

Mineração de Dados

03 - Introdução ao Software Weka

Marcos Roberto Ribeiro



Instituto Federal Minas Gerais - Campus Bambuí

2018

- Waikato Environment for Knowledge Analysis (Weka).
- Conjunto de algoritmos de aprendizagem de máquina
- Desenvolvido em Java na Universidade de Waikato, Nova Zelândia
- Software livre
- Página oficial: <http://www.cs.waikato.ac.nz/~ml/weka/>

Ferramentas do Weka

Explorer: Ambiente para explorar bases de dados (pré-processamento e mineração de dados)

Experimenter: Ambiente para realizar experimentos e testes estatísticos

KnowledgeFlow: Tem a mesma funcionalidade do Explorer com uma interface drag-and-drop

SimpleCLI: Interface em linha de comando

Workbench: Basicamente um integração de todas as interfaces

Tools > Package Manager: Possibilidade de instalar uma grande variedade de pacotes

Explorador de Dados (*Explorer*)

- Possui as seguintes abas:
 - Preprocess:** Pré-processamento dos dados
 - Classify:** Tarefas de classificação
 - Cluster:** Agrupamento de dados
 - Select Attributes:** Seleção de atributos mais relevantes
 - Visualize:** Visualização 2D dos dados

Status Box

- Caixa de status na parte inferior
- O botão direito habilita algumas opções como o uso de memória

Log

Mostra uma janela com um log de tudo que foi realizado durante a sessão

Observação

Os gráficos podem salvos com ALT+SHIFT+LEFT CLICK

Carregando Dados

Open file: Carrega os dados de um arquivo

Open URL: Carrega os dados de um endereço URL

Open DB: Lê os dados de um banco de dados

Generate: Gera dados sintéticos

Edit: Permite alterar os dados

Save: Exporta a base de dados

Arquivos ARFF

- O Weka consegue trabalhar com diversos formatos de dados
- Porém, o formato padrão é o Attribute-Relation File Format (ARFF)

Exemplo: iris.arff

```
% Comment
```

```
@RELATION iris
```

```
@ATTRIBUTE sepallength REAL
```

```
@ATTRIBUTE sepalwidth REAL
```

```
@ATTRIBUTE petallength REAL
```

```
@ATTRIBUTE petalwidth REAL
```

```
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

```
@DATA
```

```
5.1,3.5,1.4,0.2,Iris-setosa
```

```
4.9,3.0,1.4,0.2,Iris-setosa
```

```
4.7,3.2,1.3,0.2,Iris-setosa
```

```
4.6,3.1,1.5,0.2,Iris-setosa
```

```
...
```

Filtros e Informações

Filtros

- Permite transformar os dados de diferentes maneiras
- São divididos em:
 - Supervisionados:** levam em consideração as classes
 - Não supervisionados:** não consideram informações sobre as classes
- Os filtros podem ser aplicados nas instâncias ou nos atributos
- Exemplos:
 - Discretize (supervised):** Discretiza valores numéricos em nominais
 - NominalToBinary (supervised) :** Transforma valores nominais em binários

Informações

Current relation: mostra informações da base de dados

Selected attribute: mostra informações sobre o atributo selecionado

- Clicando no nome do filtro, são abertas as configurações do filtro
- No botão “More” há uma explicação mais detalhada do filtro
- Muitas ferramentas citam os artigos científicos nos quais elas foram baseadas

Classificação, Agrupamento e Associação

Classify (Classificação)

Classifier: Seleção e parametrização do classificador a ser utilizado

Test options: Metodologia de teste para validar o classificador

Classifier output: Saída do classificador

Result list: Últimos resultados

Cluster (Agrupamento)

- A aba de agrupamento de dados é semelhante à aba de classificação
- Porém, são tarefas de agrupamentos de dados

Associate (Associação)

- Na aba de associação só é possível escolher o algoritmo (*associator*) a ser usado

Seleção de Atributos e Visualização

- Permite investigar quais atributos são mais preditivos
- A seleção é feita por um avaliador de atributo (*attribute evaluator*)
- O avaliador, por sua vez, utiliza algum método de busca (*search method*) para tentar selecionar o melhor subconjunto de atributos
- Na aba de visualização (*visualize*) permite a exibição de diversos gráficos sobre os dados



(2018).

Weka 3: Data mining software in java.

<http://www.cs.waikato.ac.nz/ml/weka/>.



Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., and Scuse, D. (2017).

Weka manual for version 3-8-2.