

INE5644 – Data Mining – 20181

Aula 17 - Classificação - Conceitos básicos e Árvores de Decisão - Parte 1 - Prática

Prof. Alexandre Savaris

1) Explorar a ferramenta Weka no contexto de árvores de decisão – mais especificamente, árvores construídas com o algoritmo ID3. Para isso:

- utilizar os *datasets* `tic-tac-toe.data` e `connect-4.data.rar` disponibilizados no Moodle;
- adaptar os *datasets* para carga na ferramenta Weka, observando as restrições e exigências da mesma com relação a arquivos CSV;
- para cada *dataset*, construir uma árvore de decisão utilizando o algoritmo ID3 (dica: é possível que a instalação padrão da ferramenta não disponibilize o algoritmo; se esse for o caso, pesquisar sobre o pacote *simpleEducationalLearningSchemes*);
- após o treinamento, avaliar a árvore de decisão gerada a partir do *dataset*. Quanto tempo foi necessário para construí-la? Quais suas características? (Altura? Nós por nível?)
- Visualizar graficamente a árvore gerada. (Dica: é possível que a ferramenta Weka não possua uma interface para visualização gráfica da árvore gerada utilizando o algoritmo ID3; se esse for o caso, pesquisar sobre o pacote *prefuseTree* e repetir o treinamento utilizando o algoritmo J48 – configurado para funcionar sem poda (*pruning*)).

Observação: esta questão não precisa ser entregue.

2) Utilizando a ferramenta Weka, construa uma árvore de decisão com o algoritmo ID3 para o *dataset* a seguir (cogumelos venenosos). O resultado obtido após o treinamento/testes deverá ser enviado como resposta da questão ao Moodle.

Cor	Altura	Faixas	Textura	Venenoso?
Púrpura	Alta	Sim	Áspera	Sim
Púrpura	Alta	Sim	Suave	Sim
Vermelha	Baixa	Sim	Intermediária	Não
Azul	Baixa	Não	Suave	Não
Azul	Baixa	Sim	Intermediária	Sim
Vermelha	Alta	Não	Intermediária	Não
Azul	Alta	Sim	Suave	Sim
Azul	Baixa	Sim	Suave	Sim

Azul	Alta	Não	Intermediária	Não
Azul	Baixa	Sim	Áspera	Sim
Vermelha	Baixa	Não	Suave	Não
Púrpura	Baixa	Não	Intermediária	Sim
Vermelha	Alta	Sim	Intermediária	Não
Púrpura	Alta	Sim	Intermediária	Sim
Púrpura	Alta	Não	Áspera	Não
Púrpura	Alta	Não	Suave	Não

3) Utilizando a ferramenta Weka, construa uma árvore de decisão com o algoritmo ID3 para o *dataset* a seguir (abordar motoristas). O resultado obtido após o treinamento/testes deverá ser enviado como resposta da questão ao Moodle.

Usando telefone?	Luzes	Velocidade	Direção	Abordar?
Sim	Funcionando	Lenta	Arriscada	Não
Não	Queimadas	Rápida	Perigosa	Sim
Não	Queimadas	Lenta	Segura	Sim
Não	Funcionando	Rápida	Arriscada	Não
Sim	Funcionando	Normal	Perigosa	Não
Sim	Funcionando	Rápida	Perigosa	Sim
Não	Queimadas	Lenta	Segura	Sim
Sim	Queimadas	Rápida	Arriscada	Sim
Não	Funcionando	Normal	Segura	Não
Sim	Queimadas	Normal	Arriscada	Sim
Sim	Queimadas	Lenta	Arriscada	Sim
Não	Funcionando	Lenta	Segura	Não
Não	Queimadas	Lenta	Arriscada	Sim
Sim	Funcionando	Normal	Segura	Não
Não	Funcionando	Normal	Arriscada	Não
Não	Funcionando	Lenta	Perigosa	Sim
Não	Funcionando	Rápida	Segura	Não

4) Avaliar se a ordem dos dados no conjunto de treinamento influencia na construção de uma árvore de decisão. Para isso, carregar o *dataset* utilizado como exemplo na aula 16 na ferramenta Weka, gerando uma árvore de decisão a partir do seu conteúdo. Repetir o experimento para duas ordenações diferentes do conjunto. Responder: a alteração na ordem do conjunto implicou em alterações estruturais nas árvores de decisão? Em caso positivo, quais

foram as alterações identificadas? Percebe-se impacto nos resultados obtidos pelo uso da árvore de decisão durante os testes?