

CUIDADO PARA NÃO EDITAR CAGANDO O DOCUMENTO INTEIRO

ps: tentem colocar o slide de onde foi retirada a resposta

ps2: se encontrarem erros, editem com cor diferente!

ps3: ajudem a resolver essa bomba

ps4: 4k

situacao atual:

<https://www.youtube.com/watch?v=o3fllGdbE-w>

ALGUÉM CONSEGUE RESPONDER A MERDA DA 2.2?

o que eu marquei como importante:

0.12

A Web 3.0, anunciada como a terceira onda da Internet, projeta estruturar todo o conteúdo disponível na rede mundial de computadores dentro dos conceitos de “compreensão das máquinas” e “semântica das redes”.

0.14

O valor de um sistema de comunicação cresce na razão do quadrado do número de usuários do sistema

0.15

Tem resolvido.

1.17

Slide 56 =>

ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ea072_2s14/topico1_EA072_2s2014_Parte1.pdf

2.10

3.7

5.9

8.3

10.9 (esse ele disse que ia cair com certeza)

10.10 (esse ele disse que ia cair com certeza)

Sumário

Tópico 0 – Introdução e Motivação

- (0.1) O que é inteligência?
- (0.2) O que é inteligência artificial?
- (0.3) O que caracteriza um comportamento inteligente?
- (0.4) O que é um organismo cibernético?
- (0.5) O que é um sistema inteligente sob a perspectiva de Albus (1991)?
- (0.6) O que é mente? O que é consciência?
- (0.7) Faça uma distinção apropriada entre sentimento e pensamento.
- (0.8) Qual a motivação para que Searle apresentasse o cenário de sua sala chinesa?
- (0.9) Explique como funciona o teste de Turing.
- (0.10) Na teoria das múltiplas inteligências de Gardner, quais compõem o quociente de inteligência (QI) e quais compõem o quociente emocional (QE)?
- (0.11) O que é aprendizado de máquina?
- (0.12) O que caracteriza a Web 3.0?
- (0.13) Qual é o significado da Lei de Moore?
- (0.14) Qual é o significado da Lei de Metcalfe?
- (0.15) O que é otimização multimodal?

Tópico 1 – Redes Neurais Artificiais

- (1.1) Esboce graficamente o comportamento da densidade de sinapses no cérebro, ao longo da vida de um ser humano que atinge uma idade avançada.
- (1.2) Numa abordagem conexionista, explique como se dá o aprendizado.
- (1.3) Apresente a diferença de funções entre os hemisférios direito e esquerdo do cérebro.
- (1.4) Em que o paradigma da neurocomputação contrasta com o paradigma de um computador digital padrão von Neumann?
- (1.5) Explique como se dá o processo eletroquímico de transmissão de sinais entre duas células nervosas, na conexão sináptica, enfatizando o papel dos neurotransmissores e das bombas de sódio-potássio.
- (1.6) Por que diz-se que o caso de Phineas Gage representa o início histórico dos estudos das bases biológicas do comportamento?
- (1.7) Explique o modelo matemático de um neurônio artificial, dando destaque para (i) a existência de produto interno; (ii) a presença de saturação na forma de uma função sigmoide; e (iii) a saída na forma de uma função de expansão ortogonal (ridge function).
- (1.8) No aprendizado a partir de dados amostrados (aprendizado supervisionado) e empregando uma rede neural MLP, apresente os três mapeamentos envolvidos e indique quais são suas dimensões de entrada e de saída.
- (1.9) Explique o conceito de sobreajuste (overfitting) e apresente mecanismos capazes de contribuir para reduzir os seus efeitos. Para tanto, use os conceitos de capacidade de generalização e de regularização do mapeamento a ser aproximado.

(1.10) Qual foi a conjectura de Minsky e Papert e por que ela foi tão prejudicial para a pesquisa em redes neurais artificiais entre 1969 e 1985?

(1.11) Apresente o princípio básico de operação de um algoritmo de otimização não-linear irrestrita, empregado no ajuste de pesos de redes neurais artificiais sujeitas a processos de treinamento supervisionado.

(1.16) No caso de redes neurais RBF, apresente uma proposta para a determinação dos centros das funções de base radial, de suas dispersões e dos pesos da camada de saída.

(1.17) Compare uma máquina de aprendizado extremo (ELM, do inglês extreme learning machine) com uma rede neural MLP.

(1.18) Explique o conceito de quadrados mínimos regularizados em ELMs (Máquinas de aprendizado extremo).

(1.19) Quais são os três requisitos fundamentais para a ocorrência de auto-organização?

(1.26) O que é uma rede neural recorrente?

(1.27) Explique como se dá a operação de uma rede neural de Hopfield como memória endereçável por conteúdo, abordando também o conceito de atrator espúrio.

(1.29) No caso de um sistema dinâmico não-linear autônomo, quais são os quatro possíveis comportamentos que podem ocorrer em regime?

(1.30) Em dinâmica não-linear, o que é uma bacia de atração e qual é a influência da condição inicial na determinação do estado de regime de um sistema dinâmico com múltiplos pontos de equilíbrio?

Tópico 2 – Computação Evolutiva

(2.1) O que é uma meta-heurística?

(2.2) Dada a definição formal de uma solução candidata num problema de otimização a ser resolvido via computação evolutiva, apresente uma proposta de codificação (representação computacional) capaz de representar unicamente cada ponto do espaço de busca. Esta codificação pode ser na forma de um vetor de atributos binários, inteiros, em ponto flutuante, ou uma mistura dessas três possibilidades. Matriz, árvore e outras estruturas de dados não estão descartadas na representação computacional, em lugar do vetor mencionado acima.

(2.7) Qual é a diferença entre fenótipo e genótipo?

(2.8) Usando os conceitos de transcrição, tradução e códon, explique como o código genético determina a produção de uma proteína, sabendo que uma proteína é dada por uma sequência de aminoácidos.

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

Tópico 3 – Sistemas Nebulosos

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

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.

(4.4) A resposta à questão 4.3 envolve o conceito de vizinhança entre as partículas.
Quais são os principais tipos de vizinhança adotados em PSO?

Tópico 5 – Lógica Matemática, Representação e Inferência

(5.1) O que é uma proposição?

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

(5.3) O que é uma tautologia?

(5.4) O que é uma contradição?

(5.5) O que é um problema de satisfação?

(5.6) O que é vinculação de proposições?

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

(5.8) Dada a regra da resolução:

(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.

(5.11) Dadas as seguintes proposições:

Tópico 6 - Sistemas Baseados em Regras e Árvores de Decisão

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

(6.6) O que é um flat file?

Tópico 11 Sistemas complexos e Redes complexas

(12.9) O que foi “The DARPA Grand Challenge” e por que esse assunto é relevante no contexto de Inteligência Artificial?

(12.10) O que você entende por robótica coletiva?

(12.11) Descreva uma aplicação prática em que a robótica coletiva é vantajosa quando comparada ao emprego de um único robô.

Versão 2014

Tópico 0 – Introdução e Motivação

(0.1) O que é inteligência?

Inteligência é um termo genérico que descreve propriedades da mente, como capacidade de aprendizado, planejamento, resolução de problemas, compreensão de conceitos complexos, pensamento abstrato e comunicação.

O termo inteligência deriva da expressão em latim “intellegere” (inter-legere), que significa escolher entre, discernir.

Teoria das múltiplas inteligências (Gardner, 1983):

- vivacidade verbal
 - vivacidade matemático-lógica (QI)
 - aptidão espacial
 - gênio cinestésico
 - dons musicais
 - aptidão interpessoal (liderança e ação cooperativa)
 - aptidão intrapsíquica (modelo adequado de si mesmo)
- Os índices de inteligência não são inteiramente consistentes: o desempenho intelectual de uma certa pessoa pode variar em diferentes ocasiões, em diferentes domínios e quando avaliado por diferentes critérios

(0.2) O que é inteligência artificial?

- Área da computação voltada para a formalização e implementação de processamento inteligente de informação em máquinas, numa tentativa de construir mecanismos computacionais para atividades que supostamente requerem inteligência quando feitas por seres humanos.
- Denominação proposta em 1956, associada à ciência e à engenharia aplicadas na concepção de máquinas inteligentes, com comportamento adaptativo orientado a objetivos.

(0.3) O que caracteriza um comportamento inteligente?

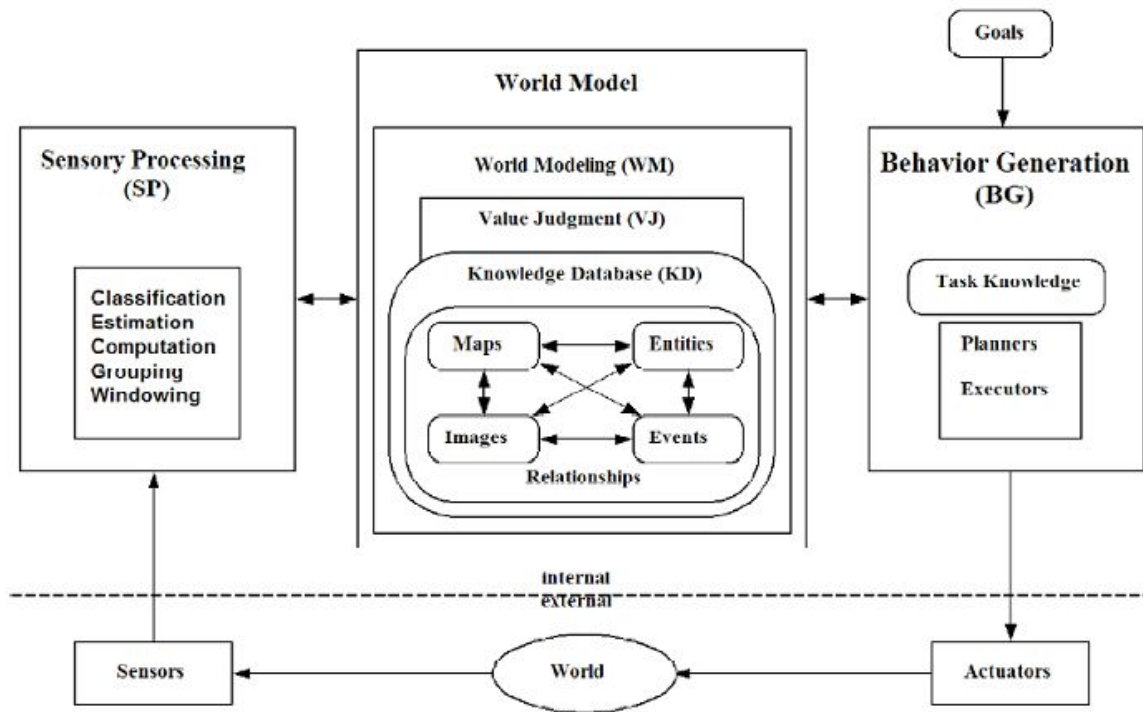
Sabe-se, de forma genérica, que um sistema inteligente é aquele que apresenta capacidade para:

- Raciocinar;
- Planejar;
- Resolver problemas;
- Realizar indução, dedução
- Lógica e abdução;
- Armazenar conhecimento;
- Comunicar-se através de uma
- linguagem;
- Perceber e adaptar-se ao meio;
- Aprender.

(0.4) O que é um organismo cibernético?

Com capacidade de controle e processamento de informações.

(0.5) O que é um sistema inteligente sob a perspectiva de Albus (1991)?



(0.6) O que é mente? O que é consciência?

A mente refere-se a algo interior e subjetivo dos seres humanos, na medida em que o nosso mundo interior, os nossos pensamentos e desejos, os nossos medos e sentimentos são algo que não se pode ver, mas que existe em cada indivíduo.

Consciência: Conhecimento imediato da sua própria atividade psíquica ou física.

Auto-consciência: Consciência que adquire a capacidade de refletir sobre si mesma, i.e., que se reconhece como o domínio da racionalidade, do pensamento, ou dos chamados estados interiores; consciência-de-si.

(0.7) Faça uma distinção apropriada entre sentimento e pensamento.

Não somos livres em nossos sentimentos, nem em nossos impulsos de vontade que vêm do nosso inconsciente. Não é possível controlar um sentimento.

(0.8) Qual a motivação para que Searle apresentasse o cenário de sua sala chinesa?

SEARLE (1991) argumenta que há uma diferença marcante entre este operador e uma pessoa que domina o idioma chinês e responda às mesmas perguntas sem usar o livro de regras. O primeiro está apenas seguindo regras sintáticas. O segundo está associando semântica (significado) ao que está fazendo e, portanto, está fazendo muito mais que o primeiro.

(0.9) Explique como funciona o teste de Turing.

O teste de Turing afirma que um computador pode ser chamado de inteligente se ele puder enganar um ser humano ao fazê-lo acreditar que o computador é um humano. O teste consiste em um ser humano realizar um interrogatório através de um terminal, visando descobrir se do outro lado do terminal se encontra um ser humano ou um computador. Caso seja um computador que esteja fornecendo as respostas e o interrogador não conseguir decidir se se trata de um ser humano ou um computador, então este computador é dito ter sido aprovado no teste de Turing.

(0.10) Na teoria das múltiplas inteligências de Gardner, quais compõem o quociente de inteligência (QI) e quais compõem o quociente emocional (QE)?

(0.11) O que é aprendizado de máquina?

Aprendizado de Máquina é uma área de IA cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática

(0.12) O que caracteriza a Web 3.0?

/(Slide: Motivação - 58)

- Organização e uso inteligente de todo o conteúdo existente na Web
- Semântica e contexto

(Embora pareçam não estar relacionados pela estrutura dos tópicos, as informações abaixo também fazem parte da web 3.0)

- Sistemas de recomendação:
 - Sugestão (filtros colaborativos)
 - Cruzamento de informações
 - Extração de conhecimento operante

(0.13) Qual é o significado da Lei de Moore?

(<http://eletronicos.hsw.uol.com.br/lei-de-moore3.htm>)

A versão mais comum da Lei de Moore é que o número de transistores em um circuito dobra a cada 18 (ou 24) meses.

(http://pt.wikipedia.org/wiki/Lei_de_Moore)

Inicialmente a lei de Moore não passava de uma observação, mas acabou tornando-se um objetivo para as indústrias de semicondutores, fazendo as mesmas desprenderem muitos recursos para poder alcançar as previsões de Moore no nível de desempenho e isso que torna a lei de Moore realmente importante, pois sem ela, talvez não tivéssemos um desenvolvimento tão acelerado em nível de hardware e com custos cada vez mais acessíveis.

(0.14) Qual é o significado da Lei de Metcalfe?

Lei de Metcalfe é uma lei formulada pelo estadunidense Robert Metcalfe, inventor do sistema Ethernet de redes locais. A lei trata do valor de sistemas de comunicação. Seu enunciado é o seguinte:

“O valor de um sistema de comunicação cresce na razão do quadrado do número de usuários do sistema”

- Envolve diretamente a introdução de novas tecnologias: muitos esforços devem ser investidos em convencer os clientes a adotarem o novo.
- Princípio da externalidade da rede: o valor de um novo serviço de comunicação depende de quantos usuários adotaram (ou vão adotar) este novo serviço.
- Os primeiros a proporem o novo têm mais dificuldades de conquistar o mercado do que aqueles que passam a trabalhar com o novo depois que este já foi aceito pelo mercado.
- Uma possível compensação aos pioneiros é o rótulo de inovador, naturalmente atribuído pelo mercado.

(0.15) O que é otimização multimodal?

multimodal -> vários ótimos locais (ótimo pode variar)

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

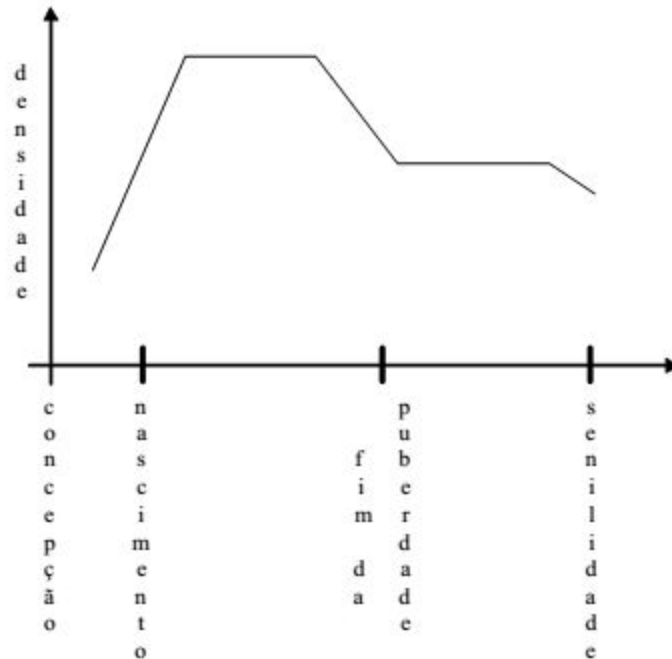
otimização. Abordagens de busca populacionais com manutenção de diversidade geralmente apresentam mecanismos para localização e preservação de múltiplos ótimos, assim como tendem a expressar uma melhor capacidade de seguimento de ótimos que variam no tempo.

(0.16) Apresente argumentos que sustentem a seguinte afirmação: “Mesmo recorrendo a técnicas avançadas e eficazes para processamento de informação, normalmente associadas à área de inteligência artificial, o principal fator que levou Deep Blue a superar Kasparov (campeão mundial de xadrez) em 1997 foi a ‘força bruta’ da máquina.”

A codificação de um jogo de xadrez é mecânica: sabendo-se as regras de movimentação e o estado do tabuleiro, é possível imaginar possíveis situações futuras e criar estratégias para alcançar algum objetivo no jogo. Essa é a forma que muitos jogadores profissionais usam, no entanto a mente humana é capaz de imaginar apenas alguns passos no futuro; já um computador está limitado apenas à sua capacidade de hardware. No caso do Deep Blue, sua vantagem para poder ganhar no jogo de xadrez foi sua capacidade de gerar um número imenso de possíveis jogadas futuras, e escolher as melhores de acordo com o estilo de jogo de Kasparov. Se limitássemos o número de estados futuros que a máquina tem acesso à mesma quantidade da mente humana, o resultado do jogo provavelmente seria diferente.

Tópico 1 – Redes Neurais Artificiais

(1.1) Esboce graficamente o comportamento da densidade de sinapses no cérebro, ao longo da vida de um ser humano que atinge uma idade avançada.



(1.2) Numa abordagem conexionista, explique como se dá o aprendizado.

Consequência de conexões repetidas da rede neural. Se caracteriza por mudanças dos padrões nessas conexões.

O número de conexões está associado à especialização da rede neural (natural ou artificial). Quanto mais sinapses, mais geral é a rede. Conforme a rede é submetida a experiências - e aprendizado - são realizadas podas de sinapses, tornando-a mais especializada e madura.

(1.3) Apresente a diferença de funções entre os hemisférios direito e esquerdo do cérebro.

Hemisfério esquerdo (paradigma sequencial): produção, lógica e compreensão da linguagem, processamento serial (considera informações pontuais a cada instante), processamento simbólico, inferência, planejamento, noção de tempo.

Hemisfério direito (paradigma paralelo): aprendizado e memória espacial, síntese da percepção, sentimentos e emoções, pensamento associativo, processamento global da informação, raciocínio por analogia, comparação e

identificação de imagens.

(1.4) Em que o paradigma da neurocomputação contrasta com o paradigma de um computador digital padrão von Neumann?

Um neurônio biológico, é ao mesmo tempo, memória e processador. Na computação digital, essas tarefas são executadas por componentes distintos (memória é separada de processamento).

(1.5) Explique como se dá o processo eletroquímico de transmissão de sinais entre duas células nervosas, na conexão sináptica, enfatizando o papel dos neurotransmissores e das bombas de sódio-potássio.

Sinapses são junções entre terminais nervosos no fim de um ramo de um neurônio e a membrana de outros. Para que a mensagem seja passada adiante, um novo potencial de ação deve ser criado no neurônio seguinte, e isso requer a liberação de substâncias denominadas neurotransmissores.

Quando o potencial de ação atinge a sinapse, as vesículas se fundem com a membrana do neurônio e liberam os neurotransmissores na fenda sináptica e os neurotransmissores geram um novo impulso nervoso na célula vizinha, o qual se propagará até a sinapse seguinte. A ligação dos neurotransmissores acarreta na abertura dos canais, o que desencadeia um potencial de ação específico no neurônio que recebe a informação. O potencial de ação é propagado ao longo da célula pela diferença de concentração entre Sódio e Potássio no citoplasma e no meio extra-celular. Para não manter o equilíbrio eletrostático, as bombas NaK gastam ATP para colocar K^+ dentro da célula e tirar Na^+ de dentro da célula, mantendo uma diferença de concentração que garante uma diferença de potencial elétrico de cerca de -70mv.

(1.6) Por que diz-se que o caso de Phineas Gage representa o início histórico dos estudos das bases biológicas do comportamento?

(pelo que entendi) Por que até então não se acreditava na compartimentação do cérebro (com regiões responsáveis por certas tarefas), e só depois desse caso é que começaram a considerar que danos cerebrais poderiam causar mudanças psicológicas e comportamentais.

[Fonte](#)

(1.7) Explique o modelo matemático de um neurônio artificial, dando destaque para (i) a existência de produto interno; (ii) a presença de saturação na forma de uma função sigmoidal; e (iii) a saída na forma de uma função de expansão ortogonal (ridge function).

Modelo matemático: Simplificações da realidade com o propósito de representar aspectos relevantes de um sistema em estudo, sendo que detalhes de menor significância são descartados para viabilizar a modelagem.

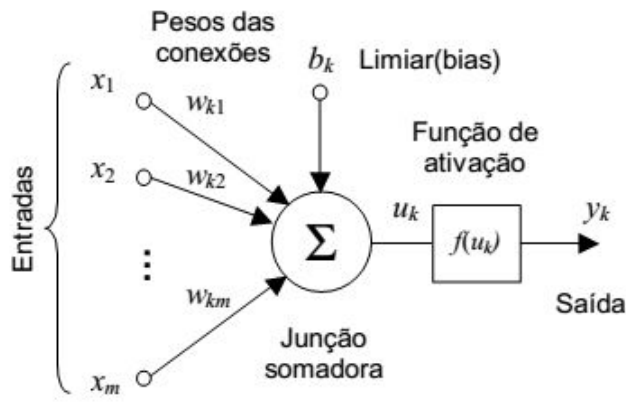


Figura 1 – Modelo matemático de um neurônio artificial

(1.8) No aprendizado a partir de dados amostrados (aprendizado supervisionado) e empregando uma rede neural MLP, apresente os três mapeamentos envolvidos e indique quais são suas dimensões de entrada e de saída.

Aprendizado supervisionado: quando agente externo indica à rede uma resposta desejada para certo padrão de entrada.

Mapeamentos envolvidos (slide 77 - Tópico 1, parte 1):

- 1) mapeamento a ser aproximado (do qual se conhece apenas dados amostrados);
- 2) mapeamento resultante do processo de aproximação;
- 3) mapeamento entre cada vetor de pesos e o erro (superfície de erro).

(1.9) Explique o conceito de sobretreinamento (overfitting) e apresente mecanismos capazes de contribuir para reduzir os seus efeitos. Para tanto, use os conceitos de capacidade de generalização e de regularização do mapeamento a ser aproximado.

O overfitting ocorre quando os pesos de uma rede adaptam-se muito bem aos dados fornecidos, mas não são eficientes na solução do problema num caso geral. (note que esta “aproximação ótima” inclui o ruído das amostras).

Capacidade de generalização: uma das técnicas é dividir o conjunto de amostras em dois conjuntos, o conjunto de treinamento e o conjunto de validação. O primeiro é empregado no ajuste dos pesos sinápticos e o segundo é utilizado para dizer quando o treinamento deve ser interrompido.

(1.10) Qual foi a conjectura de Minsky e Papert e por que ela foi tão prejudicial para a pesquisa em redes neurais artificiais entre 1969 e 1985?

Ou-exclusivo não pode ser representado.

Esses autores também acreditavam que não havia razão para supor que redes multicamadas pudessem conduzir a uma solução para o problema proposto. Esta hipótese só foi definitivamente rejeitada com o desenvolvimento do algoritmo de retro-propagação (back-propagation), já nos anos 80, o qual permite o ajuste automático de pesos para redes neurais multicamadas, arquitetura necessária para a realização de mapeamentos não-lineares.

(1.11) Apresente o princípio básico de operação de um algoritmo de otimização não-linear irrestrita, empregado no ajuste de pesos de redes neurais artificiais sujeitas a processos de treinamento supervisionado.

Neste caso, o processo de treinamento supervisionado de redes neurais artificiais multicamadas é equivalente a um problema de otimização não-linear irrestrita, em que a superfície de erro é minimizada a partir do ajuste dos pesos sinápticos.

não linear -> não necessariamente fornece um ótimo global. Espera-se pelo menos um mínimo local.

$x = x_0 + td$

X = proximo passo

x_0 = passo anterior

t = tamanho do passo

d = direção

(1.12) O que é mínimo local no treinamento supervisionado de uma rede neural artificial?

É o ponto correspondente aos pesos sinápticos que geram o menor erro, numa vizinhança, entre a saída e o conjunto de treinamento.

(1.13) Por que existem arquiteturas distintas de redes neurais artificiais?

(1.14) Descreva geometricamente como uma rede neural RBF aproxima mapeamentos multidimensionais de entrada-saída.

As redes neurais com função de ativação de base radial (RBF) apresentam três diferenças principais em relação às redes tipo perceptron multicamadas:

- Elas sempre apresentam uma única camada intermediária;
- Neurônios de saída são sempre lineares;
- Os neurônios da camada intermediária têm apenas uma função de base radial como função de ativação, ao invés de uma função sigmoideal ou outras.

Os mapeamentos são feitos encontrando-se os pesos que cada “montinho” deve ter para que, somados, reduzam o erro junto à saída.

(1.15) Descreva geometricamente como uma rede neural MLP aproxima mapeamentos multidimensionais de entrada-saída.

Os mapeamentos são feitos encontrando-se os pesos que cada função deve ter para que, somados, reduzam o erro junto à saída.

(1.16) No caso de redes neurais RBF, apresente uma proposta para a determinação dos centros das funções de base radial, de suas dispersões e dos pesos da camada de saída.

Existem, basicamente, 3 alternativas (Slide 21, Tópico 1 - Parte 2)

(1) espalhar os centros uniformemente ao longo da região em que se encontram os dados; (2) escolher aleatoriamente, ou segundo um critério específico, um subconjunto de padrões de entrada como centros; (3) auto-organizar os centros, de acordo com a distribuição dos dados de entrada (e.g., k-means).

(Slides 21-22) Dispersão: adota-se uma única dispersão para todos os componentes.

(1.17) Compare uma máquina de aprendizado extremo (ELM, do inglês extreme learning machine) com uma rede neural MLP.

Tópico 1 (Parte 2), a partir do slide 27

- MLP e a RBF, produzem a sua saída (podendo ser múltiplas saídas) como uma combinação linear das ativações dos neurônios da camada anterior.
- No caso da rede neural MLP, as funções-base são funções de expansão ortogonal (ridge functions), enquanto que, no caso da rede neural RBF, as funções-base têm um comportamento radial em relação a um centro de ativação

máxima.

- Controlar a norma dos pesos sinápticos é mais relevante para a capacidade de generalização do que controlar o tamanho da rede neural, ou seja, o número n de neurônios na camada intermediária.
- As máquinas de aprendizado extremo exploram este resultado “de forma extrema”, ou seja, jogam toda a responsabilidade por garantir uma boa capacidade de generalização aos pesos da camada de saída, permitindo que os pesos da camada intermediária, responsáveis por definir as funções-base, sejam definidos de modo aleatório.

(1.18) Explique o conceito de quadrados mínimos regularizados em ELMs (Máquinas de aprendizado extremo).

As máquinas de aprendizado extremo caracterizam-se por possuir grande número de neurônios com as mais variadas funções de ativação, delegando ao conjunto de pesos escolhido a responsabilidade de minimizar o erro entre a saída e o esperado. Tais pesos podem ser calculados utilizando-se o método dos quadrados mínimos, em que o quadrado do erro de cada ponto da solução é somado e escolhe-se o conjunto com menor soma.

(1.19) Quais são os três requisitos fundamentais para a ocorrência de auto-organização?

- Aprendizado não supervisionado;
- Adaptabilidade;

Interação, realimentação e descentralização.

(1.20) Em um mapa auto-organizável, explique por que se trabalha com duas posições relativas: a posição relativa entre os neurônios do mapa e a posição relativa entre os vetores de pesos dos neurônios do mapa.

(1.21) Defina agrupamento de dados ou clusterização.

(1.22) Como um mapa auto-organizável pode ser empregado na definição do

número de agrupamentos em conjuntos de pontos definidos em espaços multidimensionais? Para tanto, utilize o conceito de matriz-U.

3

(1.23) Que outras informações relevantes podem ser extraídas de um mapa já auto-organizado?

(1.24) Como um mapa auto-organizável unidimensional pode ser empregado na solução do problema do caixeiro viajante? Neste contexto, que tipo de regra você proporia para a inserção de neurônios no mapa e para a poda de neurônios do mapa durante o processo de auto-organização?

(1.25) O que significa capacidade de aproximação universal em redes neurais artificiais, sejam elas recorrentes ou não-recorrentes?

(1.26) O que é uma rede neural recorrente?

- Redes neurais recorrentes são estruturas de processamento capazes de representar uma grande variedade de comportamentos dinâmicos.
- A presença de realimentação de informação permite a criação de representações internas e dispositivos de memória capazes de processar e armazenar informações temporais e sinais seqüenciais.
- a presença de conexões recorrentes ou realimentação de informação pode conduzir a comportamentos complexos, mesmo com um número reduzido de parâmetros.
- como estruturas de processamento de sinais, redes neurais recorrentes se assemelham a filtros não-lineares com resposta ao impulso infinita (NERRAND et al. , 1993)

(1.27) Explique como se dá a operação de uma rede neural de Hopfield como memória endereçável por conteúdo, abordando também o conceito de atrator espúrio.

Quando a rede neural de Hopfield armazena K memórias fundamentais através do ajuste de seus pesos pela regra de Hebb generalizada, os estados estáveis presentes na superfície de energia não vão se restringir aos estados associados às memórias fundamentais armazenadas. Todos os estados estáveis não associados às memórias fundamentais armazenadas são denominados atratores espúrios.

Os atratores espúrios existem em virtude dos seguintes fatores:

1. A função de energia E é simétrica, no sentido de que os estados correspondentes ao reverso das memórias fundamentais armazenadas também são estados estáveis;

2. Toda combinação linear de um número ímpar de estados estáveis também vai ser um estado estável (A MIT, 1989).

3. Para um grande número K de memórias fundamentais, a função de energia vai produzir pontos de equilíbrio que não estão correlacionados com nenhuma das memórias fundamentais armazenadas na rede (inflexibilidade da superfície de energia).

(1.28) Explique o princípio de operação de uma rede neural com estados de eco (ESN, do inglês echo state network).

(1.29) No caso de um sistema dinâmico não-linear autônomo, quais são os quatro possíveis comportamentos que podem ocorrer em regime?

Um sistema dinâmico não linear autônomo pode exibir 4 comportamentos em estado estacionário: ponto de equilíbrio, ciclo limite, quase-periodicidade e caos.

(1.30) Em dinâmica não-linear, o que é uma bacia de atração e qual é a influência da condição inicial na determinação do estado de regime de um sistema dinâmico com múltiplos pontos de equilíbrio?

Uma bacia de atração é o caso por exemplo de um mínimo local, os pontos que são colocados próximos dele são atraídos para ele, portanto dependendo de onde colocarmos o ponto inicial ele será atraído para o mínimo mais próximo e tendo assim um estado de regime diferente de um ponto iniciado próximo a outro mínimo local.

Tópico 2 – Computação Evolutiva

(2.1) O que é uma meta-heurística?

É uma otimização matemática. É um procedimento que envolve achar ou gerar soluções suficientemente boas para um determinado problema.

[Fonte](#)

Uma meta-heurística é um método [heurístico](#) para resolver de forma genérica [problemas de otimização](#).

Wikipedia

(2.2) Dada a definição formal de uma solução candidata num problema de otimização a ser resolvido via computação evolutiva, apresente uma proposta de codificação (representação computacional) capaz de representar unicamente cada ponto do espaço de busca. Esta codificação pode ser na forma de um vetor de atributos binários, inteiros, em ponto flutuante, ou uma mistura dessas três possibilidades. Matriz, árvore e outras estruturas de dados não estão descartadas na representação computacional, em lugar do vetor mencionado acima.

(2.3) Dado que se conhece a representação computacional de uma solução candidata, proponha operadores de mutação e recombinação (crossover).

(2.4) A partir de uma descrição textual dos objetivos envolvidos numa tarefa de otimização, proponha a função de adaptação (fitness) a ser empregada por um algoritmo evolutivo.

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

(2.6) Se você for ajustar os pesos de uma rede neural MLP via computação evolutiva, qual seria a codificação a ser adotada e como ficaria a função de adaptação?

A codificação em ponto flutuante deveria ser adotada, logo que os pesos sinápticos estão no intervalo $(-1,1)$.

A função objetivo seria maximizar a proximidade do resultado ao conjunto de

validação. Sendo assim, a função de adaptação seria selecionar os melhores indivíduos, realizar crossover entre os indivíduos com maior fitness gerando assim uma nova geração.

(2.7) Qual é a diferença entre fenótipo e genótipo?

Genótipo representa o conjunto específico de genes no genoma. Neste caso, indivíduos com o mesmo genoma são ditos terem o mesmo genótipo. Em computação evolutiva, ele pode ser considerado como sendo o vetor de atributos que representa o indivíduo.

Fenótipo é a manifestação do genótipo no comportamento, fisiologia e morfologia do indivíduo, como um produto de sua interação com o ambiente. No contexto de computação evolutiva, ele pode ser considerado como sendo a interpretação do vetor de atributos.

(2.8) Usando os conceitos de transcrição, tradução e códon, explique como o código genético determina a produção de uma proteína, sabendo que uma proteína é dada por uma sequência de aminoácidos.

O DNA guarda as informações genéticas em aminoácidos e é transcrito na sequência 5' 3' gerando um RNA mensageiro (mRNA). O RNA mensageiro sofre RNA polimerase no sentido 5' 3' codificando as proteínas. Cada proteína possui um códon correspondente, que é uma sequência de 3 aminoácidos. A RNA polimerase começa na sequência AUG, que é um códon de iniciação e termina em algum dos códons de parada (UAA, UAG, UGA), gerando uma proteína para cada códon encontrado entre o start e stop códons.

(2.9) 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 de elementos é dada pelo número de elementos do conjunto.

Suponha o problema da bipartição de conjuntos: a cardinalidade do espaço de busca será 2^N .

Suponha o problema da tripartição de conjuntos: a cardinalidade do espaço de busca será 3^N .

Suponha um problema que utiliza árvores binárias com N folhas, N-1 nós internos, T topologias não isomorfas, S símbolos não-terminais candidatos e P

símbolos terminais candidatos: a cardinalidade do espaço de busca será $T \times S^{N-1} \times P^N$.

Número de topologias para árvores binárias com N folhas e raiz: $(2N-3)! / \{2^{N-2} (N-2)!\}$.

Número de topologias para árvores binárias com N folhas e sem raiz: $(2N-5)! / \{2^{N-3} (N-3)!\}$.

Busca de uma máquina de estado finito com N estados, estado inicial a determinar, P símbolos de entrada e Q símbolos de saída: $N \times (N^P \times Q^P)^N$.

Caixeiro viajante com N cidades: Numero de permutação das N cidades = N!. Como o circuito é fechado, não importa qual cidade começamos o circuito, então teremos (N-1)! e como o circuito pode ser percorrido por ambos os lados, a cardinalidade será (N-1)!/2.

(2.10) Aprenda a aplicar a Tabela de Contagem de Coleções (Seção 23, pg. 63).

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	$\left(\binom{n}{k} \right) = \binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Tópico 3 – Sistemas Nebulosos

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

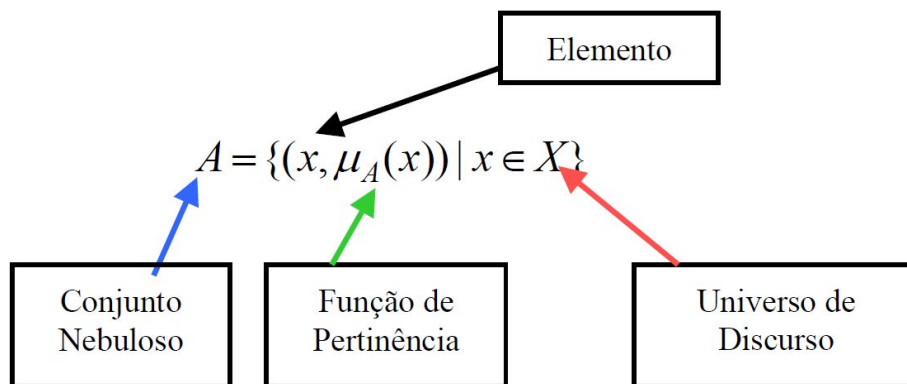
É o espaço onde estão definidos os elementos do conjunto nebuloso.

Exemplo: Conjunto fuzzy de pessoas altas.

Função de pertinência: Pessoa é considerada alta se > 1.90 .

O universo de discurso é o conjunto de todas as pessoas.

Outra definição



<http://professor.ufabc.edu.br/~ronaldo.prati/InteligenciaArtificial/AulaFuzzy.pdf>

pag. 8

ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ea072_2s13/topico3_EA072_2s2013.pdf

f

pag. 15

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

Uma variável linguística possui valores que não são números, mas sim palavras ou frases na linguagem natural.

Idade = idoso

Um valor linguístico é um conjunto fuzzy.

Todos os valores linguísticos formam um conjunto de termos:

$T(\text{idade}) = \{\text{Jovem, velho, muito jovem, ...}$

$\text{Maduro, não maduro, ...}$

$\text{Velho, não velho, muito velho, mais ou menos velho, ...}$

$\text{Não muito jovem e não muito velho, ...}$

[<https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CC0QFjAA&url=http%3A%2F%2Fwww.dimap.ufrn.br%2F~anne%2FAula%2520Fuzzy.ppt&ei=tc2SUtamGczJsQTC54KAAG&usg=AFQjCNH4VYXiBPAccBuTwQDM7Yg9zoa39w&sig2=T-ydlH68kOJsTJC67njOxw>]

(3.3) O que é conjunto nebuloso?

Um conjunto ordenado de pares $(x, F(x))$.

x é um elemento do universo amostral

$F(x)$ é a função de pertinência: Reflete o conhecimento que se tem em relação a intensidade com que o objeto pertence ao conjunto fuzzy.

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

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

(3.6) 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.

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

O modelo fuzzy Takagi-Sugeno (TS) consiste em um sistema de inferência capaz de descrever, de forma exata ou aproximada, sistemas dinâmicos não-lineares por meio de um conjunto de sistemas dinâmicos lineares, localmente válidos, interpolados de forma suave, não-linear e convexa.

Em termos de modelagem, sabe-se que os modelos TS têm a capacidade de aproximar uniformemente funções quaisquer em um domínio compacto. Em outras palavras, modelos TS têm a capacidade de aproximação universal.

Além disso, uma vantagem sobre outros modelos fuzzy é o fato dos conseqüentes serem

sistemas dinâmicos, demandando uma menor quantidade de regras Se-Então, o que facilita substancialmente a tarefa de identificação de modelos. Sob o ponto de vista de controle, modelos TS permitem conciliar técnicas de controle não-linear e, ao

mesmo tempo, estender de forma rigorosa resultados da teoria de controle de sistemas lineares.

[http://www.cpdee.ufmg.br/~palhares/dissertacao_mozelli_final.pdf]

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

(3.9) 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 30° à esquerda (SE): perto (P), médio (M), longe (L)

Sensor 30° 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)

Solução:

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.

Formigas têm a habilidade de encontrar a rota mais curta entre fontes de alimentos e a colônia: andam sem rumo, encontram alimento, retornam à colônia deixando um rastro de feromônio. Elas dão preferência a caminhos com alto nível de feromônios.

Exemplo:

Passo 1. As formigas artificiais são colocadas na cidade inicial;

Passo 2. Desconsidera cidades já visitadas, $q_0 = 0.9$, número aleatório n entre 0 e 1.

Se $n < q_0$, escolhe a melhor cidade baseado em distância e quantidade de feromônio.

Se $q_0 < n$, explora novos caminhos de maneira randômica.

Passo 3. Na nova cidade, é retirado um pouco de feromônio para evitar que um vértice com muito feromônio seja escolhido por todas as formigas. Se não for a última cidade ainda, volte ao passo 2.

Passo 4. A formiga que percorreu o menor caminho coloca feromônio em todos os vértices por onde passou. A quantidade de feromônio depositado é inversamente proporcional ao tamanho do percurso.

(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.

(4.3) Explique como se dá o cálculo das novas posições das partículas num algoritmo de enxame, como o PSO (do inglês particle swarm optimization).

[Fonte](#) (primeiro paragrafo)

(4.4) A resposta à questão 4.3 envolve o conceito de vizinhança entre as partículas. Quais são os principais tipos de vizinhança adotados em PSO?

[Fonte](#) (paragrafo **Neighbourhoods and Topologies**)

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

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

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

[Fonte](#)

Tópico 5 – Lógica Matemática, Representação e Inferência

(5.1) O que é uma proposição?

Uma proposição (ou sentença declarativa) é uma afirmação que é verdadeira ou falsa, mas nunca ambos.

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

Trata-se de uma sentença sintaticamente bem-formada, mas semanticamente sem sentido algum. sentença declarativa é também chamada de 'fórmula bem-formada'

(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?

Dedução lógica.

(5.7) 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.

(5.8) Dada a regra da resolução:

Resolução no pdf.

(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 no pdf.

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

Resolução no pdf.

(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>
- (d) Se (2) e (3) então (1). INDUÇÃO
- (b) Se (1) e (3) então (2). DEDUÇÃO
- (e) Se (2) e (4) então (1). DEDUÇÃO
- (c) Se (1) e (4) então (2). ABDUÇÃO

Explicação no pdf do tópico 5 nas páginas 42-45

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

Solução: Uma proposição abdutiva 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.

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

(5.15) O que é uma falácia lógica?

- Considerar uma proposição como sendo verdadeira simplesmente pelo fato dela não ter sido provada como falsa, ou então considerar uma proposição como sendo falsa simplesmente pelo fato dela não ter sido provada como verdadeira.
- A falta de evidência associada a uma certa proposição constitui uma prova de que uma proposição alternativa é verdadeira.

- Ano anterior

(5.12) - O que é necessário para se realizar uma prova por refutação de uma determinada proposição, dada a base de conhecimento?

Adicionar o negativo da preposição na base de conhecimento e antes de manipular essa base, as proposições tem que ser convertidas na forma normal disjuntiva.

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.

Regras e fatos. O sistema utiliza-se dos fatos tomados como verdade e deduz novos fatos até atingir um objetivo.

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

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?

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.

Antigos

(6.6) O que é um flat file?

Flat-file é uma matriz de objetos e seus respectivos atributos usada em árvores de decisão. Nela, objetivos distintos não podem requerer coleções distintas de atributos.

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

Contínuo: assumem valores numéricos em intervalos no eixo dos números reais

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

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

(6.8) 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).

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

É o ganho de informação relativo (ponderado).

Lore:

Na primeira etapa, é calculado o ganho de informação para todos os atributos. Após isso, considerar apenas aqueles atributos que obtiveram um ganho de informação acima da média, e então escolher aquele que apresentar a melhor razão de ganho. A razão de ganho supera o ganho de informação tanto em termos de acurácia quanto em termos de complexidade das árvores de decisão geradas.

(6.10) 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 ou ID3: É uma partição binária, mas 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.

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

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. Esse tipo de partição é uma das que foram implementadas para o algoritmo CART (BREIMAN et al., 1984).

C4.5:

atributos numéricos:

Testes simples ou pesquisa exaustiva.

Testes múltiplos (segmentação global e segmentação ao nível do nó)

Combinação linear de características

C4.5 para contínuo: define um limiar e então divide os exemplos de forma binária: aqueles cujo valor do atributo é maior que o limiar e aqueles cujo valor do atributo é menor ou igual ao limiar.

categóricos ordinais:

Partição(Cart): É possível fazer uma partição binária dos atributos de forma que atributos com valores dentro de uma faixa escolhida seguem para uma aresta, caso contrário, vão para outra aresta.

C4.5 também é possível para ordinais.

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

Evita que a árvore se torne muito especializada e perda a capacidade de generalização. A poda melhora a taxa de acerto do modelo para novos exemplos.

A poda também simplifica a árvore, facilitando o entendimento da mesma.

O ganho da informação pode ser utilizado como critério de poda.

Tópico 7 – Representação do Conhecimento e Solução de Problemas

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

A seleção de variáveis tenta escolher o melhor subset de preditores. Com ela, é possível mitigar os seguintes problemas:

Preditores Desnecessários: adicionam ruído à estimativa, perdendo-se graus de liberdade

Preditores Colineares: Quando existem muitas variáveis executando o mesmo trabalho

Custo: Removendo preditores redundantes (colineares) ou desnecessários, reduz-se o custo da estimativa.

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

É o processo de seleção de subsets de features relevantes à construção do modelo.

A principal premissa, é que os dados contém atributos redundantes ou irrelevantes. Os redundantes não adicionam nenhuma informação além das já selecionadas, e as irrelevantes não adicionam informação útil.

Tópico 8 - Estruturas e estratégias de busca

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

Basicamente, busca informada utiliza de estratégias para direcionar tomadas de decisão durante a busca, visando otimizar caminhos e encontrar soluções mais rapidamente. Em outras palavras, utilizam heurísticas durante o percurso da árvore.

Busca não-informada se parece com *brute force* realizado em árvore. Todos os tipos de busca não informadas (profundidade, largura, profundidade restrita) saem percorrendo todos os caminhos possíveis, não utilizando heurísticas nesse processo, apenas percorrendo todos os recursos possíveis.

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

O algoritmo A* realiza busca em profundidade baseado em uma heurística, como por exemplo, olhar nos filhos do caminho que apresenta menor custo, porém caso apareçam dois filhos com o mesmo valor de heurística (custos iguais) este realiza uma busca em largura para analisar todos os filhos desses dois filhos.

(8.5) Explique o princípio de operação do simulated annealing (recozimento simulado).

A idéia 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 **decrece ao longo do processo de busca**.

O algoritmo inicia-se com a geração de uma solução inicial s , que pode ser obtida aleatoriamente ou de acordo com alguma outra heurística. Em seguida, é inicializado um parâmetro de temperatura T , o principal parâmetro de controle do recozimento simulado. A cada iteração, uma solução vizinha s' é tomada aleatoriamente do conjunto de soluções vizinhas $N(s)$ da solução atual.

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

É uma meta-heurística de escopo local provida da habilidade de contornar pontos de ótimo local prematuros. Isso é realizado via uma memória seletiva que reflete o histórico da busca, considerando soluções obtidas no decorrer do processo de busca. Assim, restrições que delimitam o espaço de busca são sistematicamente impostas de modo que certas soluções se tornem proibidas ou tabu, no intuito de varrer regiões ainda inexploradas do espaço de soluções.

Tópico 9 – Teoria de Jogos

(9.1) Dadas as regras do jogo e as estratégias dos 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?

Num jogo de soma nula, o ganho de um jogador é representado pela perda dos demais.

Estratégia mista visa escolher uma dentre as possíveis soluções, aleatoriamente, com atribuição fixa de possibilidade.

A estratégia mista envolve a escolha aleatória de uma dentre as estratégias admissíveis, seguindo uma atribuição fixa de probabilidade para a escolha de cada estratégia.

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.

Ver exemplo da água e veneno no slide de Fuzzy

(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 evento?

Um evento é um subconjunto do espaço amostral.

Exemplos:

Experimento aleatório 1: “Jogar uma moeda três vezes e observar quantas vezes aparece cara e quantas vezes aparece coroa”.

Evento: “Aparecer 2 caras (k) e uma coroa (c) em qualquer ordem”.

Subconjunto: {(kkc), (kck), (ckk)}

(10.4) 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.

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

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

Cada nó possui uma tabela de probabilidades, dizendo as chances de ocorrer o evento representado por ele (inclusive probabilidade condicional).

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

(10.8) 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?

(10.9) 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 P1, 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:

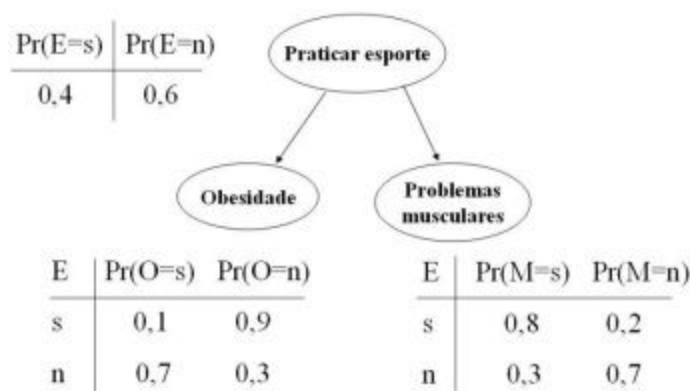
$$\Pr(P_1 | Q) = \frac{0,2 * 0,2}{0,2 * 0,2 + 0,05 * 0,3 + 0,02 * 0,5} = \frac{0,04}{0,04 + 0,015 + 0,01} = \frac{0,04}{0,065} \cong 61,54\%$$

(10.10) Um equipe da área de saúde coletou dados de campo e desenvolveu uma rede bayesiana visando explicar as relações de causa-efeito entre $E = \langle \text{Praticar Esporte} \rangle$, $O = \langle \text{Ocorrência de Obesidade} \rangle$ e $M = \langle \text{Ocorrência de Problemas Musculares} \rangle$. 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. Sugestão: aplique a seguinte formulação para o teorema de Bayes:

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



Resolução: Embora seja possível empregar também a formulação da Questão 1(a), iremos adotar a sugestão, o que nos remete a calcular $\Pr(E = s, M = s)$ e $\Pr(M = s)$.

Cálculo de $\Pr(E = s, M = s)$:

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

ou de modo alternativo:

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

Cálculo de $\Pr(M = s)$:

$$\Pr(M = s) = \Pr(E = s, M = s, O = s) + \Pr(E = s, M = s, O = n) + \Pr(E = n, M = s, O = s) + \Pr(E = n, M = s, O = n) = 0,4 \cdot 0,8 \cdot 0,1 + 0,4 \cdot 0,8 \cdot 0,9 + 0,6 \cdot 0,3 \cdot 0,7 + 0,6 \cdot 0,3 \cdot 0,3 = 0,32 + 0,18 = 0,5$$

ou de modo alternativo:

$$\Pr(M = s) = \Pr(M = s | E = s) \Pr(E = s) + \Pr(M = s | E = n) \Pr(E = n) = 0,8 \cdot 0,4 + 0,3 \cdot 0,6 = 0,32 + 0,18 = 0,5$$

Logo:

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

(10.11) Qual a diferença entre os frequentistas e os bayesianos?

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.

(10.12) 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.

Tópico 11 Sistemas complexos e Redes complexas

PS. é usado em estudo de redes sociais, interligação de pessoas em redes e muita informação é gerada com isso.

(11.0) O que é um sistema complexo?

“Um sistema no qual grandes redes de componentes sem controle central e com regras de operação simples leva a um comportamento coletivo complexo, processamento sofisticado de informação e adaptação via aprendizado ou evolução”

(Melanie Mitchell - Complexity: A guided tour)

“Formigas fazendo ponte com os próprios corpos para outras atravessarem”

(11.1) O que é reducionismo? Que tipos de vantagens e problemas podem surgir de uma análise baseada neste ponto de vista?

Consiste de particionar sistemas com alta complexidade em pedaços cada vez menores, ate que seja possível obter a solução para esses pedaços.

A solução para o sistema geral consiste na combinação das soluções parciais.

QUENTE! (11.9) Quais as três principais características não-triviais compartilhadas por várias redes complexas? Descreva-as.

Small-World: a maioria dos nós não são vizinhos porém qualquer nó pode ser alcançado a partir de outro através de poucos hops(saltos). Isso gera o conceito de mundo pequeno. Em uma rede small-world a distância L entre dois nós aleatoriamente escolhidos (número de pulos - hops) cresce proporcionalmente ao logaritmo do numero de N dessa rede: $L \propto \log N$.

O resultado disso são redes com alta clusterização e pequenas distâncias médias.

Lei de Potência: a probabilidade de um nó ter grau k é proporcional a k^γ (onde γ é uma constante). Isso implica que a maior parte dos nós tem poucas conexões, enquanto que há um conjunto pequeno de nós que tem muitas conexões (esses são os hubs).

Presença de comunidades: grupos de vértices com grande numero de conexões internas e poucas conexões externas ao grupo.

Tópico 12 - Agentes Inteligentes e Robótica Autônoma

(12.1) O que é um agente computacional?

Um agente é qualquer entidade capaz de perceber o mundo exterior através de sensores e agir neste mundo empregando atuadores.

(12.2) Qual é a diferença entre automático e autônomo?

Automático possui mecanismos pré-definidos de associação entre percepção (entrada) e ação (saída).

Um indivíduo autônomo, tem o poder de auto-definir o mecanismo de controle.

Carro com mapa, processa um caminho -> Automático

Carro sem mapa do terreno, mas que analisa o ambiente e reage a ele, criando um caminho -> Autônomo

(12.3) Quais são as diferenças e semelhanças entre adaptação e aprendizado?

Adaptação é a curto prazo. Aprendizado leva tempo. Ambos são processos de ajuste do comportamento.

(12.4) Quais são as diferenças entre agentes reativos e deliberativos (ou não-reativos)?

Reativos reagem aos estímulos do ambiente.

Não-reativos, usam também o histórico de ações.

(12.5) O que são agentes inteligentes?

Agentes inteligentes **executam** continuamente **três funções**: **percebem** as condições dinâmicas do ambiente, **agem** alterando as condições do ambiente e **raciocinam** de modo a interpretar percepções, **resolver problemas**, **fazer inferências** e **determinar ações**.

(12.6) Forneça um exemplo de um robô virtual para a Web 3.0, destacando 3 aspectos: (1) o que o robô coleta de dados (sensoriamento do ambiente); (2) que ações o robô executa (atuação no ambiente); (3) quais são os objetivos do robô.

(12.7) O que você entende por robótica evolutiva?

Metodologia que usa a **Computação Evolutiva para desenvolver robôs autônomos**. A partir de populações aleatórias de candidatos, repetidas modificações são realizadas de acordo com uma função de avaliação até que ao menos um dos novos indivíduos satisfaça requisitos mínimos.

(12.8) Dada a tarefa de síntese de um controlador autônomo para um robô móvel com estrutura física já definida, descreva todas as etapas de uma metodologia que permite sintetizar o cérebro do robô empregando uma rede neural MLP e computação evolutiva.

Define vários cérebros (MLPs com pesos distintos). Essas serão as soluções candidatas para o algoritmo evolutivo. O fitness é calculado colocando cada cérebro no robô e verificando o número de colisões. $fit(c_i) = 1 / (\# \text{ colisões})$. A partir daí empregam-se os operadores padrão da computação evolutiva (mutação, cross-over, etc).

(12.9) O que foi “The DARPA Grand Challenge” e por que esse assunto é relevante no contexto de Inteligência Artificial?

Ele instigou a criação dos primeiros robôs de navegação autônoma, e deu um grande impulso ao estudo e evolução de técnicas de aprendizado e autonomia para sistemas robóticos. Foi uma aplicação prática da teoria que vinha se desenvolvendo a muitos anos pela academia.

(12.10) O que você entende por robótica coletiva?

É tudo que se refere a robótica que se aproveita da iteração de vários indivíduos entre si para criação de uma consciência de colônia, que faz com que indivíduos simples possam ter um entendimento complexo do ambiente.

(12.11) Descreva uma aplicação prática em que a robótica coletiva é vantajosa quando comparada ao emprego de um único robô.

Varredura de terreno, vigilância, limpeza de área.