

Desarrollando modelos de Data Mining con Data Engine

Esquema general:

- (1) Crear proyecto
- (2) Crear e insertar archivos de datos al proyecto.
- (3) Configurar atributos para insertar en el proyecto.
- (4) Crear e insertar modelo al proyecto.
- (5) Crear e insertar análisis de los resultados del modelo.
- (6) Configurar modelo.
- (7) Configurar análisis de modelo.
- (8) Correr modelo.
- (9) Revisar y analizar los resultados.

Para desarrollar un nuevo proyecto se recomienda crear una nueva carpeta de modo de incluir todos los archivos y modelos en esta carpeta.

A continuación se describen los pasos fundamentales a seguir dentro del esquema general mencionado.

(1) Crear proyecto

- Ir a file>new>project
- Poner nombre
- Guardar

(2) Crear e insertar archivos de datos al proyecto.

- Para crear un archivo: file>new>data. Luego guardar.
- Para insertar un archivo guardado: ir a la carpeta input/output en el proyecto, luego insert file (con el botón derecho del mouse) y aceptar.

(3) Configurar atributos para insertar en el proyecto.

- Forma manual: ir a la carpeta features en el proyecto. Luego insert feature (botón derecho del mouse).
- Forma automática: abrir el archivo de datos insertado en el proyecto. Ir a table>add columns to project.

(4) Crear e insertar modelo al proyecto.

- Crear modelo: file>new> “modelo XXX”. Luego poner nombre y guardar.
- Insertar modelo: ir a models en el proyecto. Luego insert file (botón derecho) y aceptar.

(5) Crear e insertar análisis de los resultados del modelo.

- Crear análisis: file>new>analysis. Luego poner nombre y guardar.
- Insertar análisis: ir a análisis en el proyecto. Luego insert file (botón derecho) y aceptar.

(6) **Configurar modelo.**

Para configurar el modelo deben ir a proyecto, luego models y posteriormente seleccionar el modelo a configurar.

La configuración particular depende del tipo de modelo:

Multilayer perceptron:

- Architecture: configurar número de neuronas, función de activación, y atributos a incluir.
- Learning method: (no necesario modificar)
- Learning parameters: (no necesario modificar)
- Weight initialization: (no necesario modificar)
- Stop conditions: modificar Test>perform a test every 1 epochs.Activar “Keep best training state”
- Input /output:
 - i. Training: insertar archivo de entrada y archivo de salida¹.
 - ii. Test : insertar archivo de entrada (puede o no ser igual al archivo de training) y archivo de salida.
- Guardar.

Fuzzy c-means:

- Parameters: insertar atributos a ocupar. Elegir el número de clusters (2 o más) y el exponente de difusidad.
- Training: (no es necesario cambiar nada).
- Input /output:
 - i. Training: insertar archivo de entrada y salida del entrenamiento
 - ii. Labeling: insertar archivo que contiene nombre de las clases
 - iii. Test : insertar archivo de entrada (puede o no ser igual al archivo de training) y archivo de salida.
- Guardar.

(7) **Configurar análisis de modelo.**

Para configurar el análisis deber ir al proyecto creado, luego análisis y posteriormente seleccionar el análisis creado.

- Model: insertar modelo a analizar.
- Matrix of confusion: insertar archivos de resultados (optativo).
- Calculated errors: insertar archivos de resultado.
- Guardar.

¹ Estos archivos debes estar creados y agregados al proyecto. Ir a (2) para revisar cómo hacerlo.

(8) Correr modelo.

- Abrir modelo.
- Ejecutar new training.(botón en barra de herramientas o ir a MLP >new training)
- Luego test, labeling y recall (según modelo que corresponda y si se agregaron archivos a esta etapa en el modelo)

(9) Revisar y analizar los resultados.

- Abrir archivos de resultados. Se puede realizar los siguientes análisis según corresponda:
 - Ver correlaciones.
 - Dibujar series de tiempo.
 - Ver grados de pertenencia.

Suerte!

Francisco Cisternas V.

Problemas y comentarios a: frcister@mi.cl