

Quarto Trabalho de Inteligência Artificial / Sistemas Inteligentes

Entrega: 30/05/2016

5 de Maio de 2016

Árvores de Decisão

Usando como base o material disponível nas seções 18.1 a 18.3 do livro texto, Artificial Intelligence: a Modern Approach, de Peter Norvig e Stuart Russell (3ed), implemente um algoritmo para indução de árvores de decisão (o ID3, mostrado no livro, na Figura 18.5), utilizando como função para seleção do nó da árvore, o ganho de informação (definido na página 704).

A entrada para o seu programa é um conjunto de exemplos (observações) com atributos e valores para os atributos, incluindo uma variável de classe. A saída do seu programa deve estar no seguinte formato:

```
<attribute>
  value1:
    <attribute>
      value1: class1 (counter1)
      value2: class2 (counter2)
    value2: class3 (counter3)
  value3:
    <attribute>
      value1:
        <attribute>
          value1: class4 (counter4)
          value2: class2 (counter5)
        value2: class3 (counter6)
```

Onde, `attribute` indica o atributo escolhido para estar naquela posição da árvore, `value#` é um valor de um atributo, `class#` é a classe atribuída àquele ramo da árvore e `counter#` é um contador dos vários elementos de cada classe que seguem o padrão correspondente àquele ramo da árvore.

Exemplos de dados encontram-se nos ficheiros <http://www.dcc.fc.up.pt/~ines/aulas/1516/IA/trab4/restaurant.csv>, <http://www.dcc.fc.up.pt/~ines/aulas/1516/IA/trab4/weather.csv> e <http://www.dcc.fc.up.pt/~ines/aulas/1516/IA/trab4/iris.csv>, que são descritos a seguir.

- **restaurant:** (exemplo do livro, página 700) contém informações sobre restaurantes (tipo de comida, tempo de espera, preço etc), e o atributo de classe diz se o cliente esperou ou não para comer no restaurante. A tarefa é gerar uma árvore de decisão que possa dar apoio à decisões futuras dizendo se um cliente vai esperar ou não para comer naquele restaurante.
- **weather:** contém informações sobre condições climáticas e uma variável de classe que diz se uma pessoa vai jogar ténis ou não, dadas as condições climáticas. A tarefa é gerar uma

árvore de decisão que, dadas novas condições climáticas, decida se uma pessoa vai jogar ou não.

- **iris**: contém informações numéricas sobre plantas de três classes: iris setosa, iris virginica e iris versicolor. Os atributos são largura e comprimento da pétala e largura e comprimento da sépala. A tarefa é construir uma árvore de decisão que, dada uma nova planta, consiga categorizá-la em umas das três classes.

Além de gerar a árvore no formato indicado anteriormente, o seu programa deve ser capaz de aceitar como entrada um outro ficheiro com novas instâncias, não utilizadas durante a geração da árvore, e “prever” a classe de cada uma destas instâncias com base no modelo de árvore que gerou.

O que entregar?

1. trabalho escrito com a descrição da implementação e resultados.

Organização do trabalho escrito:

Introdução

- o que é uma árvore de decisão?
- para que serve?

Algoritmos para indução de árvores de decisão

- pesquisar alguns algoritmos
- explicar as diferenças
- discutir sobre métricas para a escolha dos atributos na construção da árvore
- ID3

Descrição da Implementação

- linguagem
- estruturas de dados utilizadas e justificativas para a sua utilização
- organização do código

Resultados

- reportar a árvore gerada para cada conjunto de dados
- mostrar erros e acertos em cada ramo da árvore

Comentários Finais e Conclusões

2. o código fonte dos programas e como compilar, isto é, um pequeno manual de como correr os programas (pode ser um 'help' ou 'readme'). Além disso, em que ambiente foi compilado (tipos e versões do SO e da linguagem). Seu programa deve correr na minha máquina (com fedora instalado). Não assuma que eu tenho uma IDE (Integrated Development Environment) de qualquer tipo. O programa deve correr na linha de comando.

A submissão será através do Moodle UP. Por favor, não inclua caracteres acentuados no nome do ficheiro de submissão.

O trabalho pode ser feito em grupo de no máximo duas pessoas. Todos os trabalhos serão apresentados em data a combinar. **Os dois componentes do grupo deverão estar presentes durante a apresentação.** Quem não estiver presente vai ter nota zero!