Summary of "Time Series Clustering: A Decade Review"

RE6124019_ 吳明軒

2024-05-14

Table of contents

1 Introduction

Time-series clustering is a technique used to classify large datasets without prior knowledge of class definitions. With the emergence of new concepts like cloud computing

and big data, research work has significantly increased to extract knowledge from this avalanche of data. This study aims to provide an updated investigation on the trend of improvements in efficiency, quality, and complexity of time-series clustering approaches over the past decade and to illuminate new paths for future works.

2 Types of Time-Series Clustering

2.1 1. Whole Time-Series Clustering

- **Shape-Based**: Matches the shapes of two time-series as closely as possible through non-linear stretching and contracting of the time axes.
- **Feature-Based**: Converts raw time-series into a lower-dimensional feature vector, then applies a clustering algorithm.
- **Model-Based**: Transforms raw time-series into model parameters.

2.2 2. Subsequence Clustering

Clustering segments extracted from a single long time-series.

2.3 3. Time Point Clustering

Clustering time points based on their temporal proximity and the similarity of the corresponding values.

3 Four Components of Whole Time-Series Clustering

3.1 1. Time-Series Representation

- **Data Adaptive Methods**: Such as Piecewise Linear Approximation, Singular Value Decomposition, etc.
- **Non-Data Adaptive Methods**: Such as Spectral Chebyshev Polynomials, Piecewise Aggregate Approximation, etc.
- **Model-Based Methods**: Such as Hidden Markov Model, ARMA, etc.
- **Data Dictated Methods**: Clipped.

3.2 2. Similarity or Distance Measures

- **Similarity in Time**: Calculates correlation-based distance or Euclidean distance.
- **Similarity in Shape**: Uses Dynamic Time Warping (DTW) method.
- **Structural Similarity**: Based on model parameters, such as Hidden Markov Models, ARMA, etc.

3.3 3. Clustering Prototypes

- Using Medoid as a prototype.
- Using Averaging as a prototype.
- Using Local Search as a prototype.

3.4 4. Time-Series Clustering

- **Hierarchical Clustering**: Does not require the number of clusters to be predetermined and has strong visualization capabilities.
- **Partitioning Clustering**: Such as k-Means and k-Medoids.
- **Model-Based Clustering**: Such as Self-Organizing Maps (SOM).
- **Density-Based Clustering**: Such as DBSCAN.

4 Evaluation

Evaluation of different methods in time-series clustering, focusing on efficiency, quality, and complexity.

5 Conclusion

Summarizes the progress in time-series clustering research over the past decade and points out potential future research directions.

6 Reference

1. Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering–a decade review. Information systems, 53, 16-38.
2. ChatGPT-4o