Nat. Volatiles & Essent. Oils, 2021; 8(4): 1320-1337

A Comparative Study on Prejudiced Measurements of Datasets in three variants of Automatic Evolutionary Clustering using Teaching-Learning-Based Optimization

Ramachandra Rao Kurada[1*], Ramu Yadavalli[2], Sunil Pattem[3], Karteeka Pavan Kanadam[4]

[1,2,3] *Department of CSE, Shri Vishnu Engineering College for Women, Bhimavaram,*

[4]*Professor & HOD, Dept. of Computer Applications, RVR & JC College of Engg. Guntur*

**Corresponding author email: ramachandrarao.kurada@gmail.com*

Abstract

This paper aims to conduct automatic non-predictive learning by dividing datasets into several subsets on their prejudiced measurements. To notarize paramount recitals in this direction, the work presented in this paper is captivated by conducting automatic evolutionary clustering with Teaching-Learning-Based Optimization (TLBO) to organize the collection of patterns into clusters based on similarity and by upholding minimum intra-cluster and maximum inter-cluster distances. This work has set a goal to determine the number of clusters automatically, enriching absolute positions of clusters with optimal size, shape, low computational time and minimum error rate. The cogitative content of this work is to investigate possibilities for the improvement in classical clustering algorithms with TLBO and evolutionary automated clustering cognitions. This article optimizes multiple objective functions simultaneously and evaluates clustering quality regarding the goodness-of-fit of the resulting clusters against the existing methods. This treatise advances a new point of view results by testing the performance of TLBO and its advancements with automatic clustering techniques across real-time and micro-array datasets.

Keywords: Automatic Clustering, Evolutionary Algorithms, Cluster Validity Indices, Meta-heuristics, TLBO.

1. Introduction

The inclination towards building an evolutionary automatic clustering framework by unifying a meta-heuristic TLBO [15-22, 9,12] with automatic clustering procedures is to obtain intuitive interpretation and automatic analysis of datasets embedded in a high dimensional space. The following challenges have motivated to use self-adaptive clustering [1, 2] over a divaricated variety of datasets in this work.

- The classical partitioning techniques are single-objective optimizers and perform a local search rather than a global search.
- Classical partitioning techniques performance is highly dependent on initialization of initial seed value and incorporates ergonomic knowledge of human experts to define a pattern referring to the number of classes and scale the features available to the clustering algorithm.
- The technique confines its applicability with spherical clusters of almost-equal volumes.

- An appropriate pattern proximity measure using pair-wise similarity is used to quantify the degree of interestingness within the group and to use it as a single objective function

These cluster analysis challenges are attempted and explored systematically and structurally in the form of cluster properties [10]. This approach is exclusively practiced in this work with a set of three algorithms by establishing an automatic evolutionary framework to attain impending outcomes. The set of three algorithms taken up in this paper are AutoTLBO [8, 11], AutoSpssTLBO [7], and AutoITLBO [6]. The comparing three algorithms are developed by the same authors in earlier works. These set of algorithms can automatically find an optimal number of the cluster in any dataset without any human intervention, relatively with low % of error rate and less CPU time. The adopted cluster validity indices (CVIs) [3, 5, 14, 23] in this work can validate a reasonably good index function value with its high mathematical and statistical function. An elegant and versatile impression from the user thought process after having life experiences over real-time, and micro-array datasets can discover automatically statistically significant and hidden patterns in datasets for meaningful groupings.

2. Algorithms for Comparative Study

The chief strategy in this work is compelled under three distinct algorithms to foster this broad agenda. The esteemed sets of three automatic clustering algorithms are base lined with the TLBO, a most recently emerged algorithm in multi-objective evolutionary methods [1]. This high trajectory is reformed with profound variants elements of TLBO as basic, elitist and improved TLBO [12]. All the three algorithms used in the present article are the previous works of the first author of this paper.

The first algorithm (AutoTLBO) [8, 11] presented in this work brings a likely change in conventional clustering by using a randomize function to initialize the centroid with k-value and thereby coalesce the teaching and learning phases of TLBO and elitism into the approach. This approach attempts to solve multiple objective functions such as distance functions, CVIs in a single run. The impact of this excogitation tested over real-time and micro-array datasets [4], and thereby the algorithm strengthening is compared with its potential rivals in automatic clustering. The results presented under these algorithms earns a favorable verdict and thereby profoundly confers accountable and superior quality self-activating clusters without human intervention.

The second algorithm (AutoSpssTLBO) [7] presented in this work replaces the randomize function to initialize the centroid in the initialization phase of AutoTLBO with Single Pass Seed Selection (SPSS) [13], an initial seed selection strategic algorithm and amalgamates the teaching and learning phases of TLBO in its consecutive steps. To maintain transparency and consistency amongst the algorithms the aforementioned multiple objectives and to compare automatic clustering were relented again. This loom confirms creditable automatic clusters and expedient favorable solution than its challenger algorithms.

The third algorithm (AutoITLBO) [6] presented in this work upholds the same initialization strategy adopted in the second algorithms and is vested with Improved TLBO. This algorithm replicates the same inclusions that were presumed to bestow visible clustering results. This decisive step is highlighted to factorize an automatic clustering framework to lodge useful results to reckon self-regulating optimal natural partitions over datasets from any domain.

Apparently, the study in this paper spotlights enormous potential algorithms that exhibit impended and remarkable outcomes when advocated over real-life and micro-array datasets. Subsequently, this work commissions an influencing opinion that these sets of three automatic clustering algorithms may be used to endorse any clustering problems in scientific or engineering applications to cluster data objects and enormously gain noteworthy domino effect automatically.

Table 1: Mean Values of Automatic Algorithms after Completion of 50 Independent Runs over Real-Time Datasets.

Datasets (row*col, k)	**Algorithm**	**No. of Auto Clusters**	**ARI**	**RI**	**SIL**	**HI**	**CS**	**DB**	**% of Error Rate**	**CPU Time (Sec)**
Iris (150*4, 3)	AutoTLBO	3.02	**0.9232**	**0.9737**	0.0763	0.9473	0.7194	0.5178	**6.000**	27.89
	AutoSpssTLBO	3.01	0.8957	**0.9737**	0.0867	**1.0207**	**0.8966**	0.6230	8.17	167.34
	AutoITLBO	**3.00**	0.9914	0.9210	**0.9410**	0.0912	0.8945	**0.7415**	10.22	**20.14**
Wine (178*13, 3)	AutoTLBO	3.50	0.6932	0.6842	0.3158	0.5683	0.3955	0.7693	51.22	37.16
	AutoSpssTLBO	3.10	**0.6957**	0.7052	0.3948	**0.8957**	**0.7105**	0.7550	**0.360**	234.96
	AutoITLBO	**3.01**	0.8412	**0.7781**	**0.6417**	0.5412	0.6038	**0.8974**	41.98	109.47
Glass (214*9, 6)	AutoTLBO	5.86	0.6553	0.7053	0.2947	0.410	0.2991	1.3009	47.20	**41.67**
	AutoSpssTLBO	**5.97**	0.6912	0.7947	0.2953	**0.6958**	0.4895	1.2667	**16.29**	250.03
	AutoITLBO	5.95	**0.7894**	**0.9100**	**0.6014**	0.6458	**0.5481**	**.9980**	39.22	104.22

Table 2: Centroid table of iris Dataset Attributes with a Proposed set of Algorithms.

Attributes of iris dataset	AutoTLBO			AutoSpssTLBO			AutoITLBO		
	Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 0**	**Cluster 1**	**Cluster 2**	**Cluster 0**	**Cluster 1**	**Cluster 2**
Sapel length	5.90	6.85	5.00	5.93	5.00	6.58	5.9	5.0	6.5
Sapel width	2.74	3.07	3.41	2.77	3.41	2.97	2.8	3.4	3.0
Patel length	4.39	5.74	1.46	4.26	1.46	5.52	4.35	1.5	5.55
Patel width	1.43	2.07	0.24	1.32	0.24	2.02	1.33	0.2	2.0

3. Empirical Approach to Identify Peak Performances

This section consolidates the results presented in research articles [8, 11], [7] and [6]. These articles render a set of three algorithms copiously devoted to the appendage of automatic clustering methods and are names as AutoTLBO, AutoSpssTLBO, and AutoITLBO. The key aspects, salient features, and results of the proposed methodologies are well elaborated in the present article. It is proven in research articles [8, 11], [7] and [6] that these original algorithms efficiently and efficiently produced high-quality cluster when targeted over real-time and micro-array datasets [4]. Measuring performance of the algorithm is a correlated factor affected by computing speed, memory usage, the operating system installed and software opted to implement the algorithm. The empirical approach in this study precisely assists the user to gauge algorithm in a statistically significant way, with self-benchmarking practices towards the sensitivity of problem, parametric settings, and other performance metrics. A considerable factor that affects the relative performance of the algorithm is in the evolutionary approach since these algorithms return a faintly unique solution each time they are run. The second issue in consideration is as these set of algorithm tenders in a heuristic approach; they may result in superior quality results in a particular performance measure or vice versa.

3.1 Interpreting Results on Real-time Datasets

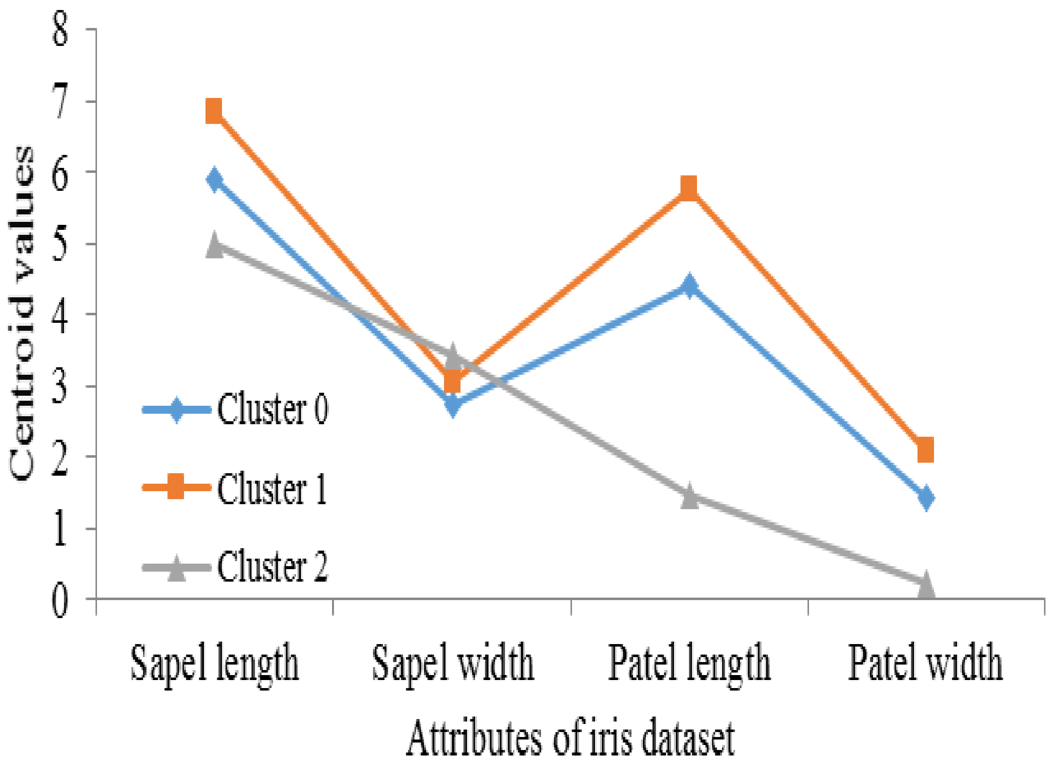

Fig. 1.Centroid values of AutoTLBO over iris Dataset.

In this experiment, the results appeared in research articles [8, 11], [7] and [6] over real-time datasets are consolidated, and a conceivable assessment of the intended methods is conducted. Table 1 represents the values occurred in the three algorithms. The best results are displayed in boldface.

In this comparing, AutoSpssTLBO exhibits peak performance in gaining an optimal number of automatic cluster, and AutoTLBO produces superior quality results in respect to the minimum error rate and CPU time. Four out of the six CVIs tenders optimal values nearer to 1. The impact of the third algorithms is relatively minimal when compared to its contenders. In concluding remarks from this section based on the outcomes of the proposed algorithm, is AutoTLBO and AutoSpssTLBO exhibit better peak performance over real-time datasets. Table 2 displays the centroid table of iris dataset comprising the set of three algorithms and the values are recorded after 50 independent runs. The other dimension of experimental study is extended towards a statistical approach to justify the performance of the algorithm. Fig. 1, 2 and 3 exhibits the same centroid values graphically of AutoTLBO, AutoSpssTLBO and AutoITLBO methods respectively. The observations from Table 2 AutoSpssTLBO gains relatively better benefits than the other two contending algorithms.

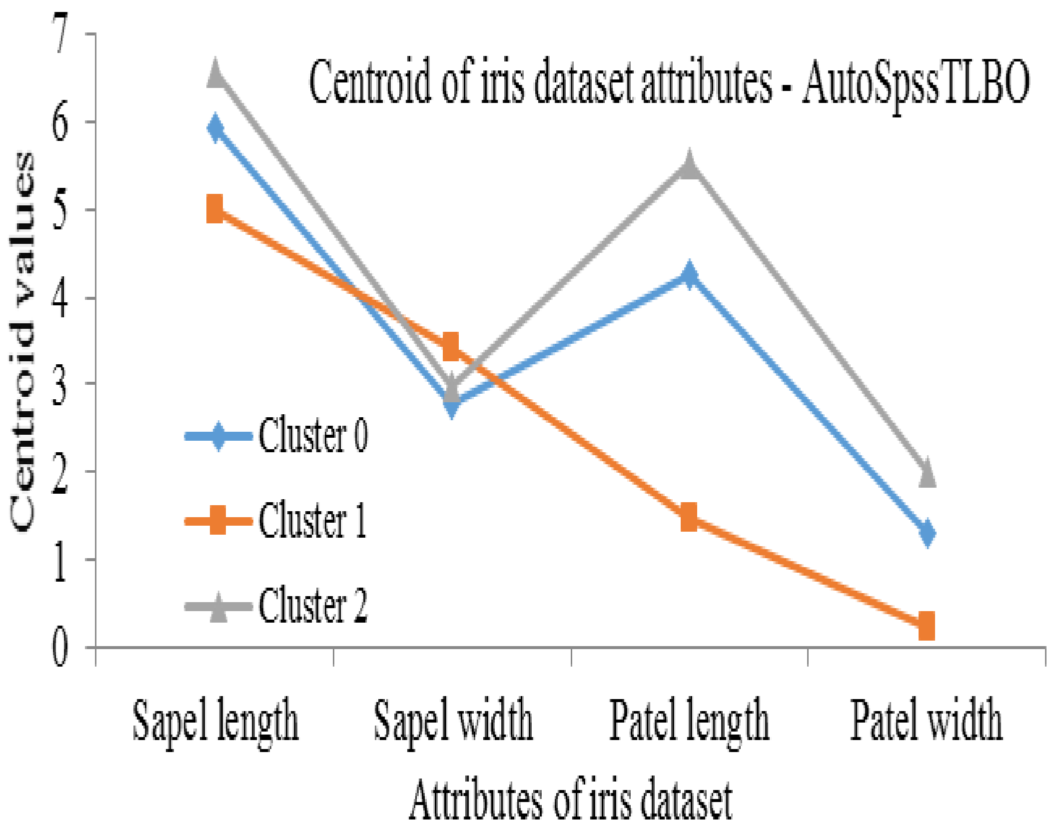

Fig. 2.Centroid values of AutoSpssTLBO over iris Dataset.

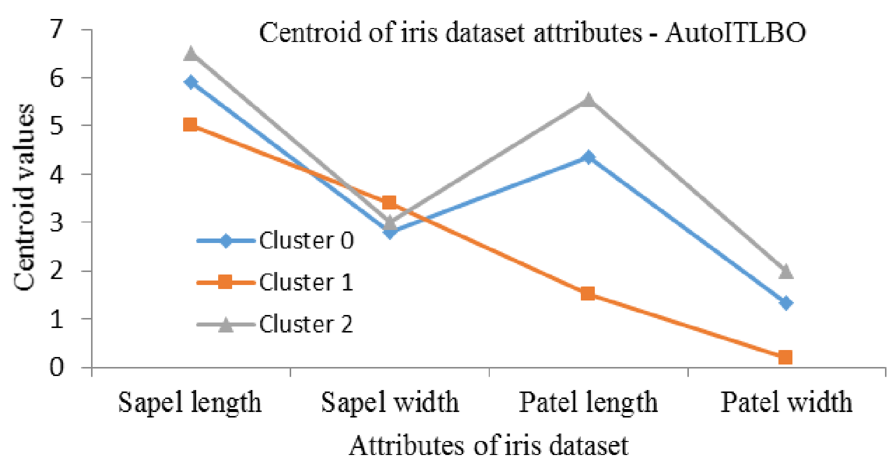

Fig. 3. Centroid values of AutoITLBO over iris dataset.

Table 3 describes the outcomes of stratified statistical evaluation methods adopted in this approach to identify the peak performers in real-time datasets, which are less influenced by errors while building the model. The 0.035 error value gained in mean absolute error (MAE) of AutoSpssTLBO, is relatively less when compared to the other two contenders. Eventually, the outcome of AutoSpssTLBO at root mean square (RMS) with an error value of 0.47, establishes a little relationship between MAE and RMS error while estimating the difference between the actual and predicted residuals. These implications thereby signal low error rate, which can be treated as a decisive factor towards estimation of AutoSpssTLBO accuracy. The errors root relative square error (RRSE) and relative absolute error (RAE) is expressed in a ratio of % in error. These metrics are together used to estimate the variance of error. AutoTLBO and AutoSpssTLBO induce a more significant variation in errors whereas the errors in AutoITLBO are of the same magnitude. The SSE is less in AutoTLBO when comparing to their algorithm, although AutoSpssTLBO exhibits low error rate in MAE, RMSE, RAE, and RRSE. The harmonic average of precision and recall, precision and recall scores accurately with a value of 0.96, by conveying an equivalent value between them. Subsequent paragraphs, however, are indented.

Table 3: Stratified Evaluation Metrics on Intended Algorithm over Iris Dataset.

Stratified evaluation metricsover iris dataset	**AutoTLBO**	**AutoSpss TLBO**	**AutoITLBO**
MAE	0.479	**0.035**	0.444
RMSE	0.157	**0.471**	0.157
RAE	9.65%	**7.87%**	100%
RRSE	32.98%	**33.63%**	100%
SSE	**7.817**	7.987	49.87
Average precision value	0.947	**0.96**	0.111
Average recall value	0.947	**0.96**	0.333
Average f-measure	0.947	**0.96**	0.167

The statistical properties studied in Table 3 justify AutoSpssTLBO is well featured to evaluate the performance of real-time datasets, and AutoTLBO follows it. Table 4 presents the centroid table of glass dataset when applied over three algorithms presented in this work. The 7 centroids attained over 9 attributes of glass dataset are recorded after 50 independent runs. Holistically, the centroid values produced in AutoSpssTLBO quotes better benefits than the other two contending algorithms in this study.

Table 5 represents the stratified evaluation methods by comparing three algorithms on glass datasets. The values considered in this table are recorded after 50 independent runs. MAE and RMSE are less influenced towards error by producing 0.0743 and 0.2374 respectively in AutoSpssTLBO as they are scoring values to lower boundary; subsequently, they are followed by 0.1026 and 0.2897 in AutoTLBO. The same set of implication is also shown in RAE and RRSE values of AutoSpssTLBO by marking a low magnitude of total errors with 35.07% and 73.13 % respectively. AutoSpssTLBO attains minimum SSE with 42.82 when compared to other contending algorithms. The performance of AutoSpssTLBO when investigated with f-measure, precision and recall values produces acceptable terms of 0.768, 0.752, and

0.751 respectively scoring higher values towards upper boundaries and more concisely followed by AutoTLBO. The overall impression after applying stratified evaluation metrics mentioned in Table 8 over glass data is AutoSpssTLBO exhibits peak performance in all the sections.

Table 4: Centroid Table of Glass Dataset Attributes after 50 Independent Runs.

Attributes of glass dataset	**AutoTLBO**						
	Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
RI	1.5184	1.5169	1.5189	1.5178	1.5173	1.519	1.525
NA	13.0645	14.5497	12.9221	13.2148	13.2341	13.3847	12.7925
MG	3.507	0.5666	0.7186	3.4261	3.4245	3.5654	0.1967
AL	1.1645	2.0369	2.0379	1.3552	1.4857	1.1263	1.1367
SI	72.7225	73.0772	72.3929	72.6526	72.6587	72.5389	72.1883
K	0.475	0.2947	1.365	0.527	0.5561	0.3946	0.215
CA	8.839	8.4513	10.1779	8.5648	8.4098	8.8333	12.97
BA	0.012	0.9425	0.1743	0.023	0.0025	0.014	0.2625
FE	0.2035	0.0122	0.0564	0.2013	0.0102	0.0093	0.1017
	AutoSpssTLBO						
	Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
RI	1.5217	1.5165	1.5197	1.5266	1.5167	1.5176	1.518
NA	13.77	14.56	13.01	12.64	13.25	13.02	13.195
MG	3.755	0	0	0	3.52	3.52	3.555
AL	0.85	2.06	1.75	1	1.52	1.28	1.395
SI	71.785	73.11	72.715	72.19	72.75	72.96	72.54
K	0.115	0	0.47	0.1	0.61	0.57	0.565
CA	9.59	8.62	11.05	13.3	8.12	8.56	8.44
BA	0	0.81	0	0	0	0	0
FE	0	0	0	0	0	0	0.175

	AutoITLBO						
	Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
RI	1.5255	1.5213	1.5175	1.5173	1.5171	1.5189	1.5175
NA	12.6318	13.7819	14.6467	13.028	14.4421	12.8277	13.204
MG	0.2145	3.7204	1.3056	3.4787	0.5383	0.7738	3.4704
AL	1.1891	0.8863	1.3667	1.2751	2.1228	2.0338	1.4539
SI	72.0709	71.8767	73.2067	72.95	72.9659	72.3662	72.6567
K	0.2345	0.1848	0	0.5522	0.3252	1.47	0.5699
CA	13.1291	9.3826	9.3567	8.516	8.4914	10.1238	8.4084
BA	0.2864	0.0352	0	0.0016	1.04	0.1877	0.0096
FE	0.1109	0.0585	0	0.0549	0.0134	0.0608	0.0743

Table 5. Stratified Evaluation Metrics on Intended Algorithm over Glass Dataset.

Stratified evaluation met-ricsover glass dataset	**AutoTLBO**	**AutoSpss TLBO**	**AutoITLBO**
MAE	0.1026	**0.0743**	0.2118
RMSE	0.2897	**0.2374**	0.3245
RAE	48.45%	**35.07%**	100%
RRSE	89.27%	**73.13%**	100%
SSE	46.71	**42.82**	49.94
Average precision value	0.67	**0.768**	0.126
Average recall value	0.66	**0.752**	0.355
Average f-measure	0.64	**0.751**	0.186

Table 6: Centroid Table of wine Dataset Attributes over Proposed Algorithms after 50 Independent Runs.

Attributes of wine dataset	AutoTLBO			AutoSpssTLBO			AutoITLBO		
	Cluster 0	Cluster 1	Cluster 2	Cluster 0	Cluster 1	Cluster 2	Clus-ter 0	Clus-ter 1	Clus-ter 2
Alcohol	13.73	13.15	12.25	13.15	13.73	12.25	13.74	13.13	12.27
Malic	2.00	3.34	1.90	3.34	2.00	1.90	1.76	3.28	1.61
Ash	2.45	2.43	2.23	2.43	2.45	2.23	2.44	2.38	2.24
Alcalinity	7.25	21.43	20.06	21.43	17.25	20.06	16.9	21	20
Magnesi-um	106.88	99.02	94.04	99.02	106.83	94.04	104.5	96.5	88
Phenols	2.84	1.67	2.25	1.67	2.84	2.25	2.82	1.64	2.2
Flavanoids	2.98	0.79	2.07	0.79	2.98	2.07	2.97	0.69	2.03
Nonfla-vanoids	0.28	0.45	0.36	0.48	0.28	0.36	0.28	0.47	0.36
Proantho-cyanins	1.90	1.16	1.62	1.16	1.90	1.62	1.85	1.10	1.58
Color	5.49	7.34	3.05	7.34	5.49	3.05	5.4	7.4	2.9
Hue	1.06	0.68	1.05	0.68	1.06	1.05	1.07	0.67	1.04
Dilution	3.16	1.69	2.78	1.69	3.16	2.78	3.17	1.68	2.83
Proline	1113.53	627.55	512.82	627.55	1113.53	512.82	1087.5	622.5	491.5

Table 6 shows the centroid table of wine dataset attributes when applied over the proposed set of algorithms. The observation from this Table 6 is AutoSpssTLBO relatively has better centroid values than AutoTLBO or AutoTLBO in all the attributes of wine dataset. Table 7 exhibits the stratified evaluation metrics imposed on wine dataset. The values posted in this table are the outcomes after 50 independent runs. It was a bright indication from the table that, the MAE and RMSE values denominated for AutoSpssTLBO are less influence towards error.

The other fact considered from this table is the value of RAE and RRSE metrics applied over AutoSpssTLBO are scaled down towards lower values. The RAE and RRSE values sense all the errors are of the same magnitude. The SSE values are shown in AutoTLBO, AutoSpssTLBO is equal and is towards the lower end. The standpoint from all these observations incriminates AutoSpssTLBO outperforms in all the stratified evaluation metrics. Consequently, AutoTLBO is the front-runner in the comparing algorithms. Fig. 4 represents the automatic clusters imparted by AutoSpssTLBO in real-time datasets.The figure depictedin Fig. 4 shows the obtained clusters which were well separatedrepresenting data points into optimal number of groups. The imparted clusters holds the property that data points in the same sets are more alike to other data points in the same set than those in other sets.

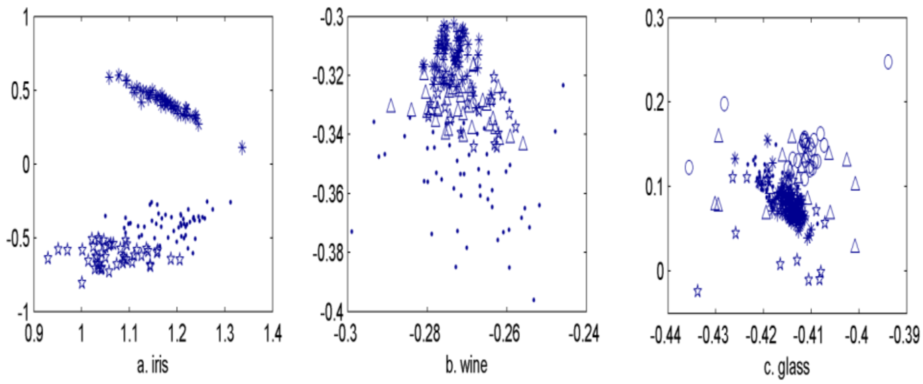

Fig. 4. Automatic Clusters Imparted by AutoSpssTLBO in Real-Time Datasets.

Table 7: Stratified Evaluation Metrics on Intended Algorithm over Wine Dataset.

Stratified evaluation metrics over wine dataset	AutoTLBO	AutoSpss TLBO	AutoITLBO
MAE	153.98	**121.97**	260.30
RMSE	151.06	**205.87**	315.21
RAE	59.15%	**46.85%**	100%
RRSE	65.31%	**47.92%**	100%
SSE	49.84	**49.84**	251.01

1.2 Interpreting Results on Micro-array Datasets

In this experiment, the results appeared in research articles [8, 11], [7] and [6] over micro-array datasets are consolidated, and a conceivable assessment of the intended methods is conducted. Table 8 represents the values occurred in the three algorithms.

Table 8 summarizes the results of all the three themes proposed in this work. Yeast datasets of different sizes are considered for observation. Interestingly AutoSpssTLBOexhibits better performance in attaining some automatic clusters, % of error rate and minimum error rate. The same was reciprocated in CVIs, i.e., in most of the cases, almost four CVIs favored AutoITLBO and is forefronted by AutoSpssTLBO. Table 9 presents centroid table of AutoITLBO and AutoSpssTLBO over yeast 238 dataset.

The centroids represented in this table are recorded after executing the algorithms after 50 independent runs. The centroid values are given as negatively scaled values in this yeast 238 datasets for all its 17 attributes. Table 10 exhibits the stratified evaluation metrics imposed on intended algorithm over yeast 238 dataset

Table 8: Mean Values of Automatic Algorithms after Completion of 50 Independent Runs over Micro-Array Datasets.

Datasets (dim, n)	**Algorithms**	**No. of Auto Clusters**	**ARI**	**RI**	**SIL**	**HIM**	**CS**	**DB**	**% of Error Rate**	**CPU Time (MSec)**
Yeast238 (238*19, 4)	AutoTLBO	4.20	**0.9210**	0.9720	0.9510	0.9440	1.6460	**1.0690**	**5.80**	25.07
	AutoSpssTLBO	**4.00**	0.6879	0.8537	0.6963	0.9702	0.8075	1.4912	83.07	355.56
	AutoITLBO	4.10	0.9140	**0.9821**	**0.9874**	**0.9948**	**0.9314**	1.1040	3.68	**22.14**
Yeast384 (348*19, 5)	AutoTLBO	5.40	**0.9520**	0.9070	0.8290	**0.9140**	1.6960	1.0880	8.937	40.84
	AutoSpssTLBO	**5.00**	0.5228	0.7930	0.4170	0.7461	0.5660	1.2424	65.20	593.40
	AutoITLBO	**5.10**	0.9580	**0.9877**	**0.9632**	1.2580	**1.0470**	**0.9784**	**8.67**	**38.96**
Yeast2885 (2885*19, 6)	AutoTLBO	6.40	**0.8210**	**0.8100**	**0.8910**	**0.7210**	2.9120	1.5050	**38.41**	1700.70
	AutoSpssTLBO	**6.10**	0.6502	0.6490	0.3510	0.1207	**1.1980**	1.6310	74.15	**1552.17**
	AutoITLBO	6.24	0.7400	0.7840	0.7140	0.7190	0.7630	**1.1480**	79.4	1593.5
Yeast2946 (2946*18, 6)	AutoTLBO	5.60	**0.8510**	**0.8790**	0.7260	**0.8590**	2.3250	1.5020	**35.47**	**1100.80**
	AutoSpssTLBO	**6.20**	0.6687	0.6896	0.4104	0.6093	**0.8792**	**0.8747**	81.02	3138.78

	AutoITLBO	5.49	0.7480	0.8150	**0.8430**	0.7140	0.7400	0.81400	78.90	3258.00
Yeast4382 (4382*25, 6)	AutoTLBO	6.20	**0.7210**	**0.6630**	**0.7470**	**0.8260**	3.7020	2.0160	**44.27**	3510.24
	AutoSpssTLBO	**6.00**	0.0125	0.6326	0.3674	0.7841	1.2653	**1.5218**	81.02	**3138.7**
	AutoITLBO	**6.00**	0.4890	0.6540	0.6410	0.6790	**0.8430**	1.8460	93.40	3364.1

The diagnose report on Table 10 justifies AutoSpssTLBO less sensitive to errors. Such implication justifies that the scores are given MAE and RMSE, i.e., 0.3611 and 0.4979 since these values are relatively low when compared to other two themes given in Table 8. The forecasted error % of AutoSpssTLBO algorithm's RAE and RRSE values and SSE are comparatively small than its contenders. The commendation from this Table 8 is the proposed AutoITLBO, and AutoSpssTLBO algorithms are noteworthy to apply over microarray datasets to extract patterns from microarray datasets.

Fig. 5 shows the clustergrams imparted by AutoSpssTLBO in yeast 238 micro-array datasets.Theclustergrams given in Fig. 5 isolatesets with alike traits and disperse them into clusters by scrutinizing the cluster affiliates.

Table 9. Centroid Table of Yeast 238 Dataset over AutoITLBO and AutoSpssTLBO.

Attributes of yeast 238 dataset	AutoITLBO				AutoSpssTLBO			
c1	0.39	-0.63	-0.73	-0.66	-0.65	-0.63	-0.09	-0.34
c2	0.19	-0.40	0.01	-0.41	-0.05	-0.46	0.30	-0.31
c3	-1.16	0.35	1.62	-0.20	1.69	0.37	-0.92	-0.74
c4	-0.99	-0.26	0.36	0.48	0.91	-0.25	-0.95	-0.30
c5	-0.71	0.06	-0.28	0.85	0.42	0.06	-0.59	0.07
c6	-0.51	-0.81	-0.50	0.91	0.05	-0.82	-0.48	0.28
c7	-0.52	-0.45	-0.39	0.18	-0.37	-0.44	-0.46	0.08
c8	-0.41	-0.47	-0.41	-0.02	-0.71	-0.44	-0.60	0.31
c9	0.19	-0.81	-0.17	-0.15	-0.63	-0.81	-0.03	0.39

c10	0.41	2.88	0.81	-0.63	0.11	2.97	0.71	-0.01
c11	0.48	0.24	0.99	-0.46	0.59	0.23	0.66	-0.24
c12	0.44	-0.13	0.49	-0.20	0.50	-0.15	0.41	-0.22
c13	-0.02	-0.70	-0.03	0.27	0.15	-0.71	-0.08	0.05
c14	0.19	0.18	-0.31	0.22	-0.15	0.19	0.15	0.15
c15	0.21	0.04	-0.43	0.09	-0.41	0.04	0.12	0.22
c16	0.80	0.54	-0.64	-0.08	-0.72	0.51	0.88	0.26
c17	1.03	0.35	-0.40	-0.20	-0.71	0.35	0.99	0.34

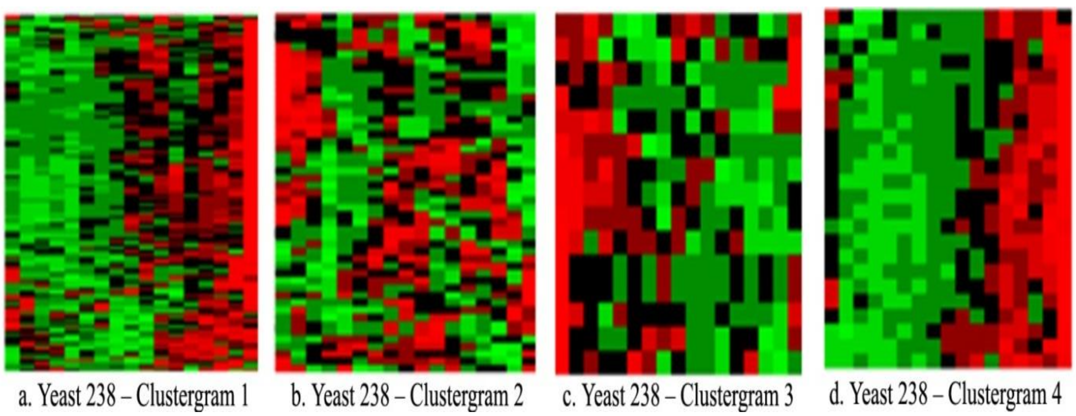

Fig. 5: Clustergramsimparted by AutoSpssTLBO in Yeast 238 Micro-Array Datasets.

Table 10: Stratified Evaluation Metrics on Intended Algorithm over Yeast 238 Dataset

Stratified evaluation metrics over yeast 238 dataset	**AutoTLBO**	**AutoSpssTLBO**	**AutoITLBO**
MAE	0.658	**0.3611**	0.5478
RMSE	0.384	**0.4979**	0.5159
RAE	65.28%	**61.36%**	69.47%
RRSE	77.94%	**67.14%**	69.32%
SSE	90.78	**88.79**	91.45

As this study is aimed to identify peak performers among the datasets shown in Table 1, a productive introspection on results of these set of three algorithms are conducted, based on the optimization metrics and stratified evaluation the following conclusion is drawn. Conceptually, all the algorithms are well coded by amalgamating partitioning clustering technique into evolutionary approaches and after that are optimized efficiently with a wide range of functions. The assumed set of three algorithms can automatically find an optimal number of the cluster in any dataset without any human intervention, relatively with low % of error rate and less CPU time. The adopted CVIs in this work were able to validate a reasonably good index function value with its high mathematical and statistical function.

An elegant and versatile impression from the user thought process after having live experiences over real-time and micro-array datasets are the second algorithms in the comparing algorithm AutoSpssTLBO evidently rationalize cluster problems to its best the first algorithms AutoTLBO. Hence the algorithms AutoSpssTLBOAutoTLBO are explicitly revealed and presented for natural interpretation to quote peak performers of this work. The methods were able to discover automatically statistically significant and hidden patterns in datasets for meaningful groupings.

3. Conclusion

With a small contribution to improving the general standard of conventional clustering towards automatic clustering, this study relates useful tips. The hope towards further recommendations and modification to the technology is to replace the baseline TLBO algorithm with any other evolutionary approach. In a nutshell, the subject of this study was to do automatic clustering by optimizing multiple objective functions in a single run with an evolutionary approach. To streamline this legacy, an automatic clustering framework is accelerated with an initial seed selection policy, into multiple variants of TLBO. This impressive recital manifested a set of three automatic clustering algorithms (AutoTLBO [8, 11], AutoSPSSTLBO [7], AutoITLBO [6]). All these possible algorithms accelerated self-learning clusters by ascertaining cluster properties, such as cohesion and separation and amides with minimum error rate over real-time and micro-array datasets. The results engendered by this work are well endorsed by CVI's and by retaining modest spike over other contending algorithms.

This study makes a philanthropic urge towards intuiting future research directions. The seductive logic in this evolutionary automatic clustering framework earmarked with an assemblage of initial seed selection algorithm, and cognitive advancements of TLBO. This proposal is adaptable to any innovative experimentation either by replacing SPSS with any other initial seed selection algorithm, nor the bed rocked TLBO evolutionary algorithm.

References

Baeck, T., D. B.Fogel, and Z. M. Michalewicz. Handbook of Evolutionary Computation New York. (1997).

Ball, Geoffrey H., and David J. Hall. ISODATA, A novel method of data analysis and pattern classification. Stanford research inst Menlo Park CA, 1965.

Davies, David L., and Donald W. Bouldin. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence 2 (1979): 224-227.

Frank A. UCI machine learning repository. http://archive. ics. uci. edu/ml. 2010, last accessed 2021/08/05.

Hubert, Lawrence, and James Schultz. Quadratic assignment as a general data analysis strategy. British journal of mathematical and statistical psychology 29.2 (1976): 190-241.

Kurada, Ramachandra Rao, and KarteekaPavanKanadam. A generalized automatic clustering algorithm using improved TLBO framework. International Journal of Applied Science and Engineering Research 4.4 (2015): 544-556.

Kurada, Ramachandra Rao, and KarteekaPavanKanadam. A Novel Evolutionary Automatic Clustering Technique by Unifying Initial Seed Selection Algorithms into Teaching–Learning-Based Optimization. Soft Computing and Medical Bioinformatics. Springer, Singapore, 2019. 1-9.

Kurada, Ramachandra Rao, and KarteekaPavanKanadam. A Novel Evolutionary Automatic Data Clustering Algorithm using Teaching-Learning-Based Optimization. International Journal of Intelligent Systems and Applications 10.5 (2018): 61.

Kurada, Ramachandra Rao, and KarteekaPavanKanadam. Automatic unsupervised data classification using Jaya evolutionary algorithm. Adv. Comput. Intell. Int. J 3 (2016): 35-42.

Kurada, Ramachandra Rao, Dr K. KarteekaPavan, and Dr AV Rao. A preliminary survey on optimized multiobjective metaheuristic methods for data clustering using evolutionary approaches.arXiv preprint arXiv:1312.2366 (2013).

Kurada, Ramachandra Rao, K. KarteekaPavan, and AllamAppa Rao. Automatic teaching–learning-based optimization: A novel clustering method for gene functional enrichments. Computational Intelligence Techniques for Comparative Genomics. Springer, Singapore, 2015. 17-35.

Kurada, Ramachandra Rao, KarteekaPavanKanadam, and R. C. Tripathi. A short study in formulation of TLBO and its synergized applications with clustering. 2015 International Conference on Advances in Computer Engineering and Applications. IEEE, 2015.

Pavan, K. Karteeka, et al. Robust seed selection algorithm for k-means type algorithms. arXiv preprint arXiv:1202.1585 (2012).

Rand, William M. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association 66.336 (1971): 846-850.

Rao, R. Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. International Journal of Industrial Engineering Computations 7.1 (2016): 19-34.

Rao, R. Venkata, and G. G. Waghmare. A comparative study of a teaching–learning-based optimization algorithm on multi-objective unconstrained and constrained functions. Journal of King Saud University-Computer and Information Sciences 26.3 (2014): 332-346.

Rao, R. Venkata, and Vivek Patel. An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems. Scientia Iranica 20.3 (2013): 710-720.

Rao, R. Venkata, V. J. Savsani, and J. Balic. Teaching–learning-based optimization algorithm for unconstrained and constrained real-parameter optimization problems. Engineering Optimization 44.12 (2012): 1447-1462.

Rao, R. Venkata, Vimal J. Savsani, and D. P. Vakharia. Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problems. Information sciences 183.1 (2012): 1-15.

Rao, R., and Vivek Patel. An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems. International Journal of Industrial Engineering Computations 3.4 (2012): 535-560.

Rao, R., and Vivek Patel. An elitist teaching-learning-based optimization algorithm for solving complex constrained optimization problems.International Journal of Industrial Engineering Computations 3.4 (2012): 535-560.

Rao, Ravipudi V., Vimal J. Savsani, and D. P. Vakharia. Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Computer-Aided Design 43.3 (2011): 303-315.

Rousseeuw, Peter J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65.

Daana, Hana Asaad. "Handling Coda Clusters Formed by Geminates in Different Arabic Dialects." International Journal of Linguistics and Literature (IJLL) 7.4 (2018): 65-80.

Ioannis, Chiotelis. "Clustering Behaviour of End-Grafted Polystyrene Chains with Zwitterionic Groups: Effect of Molecular Weight and Number of End-Groups." International Journal of Applied and Natural Sciences (IJANS) 7.6 (2018) 53-68

Sudha, T., and P. Nagendra Kumar. "Achieving Privacy Preserving Clustering in Images Through Multidimensional Scaling." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 6.1 (2016) 9-18

Sudha, T., and P. Nagendra Kumar. "Privacy Preserving Clustering by Adding Different Noise Components Generated from Progressions to Different Confidential Attributes." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 6.2 (2016) 11-16

Rajani, R., and T. Sudha. "Hybrid Kernel Fuzzy Clustering with Feed Lion Neural Network for Missing Data Imputation and Classification." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 7.5 (2017) 9-28

Padaganur, Satyanarayan K., and Jayashree Mallapur. "Improved Reposition Cluster Head in Leach Protocol for Heterogeneous Wireless Network." International Journal of Computer Networking, Wireless and Mobile Communications (IJCNWMC) 8.1 (2018) 11-18