



Trabalho Prático
Tema 2 - CBR e Inferência Difusa

Engenharia Informática

Unidade Curricular: Conhecimento e Raciocínio
Professor: Dr. Viriato M. Marques

Ano Letivo 2020/2021

Leonardo Rafael Amado Brito
2016017310

Renato Miguel Pais Freire
2018019771

Índice

Introdução.....	2
Biblioteca de Casos	3
Retrieve.....	4
Reuse.....	5
Revise.....	6
Retain.....	6
Inferência Difusa.....	7
Implementação.....	8
Conclusão.....	9

Introdução

O objetivo deste trabalho consiste na implementação de um modelo CBR para avaliação de automóveis usados, em função das suas 26 características. A seguinte tabela mostra as diversas variáveis e os domínios de cada uma delas.

Atributos	Domínio
Symboling	-3, -2, -1, 0, 1, 2, 3
Normalized-Losses	Continuous from 65 to 256
Make	alfa-romero, audi, bmw, chevrolet, dodge, honda, isuzu, jaguar, mazda, mercedes-benz, mercury, mitsubishi, nissan, peugot, plymouth, porsche, renault, saab, subaru, toyota, volkswagen, volvo
Fuel-type	Diesel, gas
Aspiration	Std, turbo
Num-of-doors	Four, two
Body-Style	Hardtop, wagon, sedan, hatchback, convertible
Drive-Wheels	4wd, fwd, rwd
Engine-Location	Front, rear
Wheel-Base	Continuous from 86.6 to 120.9
Length	Continuous from 141.1 to 208.1
Width	Continuous from 60.3 to 72.3
Height	Continuous from 47.8 to 59.8
Curb-Weight	Continuous from 1488 to 4066
Engine-Type	Dohc, dohc, l, ohc, ohcf, ohcv, rotor
Num-of-Cylinders	Eight, five, four, six, three, twelve, two
Engine-Size	Continuous from 61 to 326
Fuel-System	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi
Bore	Continuous from 2.54 to 3.94
Stroke	Continuous from 2.07 to 4.17
Compression-Ratio	Continuous from 7 to 23
Horsepower	Continuous from 48 to 288
Peak-RPM	Continuous from 4150 to 6600
City-MPG	Continuous from 13 to 49
Highway-MPG	Continuous from 16 to 54
Price	Continuous from 5118 to 45400

Para a utilização das variáveis de entrada ao longo do código Matlab e na biblioteca de casos decidimos dar os seguintes nomes às variáveis: symboling, normalizedLosses, make, fuelType, aspiration, numofDoors, bodyStyle, driveWheels, engineLocation, wheelBase, length, width, height, curbWeight, engineType, numofCylinders, engineSize, fuelSystem, bore, stroke, compressionRatio, horsepower, peakRPM, cityMPG, highwayMPG e price, correspondendo aos atributos da tabela, pela ordem que nela aparecem.

Biblioteca de Casos

De forma a criar uma biblioteca de casos, decidimos ir ler o ficheiro excel que nos foi disponibilizado pelo professor e reter toda a informação contida no mesmo formando assim a biblioteca de casos.

Para além dos 26 atributos que constituem um caso da biblioteca de casos, decidimos ainda adicionar uma outra coluna, o ID, de forma a permitir identificar assim um caso presente da biblioteca de casos.

Assim, cada caso é constituído por:

- symboling: Fator de risco de um veículo;
- normalizedLosses: É o pagamento de perda média por veículo no seguro;
- make: Nome do fabricante do veículo;
- fuelType: Indica o tipo de combustível de um veículo;
- aspiration: É um motor de combustão interna em que a entrada de ar depende exclusivamente da pressão atmosférica e não tem indução forçada por meio de um turbocompressor ou compressor;
- numOfDoors: Número de portas de um veículo;
- bodyStyle: Tipo de chassi de um veículo;
- driveWheels: Indica a tração de um veículo;
- engineLocation: Indica onde se encontra o motor de um veículo;
- wheelBase: Distância entre eixos;
- length: Indica o comprimento de um veículo;
- width: Indica a largura de um veículo;
- height: Indica a altura de um veículo;
- curbWeight: Indica o peso total de um veículo;
- engineType: Indica o tipo de motor de um veículo;
- numofCylinder: Indica o número de cilindros do motor de um dos veículos;
- engineSize: Indica o tamanho do veículo;
- fuelSystem: O sistema de combustível de um veículo compreende componentes que fornecem combustível do tanque para o motor;
- bore: Indica o tamanho do diâmetro de cada cilindro de um veículo;
- stroke: Distância percorrida pelo pistão durante cada ciclo;
- compressionRatio: Razão entre o volume do cilindro e seu espaço vazio;
- horsepower: Indica quantos cavalos tem um veículo;
- peakRPM: Indica o número máximo de rotações por minuto de um veículo;
- cityMPG: Indica o consumo que cada veículo faz em cidades;
- highwayMPG: Indica o consumo que cada veículo faz em autoestradas;
- price: Indica o preço de cada veículo;
- ID: Indica o identificador único de cada veículo;

Retrieve

A leitura da biblioteca de casos é feita no início do ficheiro CBR para facilitar o processo de enviar essa mesma biblioteca - já passada para uma lista no MATLAB - para os outros ficheiros ou funções onde seja necessária.

A recolha da descrição de um caso novo em interface texto é feita também no ficheiro CBR e é enviada posteriormente, junto com a biblioteca, para o ficheiro que trata do Retrieve.

Para a definição de medidas adequadas à determinação da semelhança global entre o novo caso e os casos passados, atendendo a que existem atributos numéricos, booleanos e nominais, fazemos um array bidimensional para cada categoria da lista de casos com as similaridades consoante os atributos e usamos funções que calculam a distância euclidiana e a distância local.

A possibilidade de filtragem opcional da biblioteca por marca do automóvel foi implementada no ficheiro CBR, tal como a apresentação dos N casos mais semelhantes ordenados por ordem decrescente de semelhança.

O desenvolvimento de uma pequena aplicação de administração foi o único ponto que não foi implementado.

No retrieve distribuímos os seguintes pesos por cada um dos atributos, os pesos são:

- symboling: 1
- normalizedLosses: 3
- make: 5
- fuelType: 5
- aspiration: 3
- numOfDoors: 4
- bodyStyle: 2
- driveWheels: 4
- engineLocation: 1
- wheelBase: 1
- length: 2
- width: 2
- height: 2
- curbWeight: 2
- engineType: 4
- numbOfCylinder: 3
- engineSize: 2
- fuelSystem: 3
- bore: 2
- stroke: 2
- compressionRatio: 2
- horsepower: 4
- peakRPM: 3
- cityMPG: 3
- highwayMPG: 3

Reuse

Nesta fase vamos calcular o preço do novo caso no qual o utilizador tem interesse. O preço desse novo caso vai depender de 8 atributos que são:

- symboling (x1);
- normalizedLosses (x2);
- bore (x19);
- compressionRatio (x21);
- horsepower (x22);
- peakRPM (x23);
- cityMPG (x24);
- highwayMPG (x25);

De seguida, de forma a calcular o preço para esse caso através de um modelo de regressão linear tradicional, vamos buscar, para uma matriz X, as variáveis independentes, para de seguida calcular os coeficientes bs, a partir da divisão entre a matriz X e os preços de todos os casos retirados. Por fim, para chegar ao valor do preço, fazemos a soma da multiplicação dos valores de b pelos respetivos X.

```
x1 = retrieved_cases(:,1); % Symboling
x2 = retrieved_cases(:,2); % Normalized Losses
x19 = retrieved_cases(:,19); %Bore
x21 = retrieved_cases(:,21); %Compression-Ratio
x22 = retrieved_cases(:,22); % hp
x23 = retrieved_cases(:,23); % peak-rpm
x24 = retrieved_cases(:,24); % City-MPG
x25 = retrieved_cases(:,25); % Highway-MPG

y = retrieved_cases(:,26); % Prices

X = [ones(size(x1)) x1 x2 x19 x21 x22 x23 x24 x25]; %[1 vars independentes]

b = X\y; %calcular coeficientes bs

new_price = b(1) + b(2) * new_case.symboling + b(3) * new_case.normalizedLosses ...
    + b(4) * new_case.bore + b(5) * new_case.compressionRatio + b(6) * new_case.horsePower ...
    + b(7) * new_case.peakRPM + b(8) * new_case.cityMPG + b(9) * new_case.highwayMPG;
```

Revise

A fase do revise tem como objetivo, após selecionarmos o caso da biblioteca que seja mais semelhante ao do utilizador, poder atualizar qualquer um dos seus atributos, inclusive o valor do preço para o novo valor calculado na fase do reuse.

Retain

Por fim, na última fase, é nos dada a possibilidade de adicionar esse caso à biblioteca de casos.

Inferência Difusa

No ponto 2, era-nos pedido para reimplementar a fase do reuse, realizando a adaptação do preço.

Começámos por definir termos linguísticos e funções de pertença.

```
new_price = mamfis;

new_price = addInput(new_price, [-3 3], 'Name', 'symboling');
new_price = addInput(new_price, [65 256], 'Name', 'normalizedLosses');
new_price = addInput(new_price, [60.3 72.3], 'Name', 'width');
new_price = addInput(new_price, [48 288], 'Name', 'horsePower');
new_price = addInput(new_price, [13 49], 'Name', 'cityMPG');
new_price = addInput(new_price, [16 54], 'Name', 'highwayMPG');
new_price = addOutput(new_price, [5118 45400], 'Name', 'price');

new_price = addMF(new_price, "symboling", "gaussmf", [1 3], 'Name', "perigo");
new_price = addMF(new_price, "symboling", "gaussmf", [-3 1], 'Name', "seguro");

new_price = addMF(new_price, "normalizedLosses", "gaussmf", [65 160.5], 'Name', "gasto_baixo");
new_price = addMF(new_price, "normalizedLosses", "gaussmf", [161 256], 'Name', "gasto_elevado");

new_price = addMF(new_price, "width", "gaussmf", [60.3 66.3], 'Name', "pequeno");
new_price = addMF(new_price, "width", "gaussmf", [66.3 72.3], 'Name', "grande");

new_price = addMF(new_price, "horsePower", "gaussmf", [48 128], 'Name', "pouco");
new_price = addMF(new_price, "horsePower", "gaussmf", [128 208], 'Name', "medio");
new_price = addMF(new_price, "horsePower", "gaussmf", [208 288], 'Name', "muito");

new_price = addMF(new_price, "cityMPG", "gaussmf", [13 22], 'Name', "baixo");
new_price = addMF(new_price, "cityMPG", "gaussmf", [22 26], 'Name', "medio");
new_price = addMF(new_price, "cityMPG", "gaussmf", [26 49], 'Name', "alto");

new_price = addMF(new_price, "highwayMPG", "gaussmf", [16 23], 'Name', "baixo");
new_price = addMF(new_price, "highwayMPG", "gaussmf", [23 30], 'Name', "medio");
new_price = addMF(new_price, "highwayMPG", "gaussmf", [30 54], 'Name', "alto");

new_price = addMF(new_price, "price", "gaussmf", [5118 15188], 'Name', "barato");
new_price = addMF(new_price, "price", "gaussmf", [15188 25258], 'Name', "medio");
new_price = addMF(new_price, "price", "gaussmf", [25258 35328], 'Name', "medio caro");
new_price = addMF(new_price, "price", "gaussmf", [35328 45400], 'Name', "caro");
```

Posteriormente, foram criadas as regras de inferência destinadas a adaptar o preço:

```
%se as normalize losses forem altas ou symboling perigo ou horsepower pouco então vale pouco
%se as normalize losses forem baixas e highway medio entao preço medio
%se widht for alto e symboling seguro e o city_mpg medio e horsepower for muito então o preço é caro
regras = [1 2 0 1 0 0 1 1 2
          0 1 0 0 0 2 2 1 1
          2 0 2 3 2 0 4 1 1];

new_price = addRule(new_price, regras);
```

Por fim, realizámos a adaptação do preço através da Inferência de Mamdani:

```
for symboling=-3:3
    for normalizedLosses=65:256
        for width=60.3:72.3
            for horsePower=48:288
                for cityMPG=13:49
                    for highwayMPG=16:54
                        entrada=[symboling normalizedLosses width horsePower cityMPG highwayMPG];
                        new_priceFuzzy = evalfis(entrada, new_price);
                        fprintf('symboling = %d\nnormalizedLosses = %d\n width = %d\nhorsePower = %d\n cityMPG = %d\n highwayMPG = %d\n price = %d\n\n', symboling, normalizedLosses, width, horsePower, cityMPG, highwayMPG, new_priceFuzzy);
                    end
                end
            end
        end
    end
end
```


Implementação

Fase	Subponto	Implementado?
Retrieve	Leitura da biblioteca de casos	Sim
Retrieve	Recolha da informação de um novo caso	Sim
Retrieve	A definição de medidas adequadas à determinação da semelhança global entre o novo caso e os casos passados, atendendo a que existem atributos numéricos, booleanos e nominais	Sim
Retrieve	A possibilidade de filtragem opcional da biblioteca por marca do automóvel, se possível;	Sim
Retrieve	A apresentação dos N casos mais semelhantes ordenados por ordem decrescente de semelhança	Sim
Retrieve	O desenvolvimento de uma pequena aplicação de administração	Não
Reuse	Se o novo caso for considerado “suficientemente próximo” de uma anterior, o preço do automóvel conhecido pode ser usado como avaliação do novo automóvel	Sim
Reuse	Caso contrário, o preço do automóvel antigo tem de ser adaptado. Para realizar essa adaptação deve atender ao significado e implicações de cada um dos atributos que considerar relevantes	Sim
Revise	Revise	Sim
Retain	Retain	Sim
Reuse (Inferência Difusa)	Para cada atributo numérico suscetível de ser <i>fuzificado</i> , defina termos linguísticos e funções de pertinência adequadas	Sim
Reuse (Inferência Difusa)	Crie regras de inferência destinadas a adaptar o preço	Sim
Reuse (Inferência Difusa)	Realize a adaptação do preço através da Inferência de Mamdani.	Sim

Testes

Fizemos dois testes, um com o Reuse, o outro com o Reuse por inferência difusa:

```
symboling: 0
normalizedLosses: 100
  make: 'audi'
  fuelType: 'diesel'
  aspiration: 'turbo'
  numOfDoors: 'four'
  bodyStyle: 'sedan'
  driveWheels: '4wd'
  engineLocation: 'front'
  wheelBase: 100.2000
  lenght: 150.7000
  width: 65.9000
  height: 49.8000
  curbWeight: 2201
  engineType: 'ohc'
  numbOfCylinders: 'six'
  engineSize: 231
  fuelSystem: '2bbl'
  bore: 3.0100
  stroke: 2.2700
compressionRatio: 18
  horsepower: 120
  peakRPM: 5234
  cityMPG: 37
  highwayMPG: 22
  price: 10000
```

O resultado calculado pelo programa com o Reuse foi: 18826.23, enquanto que o Reuse aplicado com inferência difusa foi de: 26731.

Conclusão

Analisando os testes temos que no primeiro foi o resultado estimado para o reuse por regressão linear múltipla foi de 18826.23 e por inferência difusa foi de 26731.

Por esta razão, podemos concluir que era expectável obter valores diferentes visto que no segundo caso, retiramos alguns atributos, tais como o symboling, normalizedLosses e horsePower, para retirarmos alguma complexidade ao programa e conseguirmos obter um resultado em minutos em vez de horas.

Por último, de forma geral, este trabalho permitiu-nos aprofundar o nosso conhecimento sobre certos conceitos, mais em específico, sobre CBR.