Clustering – A brief introduction
With examples from time series clustering

Presented by
Saima Aman
Agenda

- Intro to Clustering
- Different algorithms
- 3 examples – 1 general and 2 from time series clustering
Clustering

- Groups of objects that are similar to others within a group
- Dissimilar to objects in other groups

Source: Tan, Steinbach, Kumar
Issues

- How many clusters?
- Six Clusters
- Two Clusters
- Four Clusters
Types of Clustering

- **Exclusive versus non-exclusive**
  - Points belong to single / multiple clusters

- **Fuzzy versus non-fuzzy**
  - a point belongs to every cluster with some weight between 0 and 1 (probabilistic clustering)

- **Partitional versus hierarchical**
Clustering Algorithm

- K-means clustering
- Hierarchical clustering
- Density-based clustering
K-means clustering

**Basic algorithm:**
1. Select K points as initial centroids
2. **Repeat:**
   a. Form K clusters by assigning all points to the closest centroid
   b. Re-compute the centroid of each cluster
3. **Until:** The centroids don’t change (or very few points change)

**Highlights**
- Closeness is measured by Euclidean distance, cosine similarity, correlation, etc.
- Initial centroids chosen randomly – it can affect final results
- One soln: Multiple runs
- Issues: When data has outliers, clusters of different sizes and densities
- Evaluation – SSE (sum of squared errors) or external labels
- SSE can be reduced by increasing K – the number of clusters
K-means iterations
Hierarchical Clustering

- No need to specify number of clusters
- Any number of clusters can be obtained by ‘cutting’ the dendogram at the proper level
- **Agglomerative:** – more popular
  - Compute proximity matrix
  - Start with the points as individual clusters
  - At each step, merge the two closest pairs, update the proximity matrix and repeat until only one cluster left

- Inter-cluster similarity - distance based similarity
Similarity Matrix
Example - 1

- Data features:
  - Id, Age, Gender, Region, Income, Mortgage, Car
  - Demographic clusters

- Steps required: Specify # of clusters, and initial cluster centers
- Results: Different clusters – each specified by mean vectors
- 3-d Visualization – (x-axis, y-axis, color)
- Different combinations can show different relationships within a cluster
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

Problem: How to identify patterns and trends on multiple time scales (days, weeks, seasons) simultaneously.

(Patterns & Trends – energy use, product sale, employee attendance, etc.)

Goal: Merge similar day patterns into clusters

Approach: Each day pattern, \( Y_j \) \((j = 1, \ldots, M)\) consists of a sequence of pairs \((y_i, t_i)\) - where \(y_i\) denotes the measured value, and \(t_i\) denotes the time that has elapsed since midnight

Algorithm: (based on agglomerative hierarchical clustering)

- Start with \(M\) clusters, each cluster containing one day pattern.
- Compute the mutual differences between all clusters, and merge the two clusters which are most similar into a new cluster. Result: \(M-1\) clusters
- Repeat until a single cluster results
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

Distance measures:

<table>
<thead>
<tr>
<th>Distance measure</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root Mean Square distance</td>
<td>$d_{rms} = \sqrt{\frac{\sum(y_i - z_i)^2}{N}}.$</td>
<td>Robust, gives good results</td>
</tr>
<tr>
<td>Normalized Distance</td>
<td>$d_{nm} = \sqrt{\frac{\sum\left(\frac{y_i}{\max{y}} - \frac{z_i}{\max{z}}\right)^2}{N}}.$</td>
<td>Cluster patterns with similar shapes</td>
</tr>
<tr>
<td>Excluded offset (eliminates slow trends)</td>
<td>$d_{sh} = \sqrt{\frac{\sum(y_i - z_i - \Delta)^2}{N}},$ \quad \Delta = \frac{\sum(y_i - z_i)}{N}.$</td>
<td>2 patterns are equal, if they are same, except for an offset</td>
</tr>
<tr>
<td>Peak difference</td>
<td>$d_{ma} =</td>
<td>\max{y} - \max{z}</td>
</tr>
</tbody>
</table>
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

Visualization based on dendrograms:

(Full clustering tree for 365 day pattern)
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

Visualization based on calendar: (colors indicate clusters and patterns of attendance)
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

Visualization based on calendar: (colors indicate clusters and patterns of energy use)
Ex- 2: Cluster and Calendar based Visualization

Source: Jarke J. van Wijk and Edward R. van Selow. Cluster and Calendar based Visualization of Time Series Data (INFOVIS’99)

- Motivation - liberalization of the energy markets.

- In the Netherlands, customers with very high energy consumptions are allowed to choose their electricity supplier and negotiate a tailor-made tariff.

- Energy companies need to transform from utility companies into market-oriented companies.

- Consumption patterns
  - essential for the segmentation of customer markets.
  - Needed by utility and consumers for negotiation
Ex- 3: Unsupervised learning of human behavior

- Observations acquired from sensors in a smart home
- Sensor stream represented as a sequence of tokens

- Unsupervised approach based on compression and text analysis
- Use a dictionary-based compression algorithm (LZW algo)
- LZW is used to create a codebook of potential patterns
- LZW does not generalize to variations in input:

  AAAI  ➔  AAyAI  ➔  ABAI
Ex- 3: Unsupervised learning of human behavior

Source: Sook-Ling Chua and Stephen Marsland and Hans W. Guesgen in AAAI, 2011

- A new method proposed here –
  - Extend LZW method to perform lossy compression
  - First, LZW used to create a dictionary of phrases
  - Next, lossy matching is performed – allowing relatively minor changes between the input and the dictionary words
Ex- 3: Unsupervised learning of human behavior

Source: Sook-Ling Chua and Stephen Marsland and Hans W. Guesgen in AAAI, 2011

- LZW produces a large dictionary - everything that has been learnt during training, including all the substrings of each dictionary word.
- Only patterns of interest are those that are the longest frequent words in the dictionary.
- Clustering basis - the minimum number of actions required to transfer one string $p$ into another string $q$, where an action is a substitution, deletion, or insertion of a character into the string.
Thanks!