

# Guião Teórico-Prático

## Tópicos de Inteligência Artificial

Ano Lectivo de 2017/2018

©Luís Seabra Lopes

Departamento de Electrónica, Telecomunicações e Informática  
Universidade de Aveiro

Última actualização: 2018-01-02

## I Objectivos

O presente guião centra-se em exercícios que através dos quais o aluno pode testar a sua compreensão das matérias teóricas (conceitos, algoritmos). A maior parte destes exercícios podem ser realizados sem recurso ao computador.

Este guião é usado nas disciplinas de *Inteligência Artificial*, da *Licenciatura em Engenharia Informática*, e *Introdução à Inteligência Artificial*, do *Mestrado Integrado em Engenharia de Computadores e Telemática*.

## II Agentes Reactivos

1. Está a ser desenvolvido um robô para pesca submarina, o Nautilus, pedindo-se a sua colaboração no desenvolvimento do módulo de controlo. Este robô transporta um máximo de 10 arpões e um depósito com capacidade para 20 peixes. Quando sente um peixe em frente lança imediatamente um arpão (acção *Disparar*). Se o arpão atingir um peixe, o robô guarda o peixe no depósito (acção *Agarrar*), podendo neste caso recuperar o arpão. Caso contrário, o arpão perde-se. Quando perder todos os arpões, o Nautilus pode reabastecer-se (acção *Reabastecer*), ficando novamente com 10 arpões. Quando o depósito de peixes estiver cheio, estes devem ser descarregados (acção *Descarregar*), ficando o depósito novamente vazio. Quando não tem mais nada para fazer, limita-se a vaguear (acção *Vaguear*).
  - (a) Identifique as variáveis de estado, se necessárias.
  - (b) Identifique e caracterize um conjunto de predicados que possam ser usados para descrever situações em que se encontre o Nautilus.
  - (c) Especifique o conjunto de regras situação-acção que regem o comportamento do Nautilus. Pode fazê-lo na forma de uma tabela com as seguintes colunas:
    - Situação - uma conjunção de condições em lógica de primeira ordem

- Actualização - actualização das variáveis de estado, caso existam
  - Acção - acção a executar pelo agente na situação indicada
2. Considere o comportamento das formigas na sua tarefa de arrumar provisões no formigueiro. A formiga procura provisões (acção *Procurar-provisão*). Quando encontra uma provisão, agarra-a (acção *Agarrar-provisão*) e vai procurar o local (acção *Procurar-local*) de arrumação das provisões. A formiga tem sempre uma noção da distância percorrida desde que começou a procurar a arrumação. Se a formiga acha que já percorreu mais de 5 metros sem ter encontrado a arrumação, e vê outra formiga, vai atrás dela (acção *Seguir-formiga*). Quando encontra o local onde estão as outras provisões, liberta a provisão que trás consigo (acção *Libertar-provisão*). Cabe-lhe a si implementar um conjunto de regras situação-acção com base nas quais a formiga simulada se irá comportar. Com vista ao desenvolvimento de um programa de simulação do comportamento das formigas, realize os seguintes passos de análise e especificação:
- (a) Identifique as variáveis de estado, se necessárias.
  - (b) Identifique e caracterize um conjunto de predicados que possam ser usados para descrever as situações em que uma formiga se pode encontrar.
  - (c) Especifique o conjunto de regras situação-acção que regem o comportamento de uma formiga. Pode fazê-lo na forma de uma tabela como no exercício anterior.

### III Representação do Conhecimento

1. Represente as seguintes frases em lógica de primeira ordem:
  - (a) Todos em Oxford são espertos.
  - (b) Alguém em Oxford é esperto.
  - (c) Existe uma pessoa que gosta de toda a gente.
  - (d) Só um aluno chumbou a História.
  - (e) Nem todos os estudantes se inscreveram simultaneamente a Introdução à Inteligência Artificial e Sistemas de Operação.
  - (f) Só um aluno chumbou a História e a Biologia.
  - (g) A melhor nota a História foi mais elevada do que a melhor nota a Biologia.
  - (h) Todos os Portistas gostam do Pinto da Costa.
  - (i) Existe um Sportinguista que gosta de todos os Benfiquistas que não são espertos
  - (j) Existe um Barbeiro que barbeia toda a gente menos ele próprio.
2. Considere o mundo dos blocos com  $n$  blocos representados pelas constantes  $(B_1, B_2, \dots, B_n)$ , predicado  $On(x, y)$  que indica que o bloco  $x$  está em cima do objecto  $y$  e o predicado  $Clear(x)$  que indica que o bloco  $x$  não tem nenhum bloco em cima. Indique qual o número mínimo de blocos para que cada uma das fórmulas seguintes seja verdadeira:
  - (a)  $\neg Clear(B_1) \wedge \neg Clear(B_2) \wedge \exists x(On(B_3, x) \wedge x \neq Floor)$
  - (b)  $Clear(B_1) \Rightarrow Clear(B_2)$

3. Considere o seguinte mundo composto por uma torneira, dois tanques ( $T1$  e  $T2$ ) e um recipiente ( $R$ ). A torneira pode estar aberta para um dos dois tanques mas nunca para os dois ao mesmo tempo. O recipiente pode ser colocado dentro de um tanque, desde que este esteja sem água. Se a torneira estiver aberta, então o tanque respectivo ou o recipiente (se estiver dentro do tanque) ficam com água. Se o recipiente estiver dentro de um tanque, e a torneira estiver aberta para esse tanque, então o recipiente fica com água, mas o tanque não. Tanto os tanques como o recipiente podem ter água mesmo que a torneira não esteja aberta para eles.

Considere os seguintes predicados:

- $Water(x)$  indica que  $x$  (recipiente ou tanque) tem água;
- $Open(x)$  que indica que a torneira está aberta para o tanque  $x$ ;
- $Over(x, y)$  que indica que o recipiente  $x$  está colocado sobre o tanque  $y$ .

- (a) Para cada uma das observações seguintes, apresente os valores lógicos possíveis para  $Water(T1)$ ,  $Water(T2)$  e  $Water(R)$ .

- i.  $Open(T2) \wedge Over(R, T2)$
- ii.  $Open(T2) \wedge Over(R, T1)$
- iii.  $\neg Open(T2) \wedge Over(R, T1)$
- iv.  $\neg Open(T1) \wedge \neg Open(T2) \wedge Over(R, T1)$

- (b) Diga se cada uma das seguintes fórmulas é satisfatível, e caso seja, se é uma tautologia:

- |  |  |
|--|--|
| i. $\forall x (\neg(\neg Open(x) \Rightarrow Water(x)))$ | iii. $\forall x (Open(x) \Rightarrow \exists y Water(y))$          |
| ii. $\forall x (\neg(Open(x) \Rightarrow Water(x)))$     | iv. $\exists x \exists y (Open(x) \wedge Open(y) \wedge x \neq y)$ |

4. Considere o mapa apresentado na figura 1. Neste mapa estão representados algumas ruas, edifícios e o estado de algumas ruas. A rua *descanso* tem um comprimento de 6. Do mesmo modo a rua *sul* encontra-se com a rua *descanso* a 2 unidades do edifício *casa*. Tanto a rua *norte* como a *sul* têm um comprimento de 8. A rua *artes* intersecta as duas anteriores no ponto médio. A rua *norte* está fechada nos pontos 5, 6 e 7.

- (a) Proponha um conjunto de predicados que permita representar conhecimento deste tipo.
- (b) Usando os predicados que propôs, represente o conhecimento acima descrito.

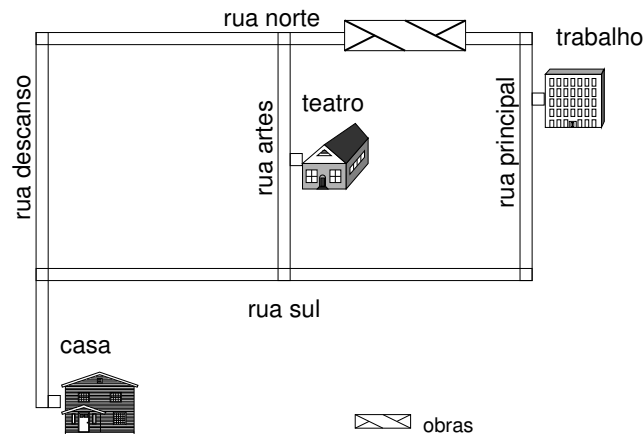


Figura 1: Mapa de uma cidade.

5. Enquadre a linguagem KIF no contexto da engenharia do conhecimento, comparando-a com outros formalismos seus conhecidos e comentando a sua relevância para a construção de agentes.
6. Represente o seguinte conhecimento através de uma rede semântica:
 

“Os robôs são máquinas. Há robôs com pernas, que podem ou não ser humanóides, e robôs que se movem sobre rodas ou até usando lagartas. O Nautilus é um robô com 3 rodas que obtém energia de 4 baterias de 12V / 7Ah. Os robôs humanóides têm 2 pernas e 2 braços.”
7. Considere o circuito electrónico apresentado na Figura 2, no qual pode encontrar uma porta AND (a1), uma porta OR (o1) e uma porta XOR (x1). O circuito tem três entradas (e1, e2, e3) e uma saída (s1). Para se poder calcular a saída em função das entradas, é necessário levar em conta o seguinte conhecimento geral sobre circuitos electrónicos:
  - O sinal em cada terminal é On ou Off
  - Dois terminais que estejam ligados um ao outro têm o mesmo sinal
  - A relação de ligação entre terminais é comutativa
  - A saída de uma porta OR é On se pelo menos uma das entradas for On
  - A saída de uma porta AND é On se todas as entradas forem On
  - A saída de uma porta XOR é On se as suas entradas forem diferentes
  - A saída de uma porta NOT é diferente da sua entrada
  - (a) Identifique os tipos de objectos presentes no domínio dos circuitos electrónicos, bem como as funções e relações relevantes
  - (b) Represente em lógica de primeira ordem o conhecimento geral do domínio
  - (c) Represente em lógica de primeira ordem o circuito da Figura 2.
8. Considere a rede de Bayes identificada pela seguinte atribuição de probabilidades:  $p(a) = 0.2$ ,  $p(b|a) = 0.3$ ,  $p(b|\neg a) = 0.2$ ,  $p(c|b) = 0.2$ ,  $p(c|\neg b) = 0.9$ ,  $p(d|b) = 0.1$ ,  $p(d|\neg b) = 0.2$ . Calcule a probabilidade conjunta  $p(a \wedge b \wedge \neg c \wedge \neg d)$ .

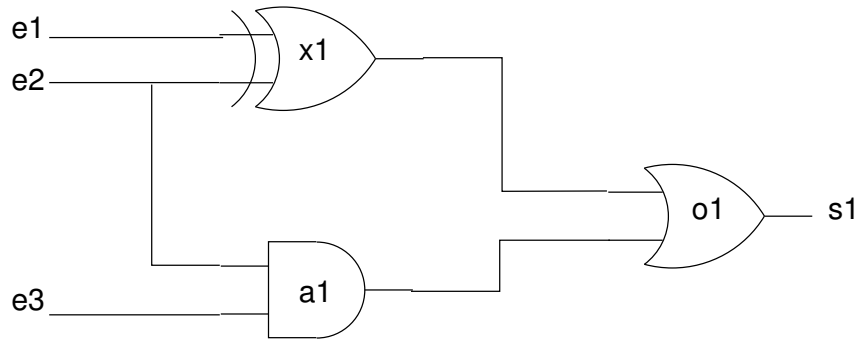
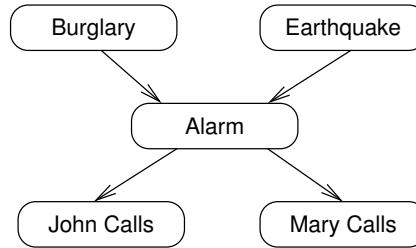


Figura 2: Exemplo de circuito electrónico.

9. Considere o cenário em que um alarme de uma casa pode disparar por causa de um assaltante, mas também por ocorrência de um terramoto. Se o alarme dispara, os ocupantes da casa podem efectuar uma chamada de telefone. Este cenário é exemplificado pela seguinte Rede Bayesana:



Os eventos *Burglary* e *Earthquake* não dependem de nenhum outro evento: são independentes de todos os outros. Assim, só é necessário especificar as suas probabilidades:

$$P(\text{Burglary}) = 0.001$$

$$P(\text{Earthquake}) = 0.02$$

O evento *Alarm* depende da ocorrência dos eventos *Burglary* e *Earthquake*: os eventos *Alarm* e *Burglary* não são independentes (tal como os eventos *Alarm* e *Earthquake*). É necessário especificar a probabilidade condicional de *Alarm* dado as várias combinações de *Burglary* e *Earthquake*:

$$P(\text{Alarm} | (\text{Burglary} \wedge \text{Earthquake})) = 0.9$$

$$P(\text{Alarm} | (\text{Burglary} \wedge \neg \text{Earthquake})) = 0.9$$

$$P(\text{Alarm} | (\neg \text{Burglary} \wedge \text{Earthquake})) = 0.1$$

$$P(\text{Alarm} | (\neg \text{Burglary} \wedge \neg \text{Earthquake})) = 0.001$$

Quanto aos eventos *MaryCalls* e *JohnCalls* são ambos dependentes do evento *Alarm*. As suas probabilidades condicionais são:

$$P(\text{MaryCalls} | \text{Alarm}) = 0.95$$

$$P(MaryCalls|\neg Alarm) = 0.001$$

e

$$P(JohnCalls|Alarm) = 0.9$$

$$P(JohnCalls|\neg Alarm) = 0.0$$

Calcule as seguintes probabilidades:

- (a)  $P(A)$
- (b)  $P(M)$
- (c)  $P(J)$

10. Considere um domínio composto por animais, espécies e intervalos de tempo, no qual o conhecimento pode ser descrito através dos seguintes predicados:

- $Animal(a)$ :  $a$  é um animal
- $Espécie(a,e)$ : o animal  $a$  é da espécie  $e$
- $Vivo(a,t)$ : o animal  $a$  está vivo no intervalo  $t$
- $Extinta(e,t)$ : a espécie  $e$  está extinta no intervalo  $t$
- $Progenitor(p,a)$ : o animal  $p$  é progenitor do animal  $a$

(a) Represente as seguintes frases em lógica de primeira ordem:

- Qualquer animal tem um progenitor
- Qualquer animal pertence a uma espécie
- Apenas os animais pertencem a espécies
- Se  $p$  é o progenitor de  $a$  e  $a$  pertence a uma espécie  $e$ , então  $p$  também pertence a  $e$
- Uma espécie  $e$  está extinta no intervalo  $t$  se nenhum animal dessa espécie está vivo nesse intervalo
- Não existiam mamutes vivos no ano de 1918

(b) Demonstre que os mamutes estavam extintos em 1918 a partir das fórmulas que escreveu.

11. A nova empresa “SOF – Sistemas Operativos do Futuro”, sediada na Costa Nova, comercializa actualmente o sistema operativo SOF2018h, mas este sistema ainda tem alguns problemas. A empresa pretende desenvolver um assistente que determina automaticamente se o utilizador precisa de ajuda, e, quando tal acontece, toma a iniciativa de fazer alguns sugestões ao utilizador. Após análise exaustiva dos problemas sentidos pelos utilizadores, verificou-se que há essencialmente dois sintomas da necessidade de ajuda. Um deles é o utilizador fazer uma “cara preocupada”, o que pode ser detectado por um sistema de reconhecimento de expressões faciais previamente desenvolvido. O outro sintoma é o utilizador aumentar a frequência de utilização do rato, por estar a navegar através de diferentes menus à procura da solução para algum problema.

Entretanto, após análise, sabe-se que 60% dos utilizadores têm sobrecarga de trabalho, o que pode também causar uma cara preocupada. Cerca de 1% dos utilizadores sobrecarregados mostram cara preocupada, caso não precisem de ajuda. Já no caso de precisarem de ajuda, essa percentagem sobe para 2%. Os utilizadores não sobrecarregados, que não precisam de ajuda, apenas em 0.1% dos casos mostram cara preocupada. Este valor sobe para 1.1% caso precisem de ajuda.

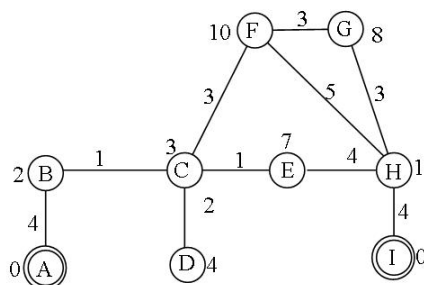
Sabe-se também que utilizadores com sobrecarga de trabalho tendem a acumular correio electrónico não lido. Apenas 0.1% dos utilizadores sem sobrecarga de trabalho acumulam correio não lido. Pelo contrário, 90% dos utilizadores com sobrecarga de trabalho acumulam correio electrónico não lido.

Há uma aplicação no SOF2018h especialmente causadora de problemas, o processador de texto SOF2018h Pal, no qual os utilizadores passam 5% do seu tempo de utilização do SOF2018h. Na verdade, quando o utilizador está a usar esta aplicação, tenderá a precisar de mais ajuda, o que acontece em 25% dos casos. Já quando não usa o processador de texto, a probabilidade de precisar de ajuda é 0.4%. Se o utilizador não está a usar o SOF2018h Pal, a existirá uma frequência exagerada de utilização do rato em 10% dos casos em que o utilizador precisa de ajuda e em 1% dos restantes casos. Quando o utilizador está a usar o SOF2018h Pal, ele fará uma utilização exagerada do rato em 90% dos casos, independentemente de precisar de ajuda ou não.

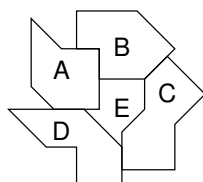
Com vista ao desenvolvimento do assistente de ajuda, pretende-se representar este conhecimento através uma rede de Bayes, tarefa que acaba de lhe cair em cima!... Identifique as variáveis da rede, desenhe a rede e apresente a tabela de probabilidades condicionadas.

## IV Técnicas de resolução automática de problemas

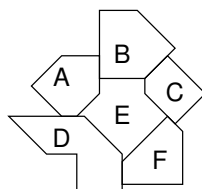
1. Considere o espaço de estados apresentado na Figura 4, em que os valores nas ligações correspondem ao respectivo custo e os valores nos nós são os da função heurística. Nos exercícios, considere os nós *A* e *I* como soluções possíveis.



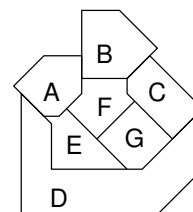
- (a) A heurística apresentada é admissível? Justifique a sua resposta e, em caso negativo, faça as alterações necessárias por forma a que passe a sê-lo.
  - (b) Desenhe a árvore de pesquisa gerada pela estratégia A\* tomando com estado inicial o estado *F*. Indique o valor da função de avaliação em cada nó e numere os nós pela ordem em que são criados. Considere que a pesquisa em árvore se faz sem repetição de estados. Em caso de empate no valor da função de avaliação, o nó escolhido para expansão será o que vem antes na ordem alfabética dos estados. Use os valores originais da heurística.
  - (c) Calcule o factor de ramificação médio da árvore gerada.
  - (d) Calcule o factor de ramificação efectivo da árvore gerada.
2. Considere uma árvore de pesquisa com factor de ramificação  $r$ . Suponha que a solução mais próxima da raiz se encontra a uma profundidade  $g$ . Qual o número mínimo e máximo de nós visitados numa pesquisa em profundidade, com limite  $d$ ?
  3. Que heurística admissível sugere que seja usada com a pesquisa A\* para planeamento de caminhos óptimos em redes viárias? Justifique.
  4. Pretende-se colorir os seguintes mapas de forma a que regiões adjacentes fiquem com cores diferentes. Apresente o grafo de restrições para cada um dos mapas e indique o número mínimo de cores necessário para cada um deles.



(a)



(b)



(c)

5. Considere o seguinte problema:



André, Bernardo e Cláudio dão um passeio de bicicleta. Cada um anda na bicicleta de um dos amigos e leva o chapéu de um dos outros. O que leva o chapéu de Cláudio anda na bicicleta de Bernardo. Que bicicleta e que chapéu levam cada um dos amigos? (Retirado de Pierre Berloquin. *100 Jogos Lógicos*. Gradiva, 1990.)

- (a) Represente o problema através de um grafo de restrições.
  - (b) Utilize o módulo `constraintsearch` para resolver o problema.
6. Considere o seguinte puzzle Su Doku em que cada linha, coluna e quadrado de 3 por 3 deve ser preenchido com os números de 1 a 9 e sem repetições. Apresente uma abordagem à resolução deste puzzle utilizando Pesquisa por Propagação de Restrições. Indique quais as variáveis, o seu domínio e as restrições a considerar.

			5		7			3
		5				8		2
	4			6				
	7	6	3		2	5		
	8							
	3	9	1		8	2		
	6			3				
		1				6		7
			8		6			9

Retirado de Yukio Suzuki. *Su Doku para especialistas e outros puzzles japoneses*. Editorial Estampa, 2005.

7. O caso particular da pesquisa por recozimento simulado (*simulated annealing*) com temperatura  $T = 0$  tem semelhanças significativas com alguma outra técnica de pesquisa sua conhecida? Nesse caso, identifique as principais semelhanças e diferenças.
8. Com vista à sua resolução através de pesquisa com propagação de restrições, formule o problema de escalonar quatro tarefas ( $A$ ,  $B$ ,  $C$  e  $D$ ) tendo em conta as seguintes informações:
  - As tarefas começam às horas certas, a partir das 8h de um dia, terminam o mais tardar às 19h desse mesmo dia.
  - A duração das tarefas é a seguinte:  $A$  - 1h,  $B$  - 2h,  $C$  - 3h,  $D$  - 4h.
  - A tarefa  $A$  deverá terminar antes das tarefas  $B$  e  $C$  começarem.
  - A tarefa  $D$  deverá começar depois de terminarem as tarefas  $B$  e  $C$ .
  - As tarefas não podem ser realizadas simultaneamente.
9. Um macaco está numa sala. Na mesma sala, pendurado num cabide e fora do alcance do macaco, está também um cacho de bananas. Se o macaco subir para cima de uma caixa, conseguirá alcançar as bananas. Inicialmente, o macaco está na posição A, as bananas na posição B e a caixa na posição C. As acções que o macaco pode executar são: deslocar-se de uma posição para outra; empurrar um dado objecto de uma posição para outra; subir para cima de um dado objecto; agarrar o cacho de bananas.
  - (a) Identifique um conjunto de condições com as quais seja possível descrever os vários estados do mundo neste problema.

- (b) Descreva o estado inicial do problema usando as condições que propôs.
  - (c) Identifique e descreva as acções possíveis de acordo com o formato STRIPS.
  - (d) Que sequência de acções deverá o macaco executar?
  - (e) Apresente uma estimativa para o tamanho aproximado que a árvore de pesquisa poderá atingir. Justifique.
10. O robô VG-10 deixado recentemente em Marte pela Agência Espacial Portuguesa (AEP) precisa de planejar as suas expedições entre várias estações anteriormente construídas pela AEP nesse planeta do sistema solar. Para cada estação, o VG-10 sabe quais as estações adjacentes, ou seja, aquelas para as quais se pode deslocar com o equivalente a um depósito de combustível. Além de encher o depósito, o VG-10 pode transportar dois bidões de combustível para recarga.

Os tipos de acções que o VG-10 pode executar são:

- $ir(E_1, E_2)$  - ir da estação  $E_1$  para a estação adjacente  $E_2$ .
- $carregar(E, B, X)$  - carregar o bidão  $B$  da estação  $E$  para o espaço  $X$  no robô.
- $encher(B, X)$  - encher o depósito com o combustível do bidão  $B$  que está carregado no robô na posição  $X$ .
- $descarregar(E, B, X)$  - descarregar o bidão  $B$  do espaço  $X$  para a estação  $E$ .

Realize os seguintes exercícios:

- (a) Proponha um conjunto de condições que permitam descrever os estados de planeamento das expedições do VG-10. Explique o seu significado.
  - (b) Especifique os operadores de planeamento necessários para representar as acções do VG-10.
  - (c) Se pretender que um planeador das missões do VG-10, baseado na estratégia A\*, encontre uma solução óptima quando ela existe, que função de avaliação/estimação de custos sugere que seja utilizada? Justifique.
  - (d) Considere agora que uma *estação adjacente* é uma estação para a qual existe uma ligação (trilho ou caminho) que o robô pode seguir. As distâncias entre todos os pares de estações adjacentes são conhecidas. Além disso, sabe-se que distância o VG-10 pode percorrer com o combustível de um depósito. Que adaptações à representação das acções seria necessário introduzir para levar em consideração estas restrições. Ilustre para o caso do operador  $ir(E_1, E_2)$ . Assuma que o combustível do depósito dá para chegar a qualquer estação adjacente, embora possa sobrar.
11. No contexto da resolução automática de problemas usando técnicas de pesquisa, defina os seguintes termos por palavras suas: *estado*, *espaço de estados*, *acção*, *árvore de pesquisa*, *restrição*.
12. Considere o problema do Caixeiro Viajante, que consiste em descobrir um caminho óptimo que passe por determinadas cidades  $A_1, \dots, A_n$ , partindo, por exemplo, de  $A_1$ , passando por todas as outras apenas uma vez, e regressando a  $A_1$ . Considere que são conhecidas as distâncias entre todos os pares de cidades. Como formularia este problema para o resolver através de pesquisa A\*? Indique em particular em que consistiriam os estados, qual seria o estado inicial, qual o método para gerar as transições de estados, qual a função de avaliação dos custos das transições e qual a função heurística.

13. Considere um jogo em que as 8 primeiras letras do alfabeto (A a H) são colocadas de forma aleatória numa matriz de 3x3, ficando portanto uma posição por preencher. Uma letra (verticalmente ou horizontalmente) adjacente à posição livre pode ser deslocada para essa posição. O objectivo é determinar uma sequência de movimentos para gerar uma outra configuração da matriz. No exemplo da figura incluída abaixo, são necessários 6 movimentos, mas em média são precisos muitos mais.

A	E	
C	B	D
F	G	H

Configuração inicial

	A	B
C	D	E
F	G	H

Objectivo

- (a) No caso de ser utilizada pesquisa em árvore, indique uma estimativa para o factor de ramificação médio das árvores de pesquisa neste domínio.
- (b) Tendo em vista a possível resolução de problemas deste tipo através de A\*, considere as seguintes heurísticas:
- ( $h_1$ ) Número de letras fora da sua posição final. ( 4 no exemplo acima)
  - ( $h_2$ ) Soma das distâncias horizontais e verticais das várias letras às respectivas posições finais. (6 no exemplo acima)

Estas heurísticas são admissíveis? Qual delas espera que funcione melhor?

14. Considere o problema de construir automaticamente passatempos de "palavras cruzadas". Como entrada, o processo recebe uma lista de palavras que podem ser utilizadas, e uma matriz, com informação de quais as posições a preencher (brancas) e quais as posições a não preencher (pretas). Qualquer sequência não interrompida de letras, seja na horizontal, seja na vertical, deve corresponder a uma palavra válida. O resultado é uma selecção das palavras a incluir na matriz e respectivas posições na matriz. Note que o problema aqui colocado é o da geração de uma matriz de palavras cruzadas, e não o da resolução do passatempo com base em sinónimos fornecidos como pistas.
- (a) No pressuposto de utilizar pesquisa em árvore, como representaria os estados e o que seriam as transições de estados? Indique uma estratégia de pesquisa em árvore adequada ao problema bem como, se necessário, uma heurística.
- (b) No pressuposto de utilizar pesquisa com propagação de restrições, que variáveis utilizaria, e quais os seus valores?
- (c) Qual das duas aproximações, pesquisa em árvore ou pesquisa com propagação de restrições, seria mais adequada?

## V Aprendizagem automática<sup>1</sup>

1. Sabe-se que a doença  $D$  ocorre em  $1/4$  de todos os casos. Sabe-se também que o sintoma  $S$  é observado em  $3/4$  dos pacientes que sofrem da doença  $D$  e em  $1/8$  da restante população. Suponha que, se está a iniciar a construção de uma árvore de decisão com o objectivo de diagnosticar a referida doença. Temos portanto duas classes,  $D$  e  $\neg D$ , e  $S$  é um dos atributos.

Nas alíneas seguintes, pode apresentar cálculos aproximados, desde que indique as simplificações que efectuou.

- (a) Qual o ganho de informação associado ao atributo  $S$ ?
  - (b) Qual a razão do ganho associada ao atributo  $S$ ?
2. Nunca se testa um atributo mais do que uma vez ao longo de um caminho numa árvore de decisão. Porquê?
  3. Considere o seguinte conjunto de dados, com três atributos binários ( $A1$ ,  $A2$  e  $A3$ ) e uma saída também binária (duas classes).

Tabela 1: Dados de treino

A1	A2	A3	Classe
1	0	0	0
1	0	1	0
0	1	0	0
1	1	1	1
1	1	0	1

Use o algoritmo TDIDT para gerar uma árvore de decisão com base nestes dados. A escolha do atributo para cada nó deve basear-se no critério do ganho de informação.

4. Suponha que um conjunto de exemplos,  $E$ , é dividido em subconjuntos  $E_i$  de acordo com os valores de um atributo. Em cada  $E_i$  existem  $p_i$  exemplos positivos e  $n_i$  exemplos negativos. Mostre que o ganho de informação será positivo, excepto se  $p_i/(p_i + n_i)$  for igual para todos os atributos.

---

<sup>1</sup>Matéria não abrangida em Inteligência Artificial da Licenciatura em Engenharia Informática.