Expert Systems With Applications 187 (2022) 115744

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Anomaly detection based on a granular Markov model

Yanjun Zhou [a,1], Huorong Ren [a,*,2], Zhiwu Li [a,b,3], Witold Pedrycz [c,4]

[a] *School of Electro-Mechanical Engineering, Xidian University, Xi'an 710071, China*

[b] *Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China*

[c] *Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada*

ARTICLE INFO

Keywords:
Time series data
Anomaly detection
Interval information granules
Principle of justifiable granularity

ABSTRACT

Since time series are characterized by a substantial volume of data, high levels of noise and the correlation between data in the time series attributes, it becomes challenging to mine crucial information from the series and apply it to anomaly detection. In this study, inspired by the concept of information granularity being applied to the process of system modelling, a granular Markov model is proposed for time series anomaly detection. Anomalies are generally caused by the changes in amplitude and shape; in this study we take both the original time series data and their amplitude change data into consideration. First, we utilize an interval information granularity representation based on the principle of justifiable granularity to represent the original time series data in an abstract manner to arrive at the corresponding representation results– that is, interval information granules. Then, based on the results of the interval information granularity representation and the Fuzzy C-Means (FCM) clustering algorithm, a granular Markov model is developed to produce anomaly scores to quantify possible anomalies. Compared with state-of-the-art methods, experimental studies completed for a large number of datasets demonstrate that the proposed method can significantly improve the anomaly detection process with higher data anomaly resolution. The obtained results are consistent across all datasets.

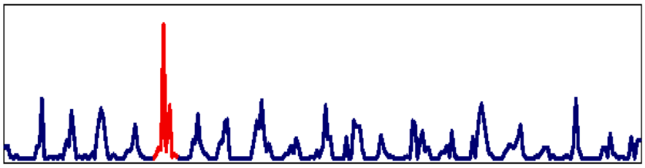
1. Introduction

Time series data describe phenomena reported over time, that is, each datum in this time series represents a value recorded at some time moment (Kumar & Toshniwal, 2016). Time series data are an important type of temporal data, which are commonly encountered in numerous application areas such as the natural sciences, engineering technology, social economics, etc. They have attracted much attention in the data mining research community (Esling & Agon, 2012). There are numerous time series data including the daily turnover of the supermarket, electrocardiogram (ECG) data, daily temperature changes, weekly sales data, daily stock prices, etc. Time series anomaly detection, also known as time series anomaly mining, refers to detecting any unexpected changes in a subsequence of a given time series data. When we compare the available patterns in the subsequence with the existing patterns in the entire time series, an "unexpected change" makes sense, namely a pattern anomaly. The detection of such anomalies is the objective we focused on in our study (Chandola et al., 2009). It is worth noting that there is a correlation between the morphology of time series data at different periods, which is generally represented by frequent change patterns and rare change patterns in time series data (Lin et al., 2012). This rare change patterns in time series data can be the result of amplitude changes or shape changes. Therefore, anomalies can be roughly categorized into two types: anomalies in shape and those in amplitude (Izakian & Pedrycz, 2013). The red segment in Fig. 1(a) is an example of the anomaly occurring in the amplitude change, while the other indicates an anomaly in shape change in Fig. 1(b). Both of these anomalies are pattern anomalies and our focus is on how to detect pattern anomalies.

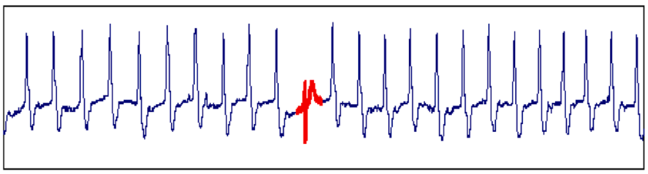
Time series anomaly detection can be briefly divided into real-time time series anomaly detection and historical time series anomaly detection (Izakian & Pedrycz, 2014). The former is generally applied to warn of the occurrence of anomalies, where the newly entered data are compared to an established anomaly model to verify whether there is an

* Corresponding author.
E-mail addresses: hrren@xidian.edu.cn (H. Ren), wpedrycz@ualberta.ca (W. Pedrycz).
[1] ORCID: https://orcid.org/0000-0002-7571-6375.
[2] ORCID: https://orcid.org/0000-0002-2058-8310.
[3] ORCID: https://orcid.org/0000-0003-1547-5503.
[4] ORCID: https://orcid.org/0000-0002-9335-9930.

https://doi.org/10.1016/j.eswa.2021.115744
Received 10 October 2020; Received in revised form 4 July 2021; Accepted 7 August 2021
Available online 29 August 2021



(a) amplitude anomaly



(b) shape anomaly

Fig. 1. Two types of data anomalies (Izakian & Pedrycz, 2013).

anomaly or not. The latter uses existing time series data to analyze and identify anomalies, which can be utilized as a basis for diagnosis or accident analysis in the future. This work focuses on historical time series anomaly detection, which can be applied to many fields of research. Timely detection of anomalies can prevent unnecessary losses and injuries. For instance, in the financial field, we can record the use of credit cards by customers, analyze the customer's consumption patterns and detect whether abnormal spending has occurred over a period of time such that misappropriations can be discovered promptly (Chandra & Chand, 2016). In the aviation field, based on the recorded data of aircraft black boxes, the causes of air crashes can be analyzed and anomalies can also be reported in advance (Baker et al., 2015). In the medical field, the status of the ECG activity can be monitored every millisecond such that abnormal situations involving a patient can be detected or diagnosed (He & Tan, 2017).

Considering high levels of noise and the volume of time series data, extracting and visualizing the available patterns in the time series data is highly beneficial for anomaly detection (Zhu et al., 2016). Information granularity helps represent data in an abstract manner by constructing information granules, which can be employed to discover meaningful and interpretable relationships between patterns (Pedrycz, Lu, Liu et al., 2014). Note here that information granules are entities composed of elements drawn together on the basis of similarity, functional closeness or spatial neighborhood (Lu et al., 2015). There are many ways to formally describe information granules, such as rough sets, shadowed sets, fuzzy sets, intervals, and so on (Zhu et al., 2016). In this study, we propose an approach called a granular Markov model for time series anomaly detection based on interval information granularity. The entire process of anomaly detection with a granular Markov model is carried out in three steps. First, we construct interval information granules using the principle of justifiable granularity. Since anomalies in time series data are generally caused by the changes in amplitude and shape, we take the amplitude change data of the original time series data into consideration and perform the same operations (interval information granularity representation) on both groups of data. Second, based on the results of the interval information granularity representation, we utilize the Fuzzy C-Means (FCM) clustering algorithm to endow different states of information granularity. In this sequel, a granular Markov model is developed to calculate the anomaly scores and is employed for anomaly detection. The proposed method exhibits the following advantages:

- The characteristics of the shape and morphology of the time series data are fully taken into consideration, and the interval information granularity representation operations are performed on both the original time series data and their amplitude change data.
- Interval information granularity is conducive to realizing data visualization and mining the internal relationships between various patterns of time series data. Furthermore, using interval information granularity to represent time series data would greatly reduce data volume and improve the performance of the method.
- Inspired by information granularity, a granular model is proposed and used in the field of anomaly detection, which significantly improves the accuracy of anomaly detection.

There are numerous studies in the literature using Markov models to detect anomalies in time series data. Nevertheless, an approach that utilizes information granules and information granularity to complete anomaly detection has not been discussed in depth and some preliminary studies are promising although very limited (Izakian & Pedrycz, 2013, 2014). In this sense, this study reaches a significant degree of originality and delivers a systematic study on the concept and granular anomaly detection and ensuing detectors. In this study, a granular Markov model for anomaly detection is proposed, which implements the process of system modelling with information granularity. It takes the time series data and its amplitude data into account for the interval information granularity representation, which mines the interpretable relationships between the data and realizes the abstraction and visualization of the data. More specifically, using interval information granules (rectangles) to represent the data in an abstract manner to achieve system modelling. In this way, the method of modelling information granularity is much more efficient than that at a numeric data level, which greatly reduces the data volume and improves the feasibility of implementation. Meanwhile, evaluation indexes have been established to quantify the performance of the proposed method for anomaly detection, which further verifies the effectiveness of the method and the accuracy of the results.

The paper is organized as follows. In Section 2, anomaly detection methods are briefly grouped into four categories and some commonly used methods are reviewed. Section 3 illustrates the method proposed in this paper in detail and briefly recalls some related knowledge. We establish the evaluation criteria and present an extensive performance evaluation on many real-world and synthetic data sets to validate the feasibility and effectiveness of the proposed method in Section 4. Finally, some conclusions are reached in Section 5.

2. Related work

There are several key and ongoing challenges to time series data mining, such as substantial volumes of data and high levels of noise (Baldán & Benítez, 2019). Moreover, most time series are collections of consecutive values varying continuously in time, which makes many methods of anomaly detection ineffective (Zhu et al., 2019). A lot of methods of time series anomaly detection have been reported in the literature, which are roughly categorized into four types: similarity-based, clustering-based, density-based, and modelling-based methods.

2.1. Similarity-based method

One straightforward method is to assign anomaly scores to each pattern by calculating the similarity between time series data and then performing anomaly detection. Generally speaking, a similarity-based method is composed of two steps, namely data representation and similarity measurement (Zhou, Ren, Li, & Pedrycz, 2021). With regard to data representation, for example, the piecewise linear representation (PLR) selects several crucial data points from the data, which are connected from head to tail with line segments to fit the data (Zhao et al., 2016). Piecewise aggregate approximation (PAA) extracts the time series data into subsequences of equal length and represents each subsequence by their mean, which may result in the loss of information about the data changing trends (Nakamura et al., 2013). Symbolic aggregate approximation (SAX) is a method of representing time series data in characters (Yahyaoui & Al-Daihani, 2019). The methods of discrete Fourier transform (DFT) and discrete wavelet transform (DWT) transform time series data from the time domain to the frequency domain and represent time series data as the features in the frequency domain (Chaovalit & Gangopadhyay, 2011). Data representation and similarity measurement are associated with each other. Similarity measurement is the basis of query, classification and prediction of time series data and is widely used for clustering, classification and segmentation of time series data (Cai et al., 2015). In terms of the approach used for similarity measurement, the Euclidean distance is the most widely applied method for expressing similarity over time series. It works only for two subsequences of the same length, and each data point in the two sequences must exhibit a one-to-one correspondence (Keogh et al., 2009). Dynamic time warping (DTW) is achieved by distorting the time axis to implement the adaptive shape matching of time series data, which is based on the minimum cost of the time curved path (Górecki & Łuczak, 2015). Longest common subsequence similarity (LCSS) can be applied to compute the similarity between two strings, which measures similarity after the data are represented by the method of SAX (Fang et al., 2015). In the literature, many scholars have used the Pearson correlation coefficient as an index to quantify the similarity. A Pearson correlation coefficient was observed between the clean class score and the level of agreement to quantify the correlation to confirm that the clean class score is closely related to the certainty/agreement of the annotators (Moeyersons et al., 2019). In the time domain, the CC (correlation coefficient) is utilized to evaluate the degree of signal similarity and the ranges from -1 to 1, where -1, 0, and $+1$ represent complete negative correlation, no correlation, and complete positive correlation, respectively (Renza et al., 2019). A multi feature decision method is proposed to detect copy-move forgery by using dual threshold voice activity detection to get the sound segments and calculate the similarity between them by means of PCCs (Pearson correlation coefficients) and average differences (Xie et al., 2018). Based on the data representation of intervals, a method of similarity measurement using the intersection and union of two intervals is presented (Ren, Li et al., 2018; Ren, Liu et al., 2018).

2.2. Clustering-based method

Clustering is another type of anomaly detection method, which clusters the time series data by using an appropriate clustering technique and assigns an anomaly score to each pattern of time series data according to the revealed cluster centers (Izakian et al., 2012). The FCM clustering method is a classical clustering method for clustering time series data where the time series data are reconstructed according to the revealed clustering centers. The reconstruction errors between the original time series data and the reconstructed data are employed to assign the corresponding anomaly scores to each pattern in time series data (Izakian & Pedrycz, 2013). Time series data can also be clustered using the method of k-medoid clustering in which anomaly scores are captured in light of the distance to the closest medoid. Training data containing unlabeled flow records are separated into clusters of normal and anomalous data (Park & Jun 2009). The corresponding cluster centroids obtained by the K-means clustering method are applied as patterns for computationally efficient distance-based detection of anomalies in new monitoring data (Münz et al., 2007).

2.3. Density-based method

The general framework for using this method is to calculate the distance between data points to characterize the density between data points and quantify the degree of anomaly for each data point. Local outlier factor (LOF) has already been proven as the most promising anomaly detection method for detecting time series data. An adaptive anomaly detection scheme for cloud computing based on LOF is presented, which learns the behavior of applications both in the training and detecting phase, and its adaptability reduces the demands of efforts on collecting training data before detecting (Huang et al., 2013). In (Smith et al., 2014), a dimensionality reduction package and a kernel density estimation function are employed as non-conformity measures to detect anomalies, and the average p-value is proposed as an efficiency criteria for conformal anomaly detection. A two-stage online anomaly detection algorithm is introduced, that is, a probability density function is constructed to model normal time series data, and then a threshold is estimated for the density of newly observed data to detect anomalies. Furthermore, a minimax optimal scheme for both stages to creates an optimal anomaly detection algorithm in a strong deterministic sense (Gokcesu & Kozat, 2017).

2.4. Modelling-based method

The modelling-based method is another variant of the anomaly detection approach, which has been reported in many works. ARMA is a mixture model of the auto regression (AR) model and the moving average (MA) model, which is assumed to have no anomalies in the training data and then a model is trained to establish an ARMA model and threshold value (Van Der Voort et al., 1996). The prediction values for each set of test data are obtained from the model and anomalies are determined according to the threshold values. The Markov-model-based method utilizes training data to create a classical model. Via training, one can pick out the initial probability, the state transition matrix and its threshold value. The Markov-model-based method can be utilized to analyze data in the process of anomaly detection, which is suitable to detect point anomalies and has high anomaly detection accuracy (Ren et al., 2017). Furthermore, deep learning has received increasing attention. In particular, deep learning-based methods are also a type of model-based methods. Anomaly detection through deep learning is mainly classified by learning, which progresses through the labeled data. Therefore, if you use a deep learning-based approach for anomaly detection, there are problems with the lack of data and imbalance between labels that arise during the learning process, which may produce misleading performance estimates. A decision boundary-based anomaly detection model using improved AnoGan is proposed, which found that the best loss balance is realized by repeatedly experimenting with the values of Discriminator and Generator. Another problem is that the decision boundary of an AnoGAN is subjective (Shin et al., 2020). Long-short term memory (LSTM) is a classical model for anomaly detection and it is prone to over-fitting of the model and is susceptible to time steps (Karim et al., 2017). An improvement scheme for LSTM prediction errors, introducing a second stage predictor that can identify the actual anomaly class from the error outputs of the first stage model, has been proposed and a successful predictor for the same is developed with good performance (Chauhan et al., 2019).

The core of a similarity-based method is data representation and similarity measurement. The data representation method can extract data features and reach to the objective of reducing the amount of data. While the similarity measurement algorithm is generally designed based on the results of the data representation method, which has a corresponding relationship with the method of data representation, and its applicability to other methods is poor. The clustering-based method allows one to cluster time series data using an appropriate clustering technique and assign an anomaly score to each pattern of time series data according to the revealed centers. However, it is sometimes difficult to design a method that can arrive at anomaly scores to map the relationship between centers and anomalies. The density-based method can be used to calculate the distance between data points to characterize the density between data points and quantify the degree of anomaly for each data point. Nevertheless, the density-based method, which is suitable for detecting point anomalies, may not be suitable for pattern anomalies caused by changes in shape and amplitude. With regard to the model-based method, in general, the original time series data are directly utilized for modelling and anomaly detection is performed according to the

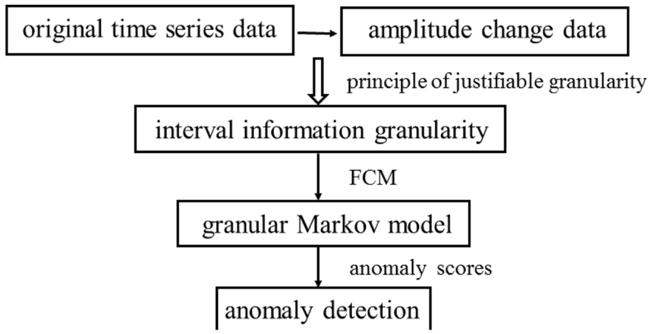


Fig. 2. The main idea of the proposed method.

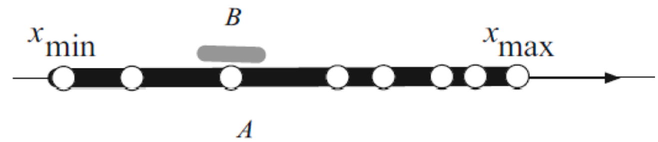


Fig. 3. An illustrative example of the principle of justifiable granularity.

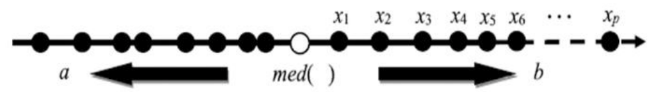


Fig. 4. A determination of a maximal value of α.

threshold value obtained by the model. It has attracted the attention of numerous scholars due to its simplicity and wide application. Whereas the model-based method usually does not conduct any processing on the original time series data, which consumes a lot of computing time and running memory. In this study, inspired by the concept of information granularity and applied to the process of modelling, we propose a granular Markov model for time series anomaly detection.

3. The proposed method

As shown in Fig. 2, the underlying idea of the proposed method, which utilizes interval information granularity representation is based on the principle of justifiable granularity to represent the original time series data and their amplitude data and arrive at the corresponding representation (that is, interval information granules) (Ren, Liu et al., 2018). Thus, in the light of the results of the interval information granularity representation and the FCM clustering algorithm, a granular Markov model is proposed to obtain the anomaly scores needed to achieve the ultimate objective of anomaly detection. A detailed illustration of the granular Markov method is given in the sequel. We start from some terms:

Time series: A time series $X = x_1, x_2, \ldots, x_n$ is a sequence of n real values ordered in time (Keogh et al., 2005; Ren, Li et al., 2018).

Subsequence: Given a time series X of length n, a subsequence Y_i of X is a series of length $w \leqslant n$ of the adjacent position from X such that $Y_i = x_i, x_{i+1}, \ldots, x_{i+w-1}$ for $1 \leqslant i \leqslant n - w + 1$ (Keogh et al., 2005).

Sliding Window: Given a time series X of length n, and a user-defined subsequence of length w, all possible subsequences can be extracted by sliding a window of size w across X.

3.1. Principle of justifiable granularity

The crux of the principle of justifiable granularity is in constructing a meaningful information granule based on available experimental data. Such a construction has to adhere to two intuitively compelling requirements: experimental evidence and specificity (Wang et al., 2018). Experimental evidence requires that the existence of an information granule is well justified as being reflective of the existing experimental data. In other words, the more data the information granule represents (involves), the more justifiable the information granule becomes. Meanwhile, the information granule should be made as specific as possible, which implies that the information granule comes with a well-defined meaning. Specificity requires that the information granule be highly detailed, which indicates that the more compact the information granule is, the better. It is apparent that these two requirements are in conflict: the increase of the criterion of experimental evidence comes with a deterioration of the specificity of the information granules. As usual, we are interested in forming a sound compromise (Ren, Liu et al., 2018).

In what follows, we elaborate on the principle of justifiable granularity by utilizing the example of intervals, which are easy to interpret and allow us to focus on the essence of the problem in the first place (Ren, Li et al., 2018). With regard to the intervals, the more data the intervals contain, and the better the experimental evidence is. On the contrary, the fewer data the intervals include, the better the specificity is. Fig. 3 shows an example of constructing an interval. Interval B contains only one datum and has high specificity, whereas it ignores other data points and is not well supported by more data. Interval A contains all data points, leading to insufficient specificity (Wang et al., 2018). Therefore, an appropriate interval results in a tradeoff between specificity and experimental evidence.

With these two criteria in mind, let us proceed with the detailed formation of the interval information granule. Let us consider a time series $X = x_1, x_2, \ldots, x_n$ and construct an information granule in the form of an interval $[a, b]$. The optimized performance index is given as the following product (Pedrycz, Lu, Liu et al., 2014; Ren, Li et al., 2018).

$$V(b) = f_1(\mathrm{card}\{x(k) \in X(n) | med(X(n)) \leqslant x(k) \leqslant b\}) \\ * f_2(|med(X(n)) - b|) \tag{1}$$

$$V(a) = f_1(card\{x(k) \in X(n) | a \leqslant x(k) < med(X(n))\}) \\ * f_2(|med(X(n)) - a|) \tag{2}$$

where X is a time series and $med()$ is its median value, which is a robust estimator. $card(x(k)|x(k) \in X(n))$ is that the count of all data $x(k)$ that fall within the bounds of X. Parameters a and b are the lower and upper boundaries of an interval, respectively. One can arrive at the upper optimal bound of the interval information granule by independently maximizing (1), which is denoted as b_{opt}, namely $V(b_{opt}) = max_{b>med(X(n))} b$. The optimal lower bound of an interval information granule can be obtained by the same means used to optimize Eq. (2), that is, $V(a_{opt}) = max_{a<med(X(n))} a$. f_1 represents an increasing function and the higher the value of f_1, the higher the experimental evidence is, while f_2 is a non-increasing function and refers to another important requirement of high specificity. The functions of f_1 and f_2 can be given as the following forms (Ren, Liu et al., 2018):

$$f_1(u) = u \tag{3}$$

$$f_2(u) = exp(-\alpha u) \tag{4}$$

where α is a positive parameter supplying some flexibility when optimizing the interval information granule. The role of α is to calibrate the processing of the interval information granule according to specificity requirements, where the higher the value of α stress, the greater the specificity. Note that if $\alpha = 0$, then $f_2(u) = 1$, which indicates that the constructed interval information granule contains all experimental data and thus the requirement for specificity is completely ruled out. We employ the optimal of $V(b)$ as an example to illustrate the role of the parameter of α. With regard to $V(a)$, it can be obtained using the method of analogous fashion. Let us assume that the maximal value of α, namely α_{max}, is reached when $b_{opt} = x_1$, where x_1 is the data point closest to the median and larger than, referred to in Fig. 4. In this case, a series of

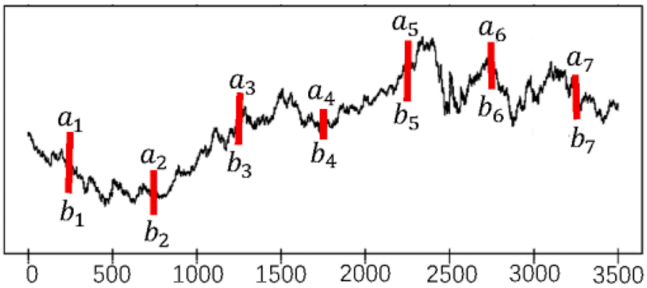


Fig. 5. Illustrative example of the intervals.

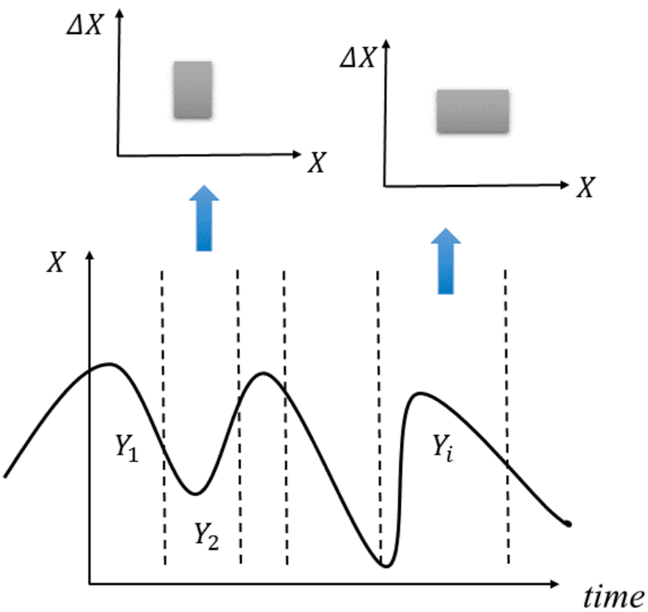


Fig. 6. Construction of building interval information granules.

values that $V(b)$ should satisfy the following expressions are (Lu et al., 2015; Pedrycz, Lu, Liu et al., 2014):

$$\begin{array}{l} 1*exp(-\alpha|med(X(n))-x_1|) > 2*exp(-\alpha|med(X(n))-x_2|) \\ 1*exp(-\alpha|med(X(n))-x_1|) > 3*exp(-\alpha|med(X(n))-x_3|) \\ 1*exp(-\alpha|med(X(n))-x_1|) > p*exp(-\alpha|med(X(n))-x_p|) \end{array} \tag{5}$$

where the data $x_1, x_2, ..., x_p$ are arranged as follows:$med(X(n))\leqslant x_1\leqslant x_2\leqslant ...\leqslant x_p$. Once the value of α_{max} is obtained, the range of these values $[0, \alpha_{max}]$ can be normalized to [0, 1]. In this study, the normalized value of $\alpha \in [0, 1]$ is set to 0.5, and then we can receive the corresponding intervals [a, b] (Ren, Li et al., 2018; Ren, Liu et al., 2018).

As shown above, all subsequences extracted across the time series X are conducted in an interval information granularity representation under the principle of justifiable granularity. As shown in Fig. 5, the data from length 3500 can be represented in multiple intervals. For instance, data points with an initial subsequence ($0 \sim 500$) are shown by an interval [a_1, b_1] (Ren, Li et al., 2018).

Anomalies in time series are caused majorly by changes in amplitude and shape. In order to extract the characteristics of the morphology and shape of the time series data, one takes into account the changes in data, namely $\Delta X = \Delta x_1, \Delta x_2, ..., \Delta x_{n-1}$, where $\Delta x_i = x_{i+1} - x_i$ (Pedrycz, Lu, & Liu, 2014; Zhu et al., 2016). In this case, we can reach the intervals in an analogous fashion under the principle of justifiable granularity; that is, each subsequence of the time series X and their corresponding amplitude change data are all represented by intervals. The intervals of each subsequence of the original time series data are regarded as the length of the rectangles, and their corresponding intervals of amplitude change data are seen as the width of the rectangles. In this way, all the two corresponding subsequences in these two groups of data can be abstracted and expressed with an interval information granule (rectangle). Fig. 6 visualizes the entire process of interval information granules.

4. FCM

The Fuzzy C-Means (FCM) clustering algorithm proposed by Dunn and Bezdek is one of the most commonly applied clustering techniques. We briefly recall the FCM method for clustering the interval information granules (rectangles), which come as the results of the principle of justifiable granularity. FCM aims to describe these subsequences extracted from the time series data (or the results of interval information granularity representation) by employing a collection of c ($c \leqslant n$) cluster centers (prototypes), that is $v_1, v_2, ..., v_c$, and a fuzzy partition matrix $U = [u_{ik}], i = 1, 2, ..., c, k = 1, 2, ..., n$, to minimize of the following objective function(Izakian et al., 2015):

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\|v_i - x_k\|^2 \tag{6}$$

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}} \quad u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|v_i - x_k\|}{\|v_j - x_k\|}\right)} \tag{7}$$

where m is a fuzzification coefficient and $||.||$ refers to the Euclidean distance function. The optimization of (6) is completed in the presence of constraints imposed on the partition matrix $\sum_{i=1}^{c} u_{ik} = 1$ with $0 < \sum_{k=1}^{n} u_{ik} < n$. The initialization of the partition matrix is completed in a random way. The minimization procedure producing a minimum of J is completed through an iterative process by adjusting the values of the partition matrix and the prototypes (Izakian et al., 2012).

4.1. Markov model

For the Markov model, the primary process of the model is to establish a structure for the model with training data and the transition probability of the data points to be determined. However, using the Markov-model-based method for anomaly detection, state transition can only be performed between various data points, which is effective for the detection of point anomalies (an individual datum has an obvious deviation from other data). One can obtain the anomaly scores for each data point and threshold value using a Markov-model-based method. When the anomaly score of a datum is greater than some threshold value, it is classified as an anomaly. In this study, we utilize the FCM clustering algorithm to cluster the results of subsequences of the above operations–interval information granules (rectangles), where a granular Markov model is proposed and applied to endow the anomaly scores. The pseudo-code of the granular Markov model is presented in Algorithm 1. Once the anomaly score for an interval information granule is greater than some threshold value, it is regarded as an anomaly, and its corresponding subsequence is also anomalous.

Algorithm 1. Anomaly detection.

Input: all the subsequences of the original time series data and their amplitude change data
Output: anomaly scores
for each subsequences
  using (1) and (2) to obtain the interval information granules
end for
for $i = 1$
  using (7) to calculate (6) by an iterative fashion
  until J_{min}
end for
using (8) to establish the model for anomaly detection

A granular Markov model based on interval information granules (rectangles) for anomaly detection is proposed, which implements a

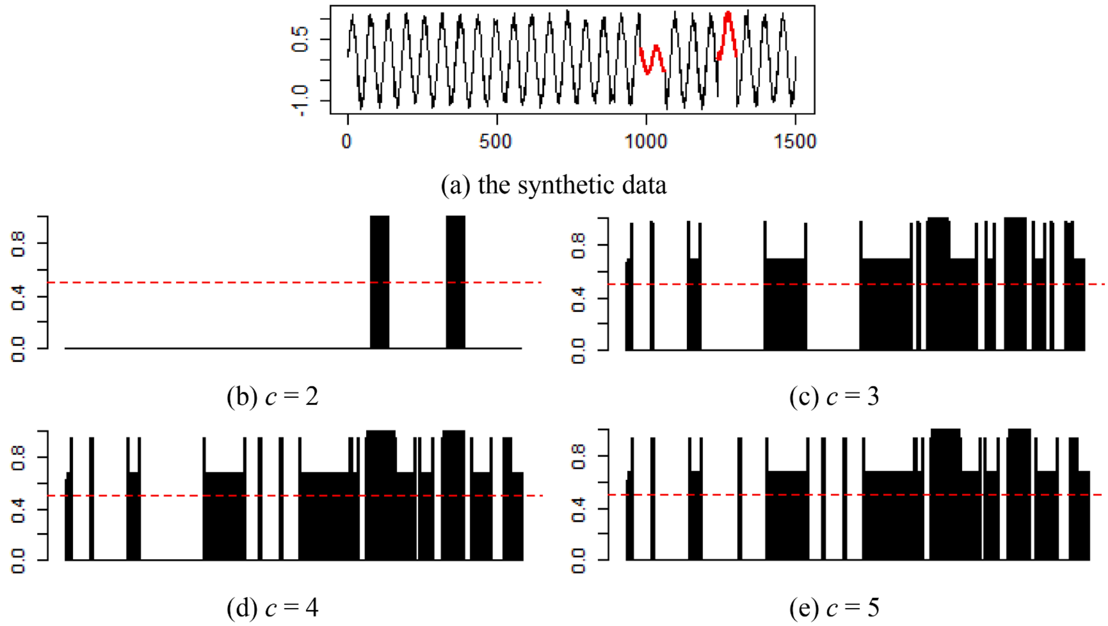


Fig. 7. Selection of parameter c.

system modelling process with information granularity. Information granularity helps represent data in an abstract manner by constructing information granules, which can be employed to discover meaningful and interpretable relationships between patterns. In this study, the interval information granules (rectangles) are constructed based on extracting the shape and morphological characteristics of the original time series data and their amplitude data. Thus, the interval information granules (rectangles) can fully represent most of the important information contained in the time series data and also greatly reduce the data volume. Furthermore, the probability transition of the granular Markov model is carried out between interval information granules (rectangles), which can be employed to detect pattern anomalies, while the traditional Markov model is to perform probability transitions between data points, which is feasible for detecting point anomalies and not working well for pattern anomalies. Here, the complete parameter set of a Markov model can be represented by the following triple (Zhou, Ren, Li, & Wu, 2021):

$$\lambda = \{S, P, Q\} \tag{8}$$

where

(1) λ represents a granular Markov model;
(2) S is the state space of interval information granules (rectangles), including all the possible states of interval information granules (rectangles), i.e., S can be represented by $S = \{1, 2, ..., w\}$, where w is the number of states present in interval information granules (rectangles). Note that we use s_t ($s_t \in S$) to denote that the interval information granules (rectangles) x_t is in the state s_t at time t.
(3) $Q = \{q_1, q_2, ..., q_w\}$ is the initial probability distribution set of interval information granules (rectangles) with $\sum_{i=1}^{w} q_i = 1$.
(4) $P = |p_{s_t s_{t+1}}|_{w \times w}$ indicates the state transition probability matrix, and $p_{s_t s_{t+1}}$ refers to the probability of the transition from state s_t to s_{t+1} with $\sum_{s_t=1}^{w} p_{s_t s_{t+1}} = 1$.

4.2. Evaluation criteria

In order to validate the proposed method and assess its quality, it is essential to establish evaluation criteria. They are introduced as follows (Ren, Li et al., 2018; Ren, Liu et al., 2018):

$$CI = \frac{mean(\sum AS_{anomaly})}{mean(\sum AS_{all})} \tag{9}$$

$$AR = \frac{M_a}{M} \tag{10}$$

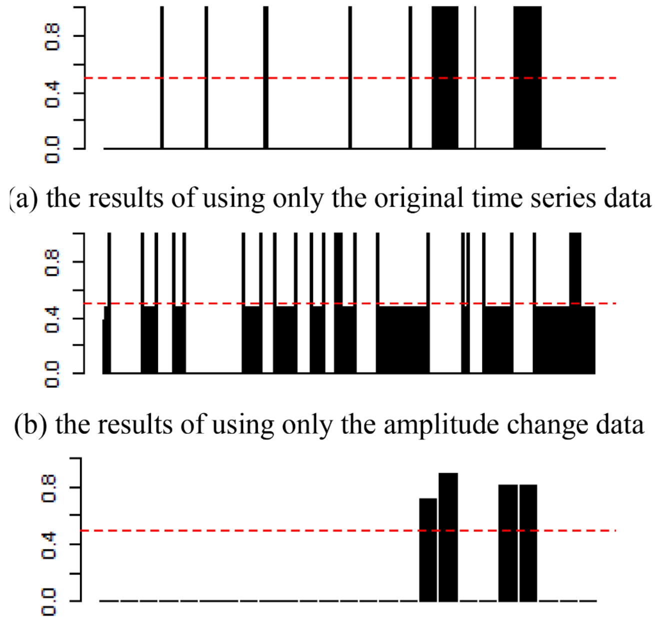
where CI is a confidence index, used to evaluate the performance of the methods for data anomaly resolution. $mean(\sum AS_{anomaly})$ is the mean value of all the anomaly scores of all the anomaly patterns, while $mean(\sum AS_{all})$ represents the mean value of all the anomaly scores of all the patterns. AR is the accuracy rate of the ratio of the count of correctly detected anomalies to all known anomalies. M_a represents the number of anomalies that have been correctly detected and M refers to the number of all known anomalies. In response to different detection sensitivity requirements, a threshold value can be obtained to determine whether a pattern is an anomaly. The larger the anomaly score the anomaly pattern is, the smaller the anomaly score of the normal pattern is, and the better the data anomaly resolution is. In other words, when all anomalies are accurately detected, the higher the CI is, the better the performance of anomaly detection accuracy and data anomaly resolution are.

5. Experimental studies

In this section, we apply a large amount of synthetic data and real-world data to evaluate the performance of the proposed method. Furthermore, the methods of PAA, SAX (similarity-based method), K-means (clustering-based method), LOF (density-based method) and dynamic Markov model (modelling-based method) are employed to conduct comparative experiments to elaborate upon the superiority of the granular Markov model in this study.

5.1. Synthetic data

We first utilize some synthetic data for parameter analysis, following the procedure in (Zhou, Ren, Li, & Pedrycz, 2021; Zhou, Ren, Li, & Wu, 2021):



(a) the results of using only the original time series data

(b) the results of using only the amplitude change data

Fig. 8. Experimental results obtained for synthetic data with and without the amplitude change data.

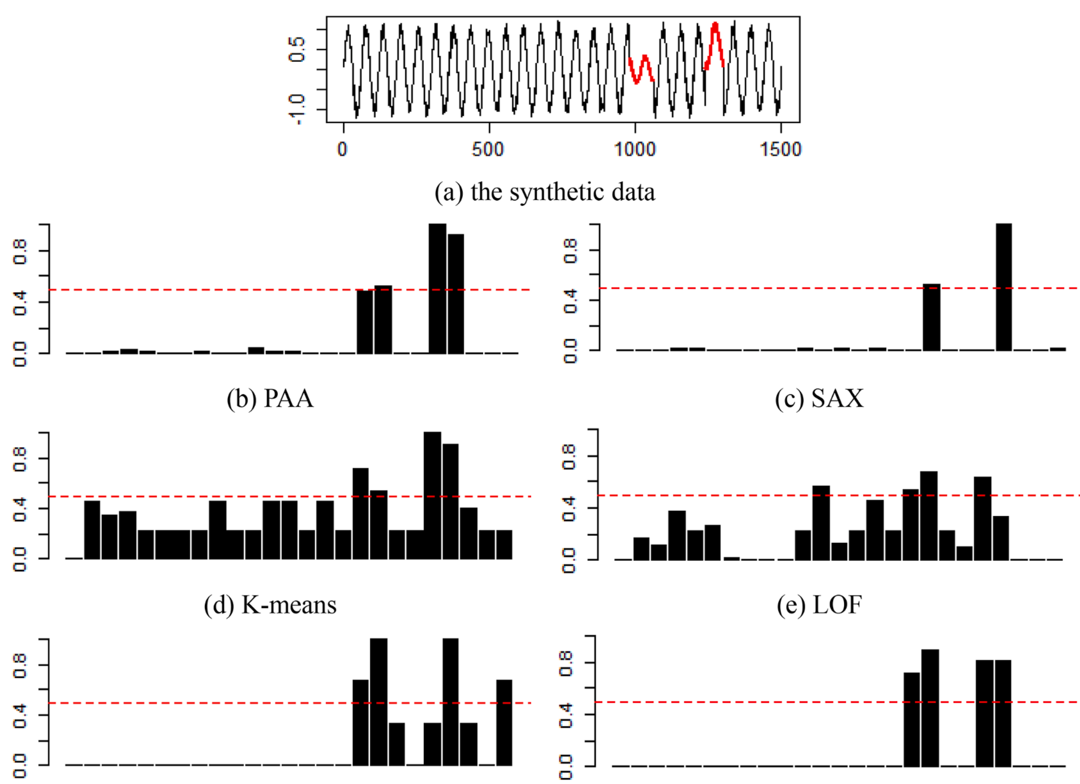
$$
\begin{aligned}
X(t) &= X_1(t) + e_1(t) + e_2(t) \\
X_1(t) &= sin\left(\frac{40\pi t}{K}\right) + n(t)
\end{aligned}
\tag{11}
$$

where $X(t)$ is the original time series data with noise and artificially added anomalies; the number of time instances is $K = 1200$. We model $n(t)$ as a Gaussian noise whose mean is 0 and standard deviation is 0.1; $e_1(t)$ and $e_2(t)$ are artificially injected anomalies, given by:

$$
\begin{aligned}
e_1(t) &= \begin{cases} 0.3{*}X_1(t), & t \in [980, 1060] \\ 0, & otherwise \end{cases} \\
e_2(t) &= \begin{cases} 0.5{*}X_1(t) + 0.6, & t \in [1240, 1300] \\ 0, & otherwise \end{cases}
\end{aligned}
\tag{12}
$$

Time series data often have a periodic phenomenon on the time axis of sufficient length, in which the time length of the periodic phenomenon or the size of the data is the value of the parameter w. In general, the coefficient w is determined by the end-user for different time series data according to different demands. In other words, since the time series data have the attribute of time, according to different requirements, if one wants to detect whether there is an anomaly in the data per minute, hour, day, etc., then the amount of data sampled per minute, hour, day, etc., is the value of parameter w. With regard to the synthetic data, Fig. 7 (a) shows the synthetic data, with the red parts being two artificially added anomalies. One of these two anomalies is the shape anomaly and the other is the amplitude anomaly. In this study, the parameter w of the synthetic data is user-defined, where $w = 60$.

Parameter c is the cluster center of the FCM method. As shown in Fig. 7, it is apparent that when $c = 2$, both anomalies in the synthetic data are accurately detected and have better data anomaly resolution. Nevertheless, when $c = 3$, 4, or 5, there are too many false alarms and the ability to detect anomalies is lost. Additionally, their *CI* values are 1.97, 1.77, and 1.74 respectively. Meanwhile, we can note that as the value of c increases, the performance of data anomaly resolution remains almost unchanged, at around 1.7. With regard to the selection of parameter c, the same conclusion is given in the literature (Izakian & Pedrycz, 2013). What needs to be supplemented here is that the corresponding threshold value can be obtained based on different detection targets and detection sensitivity requirements. The red dotted line in Fig. 7(b)–(e) stands for the threshold value in this study, which is the basis for calculating the evaluation index *CI*, and it should be emphasized that all the experiments in this paper are conducted under the same



(a) the synthetic data

(b) PAA

(c) SAX

(d) K-means

(e) LOF

(f) dynamic Markov model

(g) granular Markov model

Fig. 9. Results of comparative experiments on the synthetic data.

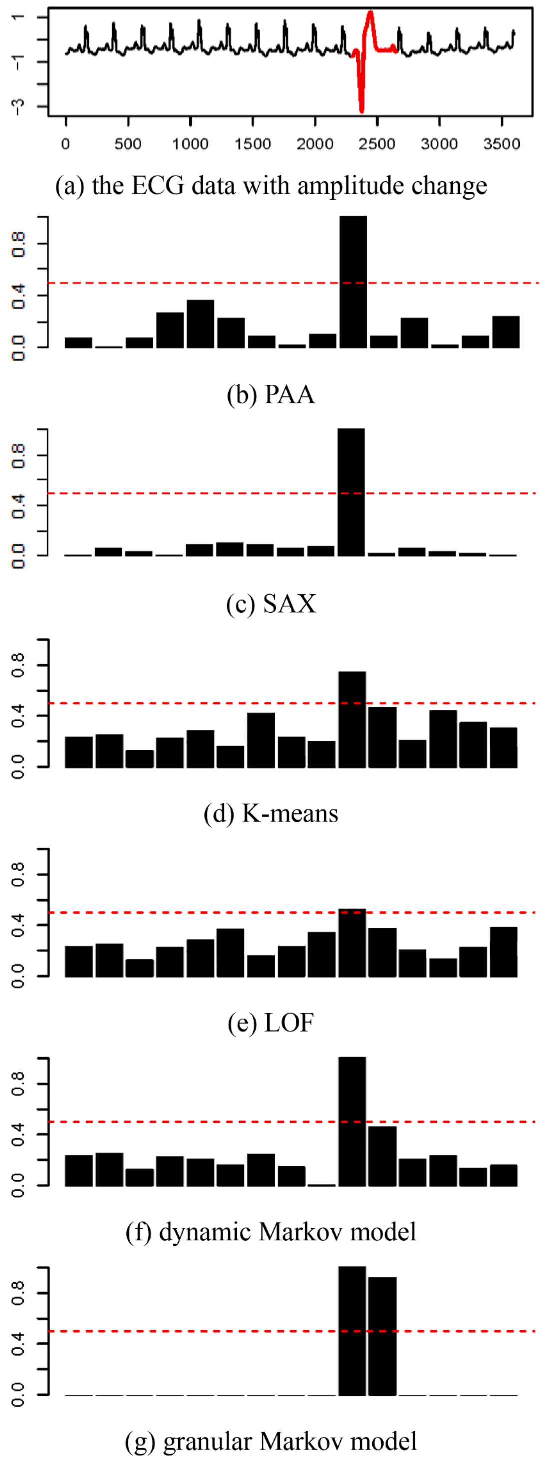


Fig. 10. Results of comparative experiments on ECG data with amplitude change.

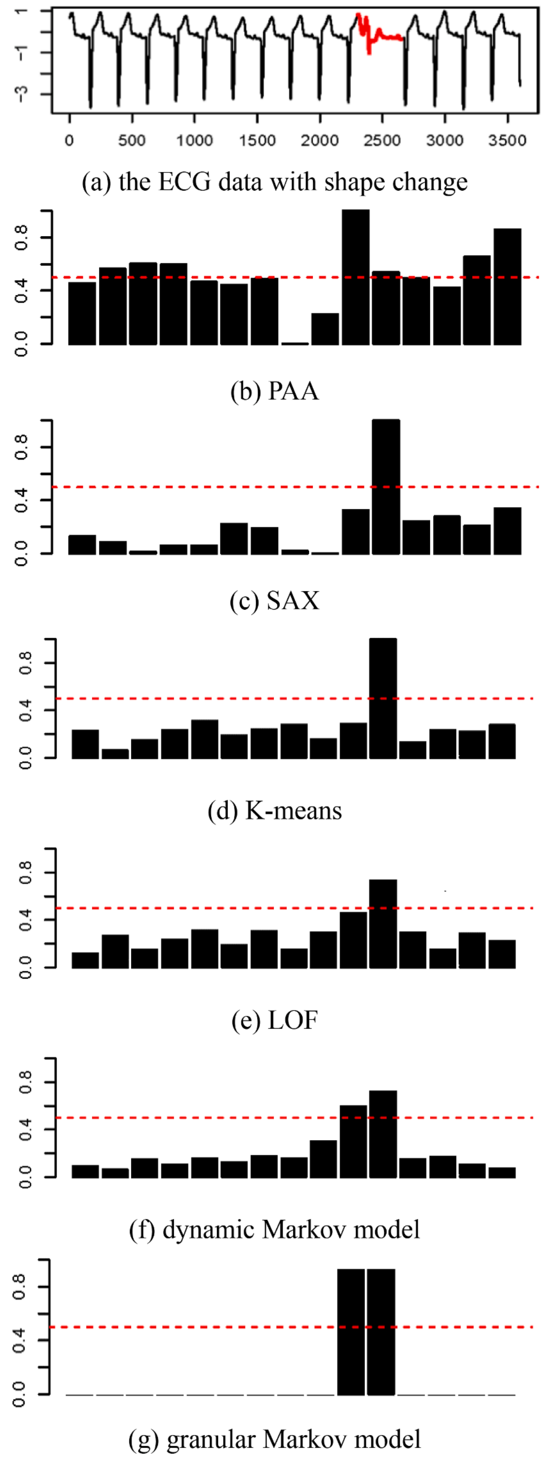


Fig. 11. Results of comparative experiments on ECG data with shape change.

threshold value.

The focus of this study is on pattern anomalies, which are generally caused by changes in amplitude and shape. In order to better perform anomaly detection, we take the amplitude change data into account to extract the characteristics of shape and morphology from both the original time series data and their amplitude change data. Fig. 8 is utilized to illustrate the effectiveness of considering the amplitude change data on the performance of anomaly detection using experiments with synthetic data. One can see that only the original time series data are utilized for anomaly detection, and there are several false alarms in the detection results ($CI = 7.68$). Furthermore, when only the amplitude change data are employed for detection, the anomaly detection function is lost ($CI = 3.33$). However, it can be noticed that when both groups of data are taken into account, the method proposed in this paper not only accurately detects all anomalies but also has better performance of data anomaly resolution ($CI = 11.56$) than others.

We utilize synthetic data to conduct a comparative experiment, as shown in Fig. 9, where one can see that the K-means can detect anomalies at poor data anomaly resolution. In contrast, the dynamic Markov

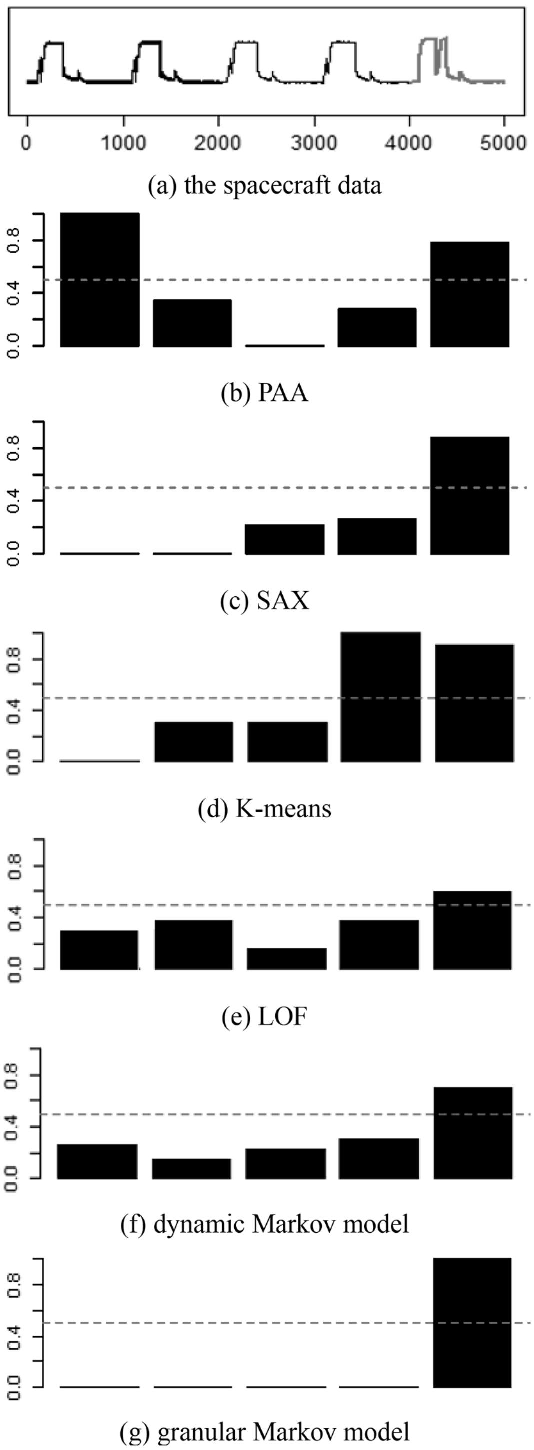


Fig. 12. Results of comparative experiments on the spacecraft data.

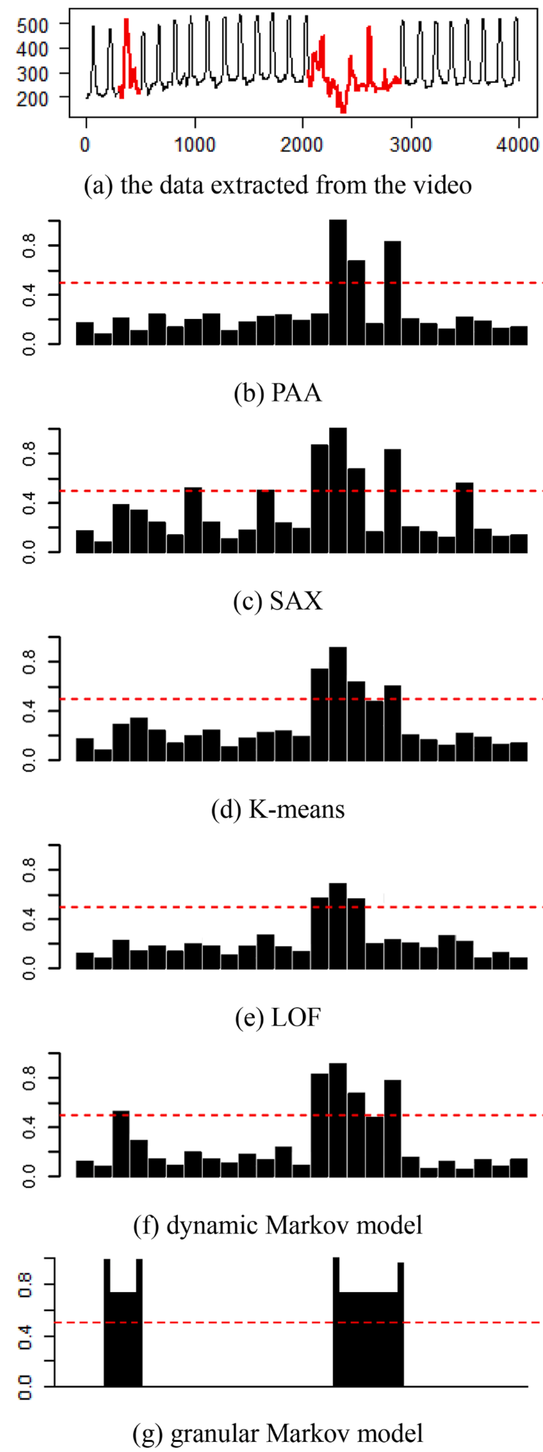


Fig. 13. Results of comparative experiments on the data extracted from the video.

model can detect all known anomalies with better performance of data anomaly resolution, yet it comes with a false alarm. Furthermore, the methods of PAA and SAX have undetected anomalies and the method of LOF has several false alarms. While the method of granular Markov model has a significant improvement over the other five methods in terms of detection accuracy and the performance of data anomaly resolution.

5.2. Real-world data

We employ five real-world data sets for the comparative experiment, which are publicly available and can be obtained from the website (https://www.cs.ucr.edu/~eamonn/discords/). These five data sets are all time series data, which have the characteristics of high levels of noise, dynamic changes, etc. Furthermore, the anomalies in these data are acknowledged by academic researchers, which are complicated and caused by changes in shape and amplitude (Izakian & Pedrycz, 2013; Keogh et al., 2005).

As shown in Figs. 10 and 11, they are both Electrocardiography (ECG) data with anomalies caused by the changes in amplitude and shape, respectively. One can notice that the methods of K-means, LOF

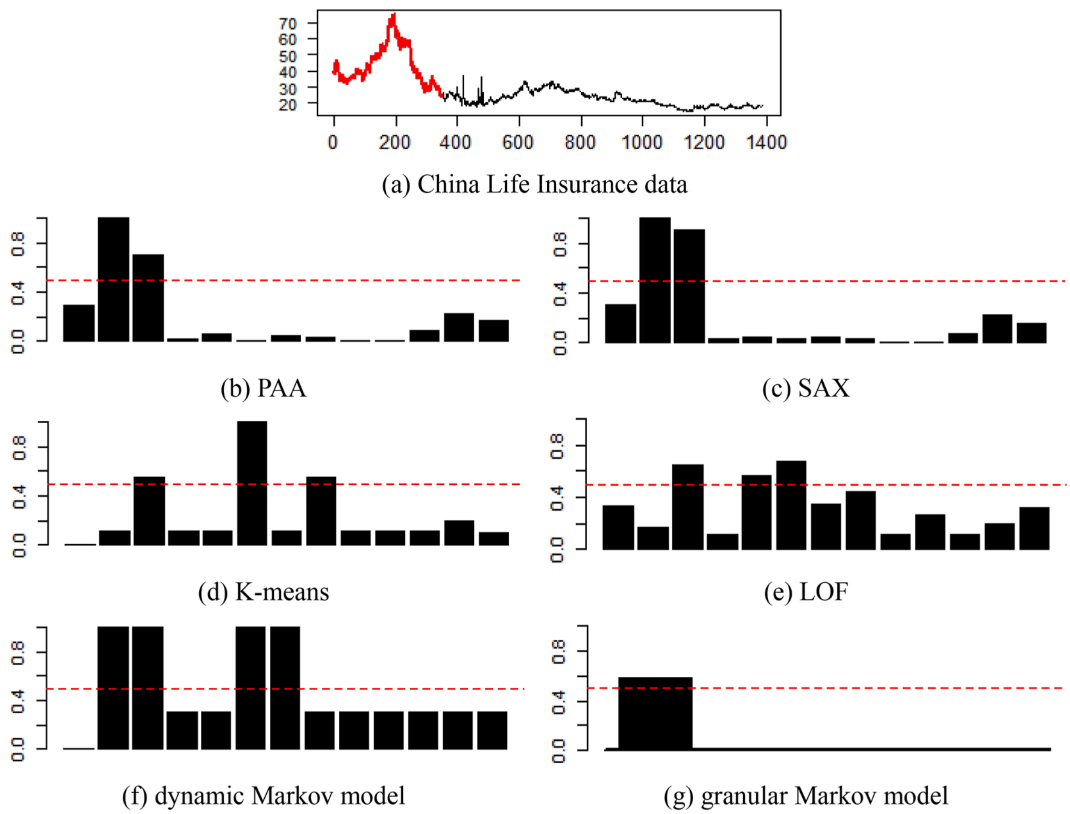


Fig. 14. Results of comparative experiments on China Life Insurance data.

and dynamic Markov model have similar detection results, that is, one anomaly is accurately detected, the other is not detected and their anomaly scores are very close to the threshold value. Both the methods of PAA and SAX have an undetected anomaly in Fig. 10(b) and 10(c). The methods of SAX, K-means and LOF have anomalies that are not detected in Fig. 11(c), 11(d) and 11(e), while both the dynamic Markov model and granular Markov model can accurately detect all anomalies in Fig. 11(f) and 11(g). Fig. 10(g) and 11(g) show that the granular Markov model can accurately detect all anomalies in both these two types of data with better performance of the data anomaly resolution, that is, the anomaly scores of all normal patterns are close to zero.

Figs. 12 and 13 are the spacecraft data and data extracted from the video, respectively. The red part in Fig. 12(a) is an anomaly, which is obviously caused by the shape change; while the anomalies in Fig. 13(a) are a bit complicated and are the results of the combined action of these two causes, namely changes in amplitude and shape. Both the methods of PAA and K-means produce a false alarm in Fig. 12(b) and 12(d), while the other four methods can accurately detect anomalies, and the granular Markov model has the best performance of data anomaly resolution. In Fig. 13, one can note that the SAX method has three false alarms, while the method of granular Markov model can accurately detect all anomalies. Furthermore, the granular Markov model offers better data anomaly resolution than the other methods. The other four methods have several undetected anomalies in Fig. 13(b) and 13(d)–13(f).

Fig. 14 presents the data from China Life Insurance, which have obvious tendencies and non-periodicity and are accompanied by large dynamic fluctuations with anomalies caused by amplitude changes. In Fig. 14, all of the methods of K-means, dynamic Markov model and LOF have several false alarms. The PAA and SAX methods have very similar detection results, that is, there are undetected anomalies and the anomaly scores of the normal patterns are also very close. The method of granular Markov model has better detection results both on passenger traffic data and China Life Insurance data. Meanwhile, the granular Markov model has better performance in data anomaly resolution.

The comparative experimental results from these five groups of real-world data are collected in Table 1, where expressions (1)/1 and 0/1 indicate that there is an anomaly in which one is accurately detected and the other is not. There are a total of 15 known anomalies in these five groups of real-world data, of which 15 anomalies are all accurately detected by the granular Markov model. The anomaly detection

Table 1
Summary of experimental results of the real-world data.

Data	Length	*w*	PAA		SAX		K-means		LOF		dynamic Markov model		granular Markov model	
			AR	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*
Fig. 10	3751	270	1/2	6.31	1/2	6.72	1/2	5.34	1/2	5.89	1/2	7.83	2/2	14.75
Fig. 11	3751	200	2/2	2.25	1/2	5.03	1/2	5.67	1/2	5.28	2/2	7.19	2/2	16.71
Fig. 12	5000	1300	1/1	3.32	1/1	4.10	1/1	3.03	1/1	4.17	1/1	3.85	1/1	5.21
Fig. 13	4000	150	3/7	2.14	4/7	2.56	4/7	1.89	3/7	1.65	5/7	2.24	6/7	3.20
Fig. 14	1380	120	2/3	4.27	2/3	4.35	1/3	4.08	1/3	3.31	2/3	3.26	3/3	4.87
total			10/15	18.29	10/15	22.76	9/15	20.01	8/15	20.30	12/15	24.37	15/15	44.74
average			67%	3.66	67%	4.55	60%	4.00	53%	4.06	80%	4.87	100%	8.95

Table 2
The results of comparative experiments.

data	length	*w*	PAA		SAX		K-means		LOF		dynamic Markov model		granular Markov model	
			AR	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*	*AR*	*CI*
chfdb_chf13_1	3750	130	1/2	2.74	1/2	2.99	1/2	3.25	1/2	2.72	1/2	3.41	2/2	4.62
chfdb_chf13_2	3750	100	1/2	2.15	2/2	1.71	2/2	3.34	1/2	3.23	1/2	2.33	2/2	4.15
chfdb_chf01_275_1	3751	210	1/1	4.42	1/1	2.98	1/1	3.50	1/1	5.17	1/1	3.17	1/1	4.28
chfdb_chf01_275_2	3751	250	1/2	1.31	2/2	1.44	2/2	1.81	1/2	1.90	2/2	1.79	2/2	2.22
mitdb_100_180	5400	110	2/3	2.00	2/3	1.57	2/3	2.53	2/3	2.23	2/3	2.92	2/3	3.13
xmitdb_x108_1	5400	150	1/2	1.65	1/2	2.83	2/2	3.44	2/2	2.16	2/2	3.34	2/2	3.77
xmitdb_x108_2	5400	170	1/3	3.35	2/3	3.19	2/3	1.72	1/3	2.74	2/3	2.83	3/3	3.56
mitdbx108_2	12,000	320	1/3	2.80	2/3	2.89	1/3	2.86	1/3	2.98	2/3	3.42	2/3	3.41
mitdbx108_3	12,000	280	1/3	2.23	2/3	2.39	2/3	2.31	1/3	3.25	3/3	2.61	2/3	2.45
qtbse1102_1	15,000	200	1/1	1.99	1/1	2.02	1/1	2.16	1/1	2.74	1/1	3.26	1/1	2.80
qtbse1102_2	15,000	200	1/1	1.62	0/1	1.54	0/1	1.63	1/1	2.45	1/1	1.32	1/1	1.97
qtbsele0606_1	15,000	150	1/3	3.17	2/3	4.80	1/3	3.95	2/3	3.56	2/3	4.69	2/3	5.50
qtbsele0606_2	15,000	160	2/6	4.39	4/6	2.43	3/6	4.35	5/6	2.93	4/6	2.71	5/6	4.49
chfdbchf15	15,000	230	3/5	3.03	2/5	1.85	2/5	2.49	2/5	1.89	4/5	2.16	4/5	3.32
mitdbx_mitdbx_108	20,000	340	7/13	4.48	6/13	4.94	5/13	4.41	8/13	2.32	9/13	4.67	11/13	5.19
total			25/50	41.33	30/50	39.54	27/50	43.75	30/50	42.27	37/50	44.63	42/50	54.94
average			50%	2.76	60%	2.64	54%	2.92	60%	2.82	74%	2.98	84%	3.66

accuracy rate reaches 100%, and its performance far exceeds the other five advanced methods. In general, one can note that the method proposed in this paper produces significant improvements in both the accuracy of anomaly detection and the performance of data anomaly resolution.

In the above experiments, the granular Markov method has higher detection accuracy and better data anomaly resolution than the methods of PAA, SAX, K-means, LOF, and dynamic Markov model. More real-world data are utilized for comparative experiments to support these conclusions that the granular Markov model method has significant advantages in anomaly detection. The comparative experiment results are collected in Table 2, where we note that the anomaly detection accuracy of the granular Markov model method is 84%, which is an improvement of about 13.51% over the other five methods. Furthermore, in terms of data anomaly resolution, the granular Markov model improves at least around 22.82%. Overall, the granular Markov model shows significant improvements in the detection accuracy and performance of data anomaly resolution and has apparent advantages over other methods.

6. Conclusions

This paper presents a novel anomaly detection method, which utilizes interval information granularity representation based on the principle of justifiable granularity to represent the original time series data and their amplitude data to arrive at the corresponding representation results, that is, interval information granules. Following the results of the interval information granularity representation, a granular Markov model is proposed for anomaly detection. The feasibility and superiority of the proposed method have been supported by a large amount of real-world time series data. Furthermore, the proposed method is comparable to the methods of PAA, SAX, K-means, LOF, and the dynamic Markov model. We conclude that the granular Markov model comes with higher anomaly detection accuracy and better data anomaly resolution. It also demonstrates the superiority of the proposed method in capturing weak anomalies. In the future, it is worth focusing on studies of various methods of abstract representation of different information granularities under different requirements, and employing them in the field of time series anomaly detection to realize multi-scale time series anomaly detection.

CRediT authorship contribution statement

Yanjun Zhou: Investigation, Software, Validation, Writing – original draft. **Huorong Ren:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Zhiwu Li:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Witold Pedrycz:** Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant K50510040013 and Grant K5051304007, in part by the Natural Science Foundation of China under Grant 61374068, and in part by the Science and Technology Development Fund, MSAR, under Grant 078/2015/A3, and the Doctoral Students' Short-Term Study Abroad Scholarship Fund of Xidian University.

References

Baker, D., Merkert, R., & Kamruzzaman, M. (2015). Regional aviation and economic growth: Cointegration and causality analysis in Australia. *Journal of Transport Geography, 43*, 140–150.
Baldán, F. J., & Benítez, J. M. (2019). Distributed FastShapelet Transform: A Big Data time series classification algorithm. *Information Sciences, 496*, 451–463.
Cai, Q., Chen, L., & Sun, J. (2015). Piecewise statistic approximation based similarity measure for time series. *Knowledge-Based Systems, 85*, 181–195.
Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR), 41*(3), 1–58.
Chandra, R., & Chand, S. (2016). Evaluation of co-evolutionary neural network architectures for time series prediction with mobile application in finance. *Applied Soft Computing, 49*, 462–473.
Chaovalit, P., & Gangopadhyay, A. (2011). Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys (CSUR), 43*, 1–37.
Chauhan, S., Vig, L., & Ahmad, S. (2019). ECG anomaly class identification using LSTM and error profile modeling. *Computers in Biology and Medicine, 109*, 14–21.
Esling, P., & Agon, C. (2012). Time-series data mining. ACM Computing Surveys (CSUR), 45, 1-34.
Fang, L., Wan, M., Yu, M., Yan, J., Liu, Z., & Wang, P. (2015). Analysis of similarity measure in the longitudinal study using improved longest common subsequence method for lung cancer. *Biomedical Signal Processing and Control, 15*, 60–66.
Gokcesu, K., & Kozat, S. S. (2017). Online minimax optimal density estimation and anomaly detection in nonstationary environments. IEEE Transactions on Signal Processing, 1-1.
Górecki, T., & Łuczak, M. (2015). Multivariate time series classification with parametric derivative dynamic time warping. *Expert Systems with Applications, 42*, 2305–2312.
He, H., & Tan, Y. (2017). Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering. *Applied Soft Computing, 55*, 238–252.

Huang, T., Zhu, Y., Zhang, Q., & Zhu, Y. (2013). An LOF-based adaptive anomaly detection scheme for cloud computing. In *Proceedings - International Computer Software and Applications Conference* (pp. 206–211).

Izakian, H., & Pedrycz, W. (2013). Anomaly detection in time series data using a fuzzy c-means clustering. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)* (pp. 1513–1518). IEEE.

Izakian, H., & Pedrycz, W. (2014). Anomaly detection and characterization in spatial time series data: A cluster-centric approach. *IEEE Transactions on Fuzzy Systems, 22*, 1612–1624.

Izakian, H., Pedrycz, W., & Jamal, I. (2012). Clustering spatiotemporal data: An augmented fuzzy C-means. *IEEE Transactions on Fuzzy Systems, 21*, 855–868.

Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence, 39*, 235–244.

Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM fully convolutional networks for time series classification. *IEEE Access, 6*, 1662–1669.

Keogh, E., Lin, J., & Fu, A. (2005). Hot SAX: Efficiently finding the most unusual time series subsequence. In Fifth IEEE International Conference on Data Mining (ICDM'05) (pp. 8 pp.): IEEE.

Keogh, E., Wei, L., Xi, X., Vlachos, M., Lee, S.-H., & Protopapas, P. (2009). Supporting exact indexing of arbitrarily rotated shapes and periodic time series under Euclidean and warping distance measures. *The VLDB Journal, 18*, 611–630.

Kumar, S., & Toshniwal, D. (2016). A novel framework to analyze road accident time series data. *Journal of Big Data, 3*, 1–11.

Lin, J., Khade, R., & Li, Y. (2012). Rotation-invariant similarity in time series using bag-of-patterns representation. *Journal of Intelligent Information Systems, 39*, 287–315.

Lu, W., Chen, X., Pedrycz, W., Liu, X., & Yang, J. (2015). Using interval information granules to improve forecasting in fuzzy time series. *International Journal of Approximate Reasoning, 57*, 1–18.

Moeyersons, J., Smets, E., Morales, J., Villa, A., De Raedt, W., Testelmans, D., … Varon, C. (2019). Artefact detection and quality assessment of ambulatory ECG signals. *Computer Methods and Programs in Biomedicine, 182*, 105050.

Münz, G., Li, S., & Carle, G. (2007). Traffic anomaly detection using k-means clustering. *GI/ITG Workshop MMBnet*, 13–14.

Nakamura, T., Taki, K., Nomiya, H., Seki, K., & Uehara, K. (2013). A shape-based similarity measure for time series data with ensemble learning. *Pattern Analysis and Applications, 16*, 535–548.

Park, H.-S., & Jun, C.-H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications, 36*, 3336–3341.

Pedrycz, W., Lu, W., Liu, X., Wang, W., & Wang, L. (2014). Human-centric analysis and interpretation of time series: A perspective of granular computing. *Soft Computing, 18*, 2397–2411.

Ren, H., Li, X., Li, Z., & Pedrycz, W. (2018). Data representation based on interval-sets for anomaly detection in time series. *IEEE Access, 6*, 27473–27479.

Ren, H., Liu, M., Liao, X., Liang, L., Ye, Z., & Li, Z. (2018). Anomaly detection in time series based on interval sets. IEEJ Transactions on Electrical and Electronic Engineering, 13, 757-762.

Ren, H., Ye, Z., & Li, Z. (2017). Anomaly detection based on a dynamic Markov model. *Information Sciences, 411*, 52–65.

Renza, D., Mendoza, S., & Ballesteros L., D. M. (2019). High-uncertainty audio signal encryption based on the Collatz conjecture. Journal of Information Security and Applications, 46, 62-69.

Shin, D.-H., Park, R. C., & Chung, K. (2020). Decision boundary-based anomaly detection model using improved AnoGAN from ECG data. *IEEE Access, 8*, 108664–108674.

Smith, J., Nouretdinov, I., Gammerman, A., Craddock, R., & Offer, C. (2014). Anomaly detection of trajectories with Kernel density estimation by conformal prediction. *IFIP International Conference on Artificial Intelligence Applications and Innovations*.

Van Der Voort, M., Dougherty, M., & Watson, S. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies, 4*, 307–318.

Wang, D., Pedrycz, W., & Li, Z. (2018). Granular data aggregation: An adaptive principle of the justifiable granularity approach. *IEEE Transactions on Cybernetics, 49*, 417–426.

Xie, Z., Lu, W., Liu, X., Xue, Y., & Yeung, Y. (2018). Copy-move detection of digital audio based on multi-feature decision. *Journal of Information Security and Applications, 43*, 37–46.

Yahyaoui, H., & Al-Daihani, R. (2019). A novel trend based SAX reduction technique for time series. *Expert Systems with Applications, 130*, 113–123.

Zhao, H., Dong, Z., Li, T., Wang, X., & Pang, C. (2016). Segmenting time series with connected lines under maximum error bound. *Information Sciences, 345*, 1–8.

Zhou, Y., Ren, H., Li, Z., & Pedrycz, W. (2021). An anomaly detection framework for time series data: An interval-based approach. *Knowledge-Based Systems, 228*, 107153.

Zhou, Y., Ren, H., Li, Z., Wu, N., & Al-Ahmari, A. M. (2021). Anomaly detection via a combination model in time series data. *Applied Intelligence, 51*, 4874–4887.

Zhu, X., Pedrycz, W., & Li, Z. (2016). Granular encoders and decoders: A study in processing information granules. *IEEE Transactions on Fuzzy Systems, 25*, 1115–1126.

Zhu, Y., Imamura, M., Nikovski, D., & Keogh, E. (2019). Introducing time series chains: A new primitive for time series data mining. *Knowledge and Information Systems, 60*, 1135–1161.