www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)

Time-Series Clustering Benchmark on Regional Economic Indicator

Yudhistira Dharma Putra
Department of Accounting, Universitas Indonesia
y.dharma@ui.ac.id

Abstract

This paper presents a benchmark study on time-series clustering using regional economic data from the World Bank Open Data (WBOD) repository. It serves as a paradigm for future researchers. This study compares the effectiveness of twenty different techniques for time series grouping. They combine three clustering algorithms (partitional, hierarchical, and fuzzy), two centroids (K-means and K-medoids), and four distance measurements (distance between two points on a graph) (Dynamic time warping, Euclidean, shape-based distance, and global triagonal kernel alignment). The internal clustering validation index will be used to compare the performance of various techniques. Additionally, statistical tests are run on the performance of the pair of approaches to establish whether they can be compared. Across all clustering algorithms evaluated, it was discovered that utilizing K-means as centroids outperformed using K-medoids. When it comes to distance measurements, all clustering algorithms perform optimally, but the Triagonal Global Alignment Kernel is the best of these (except for the fuzzy C-means). Another conclusion reached in this study is that no solution utilizing Dynamic Time Warping and Euclidean distance measures can be compared to another (insignificant Wilcoxon test result). Simultaneously, Shape-Based Distance consistently beats all other approaches to clustering in terms of consistency.

Keywords
time-series clustering; benchmark; unsupervised learning; regional economy

I. Introduction

In general, data is classified into two types: static and transitory. The basic difference between these two types of data is a shift in time. When the value of a variable does not vary over time, it is classified as static data. If the value of a variable in data changes over time, the data is said to be temporal or time-series data (Özkoç, 2020). Statistical techniques such as regression, forecasting, and classification are used to evaluate time series data. However, due to the rapid advancement of technology and the advent of various new ideas such as big data and cloud computing, researchers must now employ unsupervised learning approaches such as clustering to derive information from time series data (Aghabozorgi et al., 2015).

Additionally, the WBOD archive has open data that is current and accurately represents the economic and financial situations in each country worldwide. Economic growth is still an important goal in a country's economy, especially for developing countries like Indonesia (Magdalena and Suhatman, 2020). Additionally, it emphasizes the need of conducting a benchmark approach on economic data, which is predicted to aid future studies. Apart from the dataset, the distinctions in this study include the clustering technique to be compared and the validation performance evaluation for each clustering method. The various approaches and validation performance are primarily determined by

Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Volume 5, No 1, February 2022, Page: 6453-6469
e-ISSN: 2615-3076 (Online), p-ISSN: 2615-1715 (Print)
www.bircu-journal.com/index.php/birci
email: birci.journal@gmail.com

the data itself. Because the WBOD data lacks ground truth labels, it is impossible to quantify validation performance natively, and some clustering techniques, such as density-based clustering, are inapplicable.

II. Review of Literature

2.1 Clustering Algorithm

a. Partitional and Fuzzy Clustering

Partitional clustering may also be classified into two types based on the existence or lack of overlapping partitions, called hard and soft clustering. In hard clustering, each data item is given to a single cluster. In comparison, with soft clustering (also known as fuzzy clustering), each data item can be allocated to many clusters. The application of K-means to a hard algorithm performs somewhat better than the application to a soft algorithm (Javed et al., 2020). However, the performance of k-medoids has never been tested between hard and soft algorithms. As a result, K-medoids and K-means will be used in both hard and soft clustering in this work. In fuzzy clustering, K-means are referred to as Fuzzy C-means (FCM), whilst K-medoids are referred to as Fuzzy C-medoids (FCMdd).

b. Hierarchical Clustering

Hierarchical clustering algorithms are centered on either segmenting clusters into subgroups that are handled sequentially as a whole or on aggregating individual clusters into clusters sequentially (Özkoç, 2020). Hierarchical clustering techniques are classified into two types: agglomerative clustering techniques and divisive hierarchical clustering techniques centered on dendrogram creation. Due to their popularity (Javed et al., 2020), agglomerative clustering techniques will be employed in this study.

2.2 Distance Measures

a. Dynamic Time Warping (DTW) Distance

DTW is a technique for calculating non-linear distances. Unlike Euclidean Distance, which calculates distances based on identical data positions and adding them up, DTW calculates distances by minimizing the overall distance while computing each data location. When the result of the calculation on the same index is larger than the result of the calculation on a different index, the result of the calculation on the different index is utilized.

b. Euclidean Distance

The Euclidean distance, often known as the L2 metric, is the “standard” straight-line distance between two places computed using a ruler. Euclidean distance was chosen because it is widely used in economics, is convenient to use, and is noticed by a large number of people in everyday life, which is why it is the top candidate in economics (Li and Wang, 2022; Buckzkowska et al., 2019). Shape-based Distance

c. Triagonal Global Alignment Kernel (TGAK) Distance

TGAK is a more accurate version of the GAK distance metric. It runs into GAK constraints, such as diagonal dominance and complexity P. (nm). Additionally, by employing the triangular local integral kernel (Equation 5), the intricate structure of the GA kernel, which represents the kernel's order, may be minimized. TGAK (Equation 7) is derived by combining the kernel k (Equation 6a & 6b) based on the Gaussian kernel $\mathbf{k}_\sigma$.

This kind of kernel can be measured as P(T min(n, m)), and the triangular limit T and Gaussian kernel width σ parameterized.

2.3 Assessment Metrics

There are two main forms of measurement in the clustering algorithm: extrinsic and intrinsic. The distinction between these two categories is whether or not the supplied dataset contains ground truth labels. Due to the absence of ground truth labels in the dataset utilized in this study, the type of measurement used is intrinsic, or what is known as an internal clustering validation score.

2.4 Experimental Setups

The World Bank Open Data (WBOD) dataset was used in this research effort. WBOD is one of the world's largest platforms for regional economic data. WBOD is renowned for providing fast and statistically accurate statistics.

On the existing datasets, some pre-processing was performed. This procedure is more concerned with the selection of nations and the time period covered by the analysis. Each dataset in this research has a minimum of 33% of all nations, or 70 countries. Meanwhile, each dataset requires a minimum of 20 consecutive years. In other words, the data utilized is the only time series data with a constant length. Following this pre-processing, Table 2 shows the number of countries and years included in each dataset.

Table 1. Datasets Used From World Bank Open Data Archive

No.	Datasets	Explanation
1	AFFV	Agriculture, forestry, and fishing,
2	BRDM	Broad money
3	CAB	Current account balance
4	CONS	Final consumption expenditure
5	CPI	Consumer price index
6	DEPR	Deposit interest rate
7	EXCR	Official exchange rate
8	FDI	Foreign direct investment, net
9	GDP	Gross domestic product
10	GDPC	GDP per capita
11	GNI	Gross national income, Atlas method
12	GNIC	GNI per capita, Atlas method
13	INF	Inflation, GDP deflator
14	INVC	Changes in inventories
15	LENR	Lending interest rate
16	MANV	Manufacturing value added
17	NNI	Adjusted net national income
18	NNIV	NNI per capita
19	NTGS	Net trade in goods and services
20	RELR	Real interest rate
21	RESG	Total reserves
22	UNMP	Unemployment, total

III. Research Method

This study is a benchmark study that aims to compare which clustering method is better than other methods from the previously selected methods. A total of 20 combined clustering algorithms and distance measures (Table 3) will be applied to each of the 22 datasets used for the research project. A package called "dtwclust" on R software will mostly be used in this research project.

Table 2. Total Countries and Periods Detail per Dataset

No.	Datasets	Total Countries	Periods	Total Year
1	AFFV	70	1970-2018	49
2	BRDM	71	1967-2018	52
3	CAB	85	1977-2018	42
4	CONS	90	1970-2018	49
5	CPI	75	1965-2018	54
6	DEPR	73	1995-2018	24
7	EXCR	129	1960-2018	59
8	FDI	71	1983-2018	36
9	GDP	91	1960-2018	59
10	GDPC	91	1960-2018	59
11	GNI	71	1963-2018	56
12	GNIC	71	1963-2018	56
13	INF	87	1961-2018	58
14	INVC	75	1980-2018	39
15	LENR	74	1995-2018	24
16	MANV	71	1978-2018	41
17	NNI	123	1971-2018	48
18	NNIV	71	1971-2018	48
19	NTGS	87	1977-2018	42
20	RELR	73	1996-2018	23
21	RESG	77	1960-2018	59
22	UNMP	186	1991-2018	28

Table 3. 20 Methods Used in the Benchmark Study

No.	Methods
1	K-means, Euclidean, Partitional
2	K-means, DTW, Partitional
3	K-means, SBD, Partitional
4	K-means, TGAK, Partitional
5	K-medoids, Euclidean, Partitional
6	K-medoids, DTW, Partitional
7	K-medoids, SBD, Partitional
8	K-medoids, TGAK, Partitional
9	Fuzzy C-means, Euclidean
10	Fuzzy C-means, DTW
11	Fuzzy C-means, SBD

No.	Methods
12	Fuzzy C-means, TGAK
13	Fuzzy C-medoids, Euclidean
14	Fuzzy C-medoids, DTW
15	Fuzzy C-medoids, SBD
16	Fuzzy C-medoids, TGAK
17	Agglomerative, Euclidean, Hierarchical
18	Agglomerative, DTW, Hierarchical
19	Agglomerative, SBD, Hierarchical
20	Agglomerative, TGAK, Hierarchical

IV. Results and Discussion

4.1 Results

a. Internal CVI Score

This section will show descriptive statistics and boxplots for each internal CVI score obtained by each clustering technique in order to gain an early understanding of the scores generated.

1. Partitional Clustering

Table 4 summarizes the internal scores of the CVI for partitional clustering. Each CVI has 176 scores resulting from the application of eight clustering algorithms to 22 datasets of regional economic data. The higher the score of the first four internal CVIs (CH, D, SF, and Sil), the better the performance of a clustering technique. The "08" approach (Dataset: UNMP; Score 550.384) produced the highest CH score, the "02" method produced the highest D and Sil scores (Dataset: BRDM; Scores 2.742 & 0.988), and the "04" method produced the greatest SF score (Dataset: NNIC; Score: 0.632).

Table 4. Statistic Descriptive of Internal CVI Score's on Eight Partitional Clustering Methods

Internal CVI	n	min	max	median	mean	sd
CH	176	2.44	550.38	57.55	95.09	93.73
COP	176	0.00	0.32	0.03	0.05	0.06
D	176	0.00	2.74	0.06	0.19	0.37
DB	176	0.03	30.22	0.69	1.05	2.34
DBstar	176	0.03	30.22	0.82	1.58	3.41
SF	176	0.00	0.63	0.19	0.29	0.30
Sil	176	0.15	0.99	0.75	0.70	0.19

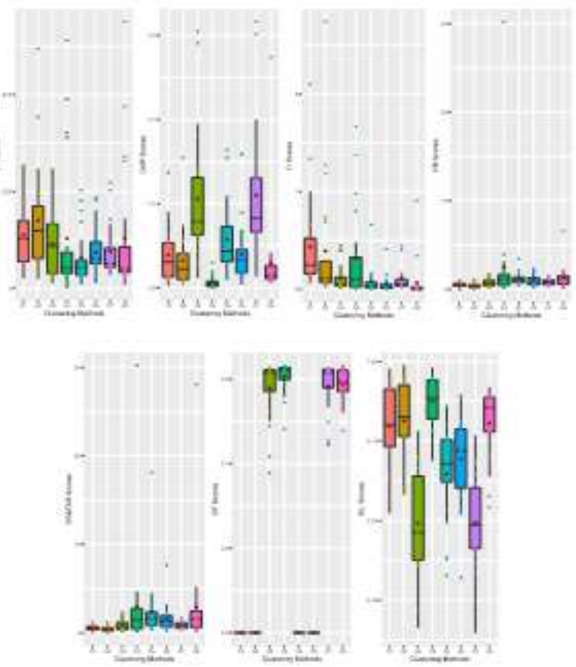

Figure 1. *Boxplot of Internal CVI Score's on Eight Partitional Clustering Methods*

2. Hierarchical Clustering

Internal CVI scores on hierarchical clustering are summarized in Table 5. Each CVI has 88 scores obtained from four clustering algorithms applied to 22 datasets of regional economic data. The clustering technique on the BDRM dataset produced the highest scores for CH, D, Sil, and SF. The application of method "20" has the maximum value in CH (Score: 28,342.026) and D (Score: 73.956). Meanwhile, method "17" has the highest SF (Score: 1) and method "18" has the highest Sil (Score: 0.988).

Table 5. Statistic Descriptive of Internal CVI Score's on Four Hierarchical Clustering Methods

Internal CVI	n	min	max	median	mean	sd
CH	88	4.67	2.E+4	56.62	410.71	3.E+3
COP	88	0.00	0.35	0.02	0.05	0.07
D	88	0.06	73.96	0.61	1.84	7.91
DB	88	0.00	0.70	0.14	0.20	0.17
DBstar	88	0.00	1.06	0.16	0.23	0.21
SF	88	0.00	1.00	0.52	0.34	0.33
Sil	88	0.36	0.99	0.85	0.81	0.15

Internal CVI ratings on fuzzy clustering are summarized in Table 6. Analogous to partition clustering, after applying eight clustering methods to 22 regional economic datasets, 176 scores were obtained. The higher the score of a clustering method's first three internal CVIs (MPC, PBMF, and SC), the better. The highest scores for MPC, PBMF, and SC are from applying the clustering algorithm to the BDRM dataset. The highest SC (Score: 193,562.80) and MPC (Score: 0.999) values are from applying method "12" to the datasets. Furthermore, method "10" has the highest PBMF value (7.78E+35 score).

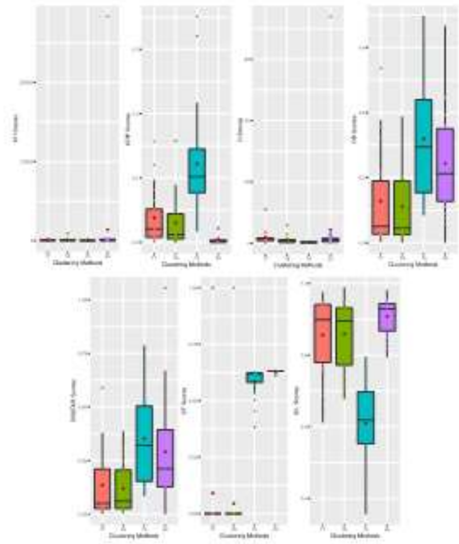

Figure 2. *Boxplot of Internal CVI Score's on Four Hierarchical Clustering Methods*

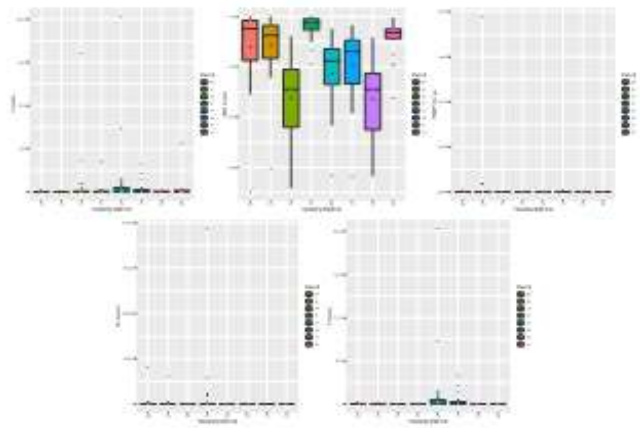

Figure 3. *Boxplot of Internal CVI Score's on Eight Fuzzy Clustering Methods*

Table 6. Statistic Descriptive of Internal CVI Score's on EIGHT FUZZY Clustering Methods

Internal CVI	n	min	max	median	mean	sd
K	176	0.44	2.E+4	20.91	430.55	2.E+3
MPC	176	0.13	1.00	0.84	0.78	0.21
PBMF	176	0.00	8.E+3	5.E+3	5.E+3	6.E+3
SC	176	-1.00	1.E+5	2.91	1.E+3	1.E+4
T	176	0.01	2.E+4	1.43	241.17	1.6

b. Statistical Test Results

This section will discuss the statistical tests (Friedman and Wilcoxon) results on each clustering algorithm.

1. Friedman Test

The Friedman test was run on the data prior to comparing the scores for each clustering method. The Friedman test is a non-parametric method for testing a number (k) of paired samples with a minimum of ordinal data scale. Pairing can refer to a combination of treatments, implying that the treatment is repeated in each condition. The Friedman test's explanation is presented in the form of a two-way table with n rows and k columns. The column denotes the patient (datasets), while the row denotes the treatment (methods).

Table 7 shows the result of the Friedman test for scores on partitional clustering and Table 8 shows the result on hierarchical clustering. It can be seen that for all scores on both algorithms, the p-value is less than alpha which H0 will be rejected. The rejection of H0 on the Friedman test means that at least one type of method in the partitional and hierarchical clustering algorithm has different scores.

Table 7. Friedman Test Results for Partitional Clustering

Internal CVI	n	statistic	df	p
SF	22	133.3	7	1.28E-25***
SIL	22	102.25	7	3.69E-19***
COP	22	99.09	7	1.66E-18***
D	22	74.71	7	1.64E-13***
DBSTAR	22	39.71	7	1.43E-06***
DB	22	35.94	7	7.44E-06***
CH	22	30.68	7	7.12E-05***

Table 8. Friedman Test Results for Hierarchical Clustering

Internal CVI	n	statistic	df	p
SF	22	52.75	3	2.08E-11***
SIL	22	49.37	3	1.09E-10***
COP	22	40.8	3	7.21E-09***
D	22	38.35	3	2.39E-08***
DBSTAR	22	20.67	3	1.20E-04***
DB	22	17.4	3	5.80E-04***
CH	22	16.25	3	1.01E-03***

For fuzzy clustering, the same result is shown in Table 9. Due to the fact that all Friedman tests for scores have a p-value less than alpha, H0 is likewise rejected. Thus, at least one method type in the fuzzy clustering algorithm has distinct scores. In conclusion, it can be stated that the data in each of the three clustering algorithms is unique.

Table 9. Friedman Test Results for Fuzzy Clustering

Internal CVI	n	statistic	df	p
PCMF	22	136.86	7	2.31E-26***
T	22	121.59	7	3.57E-23***
MPC	22	98.91	7	1.81E-18***
K	22	57.79	7	4.17E-10***
SC	22	44.74	7	1.53E-07***

2. Wilcoxon Test:

The Wilcoxon test is a non-parametric test that tests the variations' magnitude, but with an irregular distribution, between 2 paired data groups. The Wilcoxon test is an alternate test to the paired t-test where the normality expectation is not fulfilled. This test is also known as the Match Pair Test for Wilcoxon. The theories or criteria of this test include:

a) The dependent variable has an interval size, at least, but the distribution is not natural.
b) The independent variable is made up of 2 paired categories. The pairing indicates that the same observation is the subject as the source of the results.
c) The data form and distribution is symmetrical between the two paired sets.

Summary from the full result can be seen in Table 10, and it can be seen that there are 65 out of 196 pairs (33.16%) that have a significant value of Wilcoxon test for partitional clustering.

Table 10. Number of Significant Wilcoxon Test Holm-Adjusted P-Values for Partitional and Hierarchical Clustering

Internal CVI	Partitional	Hierarchical
CH	0	2
COP	17	5
D	14	3
DB	7	4
DBSTAR	8	4
SF	0	4
SIL	19	3
Significant Pairs (A)	65	25
Total Battlegrounds (B)	196	42
% (A/B)	33.16%	59.52%

In terms of hierarchical clustering, Table 10 indicates that 25 of 42 pairs (59.52 %) have a significant Wilcoxon test value. This percentage is arguably quite large. Even more significantly, one combination, “18 vs. 19,” has significant outcomes in all Internal CVI. Meanwhile, the comparison of “17” and “18” yields no significant result, implying that these methods are not comparable and will not be examined further.

Table 11. Number of Significant Wilcoxon Test Holm-Adjusted P-Values for Fuzzy Clustering

Internal CVI	Fuzzy
CH	12
COP	17
D	19
DB	15
DBSTAR	23
Significant Pairs (A)	86

Total Battlegrounds (B)	140
% (A/B)	61.34%

Finally, Table 11 summarizes the Wilcoxon test for fuzzy clustering. As can be seen, 86 of 140 pairs (61.34 %) have a significant Wilcoxon test result. This is the highest proportion of any clustering algorithm. Each Internal CVI Score has a minimum of 12 pairs with a significant p-value, with the PBMF score having the highest occurrences in 19 of 28 couples. Meanwhile, three pairings, "09 vs. 10", "11 vs. 15", and "13 vs. 14", do not have any significant Wilcoxon test in any internal CVI Score. Due to the non-significant Wilcoxon test, these three methods are incomparable and will not be further compared.

c. Benchmarking Results

In this section, results from each pair/battlegrounds will be presented. The results will be divided into three subsections based on clustering algorithms. Each subsection will provide tables that contain the ranked list for clustering methods in each algorithm.

1. Partitional Clustering

After analyzing the overall number of WCs in each battleground, eight different ranking scenarios for partitional clustering algorithms are found. Numerous situations arise as a result of the impact of several pairs/battlegrounds that do not have significant Wilcoxon test results, particularly "01 vs. 02", "03 vs. 07", and "05 vs. 06". The final ranking results from a comparison of total WCs using partitional clustering algorithms are shown in Table 12. It demonstrates that technique "04," which utilizes K-means partitional clustering with the TGAK distance measure, consistently ranks highest in all cases. Additionally, approach "08", partitional clustering of K-medoids using the TGAK distance measure, consistently scores third in the scenario. Meanwhile, because techniques "01" (K-means with Euclidean distance) and "02" (K-means with DTW distance) are not comparable (due to the Wilcoxon test being negligible), there are no scenarios that combine both approaches. These two strategies, however, regularly outperform method "04" and method "08".

Table 13 shows that TGAK as a distance measure has better performance on partitional clustering than the other three distance measures. Meanwhile, SBD is consistently ranked last for each scenario. In terms of the selection of centroids, it can be seen from Table 14 that K-means are superior to K-medoids. K-means rankings that are below K-medoids occur when SBD is used as distance measures.

2. Hierarchical Clustering

There were 550 (25 pairs with significant Wilcoxon test on 22 datasets) WCs assigned to the methods being compared in each pair/battleground in hierarchical clustering. There are no paired method scores in the battleground that had the same value (tie score).

Table 12. Partitional Clustering Methods Rank in Eight Scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	4	1	8	5	3
II	4	1	8	6	3
III	4	1	8	5	7

IV	4	1	8	6	7
V	4	2	8	5	3
VI	4	2	8	6	3
VII	4	2	8	5	7
VIII	4	2	8	6	7

Table 13. Partitional Clustering Methods Distance Measures Rank in Eight Scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	TGAK	Euclidean	TGAK	Euclidean	SBD
II	TGAK	DTW	TGAK	DTW	SBD
III	TGAK	Euclidean	TGAK	Euclidean	SBD
IV	TGAK	DTW	TGAK	DTW	SBD
V	TGAK	Euclidean	TGAK	Euclidean	SBD
VI	TGAK	DTW	TGAK	DTW	SBD
VII	TGAK	Euclidean	TGAK	Euclidean	SBD
VIII	TGAK	DTW	TGAK	DTW	SBD

Table 14. Partitional Clustering Methods Centroids Rank in Eight Scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	Means	Means	Medoids	Medoids	Means
II	Means	Means	Medoids	Medoids	Means
III	Means	Means	Medoids	Medoids	Medoids
IV	Means	Means	Medoids	Medoids	Medoids
V	Means	Means	Medoids	Medoids	Means
VI	Means	Means	Medoids	Medoids	Means
VII	Means	Means	Medoids	Medoids	Medoids
VIII	Means	Means	Medoids	Medoids	Medoids

After comparing the total number of WCs in each battleground, two different ranking scenarios for hierarchical clustering algorithms are created. These possibilities emerge as a result of one pair/battleground not having the significant Wilcoxon test results indicated in the preceding section, specifically "17 vs 18". The final ranking results from a comparison of total WCs using hierarchical clustering approaches are shown in Table 15. It demonstrates that approach "20", Agglomerative hierarchical clustering with the TGAK distance metric, consistently ranks best in both circumstances. Meanwhile, because procedures "17" (Agglomerative with Euclidean distance) and "02" (Agglomerative with DTW distance) are incompatible (due to an insignificant Wilcoxon test), no scenario includes both approaches. Additionally, both strategies outperform method "20" and method "19".

Table 15. Hierarchical Clustering Methods Rank in Two Scenarios

Scenario	Rank		
	1st	2nd	3rd
I	20	17	19
II	20	18	19

Table 16. Hierarchical Clustering Methods Distance Measures Rank in Two Scenarios

Scenario	Rank		
	1st	2nd	3rd
I	TGAK	Euclidean	SBD
II	TGAK	DTW	SBD

Table 17. Fuzzy clustering methods rank in eight scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	9	12	16	11 = 13	N/A
II	10	12	16	11 = 13	N/A
III	9	12	16	11 = 14	N/A
IV	10	12	16	11 = 14	N/A
V	9	12	16	13 = 15	N/A
VI	10	12	16	13 = 15	N/A
VII	9	12	16	14	15
VIII	10	12	16	14	15

When viewed from a distance measure perspective, the ranking results are the same as partitional clustering. Table 16 shows that the use of TGAK as a distance measure has better performance than the other three distance measures. Meanwhile, SBD is also a distance measure with the worst performance.

3. Fuzzy Clustering

There were 1892 (86 significant Wilcoxon test pairs in 22 datasets) WCs allocated to the methods being contrasted with each pair/battleground in fuzzy clustering. Same with hierarchical clustering; there are no paired method scores in the battleground with the same value (tie score).

Table 18. Fuzzy clustering methods distance measures rank in eight scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	Euc	TGAK	TGAK	SBD = Euc	N/A
II	DTW	TGAK	TGAK	SBD = Euc	N/A
III	Euc	TGAK	TGAK	SBD = Euc	N/A
IV	DTW	TGAK	TGAK	SBD = Euc	N/A
V	Euc	TGAK	TGAK	SBD = Euc	N/A
VI	DTW	TGAK	TGAK	SBD = Euc	N/A
VII	Eu	TGAK	TGAK	DTW	SBD
VIII	DTW	TGAK	TGAK	Euc	SBD

By comparing the overall number of WCs in each battleground, eight ranking scenarios for hierarchical clustering algorithms are created. These possibilities arise as a result of three pairs/battlegrounds lacking the statistical significance described in the preceding section, namely “09 vs. 10”, “11 vs. 15”, and “13 vs. 14”. The final ranking results from the comparison of total WCs using fuzzy clustering algorithms are shown in Table 17. It demonstrates that the two ways are the most effective. The Wilcoxon test result for this technique pair, “09 vs. 10,” was not significant, rendering them incomparable. Meanwhile, methods “12” (FCM) and “16” (FCMdd), which employ TGAK as the distance measure, are ranked second and third in all scenarios, respectively. Additionally, as shown in Table 15, there exist ways with tied WCs, specifically “11 vs. 13”, “11 vs. 14”, and “13 vs. 15”. Both approaches have the exact same number of WCs in each pair/battleground, 22 WCs to be precise.

When the ranking results for each method in fuzzy clustering are examined in greater detail using distance measures and centroids, it becomes clear that the first rank in partitional clustering corresponds to the second rank in fuzzy clustering and vice versa. Similarly to partitional and hierarchical clustering, Tables 18 and 19 demonstrate that SBD approaches perform poorly in comparison to the others. Although this performance has a value in six cases, it is dependent on other approaches. One of the strategies associated with SBD (both in FCM and FCMdd) is one that employs Euclidean distance measurements and FCMdd centroids. This is in contrast to the results of partitional clustering, where the Euclidean and FCMdd methods outperform the SBD approach on all centroids. Meanwhile, links between methods utilizing DTW distance and FCMdd exist exclusively in SBD and FCM. When FCMdd are used as centroids, methods employing DTW as the distance measure performs better than the one employing SBD.

3.2 Discussions

This section analyzes the outcomes of the comparison process in the benchmark study for each clustering algorithm and offers a description of the study’s findings.

a. Partitional Clustering

The first main objective of this study is to find out which partitional clustering method performs best on regional economic time-series data. The comparison of the outcomes of each technique was mentioned previously, and it can be concluded that method “04,” K-means partitional clustering with TGAK as the distance measure, is the optimal way for partitional clustering on regional economic time series data. Additionally, these results indicate that TGAK performs the best of all distance metrics. At the time of writing, no study had ever used TGAK as a distance measure in benchmarking studies.

Table 19. Fuzzy clustering methods centroids rank in eight scenarios

Scenario	Rank				
	1st	2nd	3rd	4th	5th
I	FCM	FCM	FCMdd	FCM = FCMdd	
II	FCM	FCM	FCMdd	FCM = FCMdd	
III	FCM	FCM	FCMdd	FCM = FCMdd	
IV	FCM	FCM	FCMdd	FCM = FCMdd	
V	FCM	FCM	FCMdd	FCM = FCMdd	
VI	FCM	FCM	FCMdd	FCM = FCMdd	
VII	FCM	FCM	FCMdd	FCMdd	FCMdd
VIII	FCM	FCM	FCMdd	FCMdd	FCMdd

However, if the time series data's features, such as the amount of outliers, are known, K-medoids are preferable. This is due to the fact that K-medoids are less susceptible to outliers than K-means (Malik and Tuckfield, 2019).

b. Hierarchical Clustering

The second primary purpose is to evaluate which hierarchical clustering algorithm works the best on time series data from regional economies. Four approaches were applied in 22 datasets, and internal CVI was utilized to compare the performance of each method. Hierarchical clustering was also performed using internal CVI. The comparison of the outcomes of each technique was mentioned previously, and it can be concluded that method “20,” Agglomerative hierarchical clustering with TGAK as the distance measure, is the optimal way for hierarchical clustering on regional economic time series data. This result is identical to the one previously reported for partitional clustering. Comparisons between methods that use Euclidean and DTW distance measures are not possible, and SBD's low performance is also evident in the results of hierarchical clustering.

c. Fuzzy Clustering

The study's final objective is to determine which approach performs the best on regional economic time series data using fuzzy clustering. On 22 datasets, eight algorithms were implemented, and their performance was compared using the five internal CVIs. The previous part covered the comparison of fuzzy clustering approaches, and it can be stated that two methods are the best for fuzzy clustering in time series of regional economic data. These are approaches “09,” which uses fuzzy C-means with the Euclidean distance measure, and “10,” which uses fuzzy C-means with the DTW distance measure. As is the case with partitional and hierarchical clustering, these two approaches cannot be compared because the Wilcoxon test value is not significant, resulting in each method becoming the best option for specific conditions (in total, each method ranks first in the four scenarios). The superior performance of Euclidean and DTW as distance measures in FCM clustering is in contrast to the results of partitional and hierarchical clustering.

The benchmarking result indicates that the winning margins for methods "09" and "10" over method “12” are not enormous, at 18% for method "09" and 21% for method “10”. This little margin indicates that the TGAK method on FCM outperforms methods “09” and “10” on some datasets. As a result, the implementation of this TGAK increased the accuracy, efficiency, and performance of FCM when applied to regional economic time series data. However, due to this TGAK sensitivity, the enormous gain in FCM performance on various datasets is hampered. Finally, the total number of toilets obtained using method “12” was less than those obtained using methods “09” and “10”. Meanwhile, similar to partitional and hierarchical clustering, SBD is the least performant distance metric.

Similarly to the partitional clustering finding, it is well established that K-means as centroids (FCM) performs better than K-medoids (FCMdd) in Fuzzy Clustering. While the four performance situations for FCMdd are related to FCM, this is primarily due to the fact that the methods that use FCM employ SBD as a distance measure.

V. Conclusion

The purpose of this study is to undertake a benchmark study on time series clustering using 22 regional economic datasets from the WBOD archive. The findings of this study are also expected to be utilised by other researchers in time-series clustering investigations on similar datasets and other benchmark studies. There are twenty clustering algorithms: a mix of three clustering algorithms (partitional, hierarchical, and fuzzy) and four distance measurements (Dynamic Time Warping, Euclidean, Shape-based Distance Triagonal Global Alignment Kernel). The internal cluster validation index (CVI) was used to compare these approaches, with seven for partitional and hierarchical clustering and five for fuzzy clustering. Additionally, the Friedman and Wilcoxon tests were used to compare each pair of approaches. This test was designed in such a way that only pairs with a significant result would be compared. The benchmark findings indicate that methods based on K-means outperform methods based on K-medoids as centroids.

Meanwhile, TGAK performs best for all clustering algorithms except fuzzy clustering, where approaches utilizing DTW and Euclidean as distance measures perform better. The disparate findings in fuzzy clustering are due to the fact that the performance of TGAK is diluted by the algorithm's level of sensitivity. Additionally, it was determined that the approaches utilizing DTW and Euclidean were incomparable for each clustering method. The fact that DTW and Euclidean are incomparable implies that these two approaches are substitution methods. Finally, it was discovered that approaches that employ SBD as the distance measure perform the worst of all clustering algorithms.

One of the practical ramifications of this benchmark study is that researchers who conduct further investigations using the same time-series clustering approach and data set can use and adapt these benchmark results. If more researchers have alternative methods and datasets, they can use them to conduct another benchmark study. In general, this research is expected to add to the community's and repositories' knowledge base regarding time series clustering and unsupervised learning. It is feasible to use all accessible data in the WBOD archives pertaining to regional economies in future work to supplement the benchmark research with generalizations. Additionally, by incorporating the divisive hierarchical clustering method, fresh insights will be gained.

References

Aghabozorgi, S., Shirkhorshidi, A. S., and Wah, T. Y. (2015). Time-series clustering – A decade review. *Information Systems*, vol. 53, pp. 16-38.

Chen, L., and Wan, S. (2021). Intelligent fault diagnosis of high-voltage circuit breakers using triangular global alignment kernel extreme learning machine, ISA Transactions, vol. 109, pp. 368–379.

Du, S. Wu, M., Chen, L., Cao, W., and Pedrycz, W. (2020). Operating mode recognition of iron ore sintering process based on the clustering of time series data. *Control Engineering Practice*, vol. 96, p. 104297.

Esmaili, N., Buchlak, Q. D., Piccardi, M., Kruger, B., and Girosi, F. (2021). Multichannel mixture models for time-series analysis and classification of engagement with multiple health services: An application to psychology and physiotherapy utilization patterns after traffic accidents. *Artificial Intelligence in Medicine*, vol. 111, p. 101997.

Feng, X., Zhang, X., and Xiang, Y. (2020). An inconsistency assessment method for backup battery packs based on time-series clustering. *Journal of Energy Storage*, vol. 31, p. 101666.

Franses, P. H., and Wiemann, T. (2020). Intertemporal similarity of economic time series: An application of Dynamic Time Warping. *Computational Economics*, vol. 56, no. 1, pp. 59-75.

Gorbatiuk, K., Mantalyuk, O., Proskurovych, O., and Valkov, O. (2019). Analysis of regional development disparities in Ukraine with fuzzy clustering technique. *SHS Web of Conferences*, vol. 65, p. 04008.

Graves, D., and Pedrycz, W. (2010). Proximity fuzzy clustering and its application to time series clustering and prediction. *2010 10th International Conference on Intelligent Systems Design and Applications*.

Großwendt, A., Röglin, H., and Schmidt, M. (2019) Analysis of Ward's method, in Proc. *30th Annual ACM-SIAM Symposium on Discrete Algorithms*, San Diego, PA, USA, pp. 2939-2957.

Hu, G., and Du, Z. (2019). Adaptive kernel-based fuzzy C-means clustering with spatial constraints for image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 01, p. 1954003.

Javed, A., Lee, B. S., and Rizzo, D. M. (2020). A benchmark study on time series clustering. *Machine Learning with Applications*, vol. 1, p. 100001.

Johnpaul, C., et al. (2020). Trendlets: A novel probabilistic representational structures for clustering the time series data. *Expert Systems with Applications*, vol. 145, p. 113119.

Leverger, C., Malinowski, S., Guyet, T., Lemaire, V., Bondu, A., and Termier, A. (2019). Toward a framework for seasonal time series forecasting using clustering. *Intelligent Data Engineering and Automated Learning – IDEAL 2019*, pp. 328–340.

Li, D.-D. and Wang, Z.-X. (2022). Measurement methods for relative index of Financial Inclusion. *Applied Economics Letters*, pp. 1–7.

Magdalena, S., Suhatman, R. (2020). The Effect of Government Expenditures, Domestic Invesment, Foreign Invesment to the Economic Growth of Primary Sector in Central Kalimantan. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*. Volume 3, No 3, Page: 1692-1703.

Malik, A., and Tuckfield, B. (2019). Introduction to Clustering Methods, in *Applied unsupervised learning with r: Uncover hidden relationships and patterns with K-means clustering, and PCA*, Birmingham, UK: Packt, pp. 1–49.

Niennattrakul, V., and Ratanamahatana, C. A. (2007) On clustering multimedia time series data using K-means and dynamic time warping. *2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*.

Özkoç, E. E. (2020). Clustering of Time-Series Data, in Data Mining – Methods. *Applications and Systems*, 1st ed. London, United Kingdom: IntechOpen, pp. 1-19. [Online]. Available: https://www.intechopen.com/books

Paparrizos, J. and Gravano, L. (2016). K-shape. *ACM SIGMOD Record*, vol. 45, no. 1, pp. 69–76.

Paparrizos, J., and Gravano, L. (2017) Fast and accurate time-series clustering. *ACM Transactions on Database Systems*, vol. 42, no. 2, pp. 1–49.

Putri, R. A., Rustam, Z., and Pandelaki, J. (2019). Kernel based fuzzy C-means clustering for chronic sinusitis classification. *IOP Conference Series: Materials Science and Engineering*, vol. 546, no. 5, p. 052060.

Rahman, M. A., Zaman, N., Asyhari, A. T., Al-Turjman, F., Bhuiyan, M. Z. A., and Zolkipli, M. F. (2020) Data-driven dynamic clustering framework for mitigating the adverse economic impact of covid-19 lockdown practices. *Sustainable Cities and Society*, vol. 62, p. 102372.

Sardá-Espinosa, A. (2019). Time-series clustering in R using the DTWCLUST package. *The R Journal*, vol. 11, no. 1, p. 22.

Steinmann, P., Auping, W. L., and Kwakkel, J. H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, vol. 156, p. 120052.

Wang, H., Zhou, B., Zhang, J., and Cheng, R. (2020). A novel density peaks clustering algorithm based on local reachability density. *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, p. 690.

X. Yu and S. Xiong, A dynamic time warping based algorithm to evaluate Kinect-enabled home-based physical rehabilitation exercises for older people. *Sensors*, vol. 19, no. 13, p. 2882, 2019.