

Conhecimento e Raciocínio 2015/2016

Prática 5 - Case-Based Reasoning

1. Considere um site de comércio electrónico de aluguer de apartamentos em que a selecção se faz com base numa semelhança global S , calculada entre o perfil do apartamento pretendido e os apartamentos existentes na base de casos. Os apartamentos são definidos pelos seguintes atributos:

Atributo	Tipo	Valores
Localização	Enumerado	Centro, Cidade, Arredores
Assoalhadas	Enumerado	T1, T2, T3, T4
Preço	Inteiro	0...1000 €
Garagem	Booleano	Verdadeiro / Falso

a) Defina tabelas de dupla entrada, ao seu critério, para:

- Semelhança Localização / Localização (entre apartamento pretendido e apartamentos da biblioteca)
- Semelhança Assoalhadas / Assoalhadas (entre apartamento pretendido e apartamentos da biblioteca), na seguinte base: se um apartamento na biblioteca tiver o mesmo número de assoalhadas que o pretendido, ou superior, a semelhança é 1; caso contrário a semelhança é inferior a 1, dependendo de quão diferentes são os apartamentos.

Importante a ter em conta:

1. Entre localizações iguais a semelhança deve ser 1
2. Entre diferentes a semelhança deve ser <1 e tanto menor quanto maior for a distância entre localizações

Poderia ser, por exemplo:

	Centro	Cidade	Arredores
Centro	1	0.7	0.4
Cidade	0.7	1	0.6
Arredores	0.4	0.6	1

Importante, a ter em conta:

1. Entre apartamentos iguais a semelhança deve ser 1
2. Se o pedido for menor que o da base de casos, a semelhança deve ser 1
3. Caso contrário a semelhança deve ser <1 e tanto menor quanto maior for a diferença no número de assoalhadas

Poderia ser, por exemplo:

Pedido/Base Casos	T1	T2	T3	T4
T1	1	1	1	1
T2	0.7	1	1	1
T3	0.4	0.7	1	1
T4	0.1	0.4	0.7	1

b) Considere que um cliente pretende um apartamento situado no Centro, T2, 500€, com garagem. A relevância é de 1 para Localização e Assoalhadas, 2 para Preço e 0.5 para Garagem. Calcule a semelhança entre o pedido do cliente e cada um dos seguintes apartamentos existentes na base de casos. Ordene a proposta final tal como o sistema o faria e comente o resultado obtido.

Apartamento 1	Apartamento 2	Apartamento 3
Centro	Cidade	Arredores
T1	T2	T4
600	700	700
Falso	Verdadeiro	Verdadeiro

$$D_L(q, s) = \frac{\sum_{i=1}^n d(q_i, s_i) \times w_i}{\sum_{i=1}^n w_i}$$

$$S = 1 - D$$

(Se não resolveu a) admita que para atributos enumerados a semelhança é 0 se forem diferentes e 1 se forem iguais)

Podemos arbitrar um domínio para os preços. Seja, por exemplo, [0, 1000]€.
Os preços têm agora de ser normalizados, assim:

$$600/1000=0.6$$

$$700/1000=0.7$$

Agora há que notar que as tabelas definidas em a) são de semelhança, enquanto a fórmula é de distâncias. Por isso, cada semelhança deve entrar como o seu complemento para 1. Por exemplo, $d(\text{Centro}, \text{Cidade}) = 1 - 0.7 = 0.3$. Portanto:

Aplicando a fórmula com as ponderações dadas, temos:

$$D_L(\text{pedido}, \text{Ap1}) = \frac{(1-1) \times 1 + (1-0.7) \times 1 + (0.7-0.5) \times 2 + (1-0) \times 0.5}{1+1+2+0.5} = \frac{1+0.3+0.4+0.5}{4.5} = \frac{2.2}{4.5} = 0.489$$

$$\text{Semelhança}(\text{pedido}, \text{Ap1}) = 1 - 0.489 = 0.511$$

Para os restantes apartamentos procedia-se da mesma forma apenas atendendo a que, de acordo com a segunda tabela de a) - Apartamentos / Apartamentos -, para o Apartamento 3, por ser um T4 e portanto maior que o T2 pedido, se deveria entrar com uma distância local de valor 0 na parcela relativa à distância entre tipos de apartamentos.

2. Num sistema baseado no paradigma RBC o caso A é descrito pelos atributos X=100, Y=50 e o caso B pelos atributos X=200, Y=125 todos definidos no domínio [0, 250].

a) Determine a distância e a semelhança linear entre casos considerando factores de ponderação todos de valor 1

b) Repita a) considerando agora distância e semelhança euclidiana.

c) Suponha que além de X e Y os casos possuíam ainda um terceiro atributo de nome "Cor". Como poderiam ser definidas semelhanças entre cores? Dê um exemplo.

3. Considere um sistema de diagnóstico baseado no paradigma CBR. Seja Q o *query case* (o caso actual) e K_i os *known cases* (casos passados). A função de semelhança global é $S = (\#P - \#A1 - \#A2) / \#(P \cup A1 \cup A2)$, em que P é o conjunto de sintomas presentes em Q e em K_i ; A1 é o conjunto de sintomas presentes em Q e ausentes em K_i ; A2 é o conjunto de sintomas ausentes em Q e presentes em K_i .

a) Calcule a semelhança S entre Q e K_i se $P = \{s_1, s_2, s_3\}$, $A1 = \{s_4\}$ e $A2 = \{s_5\}$ em que s_i são sintomas.

$$S=(3-1-1)/5=0.2$$

b) Estude o comportamento da função S e comente-o quanto ao grau de certeza que o seu domínio de chegada sugere. *Sugestão: considere as situações limite e (pelo menos) uma intermédia.*

Todos os sintomas estão presentes e nenhum ausente $\rightarrow S=(\#P-0-0)/\#P=1$

Todos os sintomas estão ausentes e nenhum presente (caso 1) $\rightarrow S=(0-\#A1-0)/\#(A1)=-1$

Todos os sintomas estão ausentes e nenhum presente (caso 2) $\rightarrow S=(0-0-\#A2)/\#(A2)=-1$

Todos os sintomas estão ausentes e nenhum presente (caso 3) $\rightarrow S=(0-\#A1-\#A2)/\#(A1 \cup A2)$

Como a intersecção de $A1$ e $A2$ é o conjunto vazio, também neste caso $S=-1$

Situação intermédia \rightarrow a da alínea a)

Conclusão: O grau de certeza varia entre $[-1, 1]$. No primeiro caso (-1) há uma exclusão (certa) de um diagnóstico em consideração, o que faz sentido porque todos os sintomas estão presentes num dos casos e nenhum no outro. No segundo caso $(+1)$ há uma certeza a favor de um diagnóstico o que condiz como facto de todos os sintomas estarem presentes no caso actual e no passado, não estando nenhum ausente. Em suma, a função S apresenta coerência.

c) Comente a afirmação: "se $\#A2 \neq 0$ o sistema pode sugerir procedimentos que conduzam a um diagnóstico mais fiável e simultaneamente à exclusão de vários outros diagnósticos em consideração".

Se há sintomas ausentes no caso actual e presentes no caso passado, o sistema poderá pedir ao utilizador, para tentar confirmar se se trata de facto de um caso como o passado que ele já conhece, se os sintomas em falta existem de facto ou não, actuando assim como guia de diagnóstico ao sugerir uma série de observações adicionais às originais.

d) Suponha que ao diagnóstico D_1 , relativo ao caso K_1 , corresponde a prescrição de X cápsulas do medicamento M_1 , sendo que cada cápsula contém D_1 mg do respectivo princípio activo. Escreva em pseudocódigo uma função capaz de prescrever um genérico G_i qualquer em substituição de M_1 tendo em linha de conta que a quantidade de princípio activo de um genérico é N_i vezes inferior a D_1 .

Função dosagem (parâmetros X capsulas, N vezes)

Retorna $X*N$

e) Como seria testada a validade da prescrição da terapêutica X ou G_i ?

Aplicando ao doente a medicação X ou G_i na dosagem prescrita.

f) Suponha que decorrido o período de tratamento com um genérico G_1 proposto pelo sistema, o doente não apresentou melhoras. Deveria este facto ser memorizado pelo sistema? Porquê?

Sim, porque a aprendizagem também deve ser feita por erros (exemplo CLAVIER). Ao reter esta informação o sistema poderia deixar de prescrever, no futuro, o genérico G_1 por não apresentar resultados.

g) A que fases do ciclo CBR correspondem as operações descritas em a), d) e) e f) ?

Retrieve, Reuse, Revise, Retain

h) Partindo da expressão dada, $S=(\#P - \#A1 - \#A2) / \#(P \cup A1 \cup A2)$, sugira uma nova fórmula que contemple a atribuição de um coeficiente de ponderação w_i a cada sintoma s_i de tal forma que os resultados obtidos em a) mantenham o seu significado.

$$S = \frac{\sum_1^n P_i w_i - \sum_1^k A1_j w_j - \sum_1^l A2_m w_m}{\sum_1^n w_i + \sum_1^k w_j + \sum_1^l w_m}$$

$P_i=1$ se sintoma presente em ambos os casos, 0 caso contrário

$A1_j=1$ se sintoma presente em Q e ausente em K_i

$A2_m=1$ se sintoma ausente em Q e presente em K_i

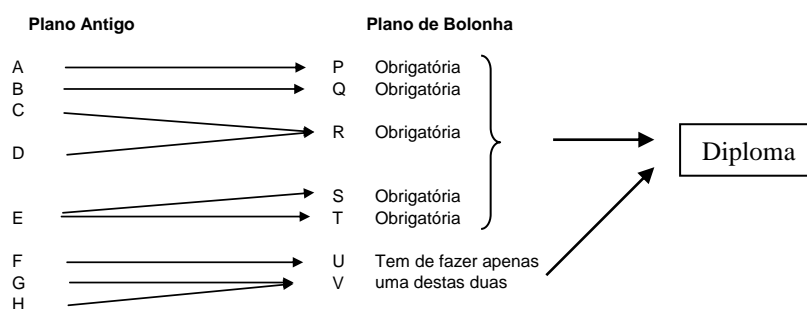
i) Em que silogismo se baseia a contribuição positiva para a semelhança global dada pela parcela #P? Formalize-o. Dê um exemplo.

Silogismo: Modus Ponens

Formalização: P
P→Q
Q

Exemplo: Tem febre alta
Se tem febre alta tem gripe
Logo tem gripe

4. Num plano de transição para o modelo de Bolonha as disciplinas do antigo plano de estudos dão equivalência às disciplinas do novo plano de acordo com o seguinte esquema:



Para alunos que reingressam - i.e., regressam à instituição depois de terem já desenvolvido actividade profissional no exterior - cursos de formação e experiência profissional são creditados, i.e., são-lhes atribuídos créditos no Plano de Bolonha. Por isso, para um aluno que reingressa, a cadeira U ou V poderá não ter de ser realizada.

A biblioteca de um sistema CBR foi inicialmente carregado com os seguintes casos:

Disciplinas feitas no antigo plano								Experiência Profissional (anos)	Cursos de Formação (horas)	Decisão
A	B	C	D	E	F	G	H	0	0	Grau de Bolonha
A	B	C	D	E				10	50	Fazer cadeira U ou V
A	B	C	D	E				20	100	Grau de Bolonha

Considerando que cada cadeira, experiência profissional e cursos de formação têm todos peso 1, qual seria a decisão para o candidato X, com o seguinte perfil?

A	B	C	D	E	F			40	200	
---	---	---	---	---	---	--	--	----	-----	--

5. A tabela seguinte representa 3 casos que descrevem as condições e o respectivo resultado relativamente ao lançamento de um foguete Ariane:

Sensores				Estado do Tempo				Lançamento
A	B	C	D	Temperatura °C	Humidade %	Trovoada	Vento Km/h	
350°C	10Kg/m2	80A	15000V	35	90	S	90	Mal sucedido
300°C	15Kg/m2	50A	15000V	30	80	N	30	Bem sucedido
280°C	12Kg/m2	30A	12000V	30	60	N	50	Contagem suspensa teste T9. Lançamento posterior bem sucedido.

a) Qual seria a decisão a tomar nas seguintes condições:

280°C	12Kg/m2	40A	12000V	35	90	N	40	
-------	---------	-----	--------	----	----	---	----	--

Factores de ponderação:

Sensores: $w_i=2$
 Temperatura, Humidade, Vento: $w_i=1$
 Trovoada: $w=5$

(Use semelhança linear)

Solução: basta aplicar a fórmula de semelhanças lineares 3 vezes (1 para cada caso) depois de normalizar todos os valores em relação aos seus máximos (já que o domínio não é conhecido). Depois disso ver qual o caso mais semelhante e decidir não lançar, lançar ou suspender a contagem.

b) Sabe-se que se o sensor C indicar mais de 100A, mesmo que as restantes condições sejam muito favoráveis, o lançamento corre um sério risco de falhar. Proponha uma regra de adaptação do desfecho dos casos que tenha em linha de conta este factor.

IF Sensor C>100 Then “Lançamento = Mal Sucedido”

6. Considere as seguintes regras destinadas à avaliação de itens antigos para colecionistas, em que EP representa a Época:

- => Valor=0
- Se EP=1 => Valor= Valor+15
- Se EP=1 e Caixa => Valor =Valor+5
- Se EP=1 e Raro => Valor=Valor +5
- Se EP=1 e Não Caixa => Valor=Valor-3
- Se EP=1 e Comum => Valor=Valor-3
- Se EP=2 => Valor=5

a) Escolha uma de entre as respostas possíveis (errado, desconta)

- i) (0.25) Pode realizar a inferência em forward chaining ? ☒ SIM NÃO
- ii) (0.25) A prioridade da regra 7 é irrelevante ? ☒ SIM NÃO
- iii) (0.25) A regra 1 dispara sempre ? ☒ SIM NÃO

b) (1.25) Considerando os atributos época, caixa e raridade, e a solução Valor, preencha as linhas necessárias na tabela seguinte de modo a que ela constitua uma Case Library, com o número mínimo de

casos de um possível sistema CBR. Use o símbolo “X” para indicar situações em que o valor de um dado atributo (já) não interessa.

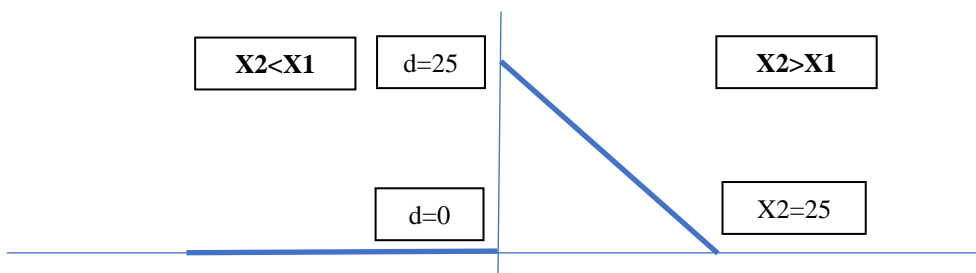
Atributos			Solução
EP	Caixa	Raro	Valor
2	X	X	5
1	S	S	25
1	S	N	17
1	N	S	17
1	N	N	9

7. Considere o paradigma CBR:

a) Considere que os dias da semana são codificados por 1 (=domingo) a 7 (=sábado). Escreva uma expressão matemática que permita calcular a distância d entre quaisquer 2 dias da semana, D1 e D2, contemplando o caso em que estes dias podem pertencer a semanas diferentes (embora contíguas)

$$d = \min(|D1-D2|, 7-|D1-D2|)$$

b) Esboce uma função de distância local entre dois atributos numéricos X1 do query case e X2 da case library, que permita obter uma distância nula quando $X1 > X2$ e igual a $X2 - X1$ quando $X2 > X1$. Os atributos estão definidos no domínio $[0, 25]$




8. Para a previsão da intensidade de fogos implementou-se um sistema baseado no paradigma CBR, em que cada caso é representado pelos atributos humidade, temperatura e vento e pela intensidade do fogo resultante. Cada um destes atributos é por sua vez representado por 2 valores que correspondem ao grau de pertença aos termos linguísticos alta, baixa (humidade e temperatura) ou forte e fraco (vento). A humidade, temperatura e vento têm ponderações de 2, 3 e 4 respetivamente. A intensidade do fogo resultante é representada por um inteiro compreendido entre 0 e 10. A figura seguinte representa a biblioteca de casos do sistema:

	Humidade		Temperatura		Vento		Intensidade do Fogo
	Baixa	Alta	Baixa	Alta	Fraco	Forte	
Ocorrência 1	0.1	0.8	0.5	0.5	0.2	0.9	6
Ocorrência 2	0.7	0.2	0.5	0.5	0.2	0.9	7

a) Usando distância linear, qual das soluções 6 ou 7 corresponde ao seguinte *query case* (caso novo)?

Escrever a expressão usada para o cálculo de D_1 apenas. Não é necessário escrever a expressão de D_2

Humidade		Temperatura		Vento	
Baixa	Alta	Baixa	Alta	Fraco	Forte
0.3	0.6	0.4	0.6	0.3	0.7



$$D_1(q, s_1) = \frac{\sum_{i=1}^n d(q, s_i) \times w_i}{\sum_{i=1}^n w_i} = \frac{2 \times (|0.3 - 0.1| + |0.6 - 0.8|) + 3 \times (|0.4 - 0.5| + |0.6 - 0.5|) + 4 \times (|0.3 - 0.2| + |0.7 - 0.9|)}{2 + 2 + 3 + 3 + 4 + 4} = \frac{0.8 + 0.6 + 1.2}{18} = 0.14$$

$$D_2 = 3.4/18 = 0.19$$

A solução do *query case* é : **6** (o que tem distância **menor**)

b) O modelo acima descrito foi posteriormente melhorado, passando a incluir uma função de adaptação que transforma a solução obtida multiplicando-a por um fator F que depende da % de eucaliptos E e da acessibilidade ao terreno A, ambas definidas em [0, 1]. Mais eucaliptos e/ou menor acessibilidade provocam fogos mais fortes. Defina a expressão de F, atendendo a que ela deve ter as seguintes propriedades:

- $F \in [0,1]$
- Quando há mais eucaliptos e/ou a acessibilidade piora, o valor de F sobe, tendendo para 1
- Contudo, o limite 1 só é possível de atingir caso E=1 ou a acessibilidade seja nula
- Se não houver eucaliptos ou se a acessibilidade for total, apenas o outro fator influencia F

$$F = E + (1-A) - E \cdot (1-A)$$

NOTA: esta expressão é simplesmente igual à da combinação de FC's do MYCIN, cujas características, amplamente discutidas e exemplificadas nas aulas teóricas (e práticas) são descritas pelo texto do enunciado: $FC = FC1 + FC2 - FC1 \times FC2$. Neste caso FC2 foi substituído por $(1-FC2)$ apenas porque a humidade funciona inversamente.

c) Após a ocorrência de um fogo cuja intensidade foi predita pelo sistema, qualquer solução por ele proposta

(Errado desconta a totalidade de cada questão)

Por vezes tem de ser confirmada ou alterada por um professor externo
 Tem de ser sempre confirmada quando corresponder à intensidade do fogo real
 Tem de ser sempre alterada quando não corresponder à intensidade do fogo real

Sim	Não
<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/>

d) Quanto às ocorrências processadas pelo sistema, devem ser registadas na biblioteca:

(Errado desconta a totalidade de cada questão)

Todos os casos tratados pelo sistema
 Apenas os casos cuja solução predita foi correta
 Todos os casos cuja solução predita foi errada, se for seguido o modelo CLAVIER
 Todos os casos cuja solução predita foi correta, mas cuja descrição é diferente

Sim	Não
<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/>

e) A que fases do ciclo CBR correspondem as operações descritas em a), b), c) e d)?

Retrieve, Reuse, Revise, Retain