Discovering Similar Time-Series Patterns with Fuzzy Clustering and DTW Methods

Guoqing Chen, Qiang Wei Hong Zhang
School of Economics and Management, Tsinghua University
Beijing, 100084, China
{chengq, weiq}@em.tsinghua.edu.cn

Abstract

Data mining, as one of active fields nowadays, is to discover useful knowledge from large data sets. This paper focuses on continuous time-series data that have often been encountered in real applications (e.g., sales records, economic data, and stock transactions), and discusses how to discover the hidden relationship among time-series patterns in terms of their similarities. Fuzzy clustering and Dynamic Time Warping (DTW) methods are used to deal with fuzzy groupings of data attributes as well as with degrees of "distance" between time-series patterned attributes, respectively. An economic time series example is provided to help illustrate the ideas.

1. Introduction

Data mining, as one of active fields nowadays, is to discover useful knowledge from large data sets [1-3, 5, 6, 9, 10, 13, 14, 17, 18]. Discovering the relationships among time series data is of particular interest since the time series patterns reflect the evolution of changes in attribute values (values of variables) with sequential factors such as time, and has often been encountered in real applications in forms of sales, economics, and stock data. There may exist many types of relationships among time series data and therefore the methods used to study the relationships may be different [1, 2, 5, 8, 9, 11-14, 16]. Unlike the relationships studied by traditional approaches such as time series analysis and econometrics, the relationship represented by the similarities of the time series patterns is the focal point of this work [5, 7, 9, 12]. In other words, the value evolution of each time-series variable is viewed as a pattern over time, and the similarity between any two such patterns is measured by pattern matching.

Two major issues are involved in dealing with similar time series patterns. One is the measurement for pairwise similarities. The problems related to this issue center around how to define the difference between any two patterns, say, in terms of "distance" and how to match the series in points of time. Pattern recognition methods such as the DTW (dynamic time warping) method [11] will be used to match the series that correspond to each other at flexible points of time, including time lags. The other issue is the grouping of the similar patterns. As the similarity between any two patterns is a fuzzy concept in nature, fuzzy logic [4, 9, 15, 20], specifically fuzzy relations and clustering, plays an important role.

The rest of the paper is organized as follows. Section 2 discusses the similarity measures with primary attention paid to the DTW method, which leads to a closeness relation on the set of series patterns. Section 3 deals with fuzzy relations and clustering to group the time-series patterned attributes. Finally in section 4, the process of discovering similar time-series patterns and related algorithm aspects are discussed. Moreover, an example is provided to show how our approach is applied to a set of macro-economic time series data.

2. Time Series Data and Their Similarities

A time series pattern can be characterized by a data attribute with a series of values in time. For the sake of convenience, pattern and attribute are hereafter referred to interchangeably, otherwise indicated where necessary. The degree of similarity between any two
patterns is eventually the degree of matching between two series of values of the attributes concerned, which could be obtained by computing the "distance" pair-wisely in a fixed matching fashion as shown in Figure 1. In this case, the matching scheme for curves $a$ and $b$ cannot be applied to the matching between curves $b$ and $c$; vice versa. Thus, any pair of curves $a$, $b$ and $c$ reflects a certain matching scenario, which is static schematically.

![Figure 1. Static Matching Schemes](image)

Furthermore, the way to discover the similarities among the curves could be improved by matching the patterns dynamically. This can be done using the Dynamic Time Warping (DTW) method, a method used in speech recognition [11].

Given two series $S(s_1, s_2, \ldots, s_n)$ and $T(t_1, t_2, \ldots, t_m)$, $S$ and $T$ can be matched point to point, where $(i, j)$ represents that $s_i$ matches $t_j$, which is called a matching pair. The matching distance of $s_i$ and $t_j$ is defined as follows:

$$d(i, j) = |s_i - t_j|$$

Then, $s_i$ min-matches $t_j$, if in $s_1, s_2, \ldots, s_n$ and $t_1, t_2, \ldots, t_m$, the sum of matched distance of all the matching pairs (denoted as $r(i, j)$) is minimal. Formally,

$$r(i, j) = \min \sum_{k=1}^{p} \delta(i_k, j_k)$$

where $i_k = 1, \ldots, i, j_k = 1, \ldots, j, p = \max(i, j)$. Since $S$ and $T$ are time series, so pairs should be matched continuously, which means $0 \leq i_k - i_{k+1} \leq 1, 0 \leq j_k - j_{k+1} \leq 1$. Thus, $S$ matching $T$ means that $S_n$ min-matches $T_m$. Therefore the distance between $S$ and $T$ is:

$$DTW(S, T) = r(n, m) = \min \sum_{k=1}^{p} \delta(i_k, j_k)$$

where $i_k - 1, \ldots, i, j_k - 1, \ldots, m, p = \max(n, m), 0 \leq i_k - i_{k+1} \leq 1, 0 \leq j_k - j_{k+1} \leq 1$. Furthermore, based on the notion of dynamic optimization, we have

$$r(i, j) = d(i, j) + \min(r(i-1, j), r(i, j-1), r(i-1, j-1)).$$

The resultant match in $S$ and $T$ constitutes the warping route shown in Figure 2 (with $w_k = (i_k, j_k)$).

![Figure 2. Dynamic Time Warping](image)

Apparenty, $S$ and $T$ are not being matched in a fixed fashion. For instance, in the min-matched distance route, $s_2$ and $t_2$ is matched. Then the next matched pair is $(s_3, t_3)$ but not $(s_2, t_3)$. In addition, some applications may require considering more constraints in the matching process. An example is the time lags. If $T$ were to be related to $S$ in $t$ periods, then constraint $j - t \leq s \leq j + t$ might be added in the optimization model.

Thus, with the DTW method, any two time-series patterns are matched dynamically in distance. Further, there are a number of techniques to convert "distance" into "closeness", which is then normalized onto $[0, 1]$. As an example, the closeness between $S$ and $T$, denoted $c(S, T)$, could be obtained by $c(S, T) = e^{-\alpha DTW(S, T)}$ where $DTW(S, T) = r(n, m) \in [0, \infty]$. Notably, here the similarity of patterns is represented by $c$, which is a closeness relation (reflexive and symmetric) on the set of all time-series patterns. Let $U$ be the set of all time-series patterns concerned, then $c$ is a mapping from $U \times U$ to $[0, 1]$.

### 3. Clustering Similar Series Patterns

Clustering is one of the widely used methods in data mining [6, 7, 9, 12]. Given a set of objects (e.g., patterns or attributes), fuzzy clustering is to "divide" the set into several groups (clusters), where the objects in the same group have high degrees of similarity. Let
U be the set of all time-series patterns concerned and c be a fuzzy relation, i.e., closeness relation on U, obtained based on the DTW method, then the problem of discovering similar time series patterns is to cluster the patterns in U into groups. Each of such groups is a subset of U and the collection of all such groups covers U.

Usually, two types of clustering techniques are of consideration, namely, hard clustering and soft clustering. Hard clustering results in a partition of U, meaning that a specific attribute can only belong to one group. Formally, given U and c, hard clustering divides U into subsets A₁, A₂, ..., Aₙ such that

1)  A₁ ∪ A₂ ∪ ... ∪ Aₙ = U;
2)  Aᵢ ∩ Aⱼ = ∅, ∀i ≠ j, i, j = 1, 2, ..., n.

In so doing, a similarity relation s (reflexive, symmetric and transitive) on U is needed because its underlying crisp counterpart is an equivalent relation. Such s can be derived based on c according to the notion of transitive closure (or s = c').

However, in many cases, an attribute belonging to more than one cluster may also make sense. For example, in macroeconomics applications, GDP, Total Income and Total Investment can be clustered into one group representing a certain characteristic (e.g., monetary association), while GDP may also be clustered into another group with Total Population, representing a different type of characteristic of association. Therefore, soft clustering can be used. Concretely, given U and c, soft clustering can derive a number of subsets A₁, A₂, ..., Aₙ of U such that

1)  A₁ ∪ A₂ ∪ ... ∪ Aₙ = U;
2)  Aᵢ ∉ Aⱼ, ∀i ≠ j, i, j = 1, 2, ..., n.

Consider an example where U = {I, II, III, IV, V}, and c can be represented in a matrix form as shown in R below, which can then be equivalently represented by a complete graph correspondingly.

\[
R = \begin{bmatrix}
1 & 0.1 & 0.8 & 0.5 & 0.3 \\
0.1 & 1 & 0.1 & 0.2 & 0.4 \\
0.8 & 0.1 & 1 & 0.3 & 0.1 \\
0.5 & 0.2 & 0.3 & 1 & 0.6 \\
0.3 & 0.4 & 0.1 & 0.6 & 1
\end{bmatrix}
\]

For Figure 3 and h ≥ 0.3, three clusters can be derived: A = {I, III, IV}, B = {I, IV, V}, C = {II, V}.

4. Basic Algorithm Analysis and Example

First, the steps of the mining algorithm based on the DTW method and fuzzy clustering are provided as follows, along with corresponding analysis and optimization.

**Step 1: Preprocess the original time series.**

This includes to prepare for the relative values of the attributes based on the original ones in order to calculate DTW(S, T). Various standardization measures are available. A simple one could be the current value over the initial value.

**Step 2: Compute the distance of each pair of attributes using the DTW method.**

In this step, given N attributes, N(N-1)/2 times of dynamic programming operations will be carried out. If the length of the time series is n, r(i, j) is computed with dynamic programming, thus the complexity is
O(n²). So the total complexity of computation is O(N²x n²), which is polynomial.

**Step 3:** Convert "distance" into "closeness" and normalize.

**Step 4:** Construct the fuzzy relation matrix (and the corresponding complete graph), based on c from step 3.

**Step 5:** Cluster the attributes.

Directly searching for all the complete sub-graphs and filtering with λ will result in a computational complexity of O(2²⁻¹). Fortunately, further optimization can be made based on the following property shown in Theorem 1.

**Theorem 1:** Given G on U, for any A, B ⊆ U, A ⊆ B, if Gₐ is not a complete λ-sub-graph, then B is not a cluster (clustered group).

**Proof:** Since Gₐ is not a complete λ-sub-graph, there exists an edge in Gₐ, whose degree is less than λ. Because of A ⊆ B, this edge belongs to Gₐ. Thereafter Gₐ is not a complete λ-sub-graph, and B is not a cluster. □

Incorporating the property shown in Theorem 1 into the clustering process could help improve the algorithm efficiency. This logic of thinking is in analogue to that of the Apriori algorithm for discovering association rules [1, 2].

(a) Sort all the subsets of U in the order of listing the set after all of its subsets, and store in S;

(b) Compute the current subset. If the complete sub-graph of this subset is a λ-sub-graph, then store it in memory M and search in M to delete its subsets (w.r.t. iii). If the complete sub-graph of this subset is not a λ-sub-graph, then search in S to delete its supersets (Theorem 1);

(c) While not the end of S, go to step b;

(d) Output all the sets in M, these are the resultant clusters.

Finally, in concluding this work, a brief example of clustered macro-economic data [19] is provided as shown in Table 1. The data consists of 93 economic variables each being a time series of 20 years.

<table>
<thead>
<tr>
<th>Group No.</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Total Consumption, Social Consumption</td>
</tr>
<tr>
<td></td>
<td>Commercial Businesses, Labor</td>
</tr>
<tr>
<td>Group 2</td>
<td>City Residents Consumption, GDP, GNP</td>
</tr>
<tr>
<td>Group 3</td>
<td>GDP, Fixed Asset Investment, Added Value of Transportation and Telecommunication</td>
</tr>
<tr>
<td>Group 4</td>
<td>Agricultural Labor, Population, Agricultural Population</td>
</tr>
<tr>
<td>Group 5</td>
<td>Citizen Consumption Index, Peasantry Consumption Index, Fixed Asset Investment Index, Social Goods Retail Index</td>
</tr>
</tbody>
</table>

Table 1. Clusters of Time-Series Patterns (Partly)

This preliminary result revealed reasonable groupings of time series attributes. While more results are to be examined, they could also be used to support further investigation of analytical relations or functional associations with such traditional approaches as econometric modeling, etc.

5. Concluding Remarks

This paper has presented an approach to discovering similar time series patterns. A closeness relation has been used to measure the degree of similarity between two patterns, which is derived using the DTW method in pattern matching. This method enables dynamic pattern matching. Consequently, fuzzy clustering techniques have been used to group similar patterns. Furthermore, a mining algorithm has been developed and then improved based on a certain property used as the pruning strategy. Finally, a preliminary result on economic series clustering has revealed certain reasonable outcomes. Future research is being carried out to analyze more results with the economic series data, as well as to develop other algorithm optimization strategies.
6. Reference


