

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.

Visualization

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

Knowledge Flow Layout

CSVLoader AddCluster CSVSaver Scatter Plot Matrix

dataSet dataSet dataSet

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 PlotMatrix, Attribute Summarizer, Model PerformanceChart, Text Viewer, Graph Viewer, and Strip Chart. The Scatter PlotMatrix component is selected. Below the palette is the Knowledge Flow Layout area, which contains a workflow diagram. The workflow starts with a CSVLoader component, followed by an AddCluster component (highlighted with a red box), then a CSVSaver component, and finally a Scatter PlotMatrix component. Red arrows labeled 'dataSet' connect the components in sequence. A text box with a white background and black text is overlaid on the diagram, stating: 'Filtersから AddClusterを選択し, 配置する'. At the bottom of the window, there is a Status bar with the text 'Welcome to the Weka Knowledge Flow' and a Log button.

Weka KnowledgeFlow Environment

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

Visualization

Data Visualizer | Scatter PlotMatrix | Attribute Summarizer | Model PerformanceChart | Text Viewer | Graph Viewer | Strip Chart

Knowledge Flow Layout

CSVLoader → dataSet → AddCluster → dataSet → CSVSaver → dataSet → Scatter PlotMatrix

Filtersから
AddClusterを選択し,
配置する

Status
Welcome to the Weka Knowledge Flow

Log

処理部品の配置(3)

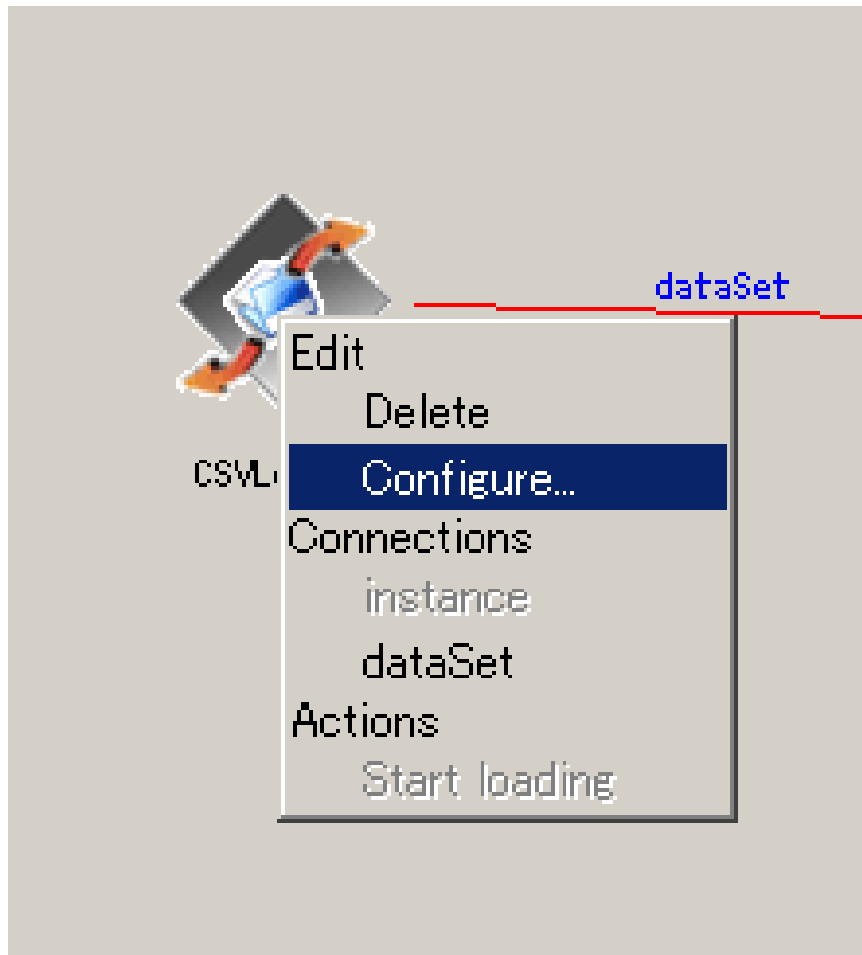
The screenshot shows the Weka KnowledgeFlow Environment interface. At the top, there are tabs for DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. The Visualization tab is active, displaying a palette of visualization components: Data Visualizer, Scatter PlotMatrix, Attribute Summarizer, Model PerformanceChart, Text Viewer, Graph Viewer, and Strip Chart. The Scatter PlotMatrix component is highlighted with a red dashed border. Below the palette is the Knowledge Flow Layout area, which contains a workflow diagram. The workflow starts with a CSMLoader component, followed by an AddCluster component, and then a CSVSaver component. Red arrows labeled 'dataSet' connect these components. A red arrow also points from the AddCluster component to a ScatterPlotMatrix component, which is highlighted with a solid red box. A text box with a white background and black border is overlaid on the diagram, containing the Japanese text: 'Visualizationから ScatterPlotMatrixを選択し、配置する'. At the bottom of the window, there is a Status bar with the text 'Welcome to the Weka Knowledge Flow' and a Log button.

Visualizationから
ScatterPlotMatrixを
選択し、配置する

Status
Welcome to the Weka Knowledge Flow

Log

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)

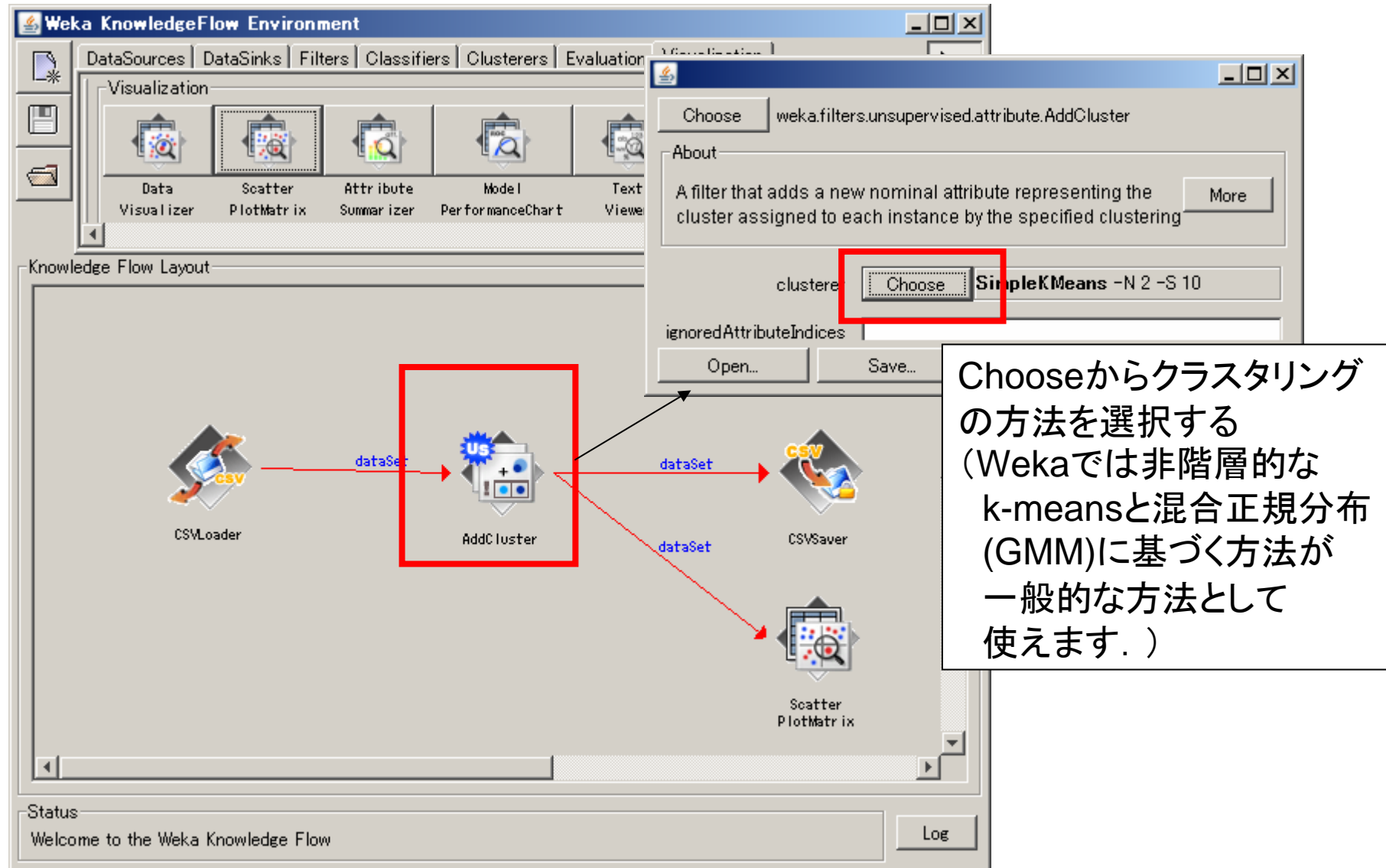
The screenshot displays the Weka KnowledgeFlow Environment interface. At the top, there are tabs for DataSources, DataSinks, Filters, Classifiers, Clusterers, Evaluation, and Visualization. The Visualization tab is active, showing a palette of visualization components: Data Visualizer, Scatter Plot Matrix,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

The main area is the Knowledge Flow Layout, which contains a workflow diagram. The workflow starts with a CSVLoader component on the left, which outputs a 'dataSet' to an AddCluster component. This 'dataSet' connection is highlighted with a red rectangular box. From the AddCluster component, two 'dataSet' connections branch out: one to a CSVSaver component and another to a Scatter Plot Matrix component.

A text box with a white background and black border is overlaid on the diagram, containing the 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

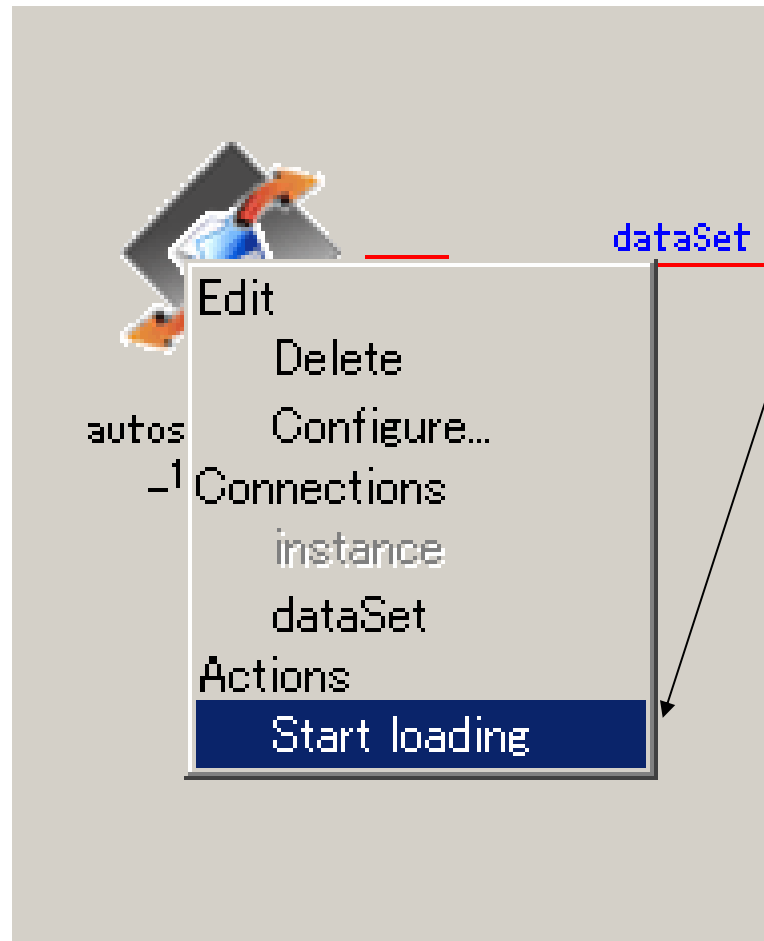
Status: Welcome to the Weka Knowledge Flow Log

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.clusterers.SimpleKMeans' filter. The 'About' section states: 'Cluster data using the k means algorithm'. The 'numClusters' field is set to 3, and the 'seed' field is set to 10. The 'clusterer' dropdown menu is open, and 'SimpleKMeans -N 2 -S 10' is selected and highlighted with a red box. The background shows a workflow diagram with 'CSVLoader' connected to 'AddCluster', which is then connected to 'Scatter PlotMatrix'. A text box with Japanese text explains the configuration process.

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 software interface. On the left, the 'Visualize' window shows a 'Plot Matrix' with columns for 'length', 'width', 'height', 'curb-weight', 'engine-type', and 'num-of-cylinders'. Below the matrix, the 'Colour' dropdown menu is highlighted with a red box and set to 'cluster (Nom)'. A text box explains that this setting colors points by cluster name. On the right, the 'Visualize' tab in the main interface shows a workflow diagram with nodes: 'AddCluster', 'CSVLoader', 'CSV Saver', and 'Scatter PlotMatrix'. The 'Scatter PlotMatrix' node is also highlighted with a red box. Arrows indicate data flow between these nodes.

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 'Visualize' window. At the top, a 'Plot Matrix' displays scatter plots for various attributes: length, width, height, curb-weight, engine-type, and num-of-cylinders. A red box highlights a plot in the 'height' vs 'width' area. Below the matrix are controls for 'PlotSize', 'PointSize', 'Jitter', and 'Colour'. A 'Class Colour' legend shows 'cluster1' (blue), 'cluster2' (red), and 'cluster3' (green).

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 in the matrix.

The 'Weka : Instance info' window is open, showing details for 'Instance: 67'. A yellow arrow points from a point in a zoomed-in plot to this window. A 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' (Right-clicking each point displays the data content) points to the zoomed-in plot.

The 'Instance info' window lists the following attributes for Instance 67:

- 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 : mphi
- bore : 3.46
- stroke : 3.9

The zoomed-in plot shows a scatter plot with points colored by cluster. The axes are labeled with values 47.8, 53.8, and 59.8. A legend at the bottom shows 'cluster1' (blue), 'cluster2' (red), 'cluster3' (green), and 'cluster4' (cyan).

The status bar at the bottom reads 'Welcome to the Weka Knowledge Flow' and includes a 'Log' button.