

Weka Knowledge Flowによる クラスタの付与と可視化

阿部 秀尚

処理部品の配置(1)

The screenshot displays the Weka KnowledgeFlow Environment interface. At the top, there is a menu bar with categories: DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. Below this, a 'Visualization' toolbar contains icons for Data Visualizer, Scatter Plot Matrix, Attribute Summarizer, Model Performance Chart, Text Viewer, Graph Viewer, and Strip Chart. The main workspace, titled 'Knowledge Flow Layout', shows a workflow with four components: CSVLoader, AddCluster, CSVSaver, and Scatter Plot Matrix. Red arrows labeled 'dataSet' indicate the flow from CSVLoader to AddCluster, and from AddCluster to both CSVSaver and Scatter Plot Matrix. The CSVLoader component is highlighted with a red rectangular box. A white text box with a black border is overlaid on the bottom left of the workspace, containing the Japanese text: 'DataSourcesから CSVLoaderを選択し, 配置する'. At the bottom of the window, a status bar reads 'Welcome to the Weka Knowledge Flow' and includes a 'Log' button.

Weka KnowledgeFlow Environment

DataSources | DataSinks | Filters | Classifiers | Clusterers | Evaluation | Visualization

Visualization

Data Visualizer | Scatter Plot Matrix | Attribute Summarizer | Model Performance Chart | Text Viewer | Graph Viewer | Strip Chart

Knowledge Flow Layout

CSVLoader → dataSet → AddCluster → dataSet → CSVSaver

AddCluster → dataSet → Scatter Plot Matrix

DataSourcesから CSVLoaderを選択し, 配置する

Status: Welcome to the Weka Knowledge Flow

Log

処理部品の配置(2)

The screenshot shows the Weka KnowledgeFlow Environment interface. At the top, there are tabs for DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. The Visualization tab is active, showing a palette of visualization components: Data Visualizer, Scatter Plot Matrix, Attribute Summarizer, Model Performance Chart, Text Viewer, Graph Viewer, and Strip Chart. The Scatter Plot Matrix component is highlighted with a dashed border.

In the Knowledge Flow Layout area, a workflow is shown with the following components and connections:

- CSLoader** (DataSink) connects to **AddCluster** (Filter) via a red arrow labeled "dataSet".
- AddCluster** (Filter) connects to **CSV Saver** (DataSink) via a red arrow labeled "dataSet".
- AddCluster** (Filter) connects to **Scatter Plot Matrix** (Visualization) via a red arrow labeled "dataSet".

The **AddCluster** component is highlighted with a red rectangular box. A white text box with a black border is overlaid on the workflow, containing the following text:

Filtersから
AddClusterを選択し、
配置する

At the bottom of the window, the Status bar displays "Welcome to the Weka Knowledge Flow" and a "Log" button.

処理部品の配置(3)

The screenshot shows the Weka KnowledgeFlow Environment interface. At the top, there are tabs for DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. The Visualization tab is active, displaying a palette of visualization components: Data Visualizer, Scatter PlotMatrix, Attribute Summarizer, Model PerformanceChart, Text Viewer, Graph Viewer, and Strip Chart. The Scatter PlotMatrix component is highlighted with a red dashed border.

The Knowledge Flow Layout area contains a workflow diagram with the following components and connections:

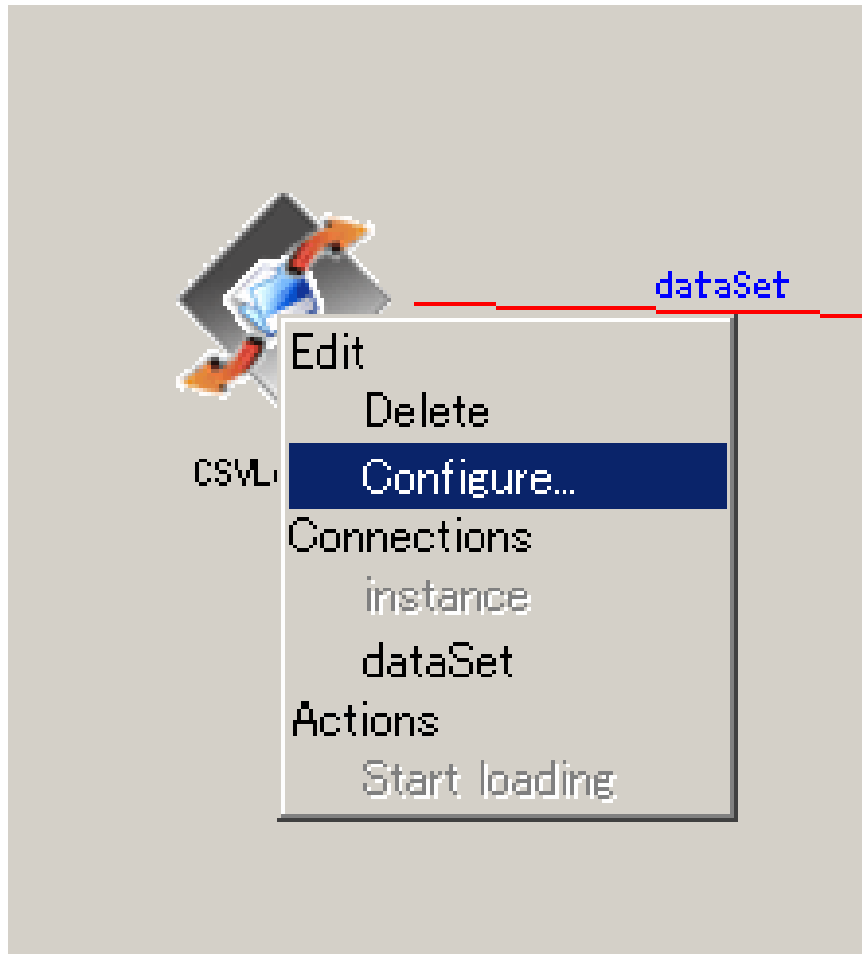
- CSVLoader** (Data Source) connects to **AddCluster** (Clusterer) via a red arrow labeled "dataSet".
- AddCluster** connects to **CSVSaver** (Data Sink) via a red arrow labeled "dataSet".
- AddCluster** also connects to **Scatter PlotMatrix** (Visualization) via a red arrow labeled "dataSet".

The **Scatter PlotMatrix** component is highlighted with a red solid border. A text box with a white background and black border is overlaid on the diagram, containing the following Japanese text:

Visualizationから
ScatterPlotMatrixを
選択し、配置する

At the bottom of the interface, there is a Status bar with the text "Welcome to the Weka Knowledge Flow" and a "Log" button.

処理部品の設定(1)



複数の項目(とりあえずは、在院日数の差と出来高との差でもOK)のCSVファイルを用意して、CSVLoaderを右クリックからConfigure...を選んで、上述のCSVファイルを指定

処理部品の設定(2)

The screenshot displays the Weka KnowledgeFlow Environment interface. At the top, there are tabs for DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. The Visualization tab is active, showing a palette of visualization components: Data Visualizer, Scatter Plot Matrix, Attribute Summarizer, Model Performance Chart, Text Viewer, Graph Viewer, and Strip Chart. The Scatter Plot Matrix component is selected.

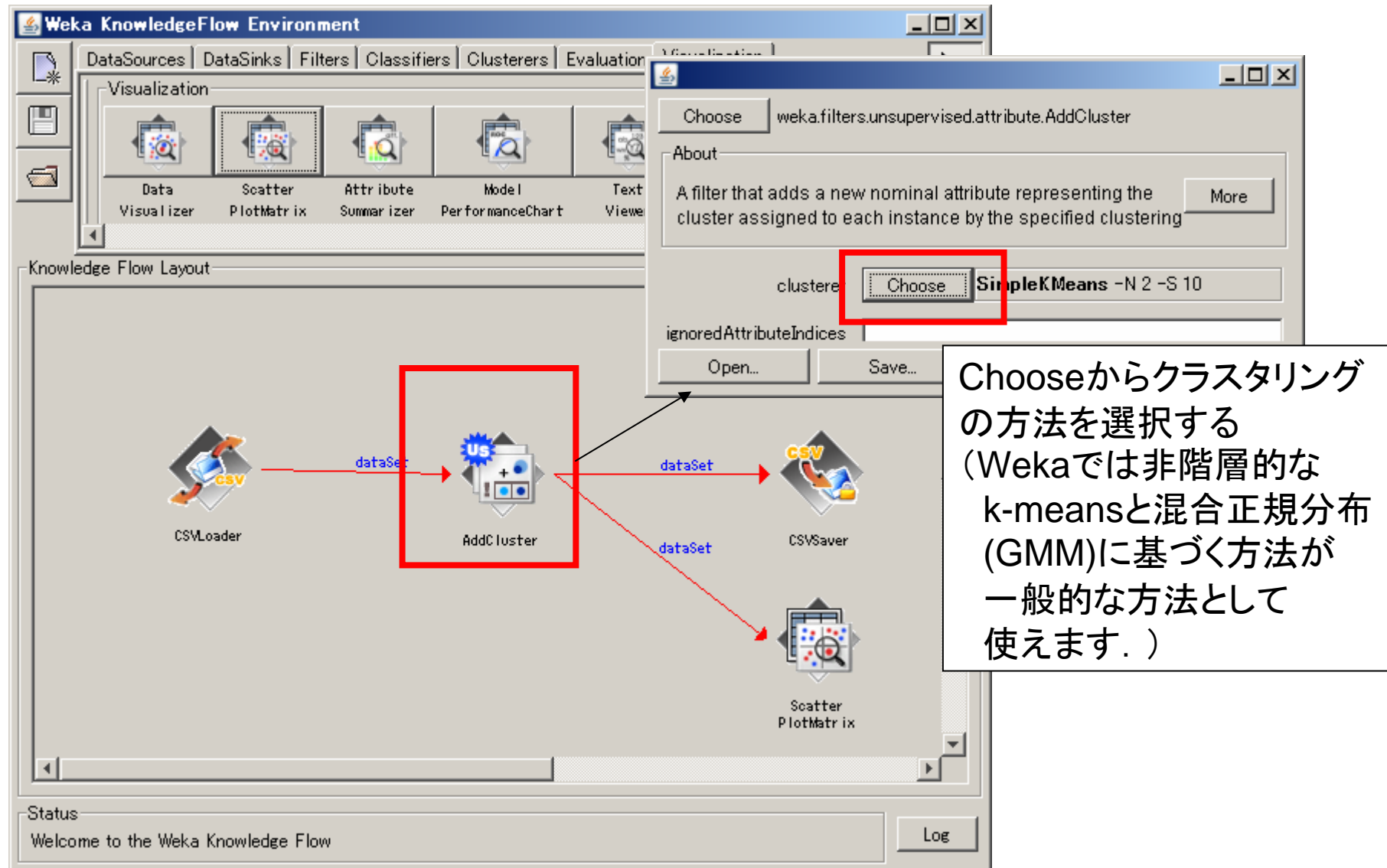
The main area, labeled 'Knowledge Flow Layout', contains a workflow diagram. It starts with a 'CSV Loader' component on the left, which connects to an 'Add Cluster' component. This connection is highlighted with a red box and labeled 'dataSet'. From the 'Add Cluster' component, two arrows labeled 'dataSet' lead to a 'CSV Saver' component and a 'Scatter Plot Matrix' component.

A text box overlaid on the diagram contains the following text:

dataSetを接続
(以下, 接続は同様)

At the bottom of the window, there is a 'Status' bar with the text 'Welcome to the Weka Knowledge Flow' and a 'Log' button.

処理部品の設定(3)



Choose weka.filters.unsupervised.attribute.AddCluster

About

A filter that adds a new nominal attribute representing the cluster assigned to each instance by the specified clustering More

clusters Choose SimpleKMeans -N 2 -S 10

ignoredAttributeIndices

Open... Save...

Chooseからクラスタリングの方法を選択する
(Wekaでは非階層的なk-meansと混合正規分布(GMM)に基づく方法が一般的な方法として使えます.)

CSVLoader → AddCluster → CSVSaver → Scatter PlotMatrix

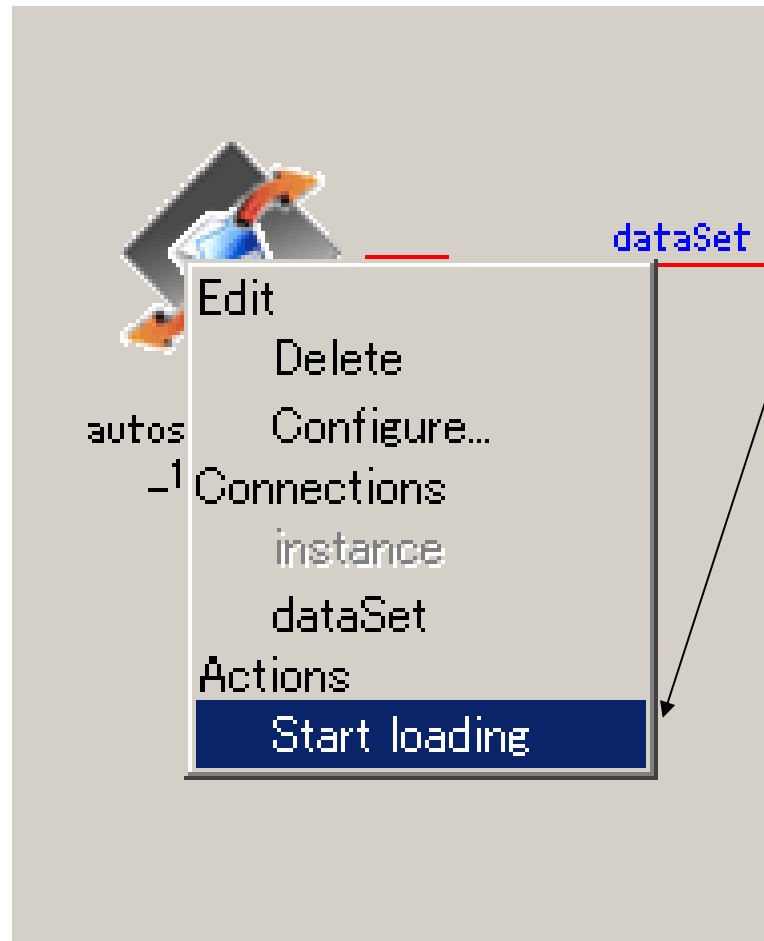
Status: Welcome to the Weka Knowledge Flow Log

処理部品の設定(4)

The screenshot displays the Weka KnowledgeFlow Environment interface. In the foreground, the 'weka.gui.GenericObjectEditor' window is open, showing the configuration for the 'weka.filters.unsupervised.attribute.AddCluster' filter. The 'About' section describes it as a filter that adds a new nominal attribute representing the cluster assigned to each instance by the specified clustering. The 'clusterer' dropdown menu is set to 'SimpleKMeans -N 2 -S 10', which is highlighted with a red box. The 'numClusters' field is set to 3 and the 'seed' field is set to 10. The background shows a workflow with 'CSVLoader' connected to 'AddCluster' via a 'dataSet' link, and 'AddCluster' connected to 'Scatter Plot Matrix'. A text box explains that the clustering method and parameters (like the number of clusters) are set here, with 'SimpleKMeans' as an example.

クラスタリングの方法について、
パラメータ(作成するクラスタ数など)
を設定する。
例はk-meansの画面。

実行



CSVLoaderを右クリックして
StartLoadingを選択

クラスタリング結果の可視化(1)

The image displays the Weka Knowledge Flow interface. On the left, a 'Visualize' window shows a plot matrix with columns for 'length', 'width', 'height', 'curb-weight', 'engine-type', and 'num-of-cylinders'. Below the plot matrix, the 'Colour' dropdown menu is highlighted with a red box and set to 'cluster (Nom)'. A text box explains that the color of the points is determined by the 'Colour' menu item, and if none is selected, the last cluster is used.

On the right, the main Knowledge Flow workspace shows a workflow: 'AddCluster' (US icon) receives a 'dataSet' and outputs a 'dataSet' to 'CSV Saver' (CSV icon). The 'CSV Saver' outputs a 'dataSet' to 'Scatter PlotMatrix' (Scatter Plot icon), which is also highlighted with a red box. The 'Scatter PlotMatrix' icon is connected to the 'Visualize' window.

The status bar at the bottom reads 'Welcome to the Weka Knowledge Flow'.

各点の色はColourのメニューで項目名を選択(何もしないと最後のcluster)

クラスタリング結果の可視化(2)

The image shows the Weka software interface for visualizing clustering results. The main window is titled "Visualize" and contains a "Plot Matrix" with several scatter plots. A red box highlights one of these plots. A callout box with the text "任意の2軸を選んで、右クリック" (Select any 2 axes and right-click) points to the highlighted plot. Below the plot matrix are controls for "PlotSize", "PointSize", "Jitter", and "Colour".

An "Instance info" window is open, displaying the details for "Instance: 67". The data is as follows:

Attribute	Value
normalized-losses	Missing
make	renault
fuel-type	gas
aspiration	std
num-of-doors	four
body-style	wagon
drive-wheels	fwd
engine-location	front
wheel-base	96.1
length	181.5
width	66.5
height	55.2
curb-weight	2579.0
engine-type	ohc
num-of-cylinders	four
engine-size	132.0
fuel-system	mpfi
bore	3.46
stroke	3.9

A second callout box with the text "各点を右クリックするとデータの内容を表示" (Clicking on each point displays the data content) points to a specific data point in a zoomed-in plot. The zoomed-in plot shows a scatter of points with axes labeled 47.8 and 59.8. The status bar at the bottom reads "Welcome to the Weka Knowledge Flow".