Information Visualization, and Winner-Takes-All

Masa Takatsuka
Introduction

Figure 1.1 shows the relationship between data, information, knowledge and wisdom. Data are processed and organized to become information. Humans gain knowledge through aggregating different pieces of information where the irrelevant parts have been filtered out. Knowledge is turned into wisdom when a person is able to recall all previously obtained knowledge to solve unanswered problems.

Data Visualization

- Scientific Visualization
- Information Visualization
Introduction

on the velocity and direction at each point in the current, or rendering a 3D shape of large complex molecules according to their structures. Two scientific visualizations are shown in Figure 1.3. The first visualization (Figure 1.3(a)) simulates the solar storm corona mass ejection blast (solar wind) and subsequent impact at Earth. The blue paths emanating from the Earth’s two poles represent some of its magnetic field lines [116]. The second visualization (Figure 1.3(b)) reconstructs the 3D shape of a protein molecule called porin [155]. The molecule contains 2219 atoms.

(a) Blasts of particles and magnetic field from the Sun that impact the magnetosphere around the Earth [116].

(b) 3D shapes of a protein molecule porin [155].

• Information Visualization focuses on abstract data which do not have inherent geometrical structure, such as financial data, collections of documents or DNA micro-arrays. Since such data sets do not have obvious spacial mappings, the research problem is how to transform them into efficient visual forms. Here, ‘efficient’ means that the viewer can correctly interpret the visualization as fast as possible. Different types of data usually require different visualization schemes. Figure 1.4 shows two examples. The first picture (Figure 1.4(a)) visualizes the file organization on a computer hard-disk [175]. The visualization method is called the Tree-Map [141] which is especially good for representing large hierarchical structures. The display space is recursively divided according to structure of the file system. Files are shown as rectangular cushions. The area of each cushion is proportional to the actual size of the corresponding file. In case of a directory, the rectangular cushion representing


Information Visualization

- Abstract data
- No explicit geometries are associated with
- No obvious pictorial/graphical representation
- Often require InfoVis expertise to create “good” visualization
Good Visualisation

• Intuitive
• See what you expected (fast)
• See what you unexpected (fast)
Example of Abstract Data

- Financial data
- Natural Languages
  - Text / Audio Analysis
- Network traffic/telecommunication data
- Bioinformatics data
- Socio-demographic data
- Kansei engineering data (psychological/categorical)
showing relationships
Introduction

Node-link diagrams are also referred to as graphs. Layouts of a graph have significant impact on the accuracy and speed for people to interpret the data. Figure 1.8 shows two drawings of the same graph. Obviously, the right one is clearer and easier to understand. Graph drawing algorithms aim to generate layouts with good readability. Various graph drawing algorithms have been designed to handle graphs of different structures. This thesis focuses on algorithms which are able to handle general graphs, i.e. the algorithms which can produce acceptable layouts for various graph structures. Related algorithms and concepts will be discussed in the next chapter.


Besides node-link diagrams, other visualization schemes have been invented to represent relational data. Most of them were designed for special purposes. For instance, the Tree-Map (see Figure 1.4(a)) specializes in visualizing graphs with hierarchical structure. Another example, the matrix view, is designed for representing very large and dense graphs. The matrix view represents a graph using an $V \times V$ adjacency matrix $A$, where $V$ is the number of vertices on the graph. If vertex $v_i$ is connected to $v_j$, then the entry $a_{ij}$ in the matrix is set to be 1 or else 0 (see Figure 1.9(b)). This method effectively saves display-space since the edges are represented as dots. However, compared to the node-link diagram, it is more difficult for viewers to follow paths between vertices.

This thesis uses the standard relational data visualization method — the node-link diagrams.

Figure 1.9: Two representations of a relational data set which contains 50 vertices and 400 edges.

(a) Node-Link Diagram  
(b) The Matrix View

• Showing attributes
1.3 Data Types

For data sets with up to three attributes, a straightforward solution is to plot the entities into a coordinate system (see Figure 1.5). Each dimension of the coordinate system represents one attribute. This method can reveal the structure of the attribute space, i.e. whether the entities are grouped into clusters, how the clusters relate to each other and whether there are anomalies/outliers. In addition, how the entities are separated into clusters along different dimensions can also be observed. Correlation of the attributes can be perceived by their common increasing/decreasing trends.

(a) One dimensional data  
(b) Two dimensional data  
(c) Three dimensional data

However, because human eyes cannot see spaces of more than three dimensions, more complex techniques are needed to represent data sets of larger dimensions. The main idea is to transform multidimensional data into low dimensional representations (two or three dimensions). Different multidimensional data visualization techniques are good at different tasks. Some of them aim to reveal the structure of the attribute space while others excel in showing the attributes' correlations. These techniques are reviewed in the next chapter.

Figure 1.5: Plotting the entities into a coordinate system according to their attribute values.

However, because human eyes cannot see spaces of more than three dimensions, more complex techniques are needed to represent data sets of larger dimensions. The main idea is to transform multidimensional data into low dimensional representations (two or three dimensions). Different multidimensional data visualization techniques are good at different tasks. Some of them aim to reveal the structure of the attribute space while others excel in showing the attributes' correlations. These techniques are reviewed in the next chapter.

Figure 1.6 gives two examples of multidimensional data visualization. The first example (see Figure 1.6(a)) represents a car simulation data set [99] produced by the BMW group, Munich, Germany. Each car has five attributes and is represented as a glyph. Details of the attributes were not described in the original paper [99] since the car models had not been officially announced. The first three attributes are used to determine the position of the glyph in a three dimensional space. The other two attributes are mapped to shape (cube, octahedron or sphere) and color of the corresponding glyphs.

The second example visualizes research papers submitted to the annual Association of American Geographers conference from 1993 to 2002. The visualization uses a metaphor of geographical map [145]. Each paper is represented by a point. Positions of the paper are determined by their attributes. The diagram illustrates how the papers are clustered and correlated in the attribute space.
Spatialisation

IN-SPIRE, PNNL
Spatialization

- find 2D or 3D position of data entries
- typically used to visualize similarities
Our Approach

• Spatialise n-dim data
• Use Self-Organizing Maps
  • non-linear mapping
  • topological mapping

• Create familiar visualisation : MAP
Self-Organizing Map?

the striate cortex of a cat

neurons respond to particular angles of light stripes

Hubel and Wiesel 1963

Fig. 4. Reconstruction of a microelectrode penetration through the postlateral gyrus of the left hemisphere. This 2½-month-old kitten had its right eye covered from birth by lid suture. Lines intersecting the electrode track represent cortical cells; directions of these lines indicate the receptive-field orientations. Crosses indicate cortical cells uninfluenced by light stimulation. Simultaneous recordings from two units, which occurred three times in this penetration, are each indicated by only one line or cross. A lesion was made while recording from the first unit, and another at the end of the penetration; these are marked by small circles. The ocular-dominance distribution of units recorded in this penetration is shown in Fig. 3. All fields positioned 5–6° to the left of the area centralis, slightly below the horizontal meridian. Scale, 0.5 mm.
Artificial Model of Self-Organization

- von der Malsburg (1973) simulated feature extracting cells organizing in a 2D space
- lateral excitatory/inhibitory connections.
Mathematical Model (for Biological SOM)

- Amari (1980) topographic organization of nerve fields
- Linsker (1986) Emergence of column orientation
- Tanaka (1990) Cortical Map formation model

Figure 1. Connection of neural fields

Amari (1980)
Kohonen’s Self-Organizing Map

- Self-Organizing Map
- Telvo Kohonen 1982

@article{KOHONEN82,
Author = {Kohonen, T.},
Journal = {Biological Cybernetics},
Pages = {59--69},
Title = {Self-Organized formation of topographically correct feature maps},
Volume = 43,
Year = 1982}

- unsupervised learning
Input Data (multidimensional)

**Table 3.6. Animal names and their attributes**

<table>
<thead>
<tr>
<th></th>
<th>d</th>
<th>o</th>
<th>h</th>
<th>g</th>
<th>a</th>
<th>e</th>
<th>w</th>
<th>t</th>
<th>i</th>
<th>l</th>
<th>h</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>goose</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>owl</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eagle</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2 legs</th>
<th>4 legs</th>
<th>has</th>
<th>likes</th>
<th>run</th>
<th>to fly</th>
<th>swim</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>hair</td>
<td>0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1</td>
<td>0 0 0 0 1 1 1 1 0 1 1 1 1 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medium</td>
<td>0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>hooves</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 1 1 1 1 0 1 1 1 1 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>big</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>mane</td>
<td>0 0 0 0 0 0 0 1 0 0 1 1 1 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>feathers</td>
<td>1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0 0 0 1 1 1 1 0 1 1 1 1 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>hunt</td>
<td>0 0 0 0 1 1 1 1 0 1 1 1 1 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0 0 0 0 0 0 1 1 0 1 1 1 1 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>run</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>to fly</td>
<td>1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>swim</td>
<td>0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 3.29. After the network had been trained with inputs describing attribute sets from Table 3.6, the map was calibrated by the columns of Table 3.6 and labeled correspondingly. A grouping according to similarity has emerged.
SOM

- Input Data
  - Multidimension
  - Raw data (Table) do not reveal much

- Output
  - mapped onto 2D (or 3D)
  - similar features are mapped onto similar locations
  - easy visual confirmation
SOM Algorithm

- Input Data
  - $X_1, X_2, ..., X_i, ..., X_n$ : Animal
  - $X_i - (x_{i1}, x_{i2}, ..., x_{ik}, ..., x_{id})$ : Attributes

- Map
  - 2D grid: rectangle or hexagonal
  - each datum is mapped to one of grids
• Allocate “weight vector: W” at each grid
• each W has the same dimension as the input
SOM Algorithm: Initialization

- Input data: $X_1, X_2, X_3, X_4$
- Data Attributes: $3\text{Dim } X_1 = (x_{11}, x_{12}, x_{13})$
- Map: $5 \times 5$

$$W_{(5,5)} = (w_{(5,5)1}, w_{(5,5)2}, w_{(5,5)3})$$
SOM Algorithm: Finding BMU

- Look for $W$ closest to $X_1$

$X_1 = (x_{11}, x_{12}, x_{13})$

- $W_{\text{win}}$ : Best Matching Unit
SOM Algorithm: Learning I

- Make $W_{\text{win}}$ closer to $X_1$

$X_1 = (x_{11}, x_{12}, x_{13})$

$$W_{\text{new}} = W_{\text{old}} + \alpha (X_1 - W_{\text{old}})$$
SOM Algorithm: Learning II

- Make $W_{\text{win}}$'s surrounding closer to $X_1$

\[
X_1 = (x_{11}, x_{12}, x_{13})
\]

\[
W_{\text{new}} = W_{\text{old}} + \alpha n (X_1 - W_{\text{old}})
\]
SOM Algorithm: Learning III

- Repeat the process I, and II
SOM Algorithm: Mapping

- similar features are mapped close together
• WEBSOM
  http://websom.hut.fi/websom
  Maps documents (> million) onto a 2D space using SOM

• Text-mining
Example
Visualising Socio-demographic behaviour

- CO2 emission levels and related attributes
  - 8 attributes
- 1980 - 2004
- 21 countries (>100 million metric tons emission)
Attributes

- Population (millions)
- Total CO2 emission
- Per capita emissions
- Emissions from petroleum
- Emissions from natural gas
- Emissions from coal
- Total primary energy production
- Total primary energy consumption
7.2.1 Attribute space visualization

Figure Xor depicts the resulting visualization when the Geodesic SOM is trained with the data set. The Wagner III cartographic projection technique is used so an entire view of the visualization can be seen. In the centre of the figure, it can be observed that there are a group of clusters. These are countries with extremely high emission levels that are in the thousands of million metric tons. The other countries have emission levels in the hundreds of million metric tons. The eight component planes are shown in figures xsnxosnxo and can be used to determine which attributes may be related to each other.

Figure xor: This figure depicts the Geodesic SOM after it has been trained with the carbon emissions data set. Large dark blue regions indicate that the Geodesic SOM has created a smooth distribution of the data. A group of clusters can also be seen where countries with high emission levels have been mapped. Threelitter country codes following the ISO trwwnr alphant standard have been used to indicate where the data for each country is generally located. The exception here is that the European Union has been abbreviated as EU.
CHAPTER 7. EXPERIMENTAL RESULTS: SOCIODEMOGRAPHIC DATA

7.2.1 Attribute space visualization

Figure 7.1: This figure displays a close-up view of the USA's location on the Geodesic SOM. The data has been placed in a cluster due to its high emission levels in trillions of metric tons.

Figure 7.2: This figure displays a close-up view of the China's location on the Geodesic SOM. The data has been placed in a cluster due to its high emission levels in trillions of metric tons.

Figure 7.3: This figure depicts the Geodesic SOM after it has been trained with the carbon emissions data set. Large dark blue regions indicate that the Geodesic SOM has created a smooth distribution of the data. A group of clusters can also be seen where countries with high emission levels have been mapped. Threethree-letter country codes following the ISO alphabetic standard have been used to indicate where the data for each country is generally located. The exception here is that the European Union has been abbreviated as EU.
7.2.1 Attribute space visualization

Figure 7x or depicts the resulting visualization when the Geodesic SOM is trained with the data set. The Wagner III cartographic projection technique is used so an entire view of the visualization can be seen. In the centre of the figure, it can be observed that there are a group of clusters. These are countries with extremely high emission levels that are in the thousands of million metric ton. The other countries have emission levels in the hundreds of million metric ton. The eight component planes are shown in figures x0s and can be used to determine which attributes may be related to each other.

Figure 7x0r: This figure depicts the Geodesic SOM after it has been trained with the carbon emissions data set. Large dark blue regions indicate that the Geodesic SOM has created a smooth distribution of the data. A group of clusters can also be seen where countries with high emission levels have been mapped. Three-letter country codes following the ISO trwwnr alphant standard have been used to indicate where the data for each country is generally located. The exception here is that the European Union has been abbreviated as EU.

Figure 7xss: This figure displays a close up view of the US' location on the Geodesic SOM. Its data has been placed in a cluster due to its high emission levels.

Figure 7xos: This figure displays a close up view of China's location on the Geodesic SOM. Its data has been placed in a cluster due to its high emission levels.
High-Performance Visualization, Communication, Computing

7.2.1 Attribute space visualization

Figure xor depicts the resulting visualization when the Geodesic SOM is trained with the data set. The Wagner III cartographic projection technique is used so an entire view of the visualization can be seen. In the centre of the figure, it can be observed that there are a group of clusters. These are countries with extremely high emission levels that are in the thousands of million metric tons. The other countries have emission levels in the hundreds of million metric tons. The eight component planes are shown in figures xosnxoz and can be used to determine which attributes may be related to each other.

Overall, the SOM distributes the data in a fairly smooth manner such that the emission levels generally decrease when moving away from the location of the US’ cluster. The smooth distribution is indicated by the generally low distances between the neurons and their direct neighbours on the Geodesic SOM. This is also indicated by the large blue regions on the Geodesic SOM.
CHAPTER XI \ EXPERIMENTAL RESULTS: SOCIODEMOGRAPHIC DATA

7.2.1 Attribute space visualization

Figure X.1 depicts the resulting visualization when the Geodesic SOM is trained with the data set. The Wagner III cartographic projection technique is used so an entire view of the visualization can be seen. In the centre of the figure, it can be observed that there are a group of clusters. These are countries with extremely high emission levels that are in the thousands of million metric tons. The other countries have emission levels in the hundreds of million metric tons. The eight component planes are shown in figures X.2 to X.5 and can be used to determine which attributes may be related to each other.

Figure X.1: This figure depicts the Geodesic SOM after it has been trained with the carbon emissions data set. Large dark blue regions indicate that the Geodesic SOM has created a smooth distribution of the data. A group of clusters can also be seen where countries with high emission levels have been mapped. Threethree-letter country codes following the ISO standard have been used to indicate where the data for each country is generally located. The exception here is that the European Union has been abbreviated as EU.

7.2.2 Temporal observation

Visual analytics allows us to perceive patterns and extract knowledge from visualizations. In the results in Figures X.6 to X.9, a close view is provided that depicts how the states of selected countries have been ordered on the SOM. These figures show that the states of each country in these clusters have also been arranged in a direction relative to the US' cluster. In other words, the closer a country is to the US' cluster on the Geodesic SOM, the higher the emission level will be. For example, the Australian data has been placed in the large dark blue regions due to its relatively low emission levels (less than 1 million metric tons). Hence, trends can be identified that allow users to make predictions on events that may occur in the future. For instance, it is evident that Australia's emission levels are increasing as indicated by the direction its data is heading toward on the Geodesic SOM, that is, toward the group of clusters containing countries with high emission levels (see Figure X.8). This information can be used to predict that Australia's emission levels in 2050 will be higher than they were in 2011. Similar predictions can be made for China, Japan, and the EU since the direction of the data inside the corresponding clusters (Figures X.7, X.6, and X.9 respectively) is heading toward the US' cluster.
7.2.1 Attribute space visualization

Figure xo depicts the resulting visualization when the Geodesic SOM is trained with the data set. The Wagner III cartographic projection technique is used so an entire view of the visualization can be seen. In the centre of the figure, it can be observed that there are a group of clusters. These are countries with extremely high emission levels that are in the thousands of million metric tons. The other countries have emission levels in the hundreds of million metric tons. The eight component planes are shown in figures xosnxoz and can be used to determine which attributes may be related to each other.

Figure xo: This figure depicts the Geodesic SOM after it has been trained with the carbon emissions data set. Large dark blue regions indicate that the Geodesic SOM has created a smooth distribution of the data. A group of clusters can also be seen where countries with high emission levels have been mapped. Three-letter country codes following the ISO trwwnr alphant standard have been used to indicate where the data for each country is generally located. The exception here is that the European Union has been abbreviated as EU.

7.2.2 Temporal observation

Visual analytics allows us to perceive patterns and extract knowledge from visualizations. In the results in Figures timl to timpg a close view is provided that depicts how the states of selected countries have been ordered on the SOM. These figures show that the states of each country in these clusters have also been arranged in a direction relative to the US' cluster. In other words, the closer a country is to the US' cluster on the Geodesic SOM, the higher the emission level will be.

Figure timw: This figure displays a close-up view of Australia's location on the Geodesic SOM. Its data has been placed in the large dark blue regions due to its relatively low emission levels (less than 1 million metric tons). Hence, trends can be identified that allow users to make predictions on events that may occur in the future. For instance, it is evident that Australia's emission levels are increasing as indicated by the direction its data is heading toward on the Geodesic SOM, that is, toward the group of clusters containing countries with high emission levels. This information can be used to predict that Australia's emission levels in the future will be higher than they were in the past. Similar predictions can be made for China, Japan and the EU since the direction of the data inside the corresponding clusters (Figures timm, timh, timo respectively) is heading toward the US' cluster.
Progressive Planning through Trajectory

These results show that the Geodesic SOM has been able to interpolate the values between data samples fairly well as the weight vector values of intermediate neurons generally lie between the weight vector values of $x_{\text{start}}$ and $x_{\text{goal}}$. Hence, the weight vectors of these neurons are meaningful in the sense that they are indicative of the progression between the two states and the intermediate states reached. In other words, if the values of $x_{\text{goal}}$ are higher than the values of $x_{\text{start}}$ as it is in these two cases, as we travel along the trajectory from $x_{\text{start}}$ to $x_{\text{goal}}$, the weight vector values of the intermediate neurons will increase as well. This behaviour can be observed by inspecting the results in the aforementioned two tables. If we look at each row, starting at the first row and go down to the last row, the values generally increase.

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$a_5$</th>
<th>$a_6$</th>
<th>$a_7$</th>
<th>$a_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.098314</td>
<td>284.5111</td>
<td>15.721142</td>
<td>104.93624</td>
<td>38.70803</td>
<td>140.86678</td>
<td>7.3129263</td>
<td>4.0355725</td>
</tr>
<tr>
<td>18.071562</td>
<td>284.27377</td>
<td>15.729849</td>
<td>104.729576</td>
<td>38.616154</td>
<td>140.9281</td>
<td>7.304681</td>
<td>4.0294037</td>
</tr>
<tr>
<td>18.148548</td>
<td>288.69827</td>
<td>15.905892</td>
<td>104.78141</td>
<td>38.880722</td>
<td>145.03613</td>
<td>7.377158</td>
<td>4.0928154</td>
</tr>
<tr>
<td>18.238848</td>
<td>293.63644</td>
<td>16.10159</td>
<td>104.833015</td>
<td>38.97587</td>
<td>149.82751</td>
<td>7.424095</td>
<td>4.1591787</td>
</tr>
<tr>
<td>18.662144</td>
<td>295.98477</td>
<td>16.017136</td>
<td>104.219315</td>
<td>38.997448</td>
<td>152.76799</td>
<td>7.4241176</td>
<td>4.186639</td>
</tr>
</tbody>
</table>
Progressive Planning through Trajectory

![Diagram showing the progression through different countries on a trajectory map.]

<table>
<thead>
<tr>
<th>Year</th>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
<th>a6</th>
<th>a7</th>
<th>a8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>18.09314</td>
<td>284.5111</td>
<td>15.721142</td>
<td>104.93624</td>
<td>38.70803</td>
<td>140.86678</td>
<td>7.3129263</td>
<td>4.0355725</td>
</tr>
<tr>
<td>1997</td>
<td>18.071562</td>
<td>284.27377</td>
<td>15.729849</td>
<td>104.729576</td>
<td>38.616154</td>
<td>140.9281</td>
<td>7.304681</td>
<td>4.0294037</td>
</tr>
<tr>
<td>1998</td>
<td>18.148548</td>
<td>288.69827</td>
<td>15.905892</td>
<td>104.78141</td>
<td>38.880722</td>
<td>145.03613</td>
<td>7.377158</td>
<td>4.0928154</td>
</tr>
<tr>
<td>1999</td>
<td>18.238848</td>
<td>293.63644</td>
<td>16.10159</td>
<td>104.833015</td>
<td>38.97587</td>
<td>149.82751</td>
<td>7.424095</td>
<td>4.1591787</td>
</tr>
</tbody>
</table>

These results show that the Geodesic SOM has been able to interpolate the values between data samples fairly well as the weight vector values of intermediate neurons generally lie between the weight vector values of $x_{\text{start}}$ and $x_{\text{goal}}$. Hence, the weight vectors of these neurons are meaningful in the sense that they are indicative of the progression between the two states and the intermediate states reached. In other words, if the values of $x_{\text{goal}}$ are higher than the values of $x_{\text{start}}$ as it is in these two cases, as we travel along the trajectory from $x_{\text{start}}$ to $x_{\text{goal}}$, the weight vector values of the intermediate neurons will increase as well. This behaviour can be observed by inspecting the results in the aforementioned two tables. If we look at each row, starting at the first row and go down to the last row, the values generally increase.
Progressive Planning through Trajectory

Table: The calculated attribute values for Australia between 1994 and 1995 using the distance transformation algorithm. The first row contains the weight vector values for the BMU corresponding to Australia's data for the year 1994. The last row contains the weight vector values for the BMU corresponding to Australia's data for the year 1995.

<table>
<thead>
<tr>
<th>a1</th>
<th>a2</th>
<th>a3</th>
<th>a4</th>
<th>a5</th>
<th>a6</th>
<th>a7</th>
<th>a8</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.093814</td>
<td>284.5111</td>
<td>15.721142</td>
<td>104.93624</td>
<td>38.70803</td>
<td>140.86678</td>
<td>7.3129263</td>
<td>4.0355725</td>
</tr>
<tr>
<td>18.071562</td>
<td>284.27377</td>
<td>15.729849</td>
<td>104.729576</td>
<td>38.616154</td>
<td>140.9281</td>
<td>7.304681</td>
<td>4.0294037</td>
</tr>
<tr>
<td>18.148548</td>
<td>288.69827</td>
<td>15.905892</td>
<td>104.78141</td>
<td>38.880722</td>
<td>145.03613</td>
<td>7.377158</td>
<td>4.0928154</td>
</tr>
<tr>
<td>18.238848</td>
<td>293.63644</td>
<td>16.10159</td>
<td>104.833015</td>
<td>38.97587</td>
<td>149.82751</td>
<td>7.424095</td>
<td>4.1591787</td>
</tr>
<tr>
<td>18.662144</td>
<td>295.98477</td>
<td>16.017136</td>
<td>104.219315</td>
<td>38.997448</td>
<td>152.76799</td>
<td>7.4241176</td>
<td>4.186639</td>
</tr>
</tbody>
</table>

These results show that the Geodesic SOM has been able to interpolate the values between data samples fairly well as the weight vector values of intermediate neurons generally lie between the weight vector values of \( x_{\text{start}} \) and \( x_{\text{goal}} \). Hence, the weight vectors of these neurons are meaningful in the sense that they are indicative of the progression between the two states and the intermediate states reached. In other words, if the values of \( x_{\text{goal}} \) are higher than the values of \( x_{\text{start}} \) as it is in these two cases, as we travel along the trajectory from \( x_{\text{start}} \) to \( x_{\text{goal}} \), the weight vector values of the intermediate neurons will increase as well. This behavior can be observed by inspecting the results in the aforementioned two tables. If we look at each row starting at the first row and go down to the last row, the values generally increase.
Entities Involved in SOM

- A group of decision makers (neurons)
- External Information (input)
- Internal Information
- (feedback from peers)
Elements of SOM

- Amount of Information Exchange
- Self-Organization require a large amount of Information Exchanges
Peer Feedback Ratio \((w_{ij})\)

\(w_{ij}\) : Information Permeability from \(u_i\) to \(u_j\)

\(u_i\) : Information Processing Unit

\(w_{ji}\) : Information flow

\(w_{kj}\) : non In-situ information flow

\(w_{kj} > 0\)

\(S_k\) : \(k\)-th Stimulus (external Input) to \(u_j\)

\(S_k (= u_k)\) : k-th Stimulus (external Input) to \(u_j\)

\(w_{kj} = 0\)

\(S_k\) : \(k\)-th Stimulus (external Input) to \(u_j\)

Copyright © ViSLAB, The University of Sydney
Neural Field

\[ S_{x_i}(t) \]: Input
\[ x_i \]: Location of i-th Neuron
\[ w_{ij} \]: Peer Feedback Ratio from i to j
Activation of Neuron

- When the net information input $n(x,t)$ exceeds a certain threshold, the neuron at $x$ will produce its output.
Active Neural Region

General Case

Active Neural Region:

$\{ \text{Active Knowledge Field/Region} \}
\begin{array}{c}
\times \\
\times \\
\times \\
\times
\end{array}
\begin{array}{c}
b_1 \\
b_2
\end{array}

Net input: $n(t)$

$r(t) = x_{b2}(t) - x_{b1}(t)$

Discrete Case

Active Neural Region:

$\{ \text{Active Knowledge Field/Region} \}
\begin{array}{c}
\times \\
\times \\
\times \\
\times
\end{array}
\begin{array}{c}
b_1 \\
b_2
\end{array}$
Peer Feedback

- A neuron receives pieces of information from other neurons, environment, etc.
- Weighted \((w_{ij})\) pieces of information will become a part of neuron’s input
Mechanics of Neural Field

\[ \tau \frac{\partial n(x_i, t)}{\partial t} = -n(x_i, t) + \int_j w_{ji} f(n(x_j, t)) \, dx_j + s(x_i, t) - h \]
Mechanics of Neural Field

$$\tau \frac{\partial n(x_i, t)}{\partial t} = -n(x_i, t) + \int \omega_{ji} f(n(x_j, t)) dx_j + s(x_i, t) - h$$

$$r(t) = x_{b2}(t) - x_{b1}(t)$$

$$\frac{dr(t)}{dt} = \frac{dx_{b2}(t)}{dt} - \frac{dx_{b1}(t)}{dt}$$

• Analyze dynamics of Neural Field by analyzing how active Neural Region changes
Distribution of Peer Feedback Ratio

\[ \frac{dr(t)}{dt} = \frac{1}{\tau z} [W(r) + s - h] \]
Analyze the changes of Active Neural Field

\[ n(x, t) \]

\[ \int_{0}^{r_0} w(x) \, dx \quad \int_{r_0}^{r_0 + \delta r} w(x) \, dx \]

\[ W(r_0 + \delta r) + s - h = W(r_0) + w(r_0) \delta r + s - h \]

\[ \frac{d(r_0 + \delta r)}{dt} = \frac{dr_0}{dt} + \frac{d(\delta r)}{dt} \]

\[ = \frac{1}{\tau z} W(r_0) + s - h + w(r_0) \delta r \]
Change of Active Region

\[
\frac{d(\delta r)}{dt} = \frac{1}{\tau z} w(r_0) \delta r
\]

\[
\delta r = A' \exp \left( \frac{1}{\tau z} w(r_0) t \right)
\]

- Depending on the distribution profile of Peer Feedback Ratio, the change of the active neural region behaves differently.

Copyright © ViSLAB, The University of Sydney
Summary

• Information Visualization
• Spatialization
• Self-Organizing Map (n-D -> 2D)
• Winner-Takes-All algorithm