

UDC 519.218

Comparison of Model-Free Algorithms For Clustering GARCH Processes

Garik L. Adamyan

Yerevan State University
e-mail: garik.adamyan@ysu.am

Abstract

In this paper, we evaluate several model-free algorithms for clustering time series datasets generated by GARCH processes. In extensive experiments, we generate synthetic datasets in different scenarios. Then, we compare K-Means (for Euclidian and dynamic time warping distance), K-Shape, and Kernel K-Means models with different clustering metrics. Several experiments show that the K-Means model with dynamic time warping distance archives comparably better results. However, the considered models have significant shortcomings in improving the clustering accuracy when the amount of information (the minimum length of the time series) increases, and in performing accurate clustering when data is unbalanced or clusters are overlapping.

Keywords: Time series clustering, GARCH process, dynamic time warping, K-Means, K-Shape.

Article info: Received 02 May 2022; received in revised form 20 July 2022; accepted 29 September 2022.

1. Introduction

Time series clustering has been used in diverse scientific disciplines to discover patterns and extract valuable information from complex and massive datasets. These algorithms have a wide range of applications in many research areas, for instance, in finance, biology, and robotics [1].

Time series clustering approaches can be classified as feature-based, shape-based, and model-based [1]. It is noteworthy that these methods are based on dissimilarity measures on time series data, according to which the time series data points are grouped by some clustering method (for instance, PAM).

In general, shape-based methods use linear and non-linear transformations to align time series samples and calculate dissimilarity measures on aligned samples. Additionally, shape-based algorithms process the time series data directly without making any statistical assumptions about the underlying data generating processes. On the contrary, model-based methods make statistical assumptions on time series generating processes. In general, model-based approaches assume that time series samples are generated from specific models (for

instance, ARIMA [2], Mixtures of ARIMAs [3]). Time series samples are transformed into fitted models, and then a suitable distance and a clustering algorithm are applied to the estimated model parameters.

Although several benchmarking results on different real-world datasets for non-parametric clustering methods can be found in ([4], [5], [6]), the comparison of non-parametric clustering methods on time series data generated from GARCH processes is not well studied. In this paper, we are interested in non-parametric models evaluation of time series data generated from the well-known GARCH process, which is the actual choice for modeling the volatility of returns on financial assets. We simulate multiple GARCH models with different data generating scenarios and compare several non-parametric time series clustering models.

Motivated by [4], for comparison we choose well-known partition-based time series clustering models: K-Means, K-Means with dynamic time warping and DTW barycenter averaging, K-Shape and Kernel K-Means models. Furthermore, we can find open-source implementations of these algorithms [7].

Although the main focus in the field of time series clustering comparison remains clustering accuracy metrics, in this work we also explore a number of other challenges of model-free methods. In particular, we study the ability of the above-mentioned model-free methods to cluster GARCH processes with imbalanced, overlapping clusters and also examine the impact of increasing information on clustering accuracy.

2. Related Work

In time series analysis research, benchmarking and numerical comparison have been recognized as integral steps to justify theoretical results. The importance of numerical comparison is emphasized in [8], where the authors reimplemented many time-series classification algorithms and compared them in 50 real-world datasets. The authors note that most reported methods have insignificant improvements regarding the variance of the evaluation metrics. This empirical evidence reclaimed the statement of the importance of the time series benchmark datasets and the empirical evaluation of the suggested methods.

Among the works that compare time series clustering models based on real-world datasets, we can mention ([4], [5], [6]) works. In [4], authors compare several partition, density, and hierarchical clustering methods to cluster all time series datasets available in the University of California Riverside (UCR) archive [9]. They conclude that the overall performance of the eight compared algorithms is quite similar with high dependence on the evaluation dataset.

The method of comparing time series clustering algorithms with synthetic, generated datasets also attracts a lot of attention among scholars. In addition to the actual clusters being known, this comparison method gives additional flexibility to examining the behavior of algorithms in different situations. In particular, scholars discussed the difference between stationary and non-stationary time series [10], the presence of noise in time series samples [11], the presence of noise clusters in time series dataset [11].

3. Clusters of GARCH

The GARCH process is introduced in [12] for statistical modeling of the volatility of returns on financial assets. The GARCH model has many extensions such as asymmetric GARCH [13], threshold GARCH [14]. The GARCH(p,q) model is defined as follows:

$$y_t = \mu_t + \epsilon_t$$

$$\epsilon_t = \sigma_t e_t, \quad \text{where } e_t \text{ i.i.d } E(e_t) = 0, \text{var}(e_t) = 1,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2,$$

where

$$\begin{aligned} \omega &> 0, \\ \alpha_i &\geq 0, i = 1, 2, \dots, p, \\ \beta_j &\geq 0, j = 1, 2, \dots, q. \end{aligned}$$

The GARCH(p, q) model admits a strictly stationary solution with a finite variance if and only if

$$\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1. \quad (1)$$

Moreover, this strictly stationary solution is also unique. [15]

For the evaluation of non-parametric models, we chose constant zero mean specification for the GARCH model because it is advised to standardize input data prior to clustering. In addition, we choose the innovations e_t as standard Gaussian innovations. So $\mu_t = 0$ and $e_t \sim \mathcal{N}(0,1)$.

In order to measure the clustering accuracy, we need to define the ground truth clusters of GARCH processes. Let $N, K, T \in \mathbb{N}$ where K is the number of clusters, N is the number of samples and T is the time sample size of each series. In this paper, we consider samples with a fixed time size T , because some of the models (ex. KM-E) support samples with fixed length. We denote by $P^i = (\omega, \alpha_1, \alpha_2, \dots, \alpha_{p_i}, \beta_1, \beta_2, \dots, \beta_{q_i})$ the vector of all parameters for the given GARCH(p_i, q_i) model.

Let $\{P^i\}_{i=1}^K$ be a family of GARCH process parameters, where K is a number of clusters. Assume that each $P^i (i = 1, 2, \dots, K)$ is unique and all the parameters satisfy (1) in order to provide a strict stationary solution of the corresponding model. We are given N samples of time series $Y_i = \{y_t^i\}_{t=1}^T$, where each sample is generated from one of the K GARCH processes.

Definition 1. *We say that Y_i and Y_j samples are from the same cluster if they are generated from the same GARCH process.*

In other words, a cluster of GARCH processes is a set of samples that are generated with the same parameters. The uniqueness of the parameters P^i and Definition 1 imply that the given sample belongs to exactly one cluster.

4. Evaluation Models

For evaluation, we choose well-known non-parametric time series clustering models such as K-Means with Euclidean (KM-E) and dynamic time warping metrics (KM-DTW), K-Shape, and Kernel K-Means with Fast Global Alignment Kernel (KKM-GAK) models. KM-E uses Euclidean distance, for cluster assignment and means averaging for the barycenter (centroid) computation. It is known that the Euclidean distance metric is not the most accurate metric for measuring time series similarities. Firstly, to use Euclidean distance, we need to take into account the order of elements in the time series; secondly, the Euclidean distance does not consider a phase shift between two curves or a length difference between the series. In this paper, we consider this model for comparison with more complex approaches.

KM-DTW uses dynamic time warping [16] for cluster assignment and DTW barycenter averaging (DBA)[17] algorithm for averaging time series within the same cluster.

k-Shape [18] is a partitional clustering algorithm that relies on an iterative refinement procedure similar to the one used in K-Means. To measure the distance between time series, K-Shape uses a normalized version of the cross-correlation measure to consider the shapes of time series while comparing them. During the iterative procedure, this model minimizes the sum of squared distances between the sequences of time series.

Kernel K-Means[19] is an alternative clustering algorithm that uses kernel functions as a nonlinear mapping from the input space to a higher dimensional space. By using kernels, Kernel K-Means can separate clusters in higher dimensional space, even if the input data is not non-linearly separable in the input space. For treating time series data, practitioners usually used Global Alignment Kernels [20]. We will refer to this algorithm KKM-GAK.

The problem is to generate synthetic datasets and evaluate non-parametric models for clustering time series processes generated by the GARCH model.

5. Assessment Metrics

In practice, the use of clustering methods is due to working with unlabeled datasets. As a result, we can find evaluation metrics that can evaluate clustering models without having labeled data. These types of metrics are called internal. By the method of our data generating process, we can use external measures, which assume that ground truth labels are available. Examples of this type of metrics are the Rand Index (RI) [21], the Adjusted Rand Index (ARI) [22], the Adjusted Mutual Information (AMI)[23].

Following the evaluation made in [4] in our study, we choose the Adjusted Rand Index, because the values of this metric are consistently low for random cluster assignments and do not depend on the number of clusters.

6. Experiments

To evaluate non-parametric models, we simulate random datasets with different setups. In the first experiment, we measure the ability of the models of clustering different numbers of clusters. For this purpose, we generate datasets for 2, 4, 8, and 10 clusters, respectively. For each number of clusters, we generate random parameter families, which satisfy (1) for guaranteeing a unique and stationary solution of processes. For the purpose of generating a family of parameters, we constrain the maximum length of p and q by 5. This constraint is inherited from the common choice of GARCH models with fewer parameters. For every

parameter vector P^i (cluster), we generate samples for the given cluster and separate them into training and testing parts (30% testing) and repeat this process for averaging purposes. In Table 1, we present the results of the first experiment evaluated with the AMI metric. We can see that the KM-DTW model outperforms other models. In the second experiment,

Table 1: AMI score for different N clusters

N clusters	KM-E	KM-DTW	k-Shape	KKM-GAK
2	0.003+-0.001	0.325+-0.403	0.004+-0.009	0.003+-0.002
4	0.004+-0.001	0.463+-0.129	0.02+-0.007	0.002+-0.001
6	0.018+-0.016	0.578+-0.151	0.043+-0.021	0.001+-0.0005
8	0.006+-0.003	0.498+-0.077	0.005+-0.011	0.001+-0.0005
10	0.005+-0.01	0.624+-0.03	0.062+-0.022	0.0001+-0.00005

we measure the clustering quality in scenarios when the amount of information increases. We generate datasets with 5 clusters and 100 samples in each cluster. We set $T = 1000$ and consider 5 intervals on the time axis. We train and evaluate models in the first interval and consequently add information. From the second experiment, we can see that the KM-DTW model outperforms other models, but we do not observe increased accuracy as a result of adding information. There is a significant increase in the accuracy of the KM-DTW model when the number of samples increases from 200 to 400, but further increases in the number of samples do not improve the accuracy of the model. The K-Shape model also shows a slight improvement in accuracy when the number of samples increases from 800 to 1000. Given that model-based methods rely on ML/Quasy ML estimates of the parameters of GARCH models and also the asymptotic properties of these estimates, this experiment may suggest that model-based methods have the potential to increase clustering accuracy as information increases. The results of the second experiment are displayed in Fig. 1.

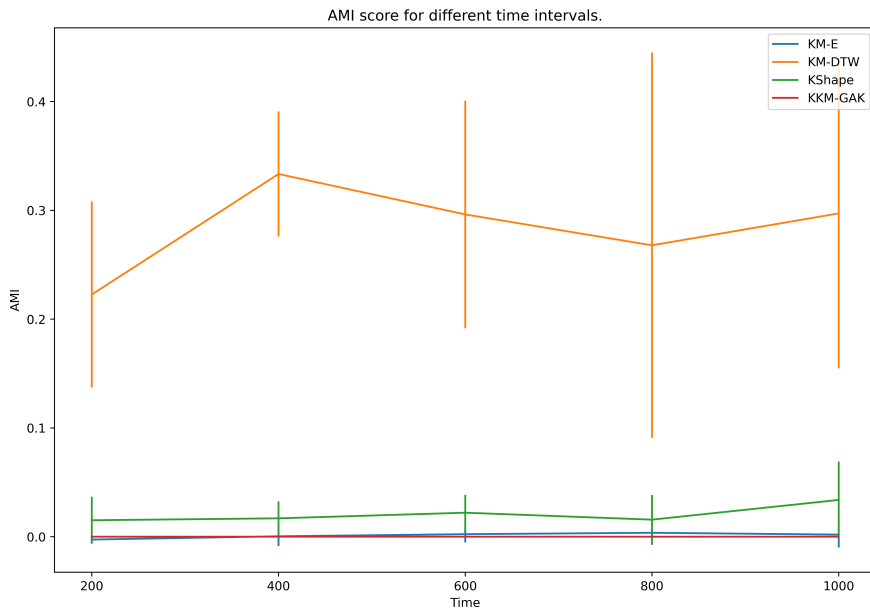


Fig. 1. AMI for different time intervals.

Fig. 2 shows the results of the third experiment. In this experiment, we measure the ability of the KM-DTW model to cluster an imbalanced dataset. For the fairness of the experiment, we generate time series samples with the GARCH(1,1) process and ensure that parameters satisfy (1). In addition, we constrain the L_2 norm of generated parameters to obtain non-overlapping clusters. We generate a dataset with different sample ratios and increase the ratio to 1. In the figure, we can observe that the best model for other experiments KM-DTW is dependent on cluster imbalance. This experiment shows that the claim made in [24] that centroid-based methods should be adapted to unbalanced scenarios also holds in the domain of time series clustering.

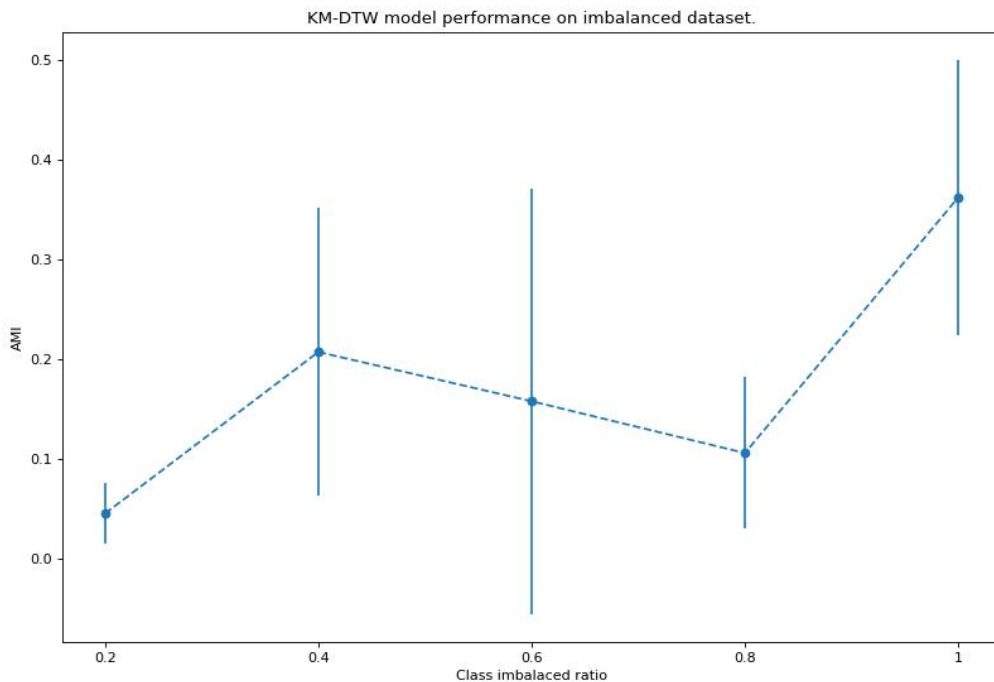


Fig. 2. Results for clustering imbalanced dataset.

Moreover, we measure the effect of the L_2 norm of generated parameters in clustering accuracy. We generate parameters for GARCH(1,1) process so that the parameters satisfy the current restriction on the L_2 norm. Throughout the experiment, we increase the bounds of the L_2 norm. During each step, we generate a balanced dataset with $T = 500$, $C = 2$, and 100 samples per cluster. We train models ten times for averaging purposes. We can observe that the KM-DTW model depends on clusters overlapping and increasing the bounds of parameters L_2 norm results in improvement of AMI. This problem is directly related to the ability of the similarity measure used in the KM-DTW algorithm to distinguish realizations of the GARCH process with parameters that are close to each other with the L_2 norm.

7. Conclusion and Future Work

In this work, several non-parametric clustering algorithms for clustering time series datasets generated by GARCH processes are evaluated. We generate multiple datasets and conduct multiple experiments to evaluate the K-Means (with Euclidean and dynamic time warping distance), K-Shape, and Kernel K-Means models. In the first experiment, we evaluate the ability of models to cluster different numbers of clusters. The results of the first experiment

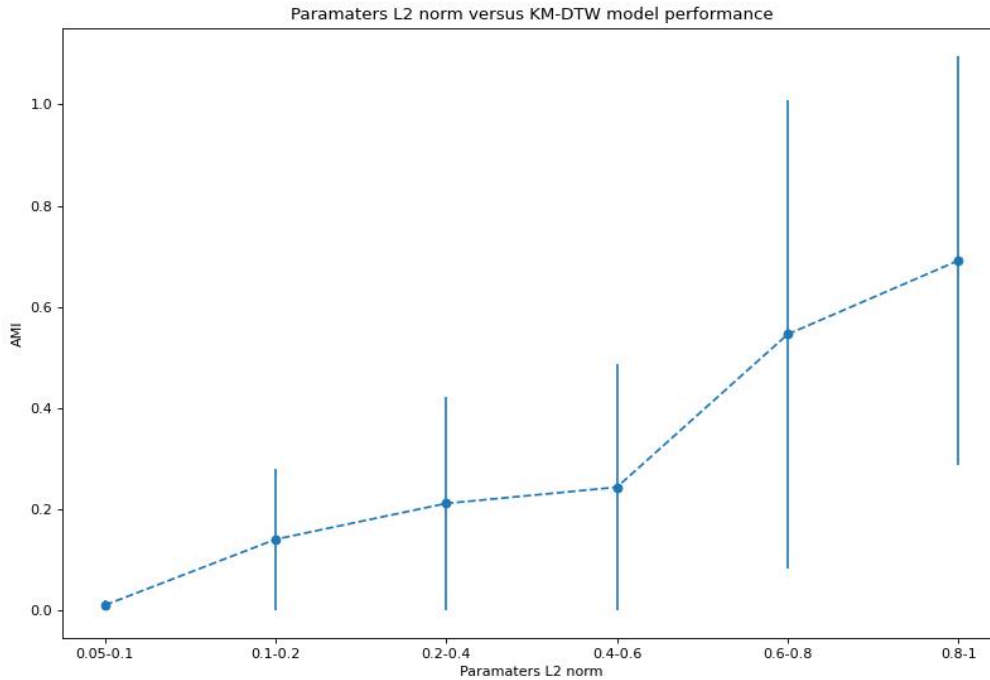


Fig. 3. GARCH parameters vector L_2 norm versus AMI score.

are displayed in Table 1. In the second experiment, we measure the clustering quality in the scenarios when the amount of information increases. We generate a dataset with 1000 time length and increase the information set. The results of the second experiment are shown in Fig. 1. During both experiments, the KM-DTW model shows better results. In the third experiment, we measure the ability of the KM-DTW model to cluster imbalanced datasets by generating multiple datasets with imbalanced samples in the cluster. The results are provided in Fig. 2. In the fourth experiment, we measure the ability of the KM-DTW model to cluster overlapping clusters. We constrain the norm of the parameters of the GARCH(1,1) model and evaluate the KM-DTW model. The experiment shows that KM-DTW is highly dependent on the norm of the generated parameters. The results of the fourth experiment are shown in Fig. 3.

We hope that our findings can motivate scholars to examine the discussed issues related to clustering accuracy, cluster overlapping, and available information effect. We think that already designed GARCH-based clustering methods have the potential to overcome these problems, so it is important to conduct similar experiments to show this. Moreover, as a direct application of our findings, it is worth applying clustering algorithms to the real-world financial dataset.

References

- [1] S. Aghabozorgi, A. Seyed Shirshorshidi, and T. Ying Wah, "Time-series clustering a decade review," *Information Systems*, vol. 53, p. 1638, 2015.
- [2] M. Corduas and D. Piccolo, "Time series clustering and classification by the autoregressive metric," *Computational Statistics amp; Data Analysis*, vol. 52, no. 4, p. 18601872, 2008.

- [3] Y. Xiong and D.-Y. Yeung, “Mixtures of arma models for model-based time series clustering,” *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*
- [4] A. Javed, B. S. Lee, and D. M. Rizzo, “A benchmark study on time series clustering,” *Machine Learning with Applications*, vol. 1, p. 100001, 2020.
- [5] N. Begum, L. Ulanova, J. Wang, and E. Keogh, “Accelerating dynamic time warping clustering with a novel admissible pruning strategy,” *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.
- [6] J. Paparrizos and L. Gravano, “Fast and accurate time-series clustering,” *ACM Transactions on Database Systems*, vol. 42, no. 2, p. 149, 2017.
- [7] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods, “Tslern, a machine learning toolkit for time series data,” *Journal of Machine Learning Research*, vol. 21, no. 118, pp. 1–6, 2020.
- [8] E. Keogh and S. Kasetty *Data Mining and Knowledge Discovery*, vol. 7, no. 4, pp. 349–371, 2003.
- [9] Y. Chen, E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, “The ucr time series classification archive,” July 2015. www.cs.ucr.edu/~eamonn/time_series_data/.
- [10] S. P. Daz and J. A. Vilar, “Comparing several parametric and nonparametric approaches to time series clustering: A simulation study,” *Journal of Classification*, vol. 27, no. 3, pp. 333–362, 2010.
- [11] P. D’Urso, L. De Giovanni, R. Massari, and D. Di Lallo, “Noise fuzzy clustering of time series by autoregressive metric,” *METRON*, vol. 71, no. 3, p. 217243, 2013.
- [12] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, vol. 31, no. 3, p. 307327, 1986.
- [13] L. Hentschel, “All in the family nesting symmetric and asymmetric garch models,” *Journal of Financial Economics*, vol. 39, no. 1, p. 71104, 1995.
- [14] J. Park, J. Baek, and S. Hwang, “Persistent-threshold-garch processes: Model and application,” *Statistics and Probability Letters*, vol. 79, no. 7, p. 907914, 2009.
- [15] A. M. Lindner, “Stationarity, mixing, distributional properties and moments of garch(p, q) processes,” *Handbook of Financial Time Series*, p. 4369, 2009.
- [16] *Dynamic Time Warping*, pp. 69–84. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
- [17] F. Petitjean, A. Ketterlin, and P. Ganarski, “A global averaging method for dynamic time warping, with applications to clustering,” *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.
- [18] J. Paparrizos and L. Gravano, “K-shape,” *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 2015.
- [19] I. S. Dhillon, Y. Guan, and B. Kulis, “Kernel k-means,” *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’04*, 2004.
- [20] M. Cuturi, “Fast global alignment kernels,” in *ICML*, 2011.
- [21] L. Hubert and P. Arabie, “Comparing partitions,” *Journal of Classification*, vol. 2, no. 1, p. 193218, 1985.

- [22] J. M. Santos and M. Embrechts, “On the use of the adjusted rand index as a metric for evaluating supervised classification,” *Artificial Neural Networks ICANN 2009*, p. 175184, 2009.
- [23] S. Romano, N. the Vinh, J. C. Bailey, and K. M. Verspoor, “Adjusting for chance clustering comparison measures,” *Journal of Machine Learning (JMLR)*, pp. 4635–4666, vol. 17, no. 1, 2016.
- [24] B. Krawczyk, “Learning from imbalanced data: Open challenges and future directions,” *Progress in Artificial Intelligence*, vol. 5, no. 4, p. 221–232, 2016.

GARCH պրոցեսների կլաստերիզացիայի համար մոդելներից անկախ ալգորիթմների համեմատություն

Գարիկ Լ. Ադամյան

Երևանի պետական համալսարան, Երևան, Հայաստան
e-mail: garik.adamyan@ysu.am

Անփոփում

Հոդվածում մենք գնահատում ենք մի քանի մոդելներից անկախ կլաստերիզացիայի ալգորիթմների GARCH պրոցեսներով գեներացված ժամանակային շարքերի տվյալների կլաստերավորման ունակությունը: Լայնածավալ փորձերի ընթացքում մենք գեներացնում ենք սինթետիկ տվյալների հավաքածուներ տարբեր սցենարներով: Այնուհետև, մենք համեմատում ենք K-Means մոդելները (Էվկլիդեսյան և ժամանակի դինամիկ փոխակերպման մետրիկաներով), K-Shape և Kernel K-Means մոդելների տարբեր կլաստերային չափիչներով: Մի քանի փորձերը ցույց են տալիս, որ K-Means մոդելը ժամանակի դինամիկ փոխակերպման մետրիկայով ցույց է տալիս համեմատաբար ավելի լավ արդյունքներ: Այնուամենայնիվ, դիտարկված մոդելներն ունեն գզայի թերություններ ինֆորմացիայի (ժամանակային շարքի նվազագույն երկարությունը) քանակի ավելացման հետ կլաստերավորման ճշգրտության բարձրացման հետ կապված, ինչպես նաև տվյալների անհավասարակշռության կամ կլաստերի համընկնման դեպքում ճշգրիտ կլաստերավորում իրականացնելու հարցում:

Բանալի բառեր` ժամանակային շարքերի կլաստերիզացիա, GARCH պրոցեսներ, ժամանակի դինամիկ փոխակերպում, K-Means, K-Shape.

Сравнение безмодельных алгоритмов кластеризации GARCH-процессов

Гарик Л. Адамян

Ереванский государственный университет, Ереван, Армения
e-mail: garik.adamyan@ysu.am

Аннотация

В этой статье мы оцениваем некоторые безмодельные алгоритмы кластеризации наборов данных временных рядов, сгенерированных GARCH процессами. В обширных экспериментах мы генерируем синтетические наборы данных для различных сценариях. Затем мы сравниваем модели K-Means (с метриками евклидовой и динамической трансформации временной шкалы), модели K-Shape и Kernel K-Means с различными метриками кластеризации. Несколько экспериментов показывают, что модель K-Means с метрикой динамической трансформации временной шкалы дает сравнительно лучшие результаты. Однако рассмотренные модели имеют существенные недостатки в повышении точности кластеризации при увеличении количества информации (минимальной длины временного ряда), а также при несбалансированности данных или перекрытии кластеров.

Ключевые слова: кластеризация временных рядов, процесс GARCH, динамическая деформация времени, K-Means, K-Shape.