

EA072 – Roteiro de Estudos para a Prova 2 – 2s2019

Tópico 2 – Computação Evolutiva

(2.1) Proponha múltiplas estratégias de seleção para a escolha de indivíduos que irão compor a próxima geração.

Algoritmo da roleta (roulette wheel): seleção proporcional ao grau de adaptação (fitness);

Torneio de p jogadores ($p \geq 2$): ajuda a preservar a diversidade para valores pequenos de p. Um aumento de p aumenta a pressão seletiva;

Rank: utiliza a roleta, mas adota apenas a ordem dos indivíduos, de acordo com o grau de adaptação, para definir a probabilidade de escolha, que é fixada a priori;

Seleção biclassista ou elitista: Na seleção bi-classista, são escolhidos os b% melhores indivíduos e os w% piores indivíduos da população. O restante $(100-(b+w))\%$ é selecionado aleatoriamente, com ou sem reposição.

(2.2) Explique como funciona a estratégia evolutiva, para busca em espaços contínuos. As estratégias evolutivas, foram utilizadas para otimizar parâmetros de valor real em sistemas dinâmicos.

Procedimento [P] = estrategia_evolutiva(,)

 inicialize μ indivíduos do tipo: $v_i = (x_i, \sigma_i, \theta_i)$, $i = 1, \dots$,

 avale os indivíduos

 t $\rightarrow 1$

 Enquanto condição_de_parada for FALSA faça,

 gere λ indivíduos (clonagem + mutação) a partir dos μ indivíduos

 avale os indivíduos gerados

 selecione os μ melhores indivíduos de λ ou $\mu + \lambda$

 Fim Enquanto

Fim Procedimento

- Sugere-se o emprego da seguinte relação: $1 < \mu < \lambda$
- Repare que cada indivíduo contém instruções sobre como aplicar a mutação aos seus descendentes e estes devem herdar estas instruções, com alguma variabilidade. A mutação emprega uma distribuição normal multivariada

(2.3) Explique como funciona o algoritmo genético, para busca em espaços binários.

GERAÇÃO ATUAL \Rightarrow Seleção de pares via Roulette Wheel \Rightarrow Determinação dos pontos de Crossover \Rightarrow Crossover \Rightarrow Mutação \Rightarrow PRÓXIMA GERAÇÃO.

(2.4) Como se dá a evolução de regras em sistemas classificadores?

Sistemas Classificadores representam metodologias para criação e atualização evolutiva de regras (denominadas classificadores) em um sistema de tomada de decisão.

Dadas as características de um ambiente em um determinado instante e levando-se em conta a “energia” de cada classificador (regra), alguns classificadores podem ser ativados. Eles codificam alternativas de ações específicas, as quais são submetidas a um processo de competição para selecionar aquela que será executada.

Dependendo do efeito de cada ação (ou sequência de ações) no atendimento dos objetivos (os quais podem estar implícitos), os classificadores responsáveis pelas ações serão recompensados ou punidos (ganhando ou perdendo “energia”).

Periodicamente, o elenco de classificadores é submetido a um processo evolutivo, que toma basicamente como medida de fitness a “energia” dos classificadores.

(2.5) Uma vez formalizado o espaço de busca, calcule a sua cardinalidade, ou seja, o número de soluções candidatas do espaço de busca.

A cardinalidade de um conjunto contável é dada pelo número de elementos do conjunto (? a pergunta é só essa mesmo?)

(2.6) Aprenda a aplicar a Tabela de Contagem de Coleções.

Aprender a utilizar essa tabela para determinar o número de coleções de **k** elementos é possível formar num universo de tamanho **n**.

- Tamanho do universo: **n**
- Tamanho da coleção: **k**
- Quantas coleções de **k** elementos existem, sabendo que existem **n** elementos candidatos?

	Repetição permitida	Repetição proibida
Ordenado	n^k	$(n)_k = \frac{n!}{(n-k)!}$
Não-ordenado	$\left(\binom{n}{k} \right) = \binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

(2.7) O que é um problema de otimização multimodal e o que é um problema de otimização dinâmica? Por que algoritmos populacionais de busca, com módulos eficazes para manutenção de diversidade na população, se mostram bastante competitivos para esses dois tipos de problemas?

Resolução: Um problema de otimização multimodal é aquele que apresenta uma função de otimização não-convexa, com múltiplos ótimos locais. Esses ótimos locais podem ou não apresentar os mesmos valores do critério de otimização. Já um problema de otimização dinâmica é aquele em que a função de otimização varia com o tempo, de modo que a(s) região(ões) do espaço de busca em que se encontra(m) o(s) ótimo(s) pode(m) variar com o tempo. Abordagens de busca populacionais com manutenção de diversidade geralmente apresentam mecanismos para localização e preservação de múltiplos ótimos.

assim como tendem a expressar uma melhor capacidade de seguimento de (ou reação a) ótimos que variam no tempo.

Tópico 3 – Sistemas Nebulosos

(3.1) Descreva o princípio da incompatibilidade, apresentado por Zadeh.

Slide 5 - À medida que a complexidade de um sistema aumenta, nossa habilidade de fazer afirmações precisas e que sejam significativas acerca deste sistema diminui até que um limiar é atingido, além do qual precisão e significância (ou relevância) tornam-se quase que características mutuamente exclusivas.

(3.2) O que é universo de discurso em lógica nebulosa?

Corresponde ao espaço onde estão definidos os elementos do conjunto.

Exemplos: ✓ Altura de seres humanos: $0 < \text{alt} < 2,5\text{m}$. Temperatura ambiente: $-70^\circ < \text{temp} < +70^\circ$

(3.3) O que é variável linguística e o que são termos linguísticos?

Entende-se por variável um identificador que pode assumir um dentre vários valores. Deste modo, uma variável linguística pode assumir um valor linguístico dentre vários em um conjunto de termos linguísticos. Exemplo, a variável poderia ser denominada *altura*, e assumir um dos seguintes valores: baixo, médio ou alto, elementos do conjunto $T=\{\text{baixo}, \text{médio}, \text{alto}\}$. Formalmente, uma variável linguística é caracterizada pela quintupla $\{X, T(X), U, G, M\}$, onde X é o nome do conjunto de termos (*altura*, no exemplo), U o universo de discurso, G uma gramática para gerar os termos $T(X)$, e M o significado dos termos linguísticos, representado através de conjuntos nebulosos.

<http://www.dainf.ct.utfpr.edu.br/~fabro/HomePage/pubantigo/cap0023.html>

(3.4) O que é conjunto nebuloso?

Slide 20 - Um conjunto nebuloso pode, então, ser definido como um conjunto ordenado de pares: $A = \{(x, \mu_A(x)) \mid x \in X\}$; A é o conj nebuloso; x é o elemento que pertence ao universo de discurso X , e $\mu_A(x)$ é a função de pertinência.

<http://www.dainf.ct.utfpr.edu.br/~fabro/HomePage/pubantigo/cap0023.html>

(3.5) O que é partição do universo de discurso?

A partição é separar/dividir seus elementos em torno de certo limiar a ser definido.

(3.6) Explique os conceitos de granularidade e de graduação em sistemas nebulosos.

A granularização, i.e. a especificação e distribuição dos termos lingüísticos define a partição nebulosa do universo correspondente. Uma partição nebulosa pode ser uniforme (com todos os termos iguais e eqüidistantes) ou não-uniforme. O número de termos lingüísticos define a granularidade das partições de cada universo. Um número pequeno de termos lingüísticos define uma partição esparsa ou grossa, ao passo que um número maior resulta numa partição fina.

(3.7) Procure se familiarizar com as formulações matemáticas das funções de pertinência triangular, trapezoidal, sigmoidal e gaussiana. O que é um singleton?

Slide 21 - Função Impulsiva

(3.8) Defina suporte e α -corte de uma função de pertinência.

Slide 27 - O suporte é formado pelos elementos do universo de discurso que apresentam pertinência não-nula ao conjunto nebuloso. O suporte é dito ser compacto se o seu tamanho é menor que o universo de discurso.

Slide 29 - O conjunto clássico A de elementos que pertencem ao conjunto nebuloso A até pelo menos o grau α $[0,1]$ é chamado de conjunto corte α (ver imagem no slide 29)

(3.9) Procure distinguir os conceitos de grau de pertinência e de probabilidade.

O grau de pertinência do elemento x ao conjunto A , isto é, o quanto é possível para o elemento x pertencer ao conjunto A . Na teoria de probabilidades, temos o evento muito bem definido e a dúvida paira sobre a ocorrência do evento.

(3.10) Dê um exemplo de uma t-norma e de uma s-norma.

t-norma -> intersecção: $x_t y = xy$ (Produto Algébrico).

s-norma -> união: $x_s y = x + y - xy$ (Soma Probabilística).

(3.11) Uma vez determinados os valores das variáveis de entrada e dado um sistema composto por regras nebulosas, aplique o método de Mamdani e o centro de gravidade para obter a saída defuzificada.

Slides 51, 55.

<http://www.inf.ufrgs.br/~engel/data/media/file/RNSF/LF.pdf> Slides 22 - 25.

(3.12) O que é um sistema nebuloso Takagi-Sugeno?

No modelo de Takagi-Sugeno, as regras são da forma: Se x_1 é A_{1i} , e x_2 é A_{2i} e . . . e x_n é A_{ni} , ENTÃO $y = f_i(x_1, x_2, \dots, x_n)$, em que $A_{1i}, A_{2i}, \dots, A_{ni}$ são conjuntos fuzzy dos antecedentes enquanto que o consequente é uma função das variáveis de entrada.

<http://www.inf.ufrgs.br/~engel/data/media/file/RNSF/LF.pdf> Slide 26

(3.13) No projeto de um controle nebuloso, dadas as variáveis de entrada e suas partições, preencha a tabela que indica os consequentes das regras nebulosas.

<http://www.inf.ufrgs.br/~engel/data/media/file/RNSF/LF.pdf> Slide 11 - 14

(3.14) Dadas as partições das variáveis a seguir, na condição de um especialista, sua tarefa é propor um conjunto de regras que formará a base de conhecimento de um robô que precisa navegar por um ambiente desconhecido, com velocidade constante e sem sofrer colisão com obstáculos.

Variáveis de entrada (leitura sensorial):

Sensor 30o à esquerda (SE): perto (P), médio (M), longe (L)

Sensor 30o à direita (SD): perto (P), médio (M), longe (L)

Variável de saída: Girar: Muito à esquerda (ME), À esquerda (E), Não girar (NG), À direita (D), Muito à direita (MD)

Resolução:

SD\SE	P	M	L
P	?	ME	ME
M	MD	E ou NG ou D	E ou NG
L	MD	D ou NG	NG

Tópico 4 – Outras Técnicas de Computação Natural

(4.1) Explique como um algoritmo baseado em colônia de formigas e uso de feromônio pode ser empregado na solução do problema do caixeiro viajante.

Dado um grafo com n vértices, colocar uma formiga artificial em cada um deles;

Cada formiga traça um caminho seguindo uma fórmula probabilística em função do feromônio presente em cada aresta do grafo;
Como os caminhos são gerados seguindo uma fórmula probabilística, cada formiga pode obter caminhos diferentes das demais;
Após a construção de todos os caminhos, a intensidade de feromônio em cada aresta é acrescida proporcionalmente à qualidade da solução gerada;

(4.2) Como você adaptaria esta proposta visando resolver outros tipos de problemas combinatórios? Observação: Para tanto, é fundamental definir quais são as decisões que cada formiga deve tomar ao construir uma solução candidata e onde deve ser depositado o feromônio, visando guiar as decisões das formigas.

(4.3) Em um problema de otimização combinatória usando colônia de formigas, cada formiga, em uma dada iteração, deve se mover para um dentre um conjunto finito de posições alternativas. De forma genérica, quais são os dois fatores que influenciam a tomada de decisão da formiga?

Resolução: Um dos fatores é a quantidade relativa de feromônio no caminho até a nova posição (ou na própria posição), de modo que posições associadas com uma maior concentração de feromônio terão uma chance proporcionalmente maior de serem escolhidas. O outro fator é um termo heurístico associado a especificidades do problema de otimização. Por exemplo, se uma posição alternativa produz o menor acréscimo junto à função-custo (supondo minimização), quando comparada com as demais posições alternativas, ela terá proporcionalmente maior chance de ser escolhida. Esses dois fatores geralmente são ponderados e, em seguida, multiplicados, visando considerá-los simultaneamente na tomada de decisão por parte de cada formiga.

(4.4) Em um problema de otimização por enxame de partículas (PSO, do inglês particle swarm optimization), quais são as três direções que contribuem para a definição da nova posição de cada partícula?

Resolução: (1) Direção adotada na última atualização de posição da própria partícula; (2) Direção que aponta da posição atual da partícula até a melhor posição dela, desde o início da busca; (3) Direção que aponta da posição atual da partícula até a melhor posição encontrada até o momento por todas as partículas vizinhas. A ponderação dessas três direções geralmente é feita de forma aleatória, a cada iteração.

(4.5) A resposta à questão 4.4 envolve o conceito de vizinhança entre as partículas. Quais são os principais tipos de vizinhança adotados em PSO? Qual é o efeito predominante de cada tipo de vizinhança no comportamento da busca?

Gbest – conecta todas as partículas entre si; (G = global);

Lbest – vizinhança composta pelas k partículas mais próximas (L = local);

(4.6) Explique como operam os princípios de seleção clonal e maturação de afinidade em sistemas imunológicos artificiais.

Quando um anticorpo possui alta afinidade a um dado antígeno, as células B que o produzem se multiplicam com maior velocidade através de clonagem;

Durante esse processo, os novos clones gerados sofrem hipermutação, com taxas de variabilidade inversamente proporcionais à sua afinidade ao antígeno em questão;

Dentre as novas células geradas, as que possuem maior afinidade com o antígeno são selecionadas e as demais suprimidas;

Este processo de expansão clonal, hipermutação e seleção das células com receptores mais adaptados é denominado Seleção Clonal

A maturação da afinidade é o processo pelo qual a afinidade dos anticorpos produzidos em resposta a um antígeno protéico aumenta com a exposição prolongada ou repetida a esse antígeno.

(4.7) Em que um algoritmo de otimização multi-modal difere de um algoritmo de otimização que busca encontrar uma única solução ótima para o problema?

(4.8) Como ocorre com o sistema nervoso, defende-se que o sistema imunológico também executa cognição, no sentido de sensoriar o seu ambiente, reagir a estímulos internos e externos, executar reconhecimento de padrões e implementar memória (por exemplo, no caso do princípio de imunização pela vacinação). Quais são as principais semelhanças e diferenças entre os componentes e a forma de operar de ambos os sistemas (nervoso e imunológico)?

(4.9) Observando a natureza, é possível constatar que existem soluções sub-ótimas ou até bem pouco otimizadas em vários contextos. Mas também existem soluções surpreendentemente eficientes e inspiradoras. Ao longo do curso, foi dada ênfase a inspirações voltadas para heurísticas e meta-heurísticas de busca e otimização. Apresente outras tecnologias inspiradas na natureza.

Resolução: Em todo lugar nos deparamos com tecnologias inspiradas diretamente na natureza, mesmo que não nos demos conta disso. Como exemplos temos velcro (inspirado em sementes de plantas que aderem ao corpo de animais, como o carrapicho), adesivos (inspirados nas patas de lagartixas), antenas e radares (inspiradas em insetos e morcegos) e nanoestrutura de superfícies, seja para reter ou dissipar água, seja para impedir a aderência de corpos indesejados (inspirada na superfície de folhas e de animais marinhos). Veja esses e outros casos em:

<http://www.bloomberg.com/news/photo-essays/2015-02-23/14-smart-inventions-inspired-by-nature-biomimicry>.

Tópico 5 – Lógica Matemática, Representação e Inferência

(5.1) O que é uma proposição?

Uma proposição (ou sentença declarativa, ou fórmula) é uma afirmação acerca de um fato básico, sendo verdadeira ou falsa, mas nunca ambos.

(5.2) O que é uma fórmula bem-formada?

Para ilustrar essa ideia, geralmente emprega-se um exemplo famoso devido a Chomsky, usando linguagem natural: “Ideias verdes descoloridas descansam furiosamente”. Trata-se de uma sentença sintaticamente bem-formada, mas semanticamente sem sentido algum. Substantivo, verbo, adjetivos e advérbio estão corretamente posicionados. Nada além disso. É por isso que uma proposição ou sentença declarativa é também chamada de ‘fórmula bem-formada’, pois deriva da gramática formal que estabelece a linguagem da lógica proposicional.

(5.3) O que é uma tautologia?

Uma proposição que é sempre verdade é uma tautologia.

(5.4) O que é uma contradição?

Uma proposição que é sempre falsa é uma contradição.

(5.5) O que é um problema de satisfação?

Um problema de satisfação em lógica proposicional é encontrar um conjunto de atribuições de valores-verdade para as variáveis que compõem uma proposição composta, de modo que esta proposição composta seja verdadeira. Isso só vai ser possível, obviamente, se a proposição composta não for uma contradição.

(5.6) O que é vinculação de proposições?

Se toda interpretação que satisfaz o conteúdo de uma base de conhecimento também satisfaz uma dada proposição, então diz-se que a base de conhecimento vincula (entails) aquela proposição, ou seja, a proposição é dedutível a partir da base de conhecimento

(5.7) Dada a regra da resolução:

$$\frac{\frac{a \vee b}{\neg b \vee r}}{a \vee r}$$

Regra da Resolução

e sua variante quando r é sempre verdadeiro:

$$\frac{\frac{a \vee b}{\neg b}}{a}$$

Varição da Regra da Resolução

o que é necessário para se realizar uma prova por refutação de uma determinada proposição, dada a base de conhecimento?

Resolução: Todas as proposições logicamente consistentes podem ser obtidas aplicando-se a regra de resolução. É por isso que se diz que a resolução é logicamente completa.

Para provar que uma proposição é verdadeira, adiciona-se o negativo dessa proposição à base de conhecimento (conjunto de proposições sabidamente verdadeiras) e busca-se chegar a uma contradição. Antes de manipular a base de conhecimento, é necessário que todas as proposições sejam convertidas para a forma normal disjuntiva.

(5.8) Em que a lógica de primeira ordem difere da lógica proposicional, em termos do valor-verdade das proposições?

A lógica proposicional simplesmente mapeia o valor-verdade de uma proposição em verdadeiro ou falso. Sendo assim, não há variáveis envolvidas, no sentido de condicionarem o valor-verdade da proposição dependendo da atribuição feita à variável. Em outras palavras, não há formas de manipular o conteúdo de uma proposição de modo a modificar seu valor-verdade.

(5.9) Exemplo de prova por inferência em lógica de primeira ordem. Não será cobrado em prova, mas pode ajudar a compreender melhor o conceito de lógica de predicados.

The law says that it is a crime for an American to sell weapons to hostile nations. The country Nono, an enemy of America, has some missiles, and all of its missiles were sold to it by Colonel West, who is an American.

Prove que Colonel West é um criminoso.

Resolução:

Representação em lógica de primeira ordem:

$\forall x,y,z \text{ American}(x) \wedge \text{Weapon}(y) \wedge \text{Nation}(z) \wedge \text{Hostile}(z) \wedge \text{Sells}(x,y,z) \Rightarrow \text{Criminal}(x)$	[1]
$\exists x \text{ Owns}(\text{Nono},x) \wedge \text{Missile}(x)$	[2]
$\forall x \text{ Owns}(\text{Nono},x) \wedge \text{Missile}(x) \Rightarrow \text{Sells}(\text{West}, \text{Nono},x)$	[3]
$\forall x \text{ Missile}(x) \Rightarrow \text{Weapon}(x)$	[4]
$\forall x \text{ Enemy}(x,\text{America}) \Rightarrow \text{Hostile}(x)$	[5]
$\text{American}(\text{West})$	[6]
$\text{Nation}(\text{Nono})$	[7]
$\text{Enemy}(\text{Nono}, \text{America})$	[8]
$\text{Nation}(\text{America})$	[9]

Sequência de Passos da Prova:

1- De [2] e eliminação do \exists : $\text{Owns}(\text{Nono},M1) \wedge \text{Missile}(M1)$	[10]
2- De [10] e And-Elimination: $\text{Owns}(\text{Nono},M1)$	[11]
$\text{Missile}(M1)$	[12]
3- De [4] e eliminação do \forall : $\text{Missile}(M1) \Rightarrow \text{Weapon}(M1)$	[13]
4- De [12], [13] e Modus Ponens: $\text{Weapon}(M1)$	[14]
5- De [3] e eliminação do \forall : $\text{Owns}(\text{Nono},M1) \wedge \text{Missile}(M1) \Rightarrow \text{Sells}(\text{West}, \text{Nono}, M1)$	[15]
6- De [15], [10] e Modus Ponens: $\text{Sells}(\text{West}, \text{Nono}, M1)$	[16]
7- De [1] e eliminação do \forall (3 \times): $\text{American}(\text{West}) \wedge \text{Weapon}(M1) \wedge \text{Nation}(\text{Nono})$ $\wedge \text{Hostile}(\text{Nono}) \wedge \text{Sells}(\text{West},\text{Nono},M1) \Rightarrow \text{Criminal}(\text{West})$	[17]
8- De [5] e eliminação do \forall : $\text{Enemy}(\text{Nono}, \text{America}) \Rightarrow \text{Hostile}(\text{Nono})$	[18]
9- De [8], [18] e Modus Ponens: $\text{Hostile}(\text{Nono})$	[19]
10- De [6,7,14,16,19] e And-Introduction: $\text{American}(\text{West}) \wedge \text{Weapon}(M1) \wedge \text{Nation}(\text{Nono})$ $\wedge \text{Hostile}(\text{Nono}) \wedge \text{Sells}(\text{West}, \text{Nono}, M1)$	[20]
11- De [17], [20] e Modus Ponens: $\text{Criminal}(\text{West})$ \square	[21]

(5.10) Dê exemplos de como expressar conhecimento em lógica de predicados.

Resolução:

- (1) Todos os cisnes são brancos. $\forall X [\text{Cisne}(X) \rightarrow \text{Branco}(X)]$
- (2) Algum cisne é branco. $\exists X [\text{Cisne}(X) \wedge \text{Branco}(X)]$
- (3) Nem todos os cisnes são brancos. $\neg \forall X [\text{Cisne}(X) \rightarrow \text{Branco}(X)]$
- Algum cisne é não-branco. $\exists X [\text{Cisne}(X) \wedge \neg \text{Branco}(X)]$
- (4) Nenhum cisne é branco. $\neg \exists X [\text{Cisne}(X) \wedge \text{Branco}(X)]$
- Todos os cisnes são não-brancos. $\forall X [\text{Cisne}(X) \rightarrow \neg \text{Branco}(X)]$
- (5) Apenas cisnes são brancos. $\forall X [\text{Branco}(X) \rightarrow \text{Cisne}(X)]$

Todas as coisas brancas são cisnes.

(6) Todos e apenas os cisnes são brancos. $X [Cisne(X) \supset Branco(X)]$

(7) Marujo é um cisne branco. $Cisne(Marujo) \supset Branco(Marujo)$

Repare que, como já informado nas notas de aula, o quantificador universal tem como principal conectivo a implicação, enquanto que o quantificador existencial tem como principal conectivo a conjunção.

(5.11) Dadas as seguintes proposições:

(1) Todos os participantes da competição são brasileiros.

(2) Todos os hóspedes do hotel são brasileiros.

(3) Todos os hóspedes do hotel são participantes da competição.

(4) Todos os participantes da competição são hóspedes do hotel.

indique o que é indução, dedução e abdução nas inferências abaixo:

(a) Se (1) e (2) então (3). ABDUÇÃO ou INDUÇÃO

(b) Se (1) e (3) então (2). DEDUÇÃO

(c) Se (1) e (4) então (2). ABDUÇÃO ou INDUÇÃO

(d) Se (2) e (3) então (1). ABDUÇÃO ou INDUÇÃO

(e) Se (2) e (4) então (1). DEDUÇÃO

Observação: Muitas vezes, é possível distinguir claramente abdução de indução (como nos exemplos apresentados nas notas de aula), mas a presença do quantificador universal em todas as proposições acima leva a uma fronteira não muito clara, pois vai depender da forma como a inferência será conduzida. Já a distinção entre dedução e as outras duas formas de inferência (indução e abdução) tende a ser sempre evidente.

(5.12) Por que, para a inferência abductiva, não basta propor uma hipótese comparativamente melhor que outras hipóteses?

Resolução: Uma proposição abductiva tem que ser forte em algum sentido absoluto. Por exemplo, quando se refere a uma hipótese que ocorre com alta probabilidade em situações parecidas com aquela em questão.

(5.13) Explique por que a ciência está fundamentada na refutabilidade de suas teorias.

Segundo a concepção tradicional de ciência, o ponto de partida do cientista está nos dados empíricos, observáveis. Esses dados da observação, acumulados, transformam-se em hipóteses que, uma vez verificadas, tornam-se leis científicas. O procedimento lógico que norteia a ciência é a indução, que parte de dados singulares para chegar ao universal. O critério de demarcação usado para definir o que é ou não ciência é a verificação. O problema da indução consiste no fato de que, por mais dados singulares que o cientista acumule, não há uma garantia lógica de que o enunciado universal daí inferido seja verdadeiro

(5.14) O que se pode dizer sobre um teoria aceita no meio científico.

Resolução:

- Ela é aceita por ser aquela que melhor resiste aos testes de refutação já propostos;
- Quanto mais testes de refutação são vencidos pela teoria, maior confiança se tem em

seu poder de explicar a verdade;

- Ela pode deixar de ser aceita, caso o seu poder de explicar a verdade seja superado por uma teoria alternativa.

(5.15) O que é uma falácia lógica?

Uma falácia é um argumento que não apresenta consistência lógica, ou seja, em que a sua conclusão não é sustentada logicamente pelas premissas. São, portanto, raciocínios inválidos, mal-estruturados e incoerentes, mas que podem ser tomados como verdadeiros, explorando a “ingenuidade” do interlocutor e empregando técnicas de persuasão.

(5.16) Supondo que os argumentos ou as premissas sejam falaciosos, o que se pode concluir sobre o valor-verdade da conclusão?

Se os argumentos ou as premissas forem falaciosos, nada se pode concluir sobre o valor-verdade da conclusão, nem que ela é verdadeira, nem que ela é falsa.

Tópico 6 – Sistemas Baseados em Regras e Árvores de Decisão

(6.1) Apresente os módulos constituintes e explique como opera um sistema especialista baseado em regras.

Componentes de um sistema especialista:

Conjunto de regras: compõe a base de conhecimento e é utilizada pela máquina de inferência para processar os dados de entrada.

Máquina de inferência: Infere conclusões a partir de fatos (dados de entrada) e da base de conhecimento.

Esses componentes formam o sistema de produção. Neste modelo, o conhecimento é separado do seu processamento.

O modelo do sistema de produção é baseado na concepção de que o ser humano resolve problemas pela aplicação de seu conhecimento (expresso na forma de regras de produção), dado um cenário que descreve uma situação ou estado de coisas (expresso na forma de fatos ou informações específicas sobre o problema)

(6.2) O que significa uma abordagem top-down para o projeto de um sistema baseado em regras?

Top-down: obtenção do modelo de classificação a partir de informações fornecidas por especialistas;

(6.3) O que significa uma abordagem bottom-up para o projeto de um sistema baseado em regras?

A construção dos modelos computacionais de classificação geralmente emprega um dentre dois paradigmas alternativos:

Top-down: obtenção do modelo de classificação a partir de informações fornecidas por especialistas;

Bottom-up: obtenção do modelo de classificação pela identificação de relacionamentos entre variáveis dependentes e independentes em bases de dados rotuladas. O classificador é induzido por mecanismos de generalização fundamentados em exemplos específicos (conjunto finito de objetos rotulados). Existem propostas também para dados não-rotulados.

(6.4) Partindo da seguinte base de conhecimento: e sabendo que a proposição I é verdade, mostre que a proposição M é verdade empregando encadeamento direto.

Resolução:

Ciclo	WM	Regras Ativas	Regra Disparada
0	I	6	6
1	I, v, r, q	6,5	5
2	I, v, r, q, s	6,5,2	2
3	I, v, r, q, s, p	6,5,2,1	1
4	I, v, r, q, s, p, M	6,5,2,1	parar

(6.5) Entenda como se aplica o encadeamento reverso.

(6.6) Quais são as principais vantagens dos sistemas especialistas? E as suas principais desvantagens?

Vantagens • Concebidos para emular a estratégia de tomada de decisão de especialistas humanos; • Tendem a ser muito precisos em suas conclusões; • Tendem a produzir uma conclusão em um tempo reduzido, embora testes exaustivos no caso de bases de conhecimento compostas por muitas regras possam ser custosos; • Apresentam uma alta capacidade de explanação do processo de inferência; • Apresentam uma estrutura uniforme, em que cada regra é um pedaço independente do conhecimento disponível; • Ao separar o conhecimento do seu processamento, permitem que um mesmo ambiente de projeto de sistema especialista possa ser empregado em diferentes aplicações; • Quando empregam regras nebulosas, são capazes de realizar um processo de inferência logicamente preciso a partir de um conhecimento inicial impreciso

Desvantagens • Atuam em um domínio restrito; • O conhecimento especialista pode não ser de fácil acesso; • Trabalham predominantemente com raciocínio simbólico apenas; • Recorrem predominantemente a processos de inferência dedutiva, em detrimento das inferências indutivas e abduativas; • Promovem pouco relacionamento entre as regras, de modo que há poucos subsídios para se definir o papel de regras individuais num processo de inferência; • O processo de inferência pode não ser sempre confiável, possivelmente motivado pela dificuldade de lidar com ambiguidades, pela presença de regras conflitantes na base de conhecimento e pela dificuldade de promover atualizações (regras podem ficar ultrapassadas / obsoletas) ou aprendizado em sua base de conhecimento;

(6.7) O que é um flat file?

Toda informação sobre cada objeto (caso) a ser classificado deve poder ser expressa em termos de uma coleção fixa de propriedades ou atributos. Dessa forma, objetos distintos não podem requerer coleções distintas de atributos. Bases de dados que atendem a este requisito são denominadas flat files.

(6.8) Quais são os três tipos de atributos que podem ser utilizados na descrição de objetos?

Contínuos: assumem valores numéricos em intervalos no eixo dos números reais;

Catagóricos ordinais: assumem um conjunto finito de valores, que podem ser ordenados;

Catagóricos não-ordinais: assumem um conjunto finito de valores que não podem ser ordenados.

(6.9) Explique como funciona o algoritmo de indução de árvores de decisão denominado Top-Down Induction of Decision Tree (TDIDT).

O TDIDT produz regras de decisão de forma implícita numa árvore de decisão, a qual é construída por sucessivas divisões dos exemplos de acordo com os valores de seus atributos preditivos. De acordo com BRAMER (2007), esse processo é conhecido como particionamento recursivo.

O esqueleto do algoritmo de TDIDT é baseado em três possibilidades sobre um conjunto de treinamento T contendo classes C_1, C_2, \dots, C_k :

1. T contém um ou mais objetos, sendo todos da classe C_j . Assim, a árvore de decisão para T é um nó folha que identifica a classe C_j .
2. T não contém objetos. A árvore de decisão também é um nó folha, mas a classe associada deve ser determinada por uma informação externa. Por exemplo, pode-se utilizar o conhecimento do domínio do problema.
3. T contém exemplos pertencentes a mais de uma classe. Neste caso, a ideia é dividir T em sub-conjuntos que são, ou tendem a se dirigir para, coleções de exemplos com classes únicas. Para isso, é escolhido um atributo preditivo A que possui um ou mais possíveis resultados O_1, O_2, \dots, O_n . T é particionado em sub-conjuntos T_1, T_2, \dots, T_n , onde T_i contém todos os exemplos de T que têm resultado O_i para o atributo A . A árvore de decisão para T consiste de um nó de decisão identificando o teste sobre o atributo A , e uma aresta para cada possível resultado, ou seja, n arestas. No lugar de um único atributo A , pode também ser considerado um subconjunto de atributos.

O mesmo algoritmo de indução de árvores de decisão (passos 1, 2 e 3) é aplicado recursivamente para cada sub-conjunto de exemplos T_i , com i variando de 1 até n .

Basicamente, o algoritmo TDIDT é um algoritmo recursivo de busca gananciosa (greedy) que procura, sobre um conjunto de atributos, aquele(s) que “melhor” dividem o conjunto de exemplos em sub-conjuntos. Inicialmente, todos os exemplos são colocados em um único nó, chamado de raiz.

A seguir, um atributo preditivo é escolhido para o teste desse nó e, assim, dividir os exemplos em sub-conjuntos. Esse processo se repete recursivamente até que todos os exemplos estejam classificados, sem erro, ou então até que todos os atributos preditivos já tenham sido testados ou não exibam mais poder discriminativo. O mesmo atributo preditivo pode ser utilizado múltiplas vezes, se ainda houver múltiplos valores desse atributo no nó, permitindo estabelecer novas partições.

(6.10) Por que não se recorre, na prática, a uma busca exaustiva pela melhor topologia de árvore de decisão, explorando o espaço formado por todas as topologias candidatas?

Há muitas maneiras de uma árvore de decisão ser estruturada a partir de um conjunto de atributos. De forma exaustiva, o número de árvores de decisão atingíveis cresce fatorialmente à medida que o número de atributos aumenta. Logo, torna-se impraticável definir a estrutura da árvore de decisão ótima para um determinado problema, devido ao elevado custo computacional da busca

(6.11) O que é ganho de informação no processo de escolha de um atributo para um nó de uma árvore de decisão?

Uma das medidas baseadas em impureza é o Ganho de Informação, o qual usa a entropia como medida de impureza. Para determinar o quão boa é uma condição de teste realizada, é necessário comparar o grau de entropia do nó-pai (antes da divisão) com o grau de entropia dos nós-filhos (após a divisão). O atributo que gerar uma maior diferença é escolhido como condição de teste

(6.12) O que é razão de ganho no processo de escolha de um atributo para um nó de uma árvore de decisão?

o ganho de informação relativo (ponderado) como critério de avaliação.

Em QUINLAN (1988), é sugerido que a razão de ganho seja realizada em duas etapas.

Na primeira etapa, é calculado o ganho de informação para todos os atributos. Após isso, considera-se apenas aqueles atributos que obtiveram um ganho de informação acima da média, e então se escolhe aquele que apresentar a melhor razão de ganho.

Dessa forma, Quinlan mostrou que a razão de ganho supera o ganho de informação tanto em termos de acurácia quanto em termos de complexidade das árvores de decisão geradas.

(6.13) Como a árvore de decisão trabalha no caso de atributos categóricos não-ordinais?

Um ramo para cada valor de atributo: É a partição mais comum, na qual é criada uma aresta para cada valor do atributo usado como condição de teste. Embora esse tipo de partição permita extrair do atributo todo o seu conteúdo informativo, possui a desvantagem de tornar a árvore de decisão mais complexa

Solução de Hunt: A partição utilizada pelo algoritmo ID3 sugere uma partição binária. Nesse caso, um dos valores é atribuído a uma das arestas e todos os outros valores à outra aresta. A desvantagem desse tipo de partição é não aproveitar todo o poder de discriminação do atributo em questão.

Agrupamento de valores em dois conjuntos: De acordo com BREIMAN et al. (1984), a divisão binária também pode ser realizada de uma forma mais complexa, onde cada um dos dois subconjuntos pode ser formado por registros com mais de um valor para o atributo utilizado como condição de teste. O grande desafio desse tipo de partição é o elevado custo computacional para encontrar a melhor divisão, pois o número de combinações possíveis é 2^{n-1} , onde n é o número de valores possíveis para o atributo em questão.

Agrupamento de valores em vários conjuntos: Visando permitir o agrupamento de valores em vários conjuntos com uma complexidade de cálculo razoável, o algoritmo C4.5 (QUINLAN, 1993) permite encontrar uma solução de boa qualidade. Para isso, inicia criando uma aresta para cada valor do atributo em questão. Após, são testadas todas as combinações possíveis de dois valores e, caso nenhuma dessas combinações produza um ganho maior que a divisão anterior, o processo é interrompido e a divisão anterior é adotada como divisão final. Senão, é repetido o processo tendo como base a melhor das soluções anteriores. Nota-se que não se pode garantir que a divisão encontrada seja a melhor possível, pois é verificado se houve melhoria apenas um passo à frente.

(6.14) Como a árvore de decisão trabalha no caso de atributos numéricos e/ou categóricos ordinais?

Atributos categóricos ordinais: Como já definido, um atributo é ordinal quando há uma relação de ordem entre os seus possíveis valores. Por exemplo, tem-se um atributo altura que pode possuir os valores baixa, média e alta. Com atributos desse tipo, é possível realizar uma partição binária do tipo altura < média, em que todos os exemplos cujo atributo altura tem valor baixa seguem por uma aresta e os outros seguem por outra aresta.

(6.15) O que se pode dizer acerca de atributos que são escolhidos para compor os nós de decisão mais próximos da raiz da árvore de decisão?

Os critérios de seleção para a melhor divisão são baseados em diferentes medidas, tais como impureza, distância e dependência. A maior parte dos algoritmos de indução busca dividir os dados de um nó-pai de forma a minimizar o grau de impureza dos nós filhos. Quanto menor o grau de impureza, mais desbalanceada é a distribuição de classes. Em um determinado nó, a impureza é nula se todos os exemplos nele pertencerem à mesma classe. Analogamente, o grau de impureza é máximo no nó se houver o mesmo número de exemplos para cada classe possível.

(6.16) Qual é a utilidade da poda em árvores de decisão?

Quando árvores de decisão são construídas, muitas das arestas ou sub-árvores podem refletir ruídos ou erros. Isso acarreta em um problema conhecido como sobreajuste, que significa um aprendizado muito específico do conjunto de treinamento, não permitindo ao modelo generalizar. Para detectar e excluir essas arestas e sub-árvores, são utilizados métodos de poda (pruning) da árvore, cujo objetivo é melhorar a taxa de acerto do modelo para novos exemplos, os quais não foram utilizados no conjunto de treinamento (HAN, 2001). Consequentemente, a árvore podada se torna mais simples, facilitando a sua interpretabilidade por parte do usuário. Junto ao método de seleção, o método de poda também varia de acordo com os diferentes algoritmos de indução de árvores de decisão

(6.17) Quais são as principais diferenças entre C4.5 e CART?

(6.18) O que são árvores de decisão multivariadas?

avalia o comportamento de $q > 1$ variáveis com base nos resultados de um conjunto de covariáveis. A análise e a interpretação conjunta de modelos univariados para cada variável resposta é inviável à medida que o número de análises aumenta e, sobretudo, possíveis correlações entre tais variáveis são ignoradas.

(6.19) O que é uma random forest?

A técnica random forest foi concebida por BREIMAN (2001) e corresponde a uma combinação de árvores de decisão simples, sendo que cada árvore de decisão é sintetizada a partir de um vetor aleatório amostrado independentemente e segundo uma mesma distribuição de

probabilidade. Esse vetor aleatório vai guiar alguma etapa do processo de síntese da árvore a partir de dados.

(6.20) O que é bagging?

Bagging é a aplicação do procedimento Bootstrap a um algoritmo de aprendizado de máquina de alta variação, geralmente árvores de decisão.

Amostragem com reposição

(6.21) O que é um comitê de máquinas (ensemble)?

comitê (ensemble) de múltiplos modelos simples que lidam com subconjuntos pequenos e aleatórios dos dados disponíveis, em lugar de um único modelo mais complexo que considera todo o conjunto de dados.

(6.22) Quais são as principais distinções entre random forest e CART?

(6.23) Apresente as motivações que sustentam o melhor desempenho de random forest sobre CART.

(6.24) Demonstre que os dados oob (out of bag) representam aproximadamente um terço das N amostras disponíveis para treinamento.

Resolução:

- Visto que *bagging* equivale a um processo de amostragem com reposição, então a chance de um dado não ser escolhido a cada amostragem é:

$$\frac{N-1}{N}.$$

- Logo, a chance de um dado não ser escolhido ao final da coleta de N amostras com reposição (compondo então o conjunto oob) é:

$$\left(\frac{N-1}{N}\right)^N.$$

- Quando N cresce, esta probabilidade converge para:

$$\lim_{N \rightarrow \infty} \left(\frac{N-1}{N}\right)^N = \lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} \cong 0,368.$$

(6.25) Apresente uma técnica capaz de estimar a relevância de atributos em random forest.

Variable Importance

Variable importance plots can be constructed for random forests in exactly the same way as they were for gradient-boosted models. At each split in each tree, the improvement in the split-criterion is the importance measure attributed to the splitting variable, and is accumulated over all the trees in the forest separately for each variable. Boosting ignores some variables

completely, while the random forest does not. The candidate split-variable selection increases the chance that any single variable gets included in a random forest, while no such selection occurs with boosting. Random forests also use the oob samples to construct a different variable-importance measure, apparently to measure the prediction strength of each variable. When the b th tree is grown, the oob samples are passed down the tree, and the prediction accuracy is recorded. Then the values for the j th variable are randomly permuted in the oob samples, and the accuracy is again computed. The decrease in accuracy as a result of this permuting is averaged over all trees, and is used as a measure of the importance of variable j in the random forest.. Although the rankings of the two methods are similar, the importances in the right plot are more uniform over the variables. The randomization effectively voids the effect of a variable, much like setting a coefficient to zero in a linear model. This does not measure the effect on prediction were this variable not available, because if the model was refitted without the variable, other variables could be used as surrogates.

Tópico 7 – Representação do Conhecimento e Solução de Problemas

(7.1) Dado um problema a ser resolvido empregando alguma técnica de inteligência artificial, que tipo de representação você adotaria para as soluções candidatas? Existe algum conhecimento a priori que pode ser empregado para simplificar o problema?

Espaço de atributos (feature space); Feature selection & Variable selection; Vetor de atributos binários; Vetor de atributos inteiros; Vetor de atributos em ponto flutuante; Vetor de atributos com representação mista; Matriz de atributos; Árvore, grafo, outras estruturas mais complexas; Isomorfismo; A representação adotada influi diretamente na definição do espaço de busca. Sua adequação depende de diversos fatores, incluindo interpretabilidade, parcimônia e facilidade de manipulação.

(7.2) Em várias situações ao longo do curso de EA072, foi necessária a adoção de modelos simplificados (aproximados) para processos que se pretendia reproduzir em computador, visando tratabilidade. Apresente exemplos disso.

Sistema baseado em regras. Árvore de busca; Máquina de estados finita; Rede semântica;

(7.3) O que é seleção de variáveis (variable selection)?

(7.4) O que é seleção de atributos (feature selection)?

(7.5) Por que se diz que o desempenho superior de deep learning, em variadas aplicações de classificação de padrões, se deve à capacidade de definição automática de um espaço de atributos com alto poder discriminatório, a cada camada da rede neural, também conhecido como aprendizado da representação?

(7.6) (Não é questão candidata para a Prova 2) Leio o livro de Polya e, em seguida, leia

o livro de Michalewicz & Fogel. Procure então discorrer acerca de como mudaram as estratégias de solução de problemas antes e depois do advento de computadores de alto desempenho.

Tópico 8 – Estruturas e Estratégias de Busca

(8.1) O que é uma busca informada e o que é uma busca não-informada?

Busca informada -> São também conhecidos como busca best-fit ou busca heurística. As buscas informadas são parecidas com a busca em amplitude, com a exceção de que a busca não procede de forma uniforme a partir do nó-raiz da árvore. De acordo com heurísticas e/ou informações específicas do problema, é possível definir uma ordem de preferência entre os caminhos possíveis a partir do nó-raiz. Logo, recorrendo a um pouco mais de informação, é possível ser mais eficiente. Em outras palavras, o melhor caminho até a solução é obtido pela adoção de decisões do tipo best-fit a cada passo. Para tanto, estimar o custo de se chegar a cada nó – $g(n)$ – e/ou o quão distante se está da solução – $h(n)$ – passam a ser de grande relevância

Não-Informada -> Busca cega, sem informações específicas do problema. Se tornam computacionalmente intratáveis quando o número de folhas é muito grande.

(8.2) Quais são os dois tipos de custo que precisam ser definidos quando se realiza busca em árvore empregando o algoritmo A* e o que caracteriza cada um deles?

Resolução: A busca empregando o algoritmo A* deve recorrer a dois índices:

- Fator de altura – $g(n)$: é o custo do caminho de custo mínimo entre o nó-raiz e o nó n .

- Fator heurístico – $h(n)$: é o custo estimado do caminho de custo mínimo entre o nó n e o nó-solução.

$f(n) = g(n) + h(n)$: é o custo do caminho de custo mínimo entre o nó-raiz e um nó-solução, considerando todos os caminhos que passam pelo nó n

O fator de altura não pode diminuir quando se aumenta a distância até a raiz, enquanto que o fator heurístico nunca pode sobre-estimar o custo real até o nó solução.

(8.3) Explique o princípio de operação do algoritmo A*.

Sempre é possível encontrar a solução ótima com a busca A*. A busca A* degenera para a busca em amplitude quando a heurística não aponta preferência por nenhum caminho ao longo da busca. A busca A* é completa, ótima e eficiente (HART et al., 1968). No entanto, sua complexidade ainda é exponencial e o seu uso de memória é intenso.

$f(n)$ nunca decresce ao longo de um caminho \rightarrow monotonicidade

n pai de n' , $f(n') \leq f(n) \rightarrow f(n') = \max[f(n), g(n') + h(n')] \rightarrow \text{pathmax}$

A^* expande todos nós com $f(n) < f^*$

Primeira solução encontrada é a solução ótima

A^* é otimamente eficiente \rightarrow nenhum outro algoritmo expande menos nós que A^*

(8.4) Explique o princípio de operação do algoritmo de busca não-populacional denominado simulated annealing (recozimento simulado).

O método é inspirado no processo de aquecimento de metais e vidros, que assumem um estado de baixa energia quando resfriados de maneira lenta e gradual. A ideia básica é contornar pontos de ótimo local permitindo a realização de movimentos que conduzam temporariamente a soluções de pior qualidade que a atual. A probabilidade de se aceitarem movimentos que levem a soluções de pior qualidade decresce ao longo do processo de busca.

Procedimento RecozimentoSimulado

Gerar solução inicial s

Inicializar temperatura T

Enquanto não atingir condição de parada faça

Selecionar solução aleatoriamente s' de $N(s)$

Se $f(s') < f(s)$ então $s \leftarrow s'$

senão aceitar s' com probabilidade $P(s, s', T)$

Atualizar temperatura T

Fim Enquanto

Fim Procedimento

(8.5) Explique o princípio de operação da busca tabu.

A busca tabu trabalha com base em uma única solução. O método de busca parte de uma solução inicial, e a cada iteração a vizinhança da solução corrente é analisada. Na sequência, uma nova solução é selecionada do conjunto de vizinhos. Esse processo é repetido até que um critério de parada seja alcançado. A principal característica desse método é que ele opera com o objetivo de transcender a otimalidade local, permitindo a degradação da solução durante o processo de busca e até a ocorrência temporária de soluções infactíveis. A busca tabu conserva um histórico do processo de busca com o objetivo de contornar pontos de ótimo local. Durante a busca, certas soluções (ou movimentos) são considerados proibidos ou tabu. Em geral, soluções ou partes de soluções recentemente visitadas ou frequentemente obtidas são

classificadas como tabu. Existem variações da busca tabu propondo melhorias, tais como estratégias de diversificação e memórias adaptativas.

(8.6) Explique os três tipos de buscas locais em espaços discretos.

Resolução:

- First improvement: respeitando a vizinhança previamente definida, pare a busca assim que for encontrada uma solução melhor que a atual.
- Best improvement: respeitando a vizinhança previamente definida, teste todas as soluções candidatas e fique com a melhor dentre elas, que pode ser a atual. Este método de busca pode ser intratável.
- Best improvement with limited resources: respeitando a vizinhança previamente definida, teste um subconjunto de soluções candidatas (por exemplo, um número fixo e razoável de soluções candidatas) e fique com a melhor dentre elas, que pode ser a atual.

(8.7) Dada uma representação matemática de soluções candidatas, defina uma vizinhança no espaço de busca.

(8.8) Dado o problema de encontrar o caminho mínimo entre o nó inicial S e o nó meta G no seguinte grafo:

- a) Apresente a árvore de busca, com todos os caminhos existentes.
- b) Apresentar a solução que seria obtida pela técnica de busca em profundidade.
- c) Apresentar a solução que seria obtida pela técnica de busca em largura.

Tópico 9 – Teoria de Jogos

(9.1) Dadas as regras do jogo e as estratégias de dois jogadores, monte a matriz de pagamentos.

(9.2) O que representa a solução de um jogo de soma nula e estratégia mista envolvendo dois jogadores?

v e as estratégias ótimas x^* e y^* compõem a solução do jogo

(9.3) Em um jogo de soma nula com dois jogadores, cada jogador tem duas estratégias

	Π_1	Π_2
I_1	a_{11}	a_{12}
I_2	a_{21}	a_{22}

(não-dominantes entre si) e a matriz de pagamento assume a forma:

Sob estas condições, responda as seguintes questões:

a) O que é estratégia, estratégia pura e estratégia mista?

Resolução: Uma estratégia deve determinar a ação do jogador frente a qualquer situação possível (estado) do jogo. Portanto, ela indica que decisão o jogador vai tomar a cada lance, mas somente quando o estado do jogo associado ao lance estiver definido. Supondo que cada jogador possua um conjunto finito de estratégias alternativas, a estratégia pura implica a adoção de uma única estratégia durante todo o jogo, enquanto que a estratégia mista envolve a escolha aleatória de uma dentre as estratégias admissíveis a cada lance, seguindo uma atribuição fixa de probabilidade para a escolha de cada estratégia.

b) Forneça valores numéricos arbitrários para a matriz de pagamento de modo que, em um cenário de estratégias puras, o Jogador I seja sempre levado a adotar a sua estratégia 1. Para tanto, use o critério maximin.

Resolução: O critério maximin garante que o Jogador I não receberá um pagamento menor do que p_{\min} , onde $p_{\min} = \max_i \min_j a_{ij}$. Para atender a solicitação do enunciado, é

necessário que o menor pagamento para quando o Jogador I utiliza a sua estratégia 1 seja maior que o menor pagamento no caso dele utilizar a estratégia 2. Dentre as infinitas possibilidades de atribuição de valores, tem-se:

	II_1	II_2	Mínimo
I_1	3	1	1
I_2	4	0	0

c) Aplicando técnicas de programação linear e supondo estratégias mistas, é possível obter a seguinte solução para o jogo:

$$✓ \text{ Valor do jogo: } v = \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{11} + a_{22} - a_{21} - a_{12}}$$

$$✓ \mathbf{x}^* = \left[\frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{21} - a_{12}} \quad \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{21} - a_{12}} \right]$$

$$✓ \mathbf{y}^* = \left[\frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{21} - a_{12}} \quad \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{21} - a_{12}} \right]$$

O que é preciso acontecer com os elementos da matriz de pagamentos para que o jogador I adote com maior frequência a sua estratégia 1?

Resolução: Basta produzir $\frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{21} - a_{12}} > \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{21} - a_{12}}$, que pode ser decomposto em:

$$\begin{cases} a_{22} - a_{21} > a_{11} - a_{12} & \text{se } a_{11} + a_{22} - a_{21} - a_{12} > 0 \\ a_{22} - a_{21} < a_{11} - a_{12} & \text{se } a_{11} + a_{22} - a_{21} - a_{12} < 0 \end{cases}$$

(9.4) Defina como você aplicaria computação evolutiva para evoluir estratégias do dilema do prisioneiro, supondo que a sua decisão de cooperar ou delatar vai depender das 3 últimas decisões suas e de seu adversário.

Tópico 10 – Raciocínio Probabilístico

(10.1) Procure distinguir as formas empregadas para se tratar incerteza em sistemas nebulosos e em sistemas baseados no raciocínio probabilístico.

(10.2) O que é um espaço amostral?

É o conjunto de todos os valores que a variável aleatória pode assumir

(10.3) O que é um experimento?

Um experimento aleatório, E, é um experimento em que o resultado não pode ser predito, mesmo que seja repetido várias vezes e nas mesmas condições

(10.4) O que é um evento?

Um evento é um subconjunto do espaço amostral.

(10.5) O que é uma variável aleatória?

É aquela que assume valores num espaço amostral e para a qual está determinada a probabilidade de ocorrência de cada um dos elementos do espaço amostral

(10.6) Como aplicar o Teorema de Bayes em situações práticas?

(10.7) Como obter probabilidades a partir de uma rede bayesiana?

(10.8) Como aplicar métodos de busca do Tópico 8 (Exemplo: First Improvement) para se obter uma topologia de rede bayesiana mais verossímil?

(10.9) Como obter uma rede bayesiana, mesmo ainda sem as tabelas de probabilidades, a partir de uma descrição em linguagem natural da relação de dependência entre eventos aleatórios?

(10.10) Suponha que um fabricante de queijo recebe o leite que utiliza na seguinte proporção:

- 20% do produtor P1;
- 30% do produtor P2;
- 50% do produtor P3.

Um órgão de fiscalização inspecionou esses produtores de surpresa e relatou que:

- 20% dos galões de leite produzido por P1 são adulterados;
- 5% dos galões de leite produzido por P2 são adulterados;
- 2% dos galões de leite produzido por P3 são adulterados.

Como os galões de leite não apresentam identificação do produtor, ao se analisar um galão

ao acaso e constatar que o seu conteúdo está adulterado (nova evidência), qual é a probabilidade de que o leite seja proveniente do produtor P1?

Sugestão: Aplique a seguinte formulação para o teorema de Bayes:

$$\Pr(A_i | B) = \frac{\Pr(B | A_i) \Pr(A_i)}{\sum_j \Pr(B | A_j) \Pr(A_j)}$$

Resolução: Chamando $\Pr(P_j)$ a probabilidade do leite ser proveniente do produtor P_j ($j=1,\dots,3$) e de Q a nova evidência, então a probabilidade $\Pr(P_1|Q)$ de que o leite seja proveniente do produtor P_1 , dada a nova evidência Q , assume a forma:

$$\Pr(P_1|Q) = \frac{\Pr(Q|P_1)\Pr(P_1)}{\sum_{j=1}^3 \Pr(Q|P_j)\Pr(P_j)}$$

A partir do enunciado da questão, extraem-se todos os valores dos termos do lado direito da equação acima, resultando:

$$\Pr(P_1|Q) = \frac{0,2 * 0,2}{0,2 * 0,2 + 0,05 * 0,3 + 0,02 * 0,5} = \frac{0,04}{0,04 + 0,015 + 0,01} = \frac{0,04}{0,065} \cong 61,54\%$$

(10.11) Uma equipe da área de saúde coletou dados de campo e desenvolveu uma rede bayesiana visando explicar as relações de causa-efeito entre E Praticar Esporte, O Ocorrência de Obesidade e M Ocorrência de Problemas Musculares. A rede bayesiana está apresentada abaixo, onde “s” significa sim e “n” significa não. Estão sendo negligenciados outros fatores que, além de E, sabidamente também influenciam O e M. Sob estas condições, encontre a probabilidade de uma pessoa praticar esporte dado que ela apresenta problemas musculares, ou seja:

$$\Pr(E = s | M = s) = \frac{\Pr(E = s, M = s)}{\Pr(M = s)}.$$

Sugestão: Aplique as formulações:

$$\Pr(E = s, M = s) = \Pr(E = s, M = s, O = s) + \Pr(E = s, M = s, O = n)$$

e

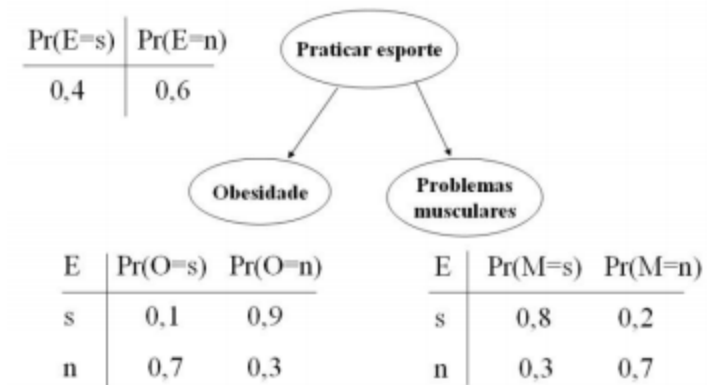
$$\begin{aligned}\Pr(M = s) &= \Pr(E = s, M = s, O = s) + \Pr(E = s, M = s, O = n) \\ &+ \Pr(E = n, M = s, O = s) + \Pr(E = n, M = s, O = n)\end{aligned}$$

ou de modo alternativo:

$$\Pr(E = s, M = s) = \Pr(M = s \mid E = s) \Pr(E = s)$$

e

$$\Pr(M = s) = \Pr(M = s \mid E = s) \Pr(E = s) + \Pr(M = s \mid E = n) \Pr(E = n)$$



Resolução: As sugestões de formulação nos remetem aos seguintes cálculos alternativos:

Cálculo de $\Pr(E = s, M = s)$:

$$\Pr(E = s, M = s) = \Pr(E = s, M = s, O = s) + \Pr(E = s, M = s, O = n) = 0,4 * 0,8 * 0,1 + 0,4 * 0,8 * 0,9 = 0,32$$

ou de modo alternativo:

$$\Pr(E = s, M = s) = \Pr(M = s | E = s) \Pr(E = s) = 0,8 * 0,4 = 0,32$$

Cálculo de $\Pr(M = s)$:

$$\begin{aligned} \Pr(M = s) &= \Pr(E = s, M = s, O = s) + \Pr(E = s, M = s, O = n) + \Pr(E = n, M = s, O = s) \\ &+ \Pr(E = n, M = s, O = n) = \\ 0,4 * 0,8 * 0,1 &+ 0,4 * 0,8 * 0,9 + 0,6 * 0,3 * 0,7 + 0,6 * 0,3 * 0,3 = 0,32 + 0,18 = 0,5 \end{aligned}$$

ou de modo alternativo:

$$\begin{aligned} \Pr(M = s) &= \Pr(M = s | E = s) \Pr(E = s) + \Pr(M = s | E = n) \Pr(E = n) = \\ 0,8 * 0,4 &+ 0,3 * 0,6 = 0,32 + 0,18 = 0,5 \end{aligned}$$

Logo:

$$\Pr(E = s | M = s) = \frac{\Pr(E = s, M = s)}{\Pr(M = s)} = \frac{0,32}{0,5} = 64\%$$

(10.12) Qual a diferença entre os frequentistas e os bayesianos? (Não é uma questão a ser cobrada em prova, mas ajuda a posicionar melhor a perspectiva bayesiana.)

Resolução:

A diferença está na interpretação do que significa probabilidade, conduzindo assim a uma forma distinta de realizar inferências sobre um todo, dado que se dispõe de amostras desse todo (população). Todos estão satisfeitos com os axiomas de probabilidade, mas como utilizar a probabilidade para realizar inferências? Os bayesianos predominaram no século 19, os frequentistas no século 20 e os bayesianos estão recebendo mais atenção nos últimos anos, particularmente devido ao apoio da computação. Considere um cenário em que foram realizados N experimentos visando medir a velocidade da luz no vácuo, sendo que erros aleatórios de medida estão presentes. Para um frequentista, existe um valor verdadeiro e fixo, mas desconhecido, para a velocidade da luz no vácuo. Ele irá tomar uma certa distribuição de probabilidade (geralmente a distribuição normal, também conhecida como gaussiana) e irá estimar os seus parâmetros (no caso, média e variância amostral, levando em conta as N amostras). A estatística resultante para a média amostral (uma estatística é uma função das amostras) será usada como uma estimativa da média da população de amostras. O frequentista, no entanto, não é capaz de dizer o quão distante a sua estatística está do valor verdadeiro da velocidade da luz.

Já um bayesiano parte de uma distribuição de probabilidade a priori. Um bayesiano

discorda de que existe um valor verdadeiro para a velocidade da luz, pois ele a toma como uma variável aleatória com uma distribuição de probabilidade. Usando o teorema de Bayes, o bayesiano utiliza as amostras para definir a distribuição de probabilidade a posteriori.

Em síntese:

- O bayesiano defende a ideia de que quantidades desconhecidas, como média populacional, possuem distribuições de probabilidade. Com isso, ele pode interpretar probabilidade como o grau de confiança (degree of belief ou medida de plausibilidade dado um conhecimento incompleto) em uma hipótese. Ele parte de uma distribuição a priori (prior knowledge ou prior belief), toma algumas poucas amostras (conhecimento adicional ou novidade), aplica a regra de Bayes para obter a distribuição a posteriori e realiza inferências. Para ele, somente as amostras são reais e a média da população é uma abstração. Para ele, alguns valores são mais prováveis que outros, baseado nos dados e na distribuição a priori. Infelizmente, nem sempre a distribuição a priori é informativa.
- O frequentista considera esse procedimento do bayesiano um uso demasiadamente liberal da teoria de probabilidade. Para ele, as probabilidades são apenas definidas quando o número de amostras independentes tende a infinito. Um frequentista acredita que a média da população é real, mas desconhecida e “desconhecível” (só conhecida para infinitas amostras), e pode ser apenas estimada a partir dos dados.

(10.13) É possível expressar o Teorema de Bayes na seguinte forma:

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^k P(B | A_j)P(A_j)}$$

Use então este teorema para resolver o Problema de *Monty Hall*.

Resolução: Definindo que A_i corresponde ao evento ⟨Carro está atrás da porta i ⟩ e B corresponde a ⟨Apresentador do programa abriu uma das portas não escolhidas⟩, então

supondo, sem perda de generalidade, que a pessoa escolheu a porta 1 e o apresentador abriu a porta 3, a probabilidade do carro estar atrás da porta 1 é dada na forma:

$$P(A_1 | B) = \frac{P(B / A_1)P(A_1)}{P(B / A_1)P(A_1) + P(B / A_2)P(A_2) + P(B / A_3)P(A_3)}$$

Interpretação: A probabilidade do carro estar atrás da porta 1, dado que o apresentador abriu a porta 3, é dada pela probabilidade do apresentador abrir a porta 3, dado que o carro está atrás da porta 1, vezes a probabilidade a priori do carro estar atrás da porta 1, dividido pela soma das seguintes probabilidades:

- Probabilidade do apresentador abrir a porta 3, dado que o carro está atrás da porta 1, vezes a probabilidade a priori do carro estar atrás da porta 1;
- Probabilidade do apresentador abrir a porta 3, dado que o carro está atrás da porta 2, vezes a probabilidade a priori do carro estar atrás da porta 2;
- Probabilidade do apresentador abrir a porta 3, dado que o carro está atrás da porta 3, vezes a probabilidade a priori do carro estar atrás da porta 3.

Resulta então:

$$P(A_1 | B) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{\frac{1}{2}}{\frac{1}{2} + 1} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1}{3}$$

De forma similar, a probabilidade do carro estar atrás da porta 2 é dada na forma:

$$P(A_2 | B) = \frac{P(B / A_2)P(A_2)}{P(B / A_1)P(A_1) + P(B / A_2)P(A_2) + P(B / A_3)P(A_3)}$$

produzindo:

$$P(A_2 | B) = \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{1}{\frac{1}{2} + 1} = \frac{1}{\frac{3}{2}} = \frac{2}{3}$$

Logo, trocar pela porta 2 dobra a probabilidade da pessoa ganhar o carro.