Welcome to Furuhashi Lab.
Nagoya University

Research Library of Furuhashi Group

http://www.emplx.cse.nagoya-u.ac.jp/~fuzzdata/
Members of Furuhashi Group

- Professor
  Takeshi Furuhashi

- Associate Professor
  Tomohiro Yoshikawa

- A student in doctor course
- 6 students in master course
- 2 undergraduate students

For contact

TEL: +81-52-789-5315
FAX: +81-52-789-3166
E-mail: furuhashi@cse.nagoya-u.ac.jp
Visualization of multivariate data is one of the key technologies in the fields of data-mining, kansei engineering, chance discovery, etc. Visualization helps a user to interact with the computer to extract useful knowledge from multi-dimensional data, to find clues for data analysis, and to discover future trends. Visualization also helps other people to know and discuss a personal view on the data.

Here we introduce our visualization methods consisting of clustering and projection. We take the following three criteria for transparency of visualized multivariate data.

- Linearity of projection
  Linear projection of clusters is easy for understanding the input-output relationships.
- Separation of clusters
  Visualized clusters should be separated.
- Parallel distribution of clusters to the projection axes
  Each visualized cluster is desirable to be distributed parallel to the projection axes for easy labeling.

We have proposed the following visualization methods that can meet these requirements: (1) A fuzzy clustering and Fuzzy Multiple Discriminant Analysis (FMDA). The FMDA incorporates a linear projection, the separation and parallel distribution criteria. (2) A new clustering method that incorporates the separation criterion of the visualized clusters.

(1) Fuzzy clustering and FMDA

The first method[1] applies a fuzzy clustering to input-output data. The obtained clusters are projected onto the input space. Then the clusters in the input space are visualized by Fuzzy MDA (FMDA). The FMDA that uses a linear projection incorporates the separation criterion of the clusters in the visible space.

This method can construct a fuzzy model on the projected input space. The structure of Box and Jenkins’s gas furnaces data is found to be almost linear from the constructed visible model as shown in the figure below:

We extended this method by employing EM-Clustering / Kernel-FCM that used a kernel function for the clustering. We, then, proposed a new discriminant analysis (FMDA2)\([2]\) that incorporated the correlation of data in each cluster to the projection axes.

The figures below show the results of visualization of test data identified by the FMDA and FMDA2.

You can see that the data in the clusters projected by the FMDA2 are well separated and distributed parallel to the axes.

We applied this method to a set of sensor data collected from a mobile robot while it was passing an aisle shown in the left hand side figure below. The robot had six inputs and one output. Our method was able to construct an interpretable visualized space as shown in the right hand side figure, and the control rules of mobile robot were extracted from the visualized data.

From the visualized data, the following four simple rules were obtained:

If (the nearest wall is on the right and it is not near) then Go Ahead
If (the nearest wall on the right is very near) then Turn Left
If (the nearest wall is on the right or fore and it is near) then Turn Small Left
If (the nearest wall on the left is near) then Turn Right

(2) A new clustering method that incorporates the separation criterion of the visualized clusters.

The above methods carry out clustering and projection independently. We proposed a new clustering method[3] that reflected the result of visualization. This method automatically identifies clusters, which maximizes the ratio of between-cluster scatter to within-cluster scatter in the projected space.

We employed iris and wine data in the UCI database (http://www.ics.uci.edu/~mlearn/MLRepository.html) for experiments. This table shows features of data.

<table>
<thead>
<tr>
<th></th>
<th># of data</th>
<th># of variables</th>
<th># of clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iris data</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Wine data</td>
<td>178</td>
<td>13</td>
<td>3</td>
</tr>
</tbody>
</table>

We applied our method and the conventional method to iris data on the assumption that the species of iris were unknown. The results are shown in the figures below. The proposed method clustered data properly, and the clusters were well separated in the projected space. And you can derive a hypothesis that three species of iris can be separated only by the size of petal.

We also applied our method to the wine data, and we obtained an interpretable visualized space.
One of the knowledge representation methods for a computer is to compose a network of symbols. A symbol means a word in human language, and the meaning of each symbol is discovered by linkages among the symbols in the composed network.

A human can construct a symbol network based on sensorimotor information obtained from interactions with the environment. We think that the computer is able to represent the relationships among symbols from quantities of features measured from the real-world-objects through its various sensors, e.g. pressure sensors, vision sensors, and so on.

The purpose of this study is to visualize the structure of the symbol network and to clarify the relationships among the symbols. The sensorimotor information is multi-dimensional data. The relationships between the objects and the symbols are visualized, and the relationships among the symbols are to be clarified.
Toward Visualization of Individuality

Design of industrial products as well as their functions and performances are very important. Impressive design leads to a boost of product in sales. The concept of new design is expressed with words. The problem here is that individual’s impressions of words differ. These differences often cause misunderstandings among the engineers and designers. If the differences of the impressions to the same words are visualized, communications among them will be smooth and productive.

We try to develop a method that visualizes the similarities/differences of the meanings of words between individuals.

Toward Knowledge Extraction from Medical Documents

In the near future, all of medical data will be stored in databases. Although analyses of numerical ones in these databases have been developed, knowledge extraction from text data has not been successful.

The purpose of this study is to discover useful knowledge from digital medical documents. We focus on “incident reports” including reports that might have caused medical errors.

We think that mapping key words selected from the reports into visible space leads to the extraction of useful knowledge such as risky behavior of nurses.
Toward Electroencephalogram-based Control

This study deals with electroencephalogram(EEG)-based control of a mobile robot. This research has started in collaboration with Prof. Tanaka’s team (at Univ. of Electro-Communications).

We try to extract the features of orders to the mobile robot from the EEG data.

Toward Efficient Search through Visualization

Performance in search for solutions by Genetic Algorithm (GA) depends on the genetic operators or their parameters, such as the number of chromosomes, probability of crossover/mutation and so on. If the solution space is less or equal to 3 dimensions, we can display it and watch the searching process. It is, however, difficult to visualize them in the real combinational optimization problems for GA, because they have a multi-dimensional search space. Therefore, the efficient genetic operators and their parameters are usually determined by trial and error.

The purpose of this research is to visualize the searching process in a multi-dimensional space to obtain the effects of genetic operations such as the diversity of chromosomes. This information could be utilized to devise and check genetic operations and/or their parameters. Self Organization Map (SOM) is used to visualize the searching process.
Toward Effective Teaching of Mobile Robot

This study is aimed at teaching robots from the viewpoint of robots themselves. The point of this research is to visualize robots’ multi-dimensional sensing data for easy understanding by a human operator.

The motivation of this research is that there are differences between operator’s own viewpoint and that of robots’ ones. Without considering the differences, teaching could not be successful. The difficulty here is that the robots have multi-sensors and this makes the operator not to be easy to grasp what the robots see and what not.

Toward Effective Display of Information

A car navigation system should avoid attracting drivers’ too much attention. This study is aimed at clarifying a guideline for effective display of information for quick recognition by people.

We have been conducting experiments called “visual search”, and measuring subjects’ reaction time in order to clarify the relationships between serial processing and grouping of stimuli. As shown in the bottom figure, in the case where white circles are paired, it is easier for us to count the number of them.