorithms and validity indices.

1 This paper was presented at the VI. International Applied Statistic Congress (UYIK)

2 Ege University, Faculty of Science, Department of Statistics, Turkey

3 Ege University, Faculty of Science, Department of Statistics, Turkey

https://doi.org/10.58830/ozgur.pub865.c3499

INTRODUCTION

As real-life datasets tend to be more complex in terms of their structure and volume, bettering and adapting an inexpensive model to be more robust seems as a viable option than using more complex and expensive models. The FCM *(*Dunn, J. C. 1973) algorithm (which is the fuzzy version of the K-Means) have been a topic of interest in this aspect, owing its simple but elegant nature. With its pitfalls came some suggestions from numerous academics as to better the algorithm and preserve its effectiveness. In this paper, two of those alternatives will be compared based on benchmark datasets, with reference to crisp and fuzzy indices. For this purpose, the algorithms evaluated will be FCM (Fuzzy C-Means), PCM (Possibilistic C-Means) (Krishnapuram, R., & Keller, J. M. 1993) and lastly UFPC (Unsupervised Fuzzy Possibilistic Clustering) (Yang, M., & Wu, K. 2005). The real-world benchmark datasets will be Glass, Seeds, German Credit and Wine Quality, respectively. Lastly the crisp indices considered will be the elbow method, silhouette score and gap statistics, alongside 12 fuzzy indices which will be named in the methods section. Hence, a comparison of evolutionary steps of an algorithm will be evaluated, with reference to both fuzzy and crisp cluster validity indices.

The investigation is significant in comprehending the evolution of an algorithm, since the latter algorithms have been suggested to better the previous approach via additional parameters. For instance, the pitfalls of the FCM (algorithm which is a fuzzy version of the K-Means) have been altered to overcome its vulnerability towards outliers and noise. In the PCM instance, two new parameters called typicality degree and typicality exponent were produced as it was expected to be a more flexible parameter than the membership degree which will be mentioned in the methods part.

As for the performance metrics, the R libraries fcvalid and ppclust will be utilized alongside others to extract the relevant cluster validity values. The datasets that have been chosen for the comparison are multidisciplinary samples that are known as benchmark datasets. These real-life datasets were chosen for two primary reasons, the first being their versatility in nature and that there is a ground truth meaning the classes are fixed in nature. This means that the datasets tend to have more noise and outliers than synthetic datasets, but also that the cluster number found could be compared with the ground truth to see if it is a match.

In this light, firstly the descriptive statistics will be shown to provide a sense of the data, and to investigate if there are peculiarities in terms of features. Secondly, the crisp validity indices will be shown to see if the classical cluster validity indices are able to spot the ground truth for the designated dataset. Thirdly, the fuzzy indices will be anlayzed for three of the algorithms which are FCM, PCM and UFPC, and the scatter plots will be colored according to cluster numbers for visualization. Finally, the comparison will be made in order to spot some sort of pattern between dataset characters, algorithms and indices. As an educated guess, it can be imagined that UFPC will perform better in rather compelx datasets, since it combines the new parameters introduced by the FCM and PCM algorithms.

METHODS

Statistical Analysis

The main aim of this paper is to compare the three algorithms FCM, PCM and UFPC on the basis of their ability to predict the ground truth with help from crisp or fuzzy validity indices.The algorithms used for this purpose can be introduced as the following.

*FCM (Fuzzy C-Means) (*Dunn, J. C. 1973)

Fuzzy C-Means mathematically works by minimizing the overall weighted distance between data points and cluster centers, where the weights reflect how strongly each point belongs to each cluster. Its aim is to group similar data together while allowing partial membership, so that each point can belong to multiple clusters to varying degrees. The latter algorithms have been suggested since it has a vulnerability towards outliers and noise, as these values tend to receive higher membership degrees that affect the clustering result.

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^{m} \, | x_j - \mu_i |^2 \qquad (1)$$

where

k shows the cluster number

n shows the sample size

u_{ij}^{m} shows the memberhsip degree

 is known as the fuzzifier parameter or the fuziness exponent. The algorithm becomes more fuzzy as increases

$| x_j - \mu_i |^2$ shows the Squared Euclidean distance

PCM (Possibilistic C-Means) (Krishnapuram, R., & Keller, J. M. 1993)

PCM (Possibilistic C-Means) works by minimizing a cost that measures how typical each data point is to clusters without forcing memberships to sum to one like in the membership degree parameter. Its goal is to group data while allowing points to have low typicality if they do not clearly belong to any cluster, making it robust to noise and outliers. A pitfall to this algorithm is that the initalization of the typicality degree is vital and that the algorithm tends to make coincidental clusters (Cebeci, Z. (2020)).

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} t_{ij}^{m} \mid x_j - \mu_i \mid^2 + \sum_{i=1}^{k} \eta_i \sum_{j=1}^{n} \left(1 - t_{ij}\right)^m \quad (2)$$

where

t_{ij} shows the typicality degree. Unlike FCM's membership degree, the typicality degree is not constrained to sum to 1. This parameter reflects how typical or compatible the point is with cluster, independently of other clusters.

ηi shows the regularization parameter for a cluster. This parameter controls how wide or tight the cluster is and influences how quickly the typicality drops with distance from the center.

$\sum_{j=1}^{n} \left(1 - t_{ij}\right)^m$ balances the influence of values on the clusters by penalizing nontypical values (like outliers). As the typicality degree is lower the equations result will be higher since it is substracted and multiplied by the typicality parameter. Since the cost function ought to be minimal, nontypical values will have a higher cost function and be penalized.

UFPC (Unsupervised Fuzzy Possibilistic C-Means) (Yang, M., & Wu, K. 2005)

UFPC combines fuzzy memberships and possibilistic typicalities to cluster data by balancing soft assignments with typicality measures. It aims to improve clustering flexibility and robustness by integrating both approaches to overcome the pitfalls of FCM and PCM. PCA (Possibilistic Clustering Algorithm) is another algorithm that has been built on top of PCM (Krishnapuram, R., & Keller, J. M. 1996).

$$J_{UFPC}(X;U,V)=\sum_{j=1}^{n}\sum_{i=1}^{c}\left(a\cdot u_{ij,FCM}^{m}+b\cdot u_{ij,PCA}^{\eta}\right)d^{2}(x_j,v_i)+\frac{\beta}{n^{2}\sqrt{c}}\sum_{j=1}^{n}\sum_{i=1}^{c}\left(u_{ij,PCA}^{\eta}\log u_{ij,PCA}^{\eta}-u_{ij,PCA}^{\eta}\right) \quad (3)$$

where

- A and b show the weighting coefficients balancing the influence for FCM and PCA (possibilistic clustering algorithm), respectively. When b is zero, the cost function turns to the FCM algorithm
- β shows the regularization parameter that balances the entropy term, helping control membership distribution

$\frac{\beta}{n^{2}\sqrt{c}}\sum u_{ij,\mathrm{PCA}}^{\eta}\left(\log u_{ij,\mathrm{PCA}}^{\eta}-u_{ij,\mathrm{PCA}}^{\eta}\right)$ shows the penalty term, inspired by entropy. The equation encourages diversity, PCA (prevents all being 0 or 1) so asd to promote meaningful typicality distributions.

While the first term combines FCM and PCA, the second term controls the shape of the possibilistic membership distribution and encourages a balanced spread of memberships.

The fuzzy cluster validity indices mentioned in Table 1 will be used with their abbrevations given below. The rationale behind every fuzzy index is intrinsict, hence, the selection of the lowest or highest value should be selected accordingly. For instance, while PC (Partition Coefficient) presents the best value with the maximum value, in PE (Partition Entropy) the minimum value must be chosen. One can analyze Table 1 to learn which value gives the best result from the different clustering values.

Table 1 Fuzzy Internal Validity Indices for FCM Algorithm for the Glass Dataset

Fuzzy Index	Full name	Optimum Cluster Value
PC	Partition Coefficient	Maximum
MPC	Modified Partition Coefficient	Maximum
PE	Partition Entropy	Minimum
XB	Xie-Beni Index	Minimum
K	Kwon Index	Minimum
TSS	Tang, Sun & Sun Index	Minimum
CL	Chen-Linkens Index	Maximum
FS	Fukuyama Sugeno Index	Minimum
PBMF	Pakhira-Bandyopadyang-Maulik	Maximum
FSIL	Fuzzy Silhouette Index	Minimum
FHV	Fuzzy Hyper Volume	Minimum
APD	Average Partition Density	Maximum

For this purpose, one ought to first pick the relevant sample, which in this case will be the real-world benchmark datasets Glass, Seeds, German Credit and Wine Quality.

*The Glass Dataset (*German, B., 1987).

The dataset belongs to the field of chemistry and consists of 214 instances of glass samples described by 9 numerical features, including refractive index (RI) and elemental compositions like Na, Mg, Al, Si, K, Ca, Ba, and Fe. The ground truth for this dataset is considered as seven glass types, varying from materials used so as to build windows or vehicle headlamps.

*The Seeds Dataset (*Charytanowicz, M., et al. 2010)

This dataset which is from biology and agriculture contains 210 instances of wheat kernels described by 7 numerical features, including area, perimeter, compactness, kernel length, kernel width, asymmetry coefficient, and groove length. The ground truth for this dataset consist of three wheets which are the following: Kama, Rosa, and Canadian.

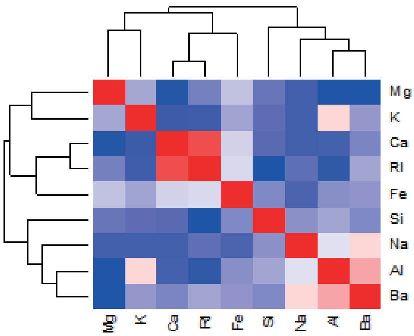
*The German Credit Dataset (*Hofmann, H. 1994).

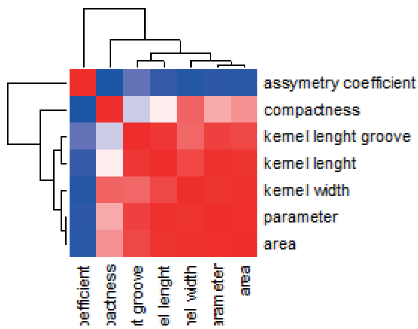
The German Credit dataset is a dataset from the field of demography and economy that contains 1,000 instances of loan applicants described by 20 attributes, including credit amount, duration, age, employment status, and housing type. The labels that are mentioned in the liteature are binary, that is good loan chance and bad loan chance.

*The Wine Quality Dataset (*Cortez, P., et al. 2009)

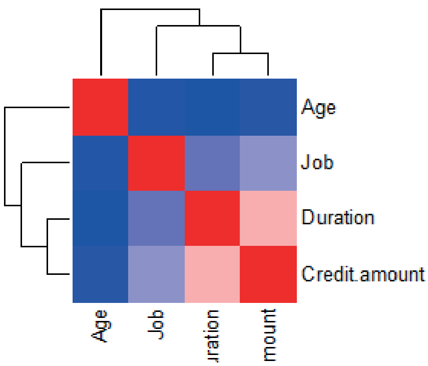
The last dataset from the field of chemistry contains 1,143 red wine samples described by 11 physicochemical attributes, such as acidity, chlorides, and alcohol content; and the label. Wine quality is rated on a scale from 0 to 10 and is grouped into three classes as a benchmark dataset: low (≤ 4), medium (5–6), and high (≥ 7) quality.

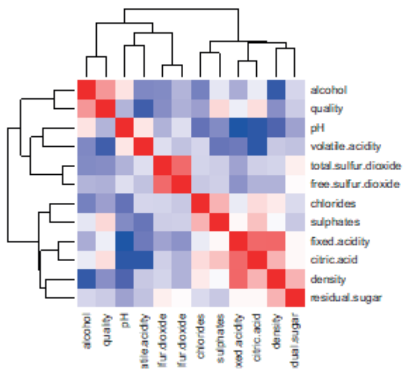
For the descriptive analysis, the following numeric features for every variables was selected. In Figure 1, one can see the heatmap for the correlation matrix for all datasets. One can see that the diagonal is red for all the intercept with the variable itself, so this would mean the correlation is 1. The blues represent negative correlation. In the Glass dataset, we can see a strong negative correlation between the variable pairs: Ca, Mg; AI, MG; Ba, Mg; RI, Si; RI, Al, alongside strong positive correlation with RI and Ca. Thus, negative correlation has dominated in the Glass dataset case. For Seeds, one can see that except for the relationship with the assymetry coefficient, the table consist of predominantly strong positive correlation which might suggest multicollinerity. In the German Credit dataset, one can see a stronger negative correlation between age and the other variables. Lastly, there does not seem to be quite as much correlation between the Wine Quality dataset, but one can spot a strong positive and negative correlation chunk between variables. This information will be useful as we are comparing the dataset in the summary section.





Glass Dataset *Seeds Dataset*





German Credit Dataset *Wine Quality Dataset*

Figure 1 the correlation is respresented as positive with red subtones, and negative with blue subtones. The stronger the color, the more strenght in correlation

RESULTS

The analysis for this paper has been done in R, with libraries such as pastecs, e1071, gridextra, factorextra, ppclust and lastly fcvalid. In Figure 2, one can view the ground truth and the crisp/ classical clustering validity indices. While all internal validity indices have failed in finding the ground truth in the datasets Glass and Wine, in the Seeds dataset, only Gap statistics could find the actual number of clusters. However, in the German Credit dataset, Gap statistics was the only index which failed in finding the accurate number of clusters according to the label variable.

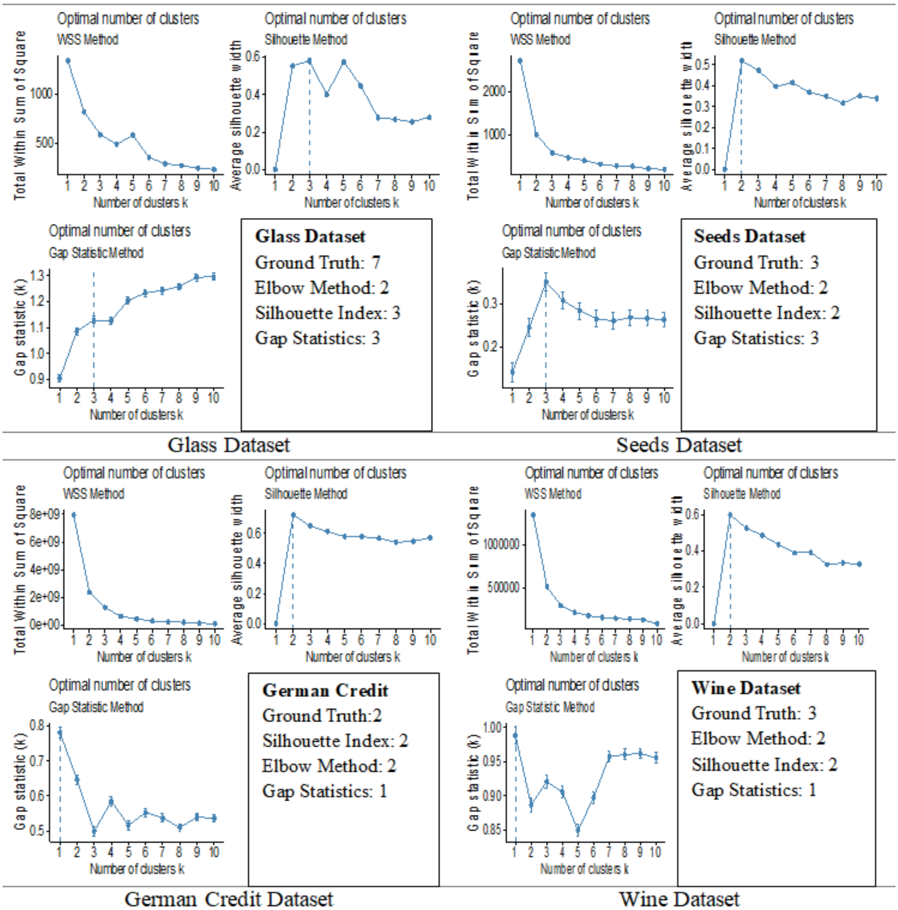


Figure 2 Results for Crisp Internal Validity Methods

As seen in Table 2, 3, 4 and 5 the relevant value according to the previous table (Table 1) was extracted for each fuzzy index for the FCM algorithm. In the Glass dataset, as the ground truth is 7, one can see that TSS (Tang, Sun & Sun Index) which searches the minimal value has matched the ground truth while neither the crisp indices or the other fuzzy indices could. In the seeds dataset it seems that PBMF and APD could find the ground truth 3. Looking at the German Credit dataset, one can see that all except for TSS, PBMF, FSIL and APD could find the ground truth, however, in the later algorithms we see that these indies also manage to find the accurate cluster number. Lastly, in the Wine Quality dataset FHV and APD have managed to find the actual cluster number which was 3.

Table 2 Fuzzy Internal Validity Indices for FCM Algorithm for the Glass Dataset

FCM	c = 3	c = 4	c = 5	c = 6	c = 7
PC	**0.756**	0.634	0.499	0.493	0.499
MPC	**0.635**	0.513	0.374	0.392	0.415
PE	**0.461**	0.690	0.948	0.996	1.008
XB	**0.129**	0.591	2.989	2.358	1.972
K	**27.920**	129.532	659.322	531.306	466.117
TSS	25.907	111.814	11.988	8.628	**1.377**
CL	**0.708**	0.564	0.390	0.395	0.408
FS	-714772.886	-599838.578	-472174.360	**-466378.017**	-471225.827
PBMF	23.935	33.632	**9.300**	13.757	23.556
FSIL	0.818	0.622	0.414	**0.401**	0.463
FHV	0.000	0.000	**0.000**	0.000	0.000
APD	**854420626.495**	661626128.419	323505171.749	148062214.934	103475658.018

Table 3 Fuzzy Internal Validity Indexes for FCM Algorithm for Seeds Dataset

F CM	c = 2	c = 3	c = 4	c = 5
PC	**0.805**	0.726	0.639	0.575
MPC	**0.610**	0.589	0.519	0.469
PE	**0.322**	0.500	0.691	0.841
XB	**0.102**	0.151	0.164	0.265
K	**21.694**	32.615	35.570	58.111
TSS	21.159	31.005	**11.396**	29.573
CL	**0.735**	0.668	0.577	0.508
FS	**-31937.055**	-29588.269	-25970.056	-23820.749
PBMF	47.012	**109.262**	45.939	95.275
FSIL	0.791	0.744	0.675	**0.623**
FHV	**0.00000171**	0.00000177	0.00000262	0.0000028
APD	111,363,680.158	**130,082,900.134**	81,572,202.790	89,941,575.179

Table 4 Fuzzy Internal Validity Indexes for FCM Algorithm for German Credit

FCM	c=2	c=3	c=4	c=5
PC	**0.900**	0.829	0.813	0.773
MPC	**0.800**	0.743	0.750	0.716
PE	**0.177**	0.311	0.354	0.444
XB	**0.053**	0.083	0.109	0.173
K	**52.718**	83.807	112.806	181.378
TSS	52.468	26.446	3.711	**3.029**
CL	**0.871**	0.787	0.776	0.732
FS	-9888988248.183	-11621096150.883	**-12951305710.689**	-12078601698.260
PBMF	44863314.899	108280617.789	134662779.782	**208025590.231**
FSIL	0.899	0.856	0.842	**0.816**
FHV	**382229.742**	544645.189	644889.918	696087.279
APD	0.002	0.003	**0.006**	0.001

Table 5 Fuzzy Internal Validity Indexes for FCM Algorithm for The Wine Dataset

FCM	c = 2	c = 3	c = 4	c = 5
PC	**0.843**	0.771	0.713	0.666
MPC	**0.687**	0.656	0.617	0.582
PE	**0.262**	0.418	0.547	0.655
XB	**0.088**	0.131	0.148	0.186
K	**100.987**	150.567	171.123	216.905
TSS	100.724	66.412	**20.299**	32.581
CL	**0.787**	0.719	0.661	0.609
FS	-1930172.422	**-2066430.093**	-2015191.039	-1974403.848
PBMF	4309.187	9469.070	**14165.947**	13527.127
FSIL	0.834	0.789	0.751	**0.715**
FHV	0.000008	0.000008	0.000009	0.000010
APD	165506620.897	275607986.305	75829566.987	**288738803.942**

One can view the different clustering results as the cluster number changes. The x and y values have been chosen as to visualize the sepeation in the best way. One can remind themselves that the ground truth for the datasets were 7, 3, 2, 3 for Glass, Seeds, German Credit and Seeds respectively.

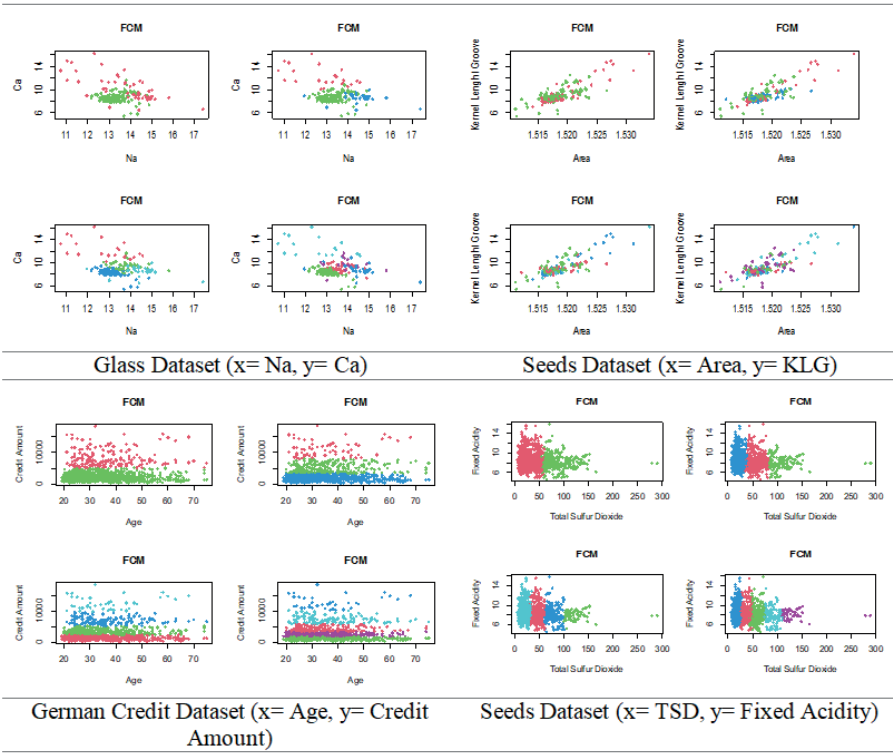
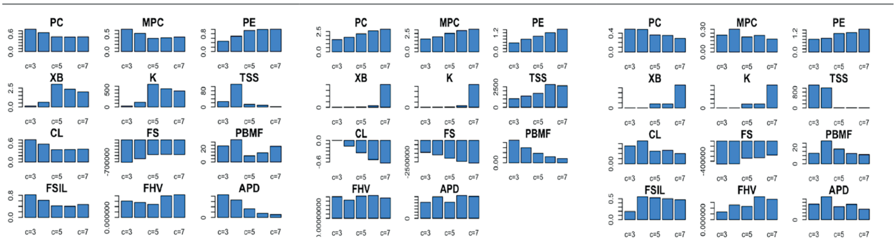


Figure 6 Fuzzy C-Means Clustering C=2, 3, 4, 5 for all dataset

The below Figures represent the bar plots for the Tables 2,3,4 and 5 which were the values for cluster numbers. From these figures, one can make the following deductions for the datasets at hand:

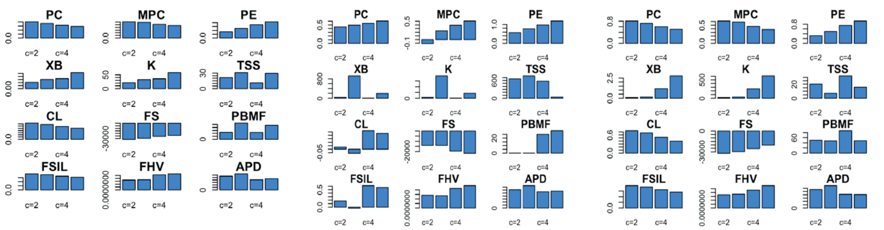
- PC and MPC tend to overestimate with the PCM algorithm (in cases of Seeds, German Credit, and Wine Quality) but they succeeded in estimating the Glass label which is 7.
- FSIL tends to act best with the PCM algorithm (according to the Seeds and German Credit datasets) but has underestimated the cluster number in Wine Quality and has printed the value NA in the Glass dataset. Furthermore, it can be seen that it has tended to overestimate the cluster number with FCM and UFPC algorithms.
- PE has managed to remain stable in its cluster estimates in all datasets, regardless of the algorithm change.
- PBMF worked best with FCM in the Seeds dataset, with UFPC in the German Credit dataset and in PCM in the Wine Quality dataset. Hence, it could be wise to also take into consideration the features of the dataset.
- APD managed to hit the mark with atleast one algorithm in all datasets, meaning it had correctly matched the ground truth 4 miss 12 thus with a 66,7 % accuracy rate. It has aced the datasets Seeds and Wine Quality. This is especially significant since the most succesful crisp index were unable to find these datasets as shown in Figure 2 (only gap statistics have correctly estimated the label for the Seeds dataset, others failed).
- Lastly FHV was succesful in all cluster estimations except for the Glass dataset.

Figure 3 Values of Fuzzy Validity Indices for FCM, PCM and UFPC of the Glass Dataset



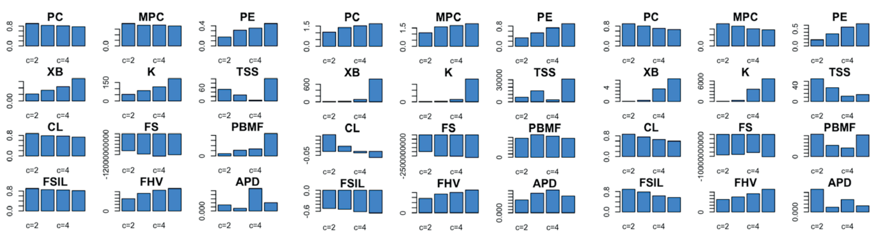
FCM	***PCM***	***UFPC***

Figure 4 Values of Fuzzy Validity Indices for FCM, PCM and UFPC of the Seeds Dataset



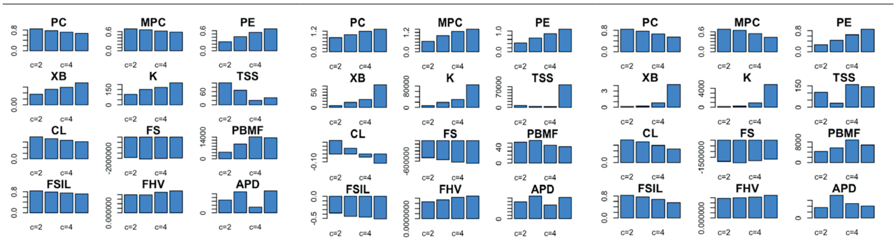
FCM	*PCM*	*UFPC*

Figure 5 Values of Fuzzy Validity Indices for FCM, PCM and UFPC of the German Credit Dataset



FCM	*PCM*	*UFPC*

Figure 6 Values of Fuzzy Validity Indices for FCM, PCM and UFPC of the Wine Dataset



FCM	*PCM*	*UFPC*

The above Figures have been summarized in the below tables for the comparison. The cluster numbers for the optimal values have been noted and those who have hit the mark of the ground truth have been written in bold, alongside the fuzzy clustering indices which have succeded with atleast one algorithm. It could be noted that APD (Average Partition Density) have performed best in all datasets by successeding to find the ground truth in numerous algorithms, showing its stability throughout algorithms. On the other hand, it could be seen that some indices performed better with particular algorithms. Such as Fuzzy Silhouette Index (FSIL) that had manage to select the accurate cluster number mostly with the PCM algorithm however, it could not be calculated with the PCM objective function in the Glass dataset (yet again it has accurately stated the cluster number with the UFPC algorithm).

Table 7 Comparison of Internal Validity Indices for Glass

Cluster Number	FCM	PCM	UFPC
PC	3	7	3
MPC	3	7	4
PE	3	3	3
XB	3	3	3
K	3	3	3
TSS	7	3	7
CL	3	7	4
FS	6	3	7
PBMF	4	3	4
FSIL	5	NA	3
FHV	5	4	3
APD	7	5	4

Table 8 Comparison of Internal Validity Indices for Seeds

Cluster Number	FCM	PCM	UFPC
PC	2	5	2
MPC	2	5	2
PE	2	2	2
XB	2	4	2
K	2	4	2
TSS	4	5	**3**
CL	2	4	2
FS	5	2	5
PBMF	**3**	5	4
FSIL	5	**3**	5
FHV	2	**3**	2
APD	**3**	**3**	**3**

Table 9 Comparison of Internal Validity Indices for German Credit

Cluster Number	FCM	PCM	UFPC
PC	**2**	5	**2**
MPC	**2**	5	**2**
PE	**2**	**2**	**2**
XB	**2**	**2**	**2**
K	**2**	**2**	**2**
TSS	4	4	4
CL	**2**	**2**	**2**
FS	**2**	**2**	4
PBMF	5	3	**2**
FSIL	5	**2**	5
FHV	**2**	**2**	**2**
APD	4	4	**2**

Table 10 Comparison of Internal Validity Indices for Wine Quality

Cluster Number	FCM	PCM	UFPC
PC	2	5	2
MPC	2	5	2
PE	2	2	2
XB	2	2	2
K	2	2	2
TSS	4	4	**3**
CL	2	2	2
FS	2	2	5
PBMF	4	**3**	4
FSIL	5	2	5
FHV	**3**	2	2
APD	**3**	**3**	**3**

DISCUSSION AND CONCLUSION

Before moving forward with the conclusion, it could be useful to view the summary of the evaluation of this paper, so as to conclude in objective remarks. One can benefit from Table 11 and Table 12 to recall the essential information concerning the algorithms and the datasets analysis results.

Table 11 Comparison of Alternative FCM Algorithms

Method	FCM	PCM	UFPC
Parameters	Number of clusters (c), fuzzifier (m)	c, fuzzifier (m), typicality degree (t), typicality exponent (η)	c, fuzzifier (m), typicality degree (t), typicality exponent (η), weights (α, β)
Novelties	Introduced fuzzy membership	Introduced typicality degree (t) and typicality exponent (η) to model noise and outliers	Combined membership and typicality parameters in one objective function
Strengths	Simple, efficient, interpretable	Robust to noise and outliers via typicality	Handles both overlapping clusters and noise
Weaknesses	Sensitive to noise and outliers	Prone to coincident clusters, sensitive to setting of t and η	Needs careful tuning of multiple parameters: m, t, η, α, β
Optimal Dataset	Well-separated, compact clusters	Datasets with overlap or noise, where fuzzy membership is insufficient	Complex datasets involving both overlapand noise simultaneously

Table 12 Comparison of Datasets

Dataset	**Glass**	**Seeds**	**German Credit**	**Wine Quality**
Sample Size	214	210	1000	1143
Variable Count	9	7	4	12
Silhouette Score	0.58	0.52	0.65	0.6
Optimal Cluster	3	2	2	2
Ground Truth	7	3	2	3
Successful Indices	FCM: TSS, APD PCM: PC, MPC, CL UFPC: TSS, FS	FCM: PBMF, APD PCM: FSIL, FHV, APD UFPC: TSS, APD	FCM: all except TSS, FS, PBMF, FSIL, APD PCM: all except PC, MPC, TSS, PBMF, APD UFPC: all except TSS, FS, FSIL	FCM: FHV, APD PCM: PBMF, APD UFPC: TSS, APD

Peculiarities	The fuzzy indices had lots of variation, Mostly negative correlation	High positive correlation and one variable has hight negative correlation. FHV was very close in FCM and UFPC while FS estimated pretty closely in PCM but they were incorrect in the end	Large dataset with few dimensions; best quality clusters, mostly negative and slightly positive correlation	Mostly negative and nearly no correlation between chosen variables
Best Method	FCM or UPFC with TSS	PCM with APD	UFPC with PE	FCM, PCM or UFPC with TSS

In conclusion, it has been investigated that some indices behave better with particular algorithms. This alligns with the argument of Krisphapuram, R., and Keller, M. (1996) mentioning an algorithm not giving the correct result does not necessarily mean that the algorithm is poorly designed, but rather that it has been missapplied in that case. Thus by this reference one can infer that as diverse algorithms have different assumptions to be fulfilled, this also means that they have pecular instances that enable them to perform their best. Therefore, just as datasets have different features that are not known beforehand, it might be suggested to apply not only multiple methods, but also multiple typologies of an algorithm that performs better in different instances, like in the case of FCM, PCM and UFPC. With domain knowledge, these complex datasets can be interpreted better. Finally, for further research, analysts and researchers can apply the algorithms in further contexts and maybe in some challenging datatypes like audio or image to test their limits.

References

Cebeci, Z. (2020). fcvalid: An R Package for Internal Validation of Probabilistic and Possibilistic Clustering. 3. 11-27. Retrieved from http://dx.doi.org/10.35377/saucis.03.01.664560

Charytanowicz, M., Niewczas, J., Kulczycki, P., Kowalski, P., & Łukasik, S. (2010). *Seeds* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5H30K

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). *Modeling wine preferences by data mining from physicochemical properties*. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/wine+quality

German, B. (1987). *Glass Identification* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5WW2P

Hofmann, H. (1994). *Statlog (German Credit Data)* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5NC77

Dunn, J. C. (1973). *A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters*. Journal of Cybernetics, 3(3), 32–57.

Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems, 1(2), 98–110. doi:10.1109/91.227387

Krishnapuram, R., & Keller, J. M. (1996). The possibilistic C-means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems, vol. 4, no. 3, pp. 385-393, Aug., doi: 10.1109/91.531779.

Yang, M., & Wu, K. (2005). Unsupervised possibilistic clustering. Pattern Recognition, 39(1), 5–21. https://doi.org/10.1016/j.patcog.2005.07.005

Acknowledgment

The authors would like to thank anonymous reviewers for valuable suggestions

Conflict of Interest

The authors declare no conflict of interest

Author Contributions

Berna ÖZBAŞARAN: writing- original draft, methodology, application

Gözde ULUTAGAY: conceptualization, writing- review & editing, supervision