22 Abril 2019 – João Paulo Santos Monteiro | Nuno José Moreira Bernardes

árvores de decisão

# Introdução

Uma árvore de decisão representa uma função que recebe como *input* uma série de atributos e variáveis e retorna uma decisão. Esta chega à sua resposta executando uma série de testes. A árvore de decisão é um dos modelos mais utilizados na aprendizagem indutiva, devido ao facto de ser simples de implementar. Todos nós em algum momento simulámos uma árvore destas para tomar alguma decisão como por exemplo, se esperaríamos por uma mesa num restaurante ou se procuraríamos outro restaurante.

**Algoritmo para árvores de decisão**

Os principais algoritmos para a indução de árvores de decisão são : O id3 (Quinlan, 1986) e o C4.5 (Quinlan, 1986). O id3 foi o algoritmo pioneiro na indução de árvores de decisão enquanto que o C4.5 foi um melhoramento do id3 que resolveu alguns dos seus problemas.

**ID3:**

O id3 é um algoritmo recursivo e baseado em busca gulosa, procurando, sobre um conjunto de atributos, aquele que tem mais importância, gerando sub-árvores. O principal problema do id3 é a incapacidade de tratar variáveis de classe binária, não sendo possível apresentar a ele conjuntos de dados com atributos contínuos, sendo assim necessário realizar previamente uma descaracterização ou *binary splitting.* Outro problema tem a ver com o possível número de arestas, sendo este um algoritmo que não tem nenhum sistema de poda, pode levar a criação de árvores demasiado complexas.

Por outro lado o C4.5 já lida com variáveis de classe binária e com atributos contínuos e , além disso, aplica uma forma de escolher os atributos diferente do id3, que permite que este algoritmo crie árvores mais simples e menos complexas em relação ao id3.

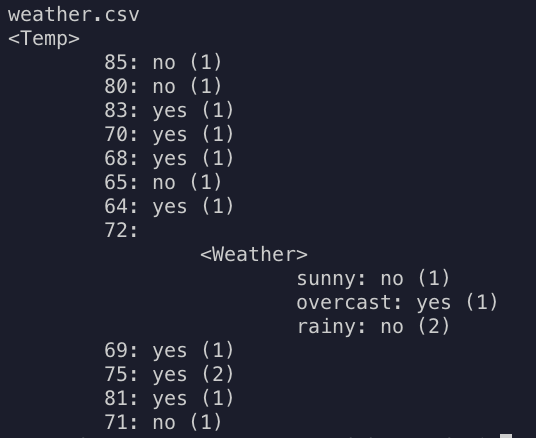
# Implementação

Para este trabalho utilizamos a linguagem Java, tivemos de criar uma classe “No”, outra “Atributos” e por ultimo uma “Exemplos” que nos permite guardar as informações necessária para a manipulação das informações recebidas nos ficheiros CSV. A classe “No” é responsável por manipular os nós da árvore, de forma a que, em cada nó, guarde informações sobre ele e sobre os filhos respetivos. No caso da classe “Atributos” esta guarda informações relativas os atributos dum ficheiro CSV, informações tais como os tipos que cada variável pode ter e o nome da variável. No caso da classe “Exemplos” esta guarda todas a informações referentes aos exemplos apresentados. Todos estes ficheiro também incluem funções que ajudam na manipulação dos dados do ficheiro e dos dados de cada variável da classe “Atributo”. Além das classes criadas utilizámos outras estruturas de dados que são fornecidas por varias bibliotecas disponíveis no Java, como por exemplo linked lists.

O nosso código está organizado em 4 ficheiros .java separados, um com a função main e os restantes com a implementação das classes criadas, bem como as funções necessárias para a sua manipulação. No ficheiro com a função main temos também outras funções necessárias para a aplicação do algoritmo proposto, assim como a função “importance” que retorna a importância de cada atributo, temos também uma função “valorMaisComum” que retorna a variável com maior recorrência nos exemplos.

# Resultados

# ÁRVORE DE DECISÃO PARA O FICHEIRO “RESTAURANT.CSV”

**ÁRVORE DE DECISÃO PARA O FICHEIRO “WEATHER.CSV”**

Comentários FINAIS E CONCLUSÃO

o NOSSO NÃO FAZ A Discretização ou *binary splitting*, SENDO ESTE O SEU PRINCIPAL PROBLEMA. Com este problema foi nos impossível mostrar a árvore gerada pelo ficheiro “iris.csv”. Além disto o nosso código não esta muito limpo, podíamos ter evitado algumas repetições.

Depois de analisarmos os resultados percebemos que o algoritmo é eficiente e que nos poderia ajudar no problema do trabalho dois, o Connect-Four.

# Bibliografia

O nosso relatório foi baseado no livro recomendado pela professora no início da unidade curricular:

- *Artificial Intelligence: A Modern Approach* (3ª Edição) *Stuart Russel & Peter Norvig;*

Todas as conclusões, estruturas de dados e algoritmos usados foram confirmados neste livro, pois têm justificações muito completas que tornam fácil o entendimento deste trabalho.