Raciocínio automático em situações de incerteza e imprecisão

Raciocínio automático em situações de incerteza ou imprecisão: principais técnicas e algumas aplicações.

Prof. Eugênio Silva

Propósito

Conhecer os conceitos e as técnicas por trás do raciocínio automático em situações de incerteza ou imprecisão é essencial para que os profissionais de Tecnologia da Informação (TI) possam desenvolver sistemas de apoio à tomada de decisão para diferentes áreas do saber.

Objetivos

- Empregar o raciocínio probabilístico em situações práticas de incerteza.
- Empregar o raciocínio nebuloso em situações práticas de imprecisão.

Introdução

O raciocínio automático, ou seja, efetuado por meio de um computador, é um recurso extremamente útil, sobretudo para a construção de sistemas computacionais de apoio à tomada de decisão em uma infinidade de áreas do conhecimento.

Para muitas delas, o conhecimento já reunido a respeito do domínio de aplicação é incerto ou impreciso. Para essas situações, as estratégias de inferência automática baseiam-se, respectivamente, em **teoria de probabilidades** e **lógica nebulosa**.

Neste conteúdo, teremos a oportunidade não apenas de conhecer os conceitos básicos por trás desses recursos teóricos, mas também de entender como podem ser empregados na construção de modelos de raciocínio automático para apoio à tomada de decisão em situações reais.

Vamos juntos em busca do conhecimento sobre esse assunto tão interessante!

Raciocínio automático em situações de incerteza

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Quando se trata de raciocínio automático, talvez a primeira abordagem que nos venha à cabeça seja aquela baseada na lógica clássica, que compreende a lógica proposicional e a lógica de predicados.

Uma proposição lógica pode ter valor verdadeiro ou falso

Na linguagem da lógica, a realidade é representada por meio de **proposições**, que são sentenças declarativas que assumem um dentre dois valores possíveis: *verdadeiro* ou *falso*.

Nesse caso, para que seja possível atribuir um valor verdadeiro ou falso a alguma afirmação sobre a realidade, é necessário ter um conhecimento pleno dessa realidade. A partir desse conhecimento, é possível fazer inferências do tipo:

Se *uma afirmação A é verdadeira,* ***então****, uma afirmação B também é verdadeira.*

O esquema de raciocínio dedutivo que está por trás da lógica é de extrema utilidade e está presente, sobretudo, em áreas como Matemática e Computação. Entretanto, essa forma de raciocínio automático está longe de ser aplicável a todas as situações práticas.

Há várias situações do cotidiano, talvez até a maioria, em que o conhecimento sobre a realidade não é pleno, e sim parcial, ou seja, há alguma incerteza associada àquele conhecimento. Em situações como essa, o mecanismo dedutivo da lógica tradicional não se aplica, pois as conclusões só podem ser inferidas quando temos certeza da veracidade das afirmações anteriores.

> Em situações que envolvem incerteza, empregamos outro tipo de esquema de raciocínio, que se baseia no conceito de probabilidade.

Lógica

Neste vídeo, vamos explorar a lógica proposicional e seus operadores fundamentais. Aprenderemos sobre a representação de proposições e como combinar e avaliar seu valor lógico usando operadores como E, OU, SE ENTÃO e SE SOMENTE SE.

Discutiremos a importância da lógica proposicional no raciocínio lógico e em sistemas de inteligência artificial.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Antes de entendermos o que é **incerteza** e como raciocinar em situações que a envolvem, vale a pena recordar rapidamente alguns conceitos básicos de lógica, especificamente de lógica proposicional.

Comecemos pelo conceito mais elementar no campo da lógica: proposição.

Atenção

Proposição é qualquer sentença declarativa à qual faz sentido atribuir um valor verdadeiro ou falso. De outro modo, a proposição é o recurso utilizado pela lógica para representar um conhecimento pleno sobre uma parte da realidade. Por exemplo, a sentença "O dia está ensolarado" é uma proposição. Afinal, faz sentido atribuir a ela um valor verdadeiro ou falso, o qual representa o conhecimento pleno que temos sobre o fato expresso na sentença.

Uma proposição simples traz um conhecimento muito restrito acerca da realidade. Para estendê-lo, podemos construir proposições compostas, que são formadas pela combinação de duas ou mais proposições simples por meio de conectivos (operadores) lógicos.

Proposições simples podem ser combinadas por meio de operadores, os quais apresentamos a seguir com suas respectivas simbologias:

"e"

"ou"

"se-então"

"se somente se"

Os operadores apresentados são todos binários, ou seja, combinam duas proposições que podem ser simples ou compostas. Há, ainda, o operador unário "não", simbolizado por $\neg$, que atua sobre apenas uma proposição (simples ou composta). Agora que conhecemos os operadores, veja como duas proposições simples são combinadas em uma proposição composta pelo operador "ou":

Proposições p ou q

Da mesma forma que uma proposição simples, uma proposição composta também é valorada como *verdadeiro* ou *falso*, mas, para isso, é preciso levar em consideração o valor de cada proposição simples que a compõe e a interpretação do(s) conectivo(s) envolvido(s).

A forma mais prática de observar como os operadores se comportam é por meio de uma **tabela-verdade**. Nela, são listadas todas as combinações possíveis de *verdadeiro* (V) e *falso* (F) para duas proposições simples A e B quaisquer. Para cada combinação, é apresentado o resultado de cada operação. Observe, a seguir, a tabela-verdade dos operadores lógicos:

A	B	¬A	A ∧ B	A ∨ B	A → B	A ↔ B
F	F	V	F	F	V	V
F	V	V	F	V	V	F
V	F	F	F	V	F	F
V	V	F	V	V	V	V

Tabela: tabela-verdade
Elaborada por: Eugênio Silva.

A mesma tabela-verdade que nos permite interpretar o comportamento dos operadores lógicos também é um instrumento bastante útil para a inferência. Em lógica, para que possamos verificar se um conhecimento pode ser inferido a partir de um conhecimento prévio, o primeiro passo é representar todo esse conhecimento por meio de proposições e essas proposições por meio de símbolos.

As proposições que representam o conhecimento prévio são denominadas premissas, simbolizadas por **P_1, P_2, ..., P_n**. Aquela que representa o que queremos inferir é denominada **conclusão** e simbolizada por **C.** O passo seguinte consiste em verificar se a proposição composta a seguir é uma **tautologia**:

$$(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \rightarrow C$$

Uma tautologia é uma proposição cuja interpretação é sempre *Verdadeira*, independentemente dos valores V e F das proposições simples que a compõem. Para fins de ilustração, consideremos as seguintes premissas:

P_1: “Se estudo, então, aprendo.”

P_2: “Se não estudo, então, sou reprovado.”

> A partir dessas premissas, será que podemos chegar à conclusão a seguir?

C: “Se não aprendo, então, sou reprovado.”

Para fazermos essa verificação, primeiramente, vamos reescrever as premissas e a conclusão em linguagem simbólica. Para isso, é necessário representar as proposições simples e, em seguida, as proposições compostas correspondentes às premissas e à conclusão:

Proposições simples	Proposições compostas
P : estudo	**P_1:P→Q**
Q : aprendo	**P_2:¬P→R**
R : reprovado	**C:¬Q→R**

Tabela: tabela-verdade
Elaborada por: Eugênio Silva.

O próximo e decisivo passo é verificar se $P_1 \wedge P_2 \rightarrow C$ é uma tautologia. Para isso, construímos a tabela-verdade a seguir:

			P_1	P_2	C	
P	**Q**	**R**	$P \rightarrow Q$	$\neg P \rightarrow R$	$\neg Q \rightarrow R$	$P_1 \wedge P_2 \rightarrow C$
F	F	F	V	F	F	V
F	F	V	V	V	V	V
F	V	F	V	F	V	V
F	V	V	V	V	V	V
V	F	F	F	V	F	V
V	F	V	F	V	V	V
V	V	F	V	V	V	V
V	V	V	V	V	V	V

Tabela: tabela-verdade
Elaborada por Eugênio Silva

Na tabela, observamos que a última coluna, correspondente à interpretação da proposição $P_1 \wedge P_2 \rightarrow C$ assume o valor V em todas as linhas, ou seja, é uma tautologia. Portanto, a partir das afirmações “Se estudo, então, aprendo” e “Se não estudo, então, sou reprovado”, podemos concluir que “Se não aprendo, então, sou reprovado”.

Outra forma de não apenas verificar se uma conclusão decorre das premissas, mas também de efetivamente derivar uma conclusão a partir das premissas, é por meio de regras de inferência e, eventualmente, de equivalências.

> As regras de inferência são tautologias conhecidas e empregadas sucessivamente, a fim de derivar proposições intermediárias, até que determinada conclusão seja alcançada.

Voltando ao exemplo anterior, podemos alcançar C a partir de P1 e P2, empregando a seguinte cadeia de raciocínio:

- A proposição $P \Rightarrow Q$ é equivalente à proposição $\neg P \Rightarrow \neg Q$.
- Considerado $\neg Q \Rightarrow \neg P$ e $\neg P \Rightarrow R$, existe uma regra de inferência denominada silogismo hipotético, que permite derivar a conclusão $\neg Q \Rightarrow R$.

A lista de regras é ampla, e a identificação da mais apropriada a cada passo do processo de inferência nem sempre é uma tarefa fácil. Às vezes, a escolha da regra mais adequada requer até mesmo certa dose de criatividade, o que, consequentemente, dificulta muito a automatização desse processo. Contudo, a automatização torna-se viável quando empregamos o **princípio da resolução**, que compreende a aplicação sucessiva de uma regra de inferência específica, até que uma conclusão seja alcançada.

Incerteza

A incerteza está presente em situações em que a realidade não é plenamente observável. Mesmo se fosse, seria impraticável considerar todas as influências que permitam construir regras capazes de concluir resultados completamente livres de quaisquer dúvidas.

Exemplo

Imagine que um dentista precise emitir um diagnóstico para um paciente com base na evidência de que ele apresenta dor de dente (RUSSEL; NORVIG, 2004). Pensando do ponto de vista da lógica tradicional, poderíamos imaginar a regra a seguir para concluir um diagnóstico: Se o paciente tem dor de dente, então, há uma Cárie naquele dente. O problema é que essa regra não é suficiente para conduzir o dentista a uma conclusão inquestionável. Afinal, nem toda dor de dente é provocada por uma Cárie, e nem toda Cárie provoca dor de dente. Nesse caso, precisaríamos incrementar um pouco mais essa regra para que as dúvidas envolvidas fossem sanadas. Assim, poderíamos pensar em uma regra do tipo: Se o paciente tem dor de dente, então, ele tem Cárie ou gengivite ou abcesso ou ... Observe que, no exemplo apresentado, para obtermos uma regra que nos conduza a uma conclusão desprovida de qualquer dúvida, precisamos listar uma quantidade enorme de possíveis causas de dor de dente. Mesmo assim, talvez não consigamos enumerar todas elas e haja causas de dor de dente que ainda nem sejam conhecidas. Afinal, dificilmente um domínio de conhecimento, em especial na área da saúde, está desvendado em sua totalidade a ponto de existir uma teoria completa sobre ele. Além disso, o paciente pode, por exemplo, ter Cárie e dor de dente, não havendo qualquer relação de causa e efeito entre elas.

O exemplo mostra que, mesmo nos esforçando ao máximo, dificilmente conseguiremos reunir conhecimento completo sobre determinado domínio. Essa falta de conhecimento sobre a realidade observada pode ocorrer por diferentes motivos, como (RUSSEL; NORVIG, 2004):

Preguiça

Enorme trabalho que teríamos caso quiséssemos considerar todas as evidências possíveis para chegar a uma conclusão indubitável.

Ignorância teórica

Dificilmente, teríamos à disposição uma teoria completa acerca do domínio em discussão.

Ignorância prática

Mesmo que todo o conhecimento do domínio esteja presente, talvez seja inviável reunir todas as evidências necessárias para chegar a uma conclusão única e totalmente correta.

Portanto, em situações como a do exemplo do diagnóstico, existe uma **incerteza**, caracterizada pela incompletude do conhecimento, que permeia o processo de raciocínio e impede que seja utilizado o mecanismo dedutivo da lógica tradicional para derivar conclusões.

Assim, com as evidências que temos em mãos, o que podemos obter como conclusão é um conjunto de hipóteses, que são as conclusões possíveis a partir daquelas evidências. A partir da análise dessas hipóteses, escolhemos aquela que parece ser a mais crível. Para medir esse grau de crença em uma hipótese, a ferramenta utilizada é a **probabilidade**.

Probabilidade

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

A probabilidade é um meio muito eficiente de medir nossa incerteza em relação a determinado fato.

Talvez, o dentista não saiba exatamente o que acomete seu paciente, mas acredite que haja uma chance de 80%, ou seja, uma probabilidade igual a 0,8 de que esse paciente tenha uma Cárie com base nas evidências apresentadas.

> E de onde veio esse valor de crença?

Em geral, vem de experiências prévias, vividas pelo especialista, que mostram que, em 80% das vezes em que um paciente apresentou os mesmos sintomas do paciente em análise, o resultado foi Cárie.

Aqui, é importante, mais uma vez, fazer uma comparação entre a abordagem lógica e a probabilística:

Lógica

Quando dizemos que o "paciente tem Cárie", essa afirmação (proposição) será verdadeira ou falsa de acordo com a observação do fato.

Probabilidade

Quando dizemos que existe uma "probabilidade de 0,8 de o paciente ter Cárie", isso não diz respeito ao fato, e sim à crença que temos em relação àquele fato. O fato (proposição) continua sendo verdadeiro ou falso. Nossa crença em sua veracidade é que assume valores que variam entre 0 e 1.

Estamos falando de probabilidade, mas ainda não dissemos exatamente o que é.

> De maneira geral, a probabilidade é uma forma de medir a chance de algo acontecer. Ela é dada pela razão entre a quantidade de eventos que nos interessam e a quantidade de eventos possíveis.

Um exemplo simples de cálculo de probabilidade está associado ao experimento de lançar um dado. Digamos que o resultado que nos interessa seja a face 4 do dado. Nesse caso, nossa probabilidade de sucesso será a razão entre 1 (apenas um evento nos interessa) e 6 (total de eventos possíveis), ou seja, aproximadamente 0,17. O conceito mais básico envolvido no estudo de probabilidades é o de variável **aleatória**: o elemento que representa a porção da realidade que ainda é desconhecida. No exemplo odontológico que analisamos, poderíamos considerar a variável aleatória *Cárie*, representando o fato de um dos dentes do paciente poder ter Cárie.

Toda variável aleatória tem um domínio, que compreende o conjunto de valores que essa variável pode assumir. De acordo com o tipo do domínio, as variáveis aleatórias podem ser classificadas como (RUSSEL; NORVIG, 2004):

Variáveis aleatórias booleanas

O domínio é composto por apenas dois valores: verdadeiro e falso. A variável Cárie, por exemplo, pode assumir o valor verdadeiro, representado por Cárie, ou falso, representado por $\neg$ Cárie.

Variáveis aleatórias discretas

O domínio é composto por um conjunto de valores enumeráveis e, portanto, tem as variáveis booleanas como um caso particular. Uma possível variável aleatória discreta seria *Tempo*, em que os valores possíveis poderiam ser *ensolarado, chuvoso, nublado* e *nevoento*.

Variáveis aleatórias contínuas

O domínio é formado por números reais compreendidos em um intervalo, de forma que entre quaisquer dois valores desse intervalo, há sempre uma quantidade infinita de valores. A variável *Altura*, por exemplo, pode assumir o valor 1,86.

Comentário

Aqui, estamos especialmente interessados em considerar apenas situações que envolvem variáveis discretas.

Como vimos, em lógica, as proposições simples, como “O paciente tem Cárie” e “O paciente tem dor dente”, podem ser combinadas por meio de conectivos lógicos para formar proposições compostas. Da mesma forma que podemos atribuir um grau de crença a uma proposição simples, também podemos fazer isso com uma proposição composta. Outro conceito importante envolvido no estudo de probabilidades é o de **evento atômico**: uma especificação completa da realidade sobre a qual não temos certeza acerca dos fatos. Em outras palavras, isso significa atribuir valores específicos a todas as variáveis envolvidas na descrição da realidade.

Recorrendo novamente ao exemplo odontológico, vamos supor que a realidade em análise seja representada por apenas duas variáveis: *Cárie* e *DorDeDente*. Suponhamos, ainda, que, assim como *Cárie, DorDeDente* também seja uma variável aleatória booleana. Nesse caso, um evento atômico seria (RUSSEL; NORVIG, 2004):

$$\text{Cárie} = \text{falso} \wedge \text{DorDeDente} = \text{verdadeiro}$$

Como a representação considera apenas duas variáveis, e ambas são booleanas, isso significa que há 4 eventos atômicos distintos, o que pode ser observado no quadro a seguir:

Cárie	*DorDeDente*
falso	*falso*
falso	*verdadeiro*
verdadeiro	*falso*

Cárie	*DorDeDente*
verdadeiro	*verdadeiro*

Quadro: Eventos atômicos para duas variáveis booleanas
Elaborada por Eugênio Silva

De posse desses conceitos mais básicos, agora temos condições de explorar os dois tipos de probabilidades: **incondicional** e **condicional**.

Probabilidade incondicional

A **probabilidade incondicional** (ou *a priori*) associada a uma proposição corresponde ao grau de crença que se atribui ao fato representado na proposição, considerando a ausência de qualquer informação adicional sobre ele.

Por exemplo, poderia haver algum tipo de acordo que permitisse atribuir o valor 0,1 à probabilidade de uma pessoa ter Cárie, o que seria representado por (RUSSEL; NORVIG, 2004):

$$P(\text{ Cárie } = \text{ verdadeiro }) = 0,1 \text{ ou } P(\text{ Cárie }) = 0,1$$

Considerando a possibilidade de associar um valor de probabilidade a cada valor que a variável aleatória pode assumir, temos, então, uma **distribuição** de **probabilidade**.

Por exemplo, no caso da variável discreta Tempo, poderíamos ter a seguinte distribuição (RUSSEL; NORVIG, 2004):

$$\begin{gathered}P(\text{ Tempo } = \text{ ensolarado }) = 0,7 \\ P(\text{ Tempo } = \text{ chuvoso }) = 0,2 \\ P(\text{ Tempo } = \text{ nublado }) = 0,08 \\ P(\text{ Tempo } = \text{ nevoento }) = 0,02\end{gathered}$$

Há, também, outro tipo de distribuição de probabilidade, denominado **distribuição de probabilidade conjunta total**, que será de grande utilidade no processo de inferência. Trata-se de atribuir um valor de probabilidade a cada evento atômico da realidade em análise. Portanto, apresenta a especificação completa da incerteza em torno da realidade.

Para o exemplo apresentado no quadro **Eventos atômicos para duas variáveis booleanas**, bastaria acrescentar uma terceira coluna ao quadro com os valores de probabilidade.

Probabilidade condicional

A partir do momento em que obtemos alguma informação adicional (evidência) que possa influenciar a probabilidade associada ao valor de uma variável aleatória, a probabilidade incondicional deixa de fazer sentido. Nesse caso, entra em cena a **probabilidade condicional**, em que a probabilidade de ocorrência de um valor leva em consideração valores associados a outras variáveis aleatórias do domínio em estudo. Considere, por exemplo, a expressão a seguir:

$$P(\text{ Cárie } \mid \text{ DorDeDente }) = 0,8$$

Atenção

Cárie é uma forma reduzida de escrever Cárie = verdadeiro, e DorDeDente também segue a mesma ideia.

Essa expressão nos diz que a probabilidade de o paciente ter Cárie, dado que ele se queixa de dor de dente, é de 0,8. De outro modo, isso significa que, diante da evidência de que o paciente tem dor de dente, existe a probabilidade de 0,8 de ele ter Cárie.

Um aspecto interessante sobre probabilidades condicionais é que podem ser calculadas a partir de probabilidades incondicionais. A equação 1.1 mostra a expressão geral para esse cálculo:

Eq 1.1

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

Em que: a e b são proposições quaisquer. Obviamente, essa expressão só se aplica a situações em que $P(b)>0$.

A equação 1.1 pode ser reorganizada e se transformar na expressão da equação 1.2, que é denominada **regra do produto**:

Eq. 1.2

$$P(a \wedge b) = P(a \mid b) \cdot P(b)$$

Como a conjunção $(a \wedge b)$ é comutativa, ou seja, $(a \wedge b)$ é igual a $(b \wedge a)$, a equação 1.2 também pode ser escrita conforme a equação 1.3:

Eq. 1.3

$$P(a \wedge b) = P(b \mid a) \cdot P(a)$$

A partir das expressões das equações 1.2 e 1.3, é possível derivar a equação 1.4, que é conhecida como **regra de Bayes**:

Eq. 1.4

$$P(b \mid a) = \frac{P(a \mid b) \cdot P(b)}{P(a)}$$

A regra de Bayes é uma espécie de "cérebro" do processo de inferência em situações de incerteza e será explorada mais adiante.

Independência

No estudo de probabilidade condicional, é importante observar que, nem sempre, uma evidência vai exercer influência sobre a probabilidade associada a uma variável aleatória. Nessas situações, dizemos que as variáveis são **independentes**.

Imagine, por exemplo, dois lançamentos consecutivos de um dado. Parece bastante óbvio que o resultado obtido no primeiro lançamento não exercerá qualquer influência sobre o segundo. Afinal, são eventos que não guardam qualquer tipo de relação, ou seja, são totalmente independentes. Para situações como essa, a probabilidade condicional é dada pela equação 1.5:

Eq. 1.5

$$P(a \mid b) = P(a)$$

Com isso, a regra do produto passa a ser definida conforme a equação 1.6:

Eq. 1.6

$$P(a \wedge b) = P(a) \cdot P(b)$$

A identificação da existência ou não de independência entre variáveis está relacionada aos conhecimentos do domínio em que essas variáveis estão inseridas.

No domínio da odontologia, não há dúvida de que uma dor de dente vai influenciar a probabilidade de existência de uma Cárie. Em contrapartida, se considerarmos a variável aleatória Tempo como parte da descrição desse domínio, fica claro, também, que o fato de o tempo estar chuvoso ou ensolarado não tem qualquer influência sobre a probabilidade de o paciente ter uma Cárie. Da mesma forma, o fato de o paciente ter uma Cárie também não exerce qualquer influência sobre o tempo.

> A identificação das independências tem uma utilidade prática muito importante. Isso contribui para tornar a especificação da distribuição de probabilidade conjunta total mais simples.

Inferência com distribuições conjuntas totais

Neste vídeo, vamos explorar o poder da inferência com distribuições conjuntas totais. Aprenderemos sobre como utilizar o raciocínio probabilístico para realizar inferências e obter conclusões a partir de distribuições conjuntas de variáveis aleatórias.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Agora, vamos conhecer um método de **inferência probabilística** bastante simples, que usa uma distribuição conjunta total como "base de conhecimento" para a obtenção de respostas a consultas acerca do domínio em análise.

Voltando mais uma vez ao exemplo odontológico, vamos imaginar que, além das variáveis *Cárie* e *DorDeDente*, outra variável importante a ser considerada nesse contexto seja *Boticão*: uma variável booleana que representa a ferramenta usada por dentistas para extrair dentes. Digamos que, para essa situação, a base de conhecimento seja representada pela distribuição conjunta total descrita no quadro a seguir:

Cárie	DorDeDente	Boticão	Probabilidade
falso	*falso*	*falso*	0,576
falso	*falso*	*verdadeiro*	0,144
falso	*verdadeiro*	*falso*	0,064
falso	*verdadeiro*	*verdadeiro*	0,016
verdadeiro	*falso*	*falso*	0,008
verdadeiro	*falso*	*verdadeiro*	0,072
verdadeiro	*verdadeiro*	*falso*	0,012
verdadeiro	*verdadeiro*	*verdadeiro*	0,108

Quadro: Base de conhecimento do exemplo odontológico
Adaptado de: Russel; Norvig, 2004, por Eugênio Silva

A partir dessa distribuição, é possível calcular tanto probabilidades incondicionais quanto condicionais relacionadas a essa situação. Vejamos alguns exemplos de probabilidades incondicionais (RUSSEL; NORVIG, 2004):

Exemplo 1

$$P(\text{Cárie})$$

$$= 0,008 + 0,072 + 0,012 + 0,108$$

$$= 0,20$$

Observação: Basta somar as probabilidades da distribuição em que *Cárie = verdadeiro*.

Exemplo 2

$$P(\text{Cárie} \vee \text{DorDeDente})$$

$$= 0,064 + 0,016 + 0,008 + 0,072$$

$$+0,012 + 0,108$$

$$= 0,28$$

Exemplo 3

$$P(\text{Cárie} \wedge \neg\text{DorDeDente})$$

$$= 0,008 + 0,072 = 0,08$$

Observação: Basta somar as probabilidades da distribuição em que *Cárie = verdadeiro* e *DorDeDente = falso*.

Na maioria das situações práticas, o que realmente interessa são as probabilidades condicionais. Para extrairmos essas probabilidades da distribuição conjunta, basta efetuarmos os cálculos usando a equação 1.1. Vejamos um exemplo (RUSSEL; NORVIG, 2004):

$$P(\text{ Cárie } \mid \text{ DorDeDente }) = \frac{P(\text{ Cárie } \wedge \text{ DorDeDente })}{P(\text{ DorDeDente })} = \frac{0,012 + 0,108}{0,064 + 0,016 + 0,012 + 0,108} = 0,60$$

Como observamos, trata-se de um método de inferência bem simples, mas há limitações importantes que inviabilizam seu uso em problemas em que a quantidade de variáveis é muito grande. Não raro, lidamos com situações práticas em que a descrição de mundo leva em consideração um conjunto bem maior de variáveis aleatórias.

Se considerarmos, por exemplo, um problema modelado com 10 variáveis booleanas, o que está longe de ser um problema realmente complexo, a distribuição conjunta total seria formada por 1024 valores. Para muitas das combinações que formam a distribuição conjunta, certamente, não haveria qualquer tipo de conhecimento que permitisse avaliar seus valores de probabilidade.

⚠

Atenção

Uma forma de simplificar a construção da distribuição conjunta total é considerar, se houver essa possibilidade, as eventuais independências que existam entre as variáveis aleatórias do domínio. Nesse caso, a distribuição conjunta pode ser fatorada, de forma que a quantidade de valores de probabilidade a serem especificados pode diminuir drasticamente.

Voltando ao exemplo odontológico, imaginemos que o domínio do problema seja representado pelas variáveis:

Cárie

Tipo: booleana

DorDeDente

Tipo: booleana

Boticão

Tipo: booleana

Tempo

Variável aleatória

Sabendo que *Cárie*, *DorDeDente* e *Boticão* são booleanas, e que Tempo pode assumir quatro valores possíveis, a tabela correspondente à distribuição conjunta total teria 32 linhas, que corresponde ao produto:

$$2 * 2 * 2 * 4 = 32$$

Apesar de não ser uma tabela muito extensa, ainda assim, não seria uma tarefa fácil especificar os valores de probabilidade para várias das combinações possíveis de valores para as variáveis. Entretanto, já sabemos que problemas odontológicos não guardam qualquer relação com a meteorologia. Assim, podemos fatorar a distribuição conjunta em duas:

Com 8 linhas

Considerando as variáveis *Cárie*, *DorDeDente* e *Boticão*.

$\times$

Com 4 linhas

Considerando a variável aleatória *Tempo*.

Com isso, seria necessário especificar apenas 12 valores de probabilidade.

Apesar dessa possibilidade de simplificação, a identificação das independências, em geral, está longe de ser uma tarefa trivial. Em contrapartida, mesmo que as independências sejam bastante claras, os subconjuntos independentes ainda podem ser bem grandes.

Pelas razões apresentadas, a inferência por meio da distribuição conjunta total tem uma importância didática, mas seu uso em situações práticas, em geral complexas, é inviável devido ao custo de construção da distribuição. Para situações práticas, a regra de Bayes é muito mais adequada.

Inferência com a regra de Bayes

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Em várias situações práticas, de diversas áreas do conhecimento, estamos em busca de um diagnóstico com base em um conjunto de evidências. Por exemplo, um médico pode estar interessado em calcular a probabilidade de um paciente estar gripado (diagnóstico), visto que está com febre (evidência). Para uma situação como essa, a regra de Bayes (equação 1.4) mostra-se bastante apropriada, porque, em geral, as outras probabilidades que precisam ser conhecidas para calcular a probabilidade desejada podem ser obtidas com mais facilidade.

A busca por diagnóstico com base em evidências é uma situação corriqueira em nossas vidas

Para o caso apresentado, temos a seguinte expressão:

$$P(\text{ gripe } \mid \text{ febre }) = \frac{P(\text{ febre } \mid \text{ gripe }) \cdot P(\text{ gripe })}{P(\text{ febre })}$$

Para esse cálculo, supostamente, os órgãos de saúde ou a literatura médica devem fornecer informações que permitam saber, de antemão, a probabilidade de uma pessoa ter febre (não importando a razão) ou gripe. Além disso, a probabilidade de um paciente gripado ter febre também deve ser uma informação que pode ser obtida com certa facilidade. Digamos que as probabilidades sejam as seguintes (COPPIN, 2010):

$$P(\text{ febre }) = 0,001$$
$$P(\text{ gripe }) = 0,0001$$
$$P(\text{ febre } \mid \text{ gripe }) = 0,8$$

Portanto:

$$P(\text{ gripe } \mid \text{ febre }) = \frac{0,8 \cdot 0,0001}{0,001} = 0,08$$

O que há de mais interessante a se observar no resultado obtido é que o fato de um paciente estar com febre não significa, necessariamente, que haja uma alta probabilidade de estar gripado. Isso acontece, porque, embora a febre seja um forte indício de gripe (probabilidade 0,8), a probabilidade de ocorrência de uma gripe é muito pequena (0,0001), e a probabilidade de ocorrência de uma febre (0,001) é muito maior que a de uma gripe.

Atenção

A regra de Bayes pode ser estendida para lidar também com situações em que o diagnóstico é baseado em mais de uma evidência.

Poderíamos estar interessados em calcular as probabilidades associadas à variável *Cárie*, dadas as evidências *DorDeDente = verdadeiro* e *Botição = verdadeiro*. Contudo, a mesma limitação que observamos no caso das distribuições conjuntas também está presente aqui. À medida que a quantidade de variáveis de evidência aumenta, a quantidade de probabilidades condicionais que precisamos calcular aumenta de forma exponencial.

Para contornar essa limitação, precisamos encontrar uma forma de diminuir a quantidade de combinações entre as evidências para que o cálculo seja praticável.

Para isso, mais uma vez, vamos nos valer do conceito de independência.

Em problemas que envolvem o cálculo da probabilidade de um diagnóstico, em geral, é bastante claro que as evidências condicionam o diagnóstico.

As evidências condicionam o diagnóstico das probabilidades

> Mas o que podemos dizer quanto à influência de uma evidência sobre outra?

O fato de várias evidências influenciarem o diagnóstico não significa, necessariamente, que haja algum tipo de influência entre elas. A prática mostra que essas independências condicionais entre as evidências são bastante comuns.

No exemplo odontológico, parece claro que as variáveis de evidência *DorDeDente* e *Boticão* condicionam a variável *Cárie*. Entretanto, *DorDeDente* e *Boticão* são independentes, não importando se *Cárie* assume o valor *verdadeiro* ou *falso*.

Atenção

As asserções de independência condicional são a chave para tornar o raciocínio probabilístico aplicável a problemas reais que envolvem muitas variáveis aleatórias.

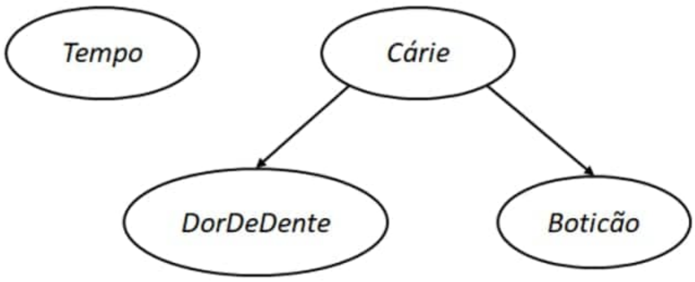
Redes bayesianas

As **redes bayesianas** são um recurso gráfico muito interessante que permite especificar, de forma bastante concisa, uma distribuição de probabilidade conjunta total.

> Uma rede bayesiana é um grafo direcionado e acíclico, em que os nós representam as variáveis aleatórias tanto de evidência quanto de hipótese (diagnóstico), e as arestas representam as dependências que existem entre as variáveis.

Uma aresta direcionada de um nó X para um nó Y indica que a probabilidade de ocorrência de Y é condicionada por X. Além disso, a cada nó também está associada uma distribuição de probabilidade condicional que especifica, para aquele nó, as influências dos nós dos quais ele depende.

Voltemos novamente ao exemplo odontológico. Considerando as variáveis *Cárie*, *DorDeDente*, *Boticão* e *Tempo*, poderíamos imaginar uma rede bayesiana com a topologia apresentada na imagem a seguir (RUSSEL; NORVIG, 2004):



Topologia da rede bayesiana do problema odontológico

Como já havíamos discutido e conforme a imagem ilustra, a variável *Tempo* é independente de todas as outras, e *Cárie* influencia diretamente *DorDeDente* e *Boticão*. Além disso, *DorDeDente* e *Boticão* são condicionalmente independentes, dada a variável *Cárie*. Portanto, a topologia da rede especifica os relacionamentos de independência condicional observados no domínio da aplicação.

Existe uma “lógica” a ser seguida para definirmos adequadamente a topologia da rede. Repare que, na rede exibida na imagem anterior, as relações existentes entre *Cárie* e *DorDeDente* e entre *Cárie* e Boticão são **causais**, ou seja, uma Cárie causa dor de dente, assim como uma Cárie causa o uso do boticão. Além de seguir essa lógica causal, também é importante identificar a ordem correta de inclusão dos nós na rede.

Inclusão das causas
O processo inicia-se com a inclusão das causas que não são influenciadas por outras variáveis.

Inclusão das variáveis
Em seguida, são incluídas as variáveis que são influenciadas por aquelas inseridas no passo anterior.

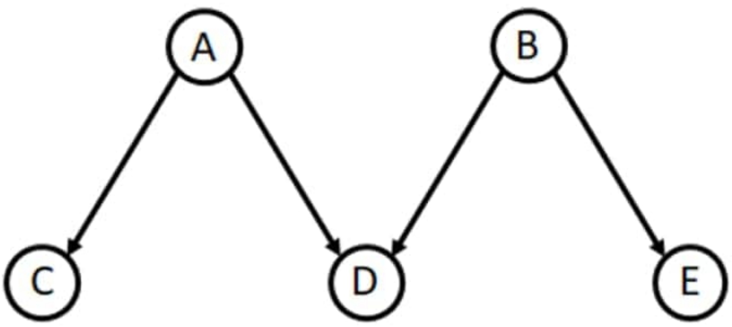
Inclusão dos nós folha
A etapa anterior se repete, até que sejam incluídos os nós folha, que representam as variáveis que não influenciam outras. Essa estratégia produz redes mais simples, ou seja, com menos dependências, o que, consequentemente, também simplifica o cálculo das distribuições de probabilidade condicionais associadas aos nós.

Especificação das probabilidades
Para completar a rede bayesiana, é necessário especificar, para cada nó da rede, a distribuição de probabilidade condicional de acordo com os nós que o condicionam. Em outras palavras, isso significa especificar as probabilidades de ocorrência da variável correspondente ao nó, considerando todas as combinações possíveis de valores para os nós superiores.

Com a combinação da topologia com as distribuições condicionais associadas a cada nó, temos, de forma implícita, a distribuição conjunta total de todas as variáveis. Assim, qualquer entrada na distribuição conjunta total pode ser calculada a partir das informações disponíveis na rede.

Já vimos que a distribuição conjunta total nos permite responder a qualquer pergunta sobre o domínio em análise. Como uma rede bayesiana é uma representação dessa distribuição, isso significa que a rede também pode responder a qualquer pergunta sobre o domínio.

Vejamos um exemplo de uma rede bayesiana com a especificação completa (COPPIN, 2010). A imagem a seguir apresenta a topologia da rede:



Topologia de uma rede bayesiana genérica

As tabelas a seguir apresentam as distribuições de probabilidades condicionais dos nós da rede:

P(A)
0,1

P(B)
0,7

A	P(C)
V	0,2
F	0,4

B	P(E)
V	0,2
F	0,1

A	B	P(D)
V	V	0,5
V	F	0,4
F	V	0,2
F	F	0,0001

Tabela: Probabilidades condicionais dos nós da rede

Observe que as distribuições de probabilidade foram especificadas levando em consideração as dependências identificadas na topologia da rede. Por exemplo, como A e B não dependem de outros nós, para eles, são definidas apenas as probabilidades incondicionais. O nó D depende dos nós A e B. Logo, sua distribuição deve considerar todas as combinações possíveis de *verdadeiro* e *falso* para A e B.

Com essas tabelas, temos toda a informação necessária para efetuar qualquer raciocínio sobre o domínio em análise, ou seja, temos a distribuição de probabilidade conjunta total, considerando todas as variáveis. Vejamos alguns exemplos de consultas que podem ser efetuadas na rede da **Topologia da rede bayesiana do problema odontológico**:

$$P(a,b,c,d,e) = P(e \mid b) \cdot P(d \mid a,b) \cdot P(c \mid a) \cdot P(b) \cdot P(a) = 0,2 \cdot 0,5 \cdot 0,2 \cdot 0,7 \cdot 0,1 = 0,0014$$

$$P(\neg a,b,\neg c,d,\neg e) = P(\neg e \mid b) \cdot P(d \mid \neg a,b) \cdot P(\neg c \mid \neg a) \cdot P(b) \cdot P(\neg a) = 0,8 \cdot 0,2 \cdot 0,6 \cdot 0,7 \cdot 0,9 = 0,0605$$

$$P(a,b,\neg c,\neg d,e) = P(e \mid b) \cdot P(\neg d \mid a,b) \cdot P(\neg c \mid a) \cdot P(b) \cdot P(a) = 0,2 \cdot 0,5 \cdot 0,8 \cdot 0,7 \cdot 0,1 = 0,0056$$

Os exemplos apresentados são estradas da distribuição de probabilidade conjunta total, mas também é possível responder a perguntas mais complexas.

Consideremos, por exemplo, o cálculo de $P(a \mid b, \neg c, d)$. Observe que se trata de uma probabilidade condicional, e que uma das evidências está oculta, uma vez que nenhum valor é atribuído à variável E. Nesse caso, o cálculo é efetuado segundo os passos a seguir:

Passo 1

Pela definição de probabilidade condicional dada pela equação 1.1, temos:

$$P(a \mid b, \neg c, d) = \frac{P(a,b,\neg c,d)}{P(b,\neg c,d)}$$

Passo 2

Para o numerador, como a variável E está oculta, é necessário considerar todos os valores possíveis que ela pode assumir:

$$
\begin{aligned}
&\text{numerador } = \\
&P(e \mid b) \cdot P(d \mid a, b) \cdot P(\neg c \mid a) \cdot P(b) \cdot P(a) \\
&+P(\neg e \mid b) \cdot P(d \mid a, b) \cdot P(\neg c \mid a) \cdot P(b) \cdot P(a)
\end{aligned}
$$

$$
\begin{aligned}
&\text{numerador } = 0,2 \cdot 0,5 \cdot 0,8 \cdot 0,7 \cdot 0,1 \\
&+0,8 \cdot 0,5 \cdot 0,8 \cdot 0,7 \cdot 0,1 \\
&= 0,0056 + 0,0224 = 0,028
\end{aligned}
$$

Passo 3

Para o denominador, como tanto a variável A quanto a variável E estão ocultas, é necessário considerar todas as combinações possíveis de V e F para essas duas variáveis:

denominador =

$$
\begin{gathered}
P(e \mid b) \cdot P(d \mid a, b) \cdot P(\neg c \mid a) \cdot P(b) \cdot P(a)+ \\
P(\neg e \mid b) \cdot P(d \mid a, b) \cdot P(\neg c \mid a) \cdot P(b) \cdot P(a)+ \\
P(e \mid b) \cdot P(d \mid \neg a, b) \cdot P(\neg c \mid \neg a) \cdot P(b) \cdot P(\neg a)+ \\
P(\neg e \mid b) \cdot P(d \mid \neg a, b) \cdot P(\neg c \mid \neg a) \cdot P(b) \cdot P(\neg a)
\end{gathered}
$$

denominador =
0,2 . 0,5 . 0,8 .0,7 . 0,1 + 0,8 . 0,8 . 0,7 . 0,1 + 0,2 .0,2 . 0,6 . 0,7 . 0,9 + 0,8 . 0,2 . 0,6 . 0,7 . 0,9
= 0,0056 + 0,0224 + 0,01512 + 0,06048 = 0,1036

Passo 4

Cálculo da consulta desejada:

$$P(a \mid b, \neg c, d) = \frac{0,028}{0,1036} = 0,27$$

Um detalhe a ser observado na rede apresentada no exemplo é que com as asserções de independência condicional, a quantidade de valores de probabilidade que precisaram ser calculados foi bem menor que os $32\left(2^5\right)$ que seriam calculados caso as independências condicionais não fossem consideradas. Isso mostra que a identificação das relações causais e das independências condicionais é fator essencial para que a quantidade de probabilidades conjuntas a serem calculadas se mantenha razoável, mesmo quando estamos lidando com problemas com muitas variáveis.

Utilizando redes bayesianas

Acompanhe a seguir algumas aplicações práticas de redes bayesianas. Vamos lá!

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

Questão 1

Considerando a distribuição de probabilidade conjunta total do problema odontológico, apresentada no quadro **Base de conhecimento do exemplo odontológico**, analise as afirmações a seguir:

I. $P(\text{ DorDeDente } \vee \text{ boticão }) = 0,20$

II. $P(\text{Cárie } \mid \text{ boticão}) = 0,53$

III. $P(\text{DorDeDente } \mid \neg \text{ Cárie}) = 0,40$

Está(ão) correta(s) apenas a(s) afirmação(ões:)

A

I e II.

B

II.

C

I e III.

D

II e III.

E

III.

A alternativa B está correta.

Item I - A afirmativa está incorreta.

$P(ext DorDeDente \vee$ boticão $)$ é dada pela soma das probabilidades das linhas da distribuição em que a variável DorDeDente = verdadeiro ou Boticão = verdadeiro, ou seja:

$P(ext$ DorDeDente $\vee$ boticão $) = 0,144 + 0,064 + 0,016 + 0,072 + 0,012 + 0,108 = 0,416$

Item II - A afirmativa está correta.

Pela definição de probabilidade condicional, temos que:

$P($ D*orDeDente* $\mid \varnothing$ Cárie $) = P($ DorDeDenteƯ $\emptyset$ Cárie $)/P(\varnothing$ Cárie)=(0,064+0,016)/(0,576+0,144+0,064+0,016)=0,08/0,8=0,10

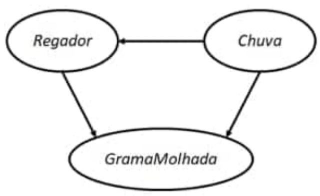
Item III - A afirmativa está incorreta.

Pela definição de probabilidade condicional, temos que:

$P($ DorDeDente $\mid \neg$ Cárie $) = P($ DorDeDente $\wedge\neg$ Cárie $)/P(\neg$ Cárie $)$ =(0,064+0,016) / (0,576+0,144+0,064+0,016)=0,08 / 0,8=0,10

Questão 2

Considere a rede bayesiana a seguir, que modela uma situação bastante específica no domínio da jardinagem, em que a grama pode ser molhada pela chuva ou por um regador:



P(Chuva)
0,2

Chuva	P(Regador)
V	0,01
F	0,4

Regador	Chuva	P(GramaMolhada)
V	V	0,99
V	F	0,9
F	V	0,8
F	F	0,0

Topologia de rede bayesiana e distribuições de probabilidade condicionais no nós.

Para a situação apresentada, qual é a probabilidade de estar chovendo, sabendo que a grama está molhada?

A

0,36

B

0,16

C

0,45

D

0,99

E

0,00

A alternativa A está correta.

O primeiro passo para calcular o resultado da consulta desejada é identificar a distribuição de probabilidade conjunta representada pela rede bayesiana. Para a situação apresentada, a distribuição conjunta é dada por:

$$P(G,R,C)=P(G \mid R,C)\cdot P(R \mid C)\cdot P(C)$$

Em que:

- G = GramaMolhada;
- R = Regador;
- C = Chuva.

Em seguida, é necessário identificar e calcular a probabilidade correspondente à pergunta:

$$P(c \mid g)=\frac{P(c \wedge g)}{P(g)}$$

Como no numerador, a variável R é oculta, temos:

$$\text{numerador} = P(g \mid r,c)\cdot P(r \mid c)\cdot P(c) + P(g \mid \neg r,c)\cdot P(\neg r \mid c)\cdot P(c) = 0{,}99\cdot 0{,}01\cdot 0{,}2+0{,}8\cdot 0{,}99\cdot 0{,}2=0{,}16038$$

Como no denominador, as variáveis R e C são ocultas, temos:

$$\begin{aligned}\text{Denominador} &= P(g \mid r,c)\cdot P(r \mid c)\cdot P(c) + P(g \mid \neg r,c)\cdot P(\neg r \mid c)\cdot P(c) \\ &+ P(g \mid r,\neg c)\cdot P(r \mid \neg c)\cdot P(\neg c) + P(g \mid \neg r,\neg c)\cdot P(\neg r \mid \neg c)\cdot P(\neg c) \\ &= 0{,}99\cdot 0{,}01\cdot 0{,}2 + 0{,}8\cdot 0{,}99\cdot 0{,}2 + 0{,}9\cdot 0{,}4\cdot 0{,}8 + 0{,}0\cdot 0{,}6\cdot 0{,}8 \\ &= 0{,}00198 + 0{,}1584 + 0{,}288 + 0{,}0 \\ &= 0{,}44838\end{aligned}$$

Portanto:

$$P(c \mid g)=\frac{0{,}16038}{0{,}44838}\approx 0{,}36$$

Raciocínio automático em situações de imprecisão

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Agora, vamos lidar com outro tipo de situação que pode se confundir com uma incerteza, mas que tem uma interpretação completamente diferente. São situações em que os eventos são imprecisos. Vamos entender a diferença:

Eventos de incerteza

São eventos que podem ou não acontecer com alguma chance, dada por uma medida de probabilidade que varia entre 0 e 1.

Eventos imprecisos

Quando se trata de eventos imprecisos, não estamos falando de algo que pode ou não acontecer, mas que acontece de fato, com determinada intensidade que varia entre 0 e 1.

Para lidar com situações que envolvem imprecisão, vamos nos valer de uma nova forma de lógica: a **lógica nebulosa**.

Conjuntos tradicionais

Neste vídeo, vamos aprender sobre os fundamentos dos conjuntos, incluindo sua definição e representação. Exploraremos as principais operações em conjuntos, como união, interseção, diferença e complemento.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Antes de conhecermos a definição e as características de um conjunto nebuloso, vale relembrarmos rapidamente o conceito de conjunto tradicional. Vamos lá!

Um conjunto tradicional, ou simplesmente um conjunto, pode ser entendido como uma coleção de elementos pertencentes a determinado domínio, ou seja, elementos que compartilham alguma(s) característica(s). Por exemplo, os alunos de determinada instituição de ensino superior formam um conjunto. Afinal, todos eles compartilham pelo menos uma característica: o fato de todos serem alunos daquela instituição.

Conjunto dos Alunos

Conjuntos dos alunos e subconjunto dos alunos do curso de Matemática

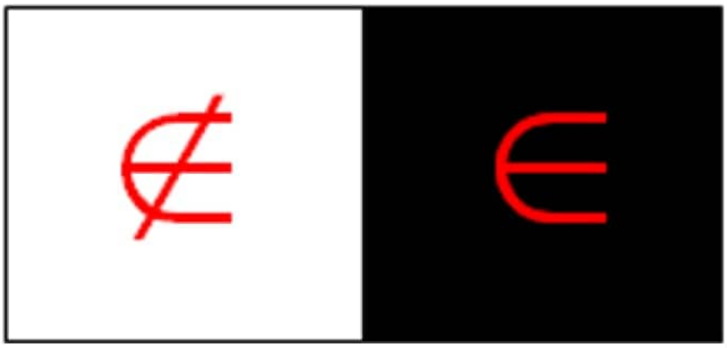
Poderíamos pensar, ainda, na possibilidade de criar subconjuntos desse conjunto de alunos, como, por exemplo, o conjunto dos alunos dessa instituição que estão matriculados em determinado curso.

Nesse caso, temos, dentro do primeiro conjunto, um conjunto com elementos que compartilham características ainda mais específicas. Esse tipo de conjunto é ideal para representar situações em que não há qualquer dificuldade de distinguir elementos que pertencem ao conjunto daqueles que não pertencem.

Vejamos outro exemplo para fixar o conceito, como o conjunto das laranjas. Aqui, temos um conjunto bastante preciso. Afinal, nesse conjunto, só há laranjas. Qualquer coisa diferente de laranja está fora do conjunto. E mais ainda: a distinção entre o que é e o que não é laranja é muito clara.

> Quando se trata de situações em que os elementos manipulados são realmente precisos, os conjuntos tradicionais são uma forma ideal de representação.

De outro modo, consideramos que, em um conjunto tradicional, existe uma fronteira muito nítida que separa os elementos pertencentes dos não pertencentes ao conjunto. A imagem a seguir ilustra essa situação:



Pertinência e não pertinência em conjuntos tradicionais

Repare que há uma transição abrupta entre a não pertinência e a pertinência ao conjunto, ou seja, o elemento está ou não está no conjunto.

De maneira mais formal, podemos definir um conjunto por sua função característica, dada pela equação 2.1:

Eq. 2.1

$$\mu : X \to \{0, 1\}$$
$$\mu_A(x) = \begin{cases} 1 & \text{se} \quad x \in A \\ 0 & \text{se} \quad x \notin A \end{cases}$$

Observe que a função μ estabelece um mapeamento entre elementos pertencentes a um conjunto X, denominado **universo de discurso**, em um dentre dois valores possíveis:

0 Quando o elemento não pertence ao conjunto.	1 Quando o elemento pertence ao conjunto.

Assim, a simbologia $\mu_A(x)$ é lida como o "grau de inclusão do elemento x no conjunto A, em que x é um elemento de $X(x \in X)$ e A é um subconjunto de $X(A \subseteq X)$. Veja o exemplo a seguir:

A= conjunto das alturas acima de 1,7m

$$X = \{x \in \mathbb{R}/0 \leq x \leq 3\}$$

$$\mu_A(x) = \begin{cases} 1 & \text{se} \quad x > 1,7 \\ 0 & \text{se} \quad x \leq 1,7 \end{cases}$$

Veja a representação gráfica:

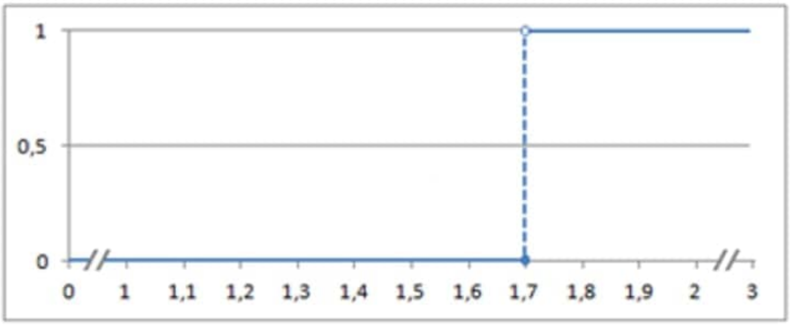
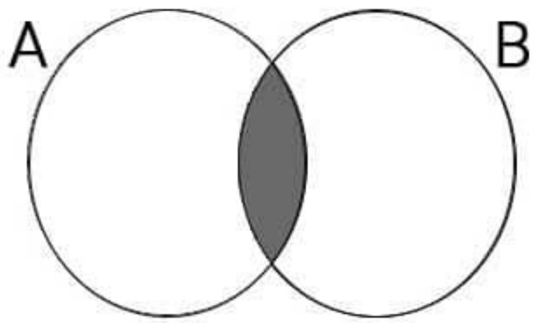


Gráfico: Conjunto das alturas acima de 1,7m

Observe a descontinuidade que existe na função exatamente onde acontece a transição entre o não pertencimento e o pertencimento.

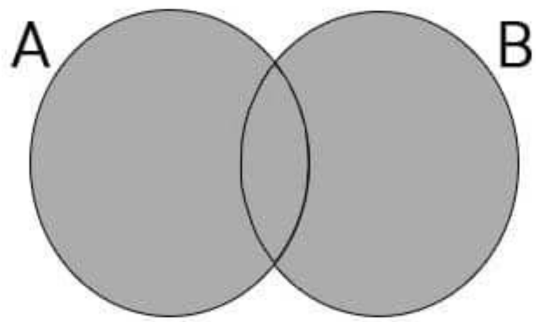
Operações com conjuntos tradicionais

Ainda com relação aos conjuntos tradicionais, vale relembrar as operações que podemos efetuar com esses conjuntos. Dados dois conjuntos A e B definidos no universo X, temos as operações resumidas a seguir:



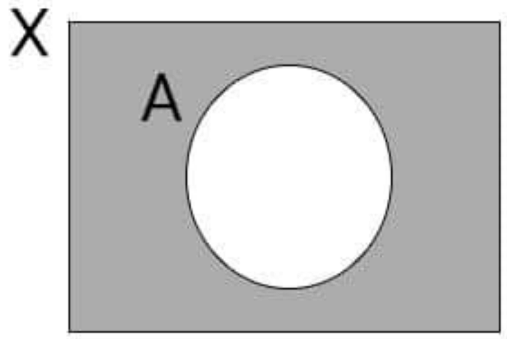
Interseção

$A \cap B = \{x/x \in A \quad \text{e} \quad x \in B\}$



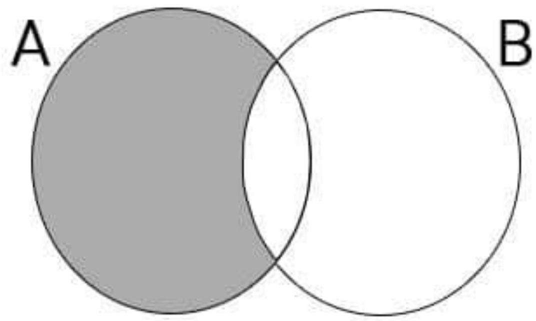
União

$A \cup B = \{x/x \in A \text{ ou } x \in B\}$



Complemento

$\bar{A} = \{x/x \in X \quad \text{e} \quad x \notin A\}$



Diferença

$A - B = \{x/x \in A \text{ e } x \notin B\}$

Assim como podemos efetuar operações com conjuntos tradicionais, também podemos efetuar operações com conjuntos nebulosos. Entretanto, elas precisam ser redefinidas para que se adéquem a esse novo tipo de conjunto.

Propriedades das operações com conjuntos tradicionais

Agora, vamos relembrar as propriedades envolvendo as operações com conjuntos tradicionais.

Dados três conjuntos A, B e C definidos no universo X, as propriedades estão resumidas no quadro a seguir:

Elemento neutro	$\begin{cases} A \cup \emptyset = A \\ A \cap X = A \end{cases}$

Idempotência	$\begin{cases} A \cup \emptyset = A \\ A \cap X = A \end{cases}$
Comutatividade	$\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \end{cases}$
Associatividade	$\begin{cases} A \cup (B \cup C) = (A \cup B) \cup C \\ A \cap (B \cap C) = (A \cap B) \cap C \end{cases}$
Dupla complementação	$\bar{A} = A$
Distributividade	$\begin{cases} A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \end{cases}$
Absorção	$\begin{cases} A \cup (A \cap B) = A \\ A \cap (A \cup B) = A \end{cases}$
Leis de De Morgan	$\begin{cases} (A \cup B) = \bar{A} \cap \bar{B} \\ (A \cap B) = \bar{A} \cup B \end{cases}$
Lei da exclusão do meio	$A \cup \bar{A} = X$
Lei da não contradição	$A \cap \bar{A} = \emptyset$

Quadro: Propriedades das operações com conjuntos tradicionais
Elaborado por: Eugênio Silva

Atenção às duas últimas propriedades:

Lei da exclusão do meio

Estabelece que um elemento está em um conjunto ou não está, ou seja, não é possível ficar “em cima do muro”. Observe que a palavra “ou” exprime a ideia de alternância.

Exemplo: O professor está ou não está em sala de aula.

Lei da não contradição

Estabelece que um elemento não pode estar e não estar em um conjunto. Nesse caso, a palavra “e” exprime a ideia de simultaneidade.

Exemplo: O professor não pode estar e não estar em sala de aula.

Atenção

A lei da exclusão do meio e a lei da não contradição retratam claramente a principal característica de um conjunto: há um limite bem definido que separa os elementos que pertencem dos que não pertencem ao conjunto.

Conjuntos nebulosos

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Os conjuntos tradicionais são ideias para a representação de situações precisas, em que a distinção entre os elementos que pertencem e os que não pertencem ao conjunto é bastante clara. Contudo, quando se trata de situações imprecisas, essa forma de representação já não se mostra tão adequada.

Para que isso fique mais claro, podemos considerar uma situação conhecida como paradoxo sorites, que consiste em uma pergunta:

> Paradoxo sorites: “Quando um monte de areia deixa de ser um monte de areia, caso seja retirado um grão de cada vez?”

Repare que, nesse caso, a transição entre o conceito de “monte de areia” para o conceito de “não monte de areia” ocorre de forma suave, e não de forma abrupta, ou seja, não existe um grão de areia decisivo que estabeleça que, com ele, temos um monte de areia, e, sem ele, não temos mais um monte de areia.

Determinar a transição de um “monte de areia” para “não monte de areia” é uma exemplo de situação imprecisa

> Para situações imprecisas, precisamos de outro tipo de representação que nos permita demonstrar melhor a imprecisão que existe na transição entre um conceito e outro. É aí que entram os chamados conjuntos nebulosos.

A maneira como os conjuntos nebulosos representam situações como essa é generalizando a noção de pertinência binária dos conjuntos tradicionais (0 ou 1) para a noção de pertinência parcial (entre 0 e 1). Com isso, é possível representar com maior fidelidade as situações que envolvem imprecisão.

Para facilitar a compreensão da distinção entre situações precisas e imprecisas, veja os exemplos a seguir:

Situação precisa: conjunto de laranjas

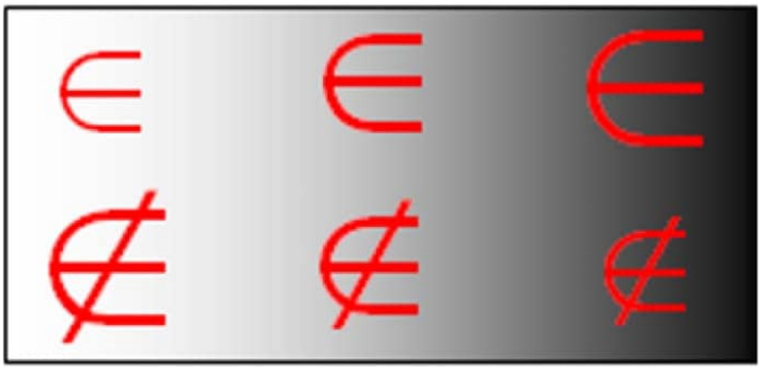
Uma laranja lima faz parte do conjunto de laranjas? Obviamente que sim. E quanto a uma maçã? Você certamente não hesitou em responder que não. Podemos dizer com precisão que uma maçã não faz parte do conjunto de laranjas. Estamos diante de uma situação precisa.

Situação imprecisa: conjunto de pessoas altas

Quando uma pessoa começa a ser alta? Podemos dizer que uma pessoa de 1,80 é alta? E se ela estiver ao lado de uma pessoa de 1,95? E uma pessoa de 1,60 pode ser considerada alta? E se essa pessoa de 1,60 tiver 10 anos de idade? A ideia de pessoa alta já não parece tão precisa quanto o conjunto de laranjas, certo?

Enquanto temos uma noção bastante clara do que é o conjunto das laranjas, não temos uma noção tão clara de quem está incluído no conjunto das pessoas altas. Nesse caso, o conceito "alto" é impreciso e, por isso, permite várias interpretações. Além disso, é fácil observar que, considerando a forma como interpretamos essas imprecisões, não parece razoável estabelecer um limite rígido que determine a fronteira entre o que é ser alto e o que não é.

De acordo com nosso senso crítico, parece que um conceito transita para o outro de forma suave. Durante essa transição, temos graus de pertinência ao conjunto que não se limitam apenas aos valores 0 e 1. Esses valores variam entre 0 e 1, permitindo, assim, que um elemento pertença parcialmente ao conjunto. A imagem a seguir ilustra uma situação como essa:



Pertinência e não pertinência em conjuntos nebulosos

Repare que a transição entre o pertence e o não pertence acontece de forma suave, de maneira que aumenta ou diminui de forma gradativa.

Assim como no caso dos conjuntos tradicionais, os conjuntos nebulosos também têm uma definição formal, por meio de sua função característica, dada pela equação 2.2:

Eq. 2.2

$$\mu : X \to [0,1] \quad \mu_A(x) = f\left(x, [a_1, a_2, \ldots, a_n]\right)$$

Observe que, diferentemente do que ocorre com os conjuntos tradicionais, aqui, a função μ mapeia elementos do universo X em algum valor compreendido entre 0 e 1. A simbologia $\mu_A(x)$ também é lida como o “grau de inclusão do elemento x no conjunto A em que x é um elemento de $X(x \in X)$ e A é um subconjunto de $X(A \subseteq X)$.

A função f, que é responsável por estabelecer o mapeamento, pode ter diferentes definições de acordo com a lista de coeficientes $[a_1, a_2, \ldots a_n]$.

Veja o exemplo a seguir:

$A =$ conjunto das alturas médias

$$X = \{x \in \mathbb{R} / 0 \leq x \leq 3\}$$

$$\mu_A(x) = \begin{cases} 0 & \text{se} \quad x \leq 1,5 \\ 5x - 7,5 & \text{se} \quad 1,51,9 \end{cases}$$

Representação gráfica:

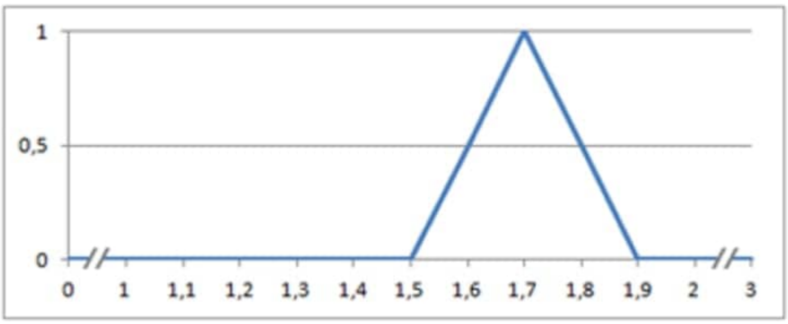


Gráfico: Conjunto das alturas médias

Observe que a pertinência aumenta suavemente para valores de x compreendidos entre 1,5 e 1,7, atingindo a pertinência máxima em x igual a 1,7 e, em seguida, diminui também suavemente entre 1,7 e 1,9.

Funções de inclusão

A função f que mapeia elementos de X no intervalo [0, 1], é denominada função de inclusão e pode assumir diferentes definições. Essas definições estabelecem a forma do conjunto nebuloso, e a forma determina como ocorrem as variações dos graus de inclusão dos valores de X no conjunto.

Há diversas funções de inclusão que podem usadas para representar conjuntos nebulosos, mas, aqui, vamos abordar apenas as mais tradicionais: as **funções triangulares** e as **trapezoidais**.

Saiba mais

Para obter detalhes sobre outras funções de inclusão, consulte as obras de Cox (1995), Oliveira Junior (1999) e Yen e Langari (1999).

Função triangular

A função triangular é definida segundo a equação 2.3:

Eq. 2.3

$$\text{triang}(x, a, b, c) = \begin{cases} 0 & \text{se} \quad x \le a \\ (x-a)/(b-a) & \text{se} \quad ac \end{cases}$$

Veja o exemplo:

$$X = \{\{x \in \mathbb{R} / 0 \le x \le 100\}$$

$$\text{triang}(x, 20, 60, 80)$$

Representação gráfica:

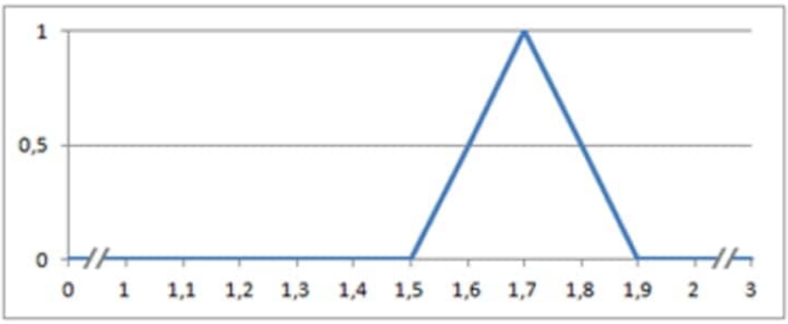


Gráfico: Função triangular

Função trapezoidal

A função trapezoidal é definida segundo a equação 2.4:

Eq. 2.4

$$\text{trap}(x, a, b, c, d) = \begin{cases} 0 & \text{se} \quad x \leq a \\ (x-a)/(b-a) & \text{se} \quad ad \end{cases}$$

Veja o exemplo:

$$X = \{x \in \mathbb{R} / 0 \leq x \leq 100\}$$

$$\text{trap}(x, 10, 20, 60, 90)$$

Representação gráfica:

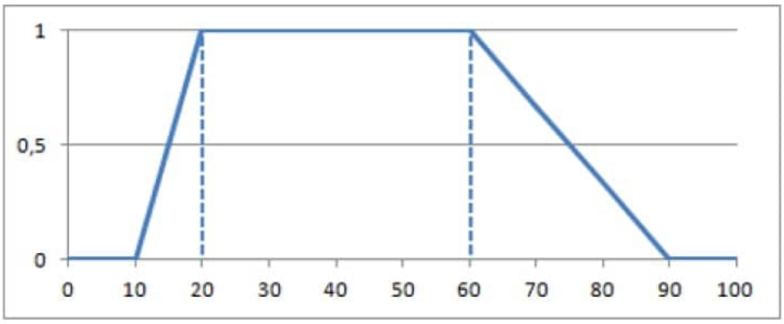


Gráfico: Função triangular.

Operações com conjuntos nebulosos

Assim como no caso dos conjuntos tradicionais, também é possível efetuar operações com conjuntos nebulosos. As operações têm as mesmas denominações daquelas aplicadas a conjuntos tradicionais, mas precisam ser redefinidas para que façam sentido no contexto dos conjuntos nebulosos.

Vejamos, então, como essas operações se comportam, considerando os conjuntos nebulosos A e B definidos em X, com as representações gráficas exibidas no gráfico a seguir:

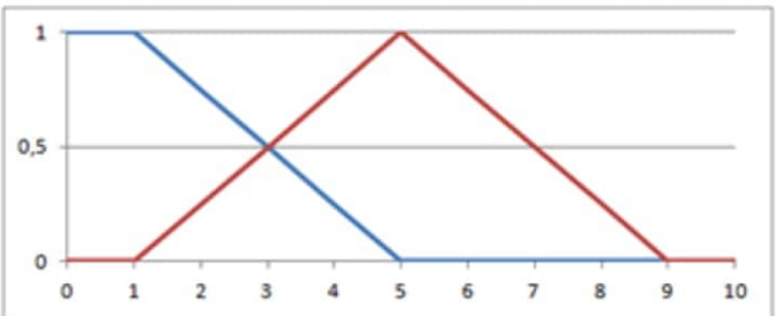
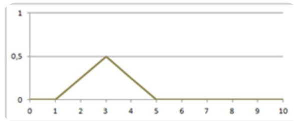


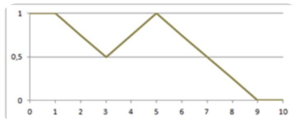
Gráfico: Conjuntos nebulosos A (mais à esquerda) e B (mais à direita)

As descrições das operações e suas respectivas representações gráficas estão resumidas a seguir:

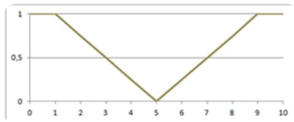
Interseção
$$\mu_{A\cap B}(x) = \min\left(\mu_A(x), \mu_B(x)\right)$$



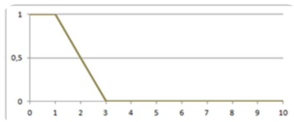
União
$$\mu_{A\cup B}(x) = \max\left(\mu_A(x), \mu_B(x)\right)$$



Complemento
$$\mu_{\bar{B}}(x) = 1 - \mu_B(x)$$



Diferença
$$\mu_{A-B}(x) = \max\left(0, \mu_A(x) - \mu_B(x)\right)$$



As operações de união e interseção entre conjuntos nebulosos são definidas, respectivamente, como o máximo e o mínimo entre os graus de inclusão dos valores de $x \in X$ nos conjuntos A e B. Esse é o modo padrão de definir tais operações em se tratando de conjuntos nebulosos.

Saiba mais

Há outros pares de funções que podem ser utilizados para representar as operações de união e interseção. Para obter detalhes sobre essas funções, consulte as obras de Cox (1995) e Yen e Langari (1999).

Propriedades das operações com conjuntos nebulosos

Quanto às propriedades das operações com conjuntos nebulosos, curiosamente, valem praticamente todas aquelas vistas para os conjuntos tradicionais. As exceções ficam por conta da lei da exclusão do meio e da lei da não contradição. É essa espécie de desobediência a essas leis que permite que um conjunto nebuloso possa representar adequadamente as situações que envolvem imprecisão.

Podemos verificar facilmente que essas duas propriedades não valem para os conjuntos nebulosos por meio de um exemplo bem simples. Imagine o conjunto nebuloso A e seu complemento $\bar{A}$ a seguir:

$$A = \{(0,3;1),(0,2;2),(0,6;3),(0,8;4),(0,3;5)\}$$

$$\bar{A} = \{(0,7;1),(0,8;2),(0,4;3),(0,2;4),(0,7;5)\}$$

Atenção

Os conjuntos e estão descritos por enumeração, de forma que, em cada par representa o grau de inclusão do elemento no conjunto.

Para esses conjuntos, os resultados das operações correspondentes às leis da exclusão do meio e da não contradição são, respectivamente:

$$A \cup \bar{A} = \max\left(\mu_A(x), \mu_{\bar{A}}(x)\right) = \{(0,7;1),(0,8;2),(0,6;3),(0,8;4),(0,7;5)\}$$

$$A \cap \bar{A} = \min\left(\mu_A(x), \mu_{\bar{A}}(x)\right) = \{(0,3;1),(0,2;2),(0,4;3),(0,2;4),(0,3;5)\}$$

A partir desses resultados, observamos que o resultado de $A \cup \bar{A} \neq X$ e que o resultado de $A \cap \bar{A} \neq \emptyset$.

Portanto, ambas as propriedades são desconsideradas no contexto dos conjuntos nebulosos.

Sistemas nebulosos

Acompanhe no vídeo o assunto que vamos estudar.

Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Até aqui, exploramos os conjuntos nebulosos, que são o alicerce do que conhecemos como lógica nebulosa e, consequentemente, do raciocínio impreciso. Contudo, para que possamos empregar o raciocínio impreciso na elaboração de soluções de apoio à tomada de decisão, precisamos entender o que são **sistemas nebulosos** e conhecer em detalhes seus componentes.

> De maneira geral, um sistema nebuloso é aquele capaz de solucionar problemas empregando alguma forma de “inteligência”.

Nesse contexto, a inteligência se caracteriza pela existência de um repositório de conhecimento acerca de determinado domínio e de um motor de inferência que é capaz de raciocinar sobre os dados que recebe como entrada, usando o conhecimento armazenado no repositório.

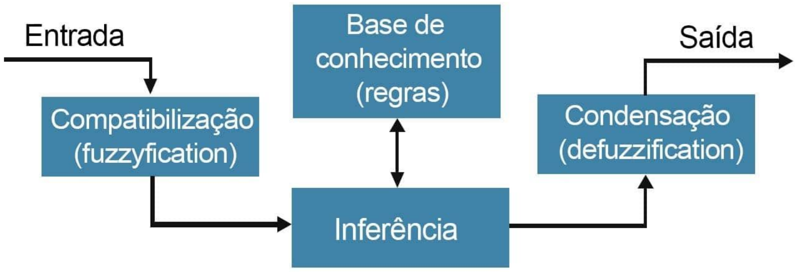
O mais interessante é que esse raciocínio se processa de maneira bastante semelhante à forma como nós raciocinamos diante de certas situações.

O raciocínio dos sistemas nebulosos se assemelha ao humano em algumas situações

Exemplo

Imagine que você precise regar um jardim, e que o tempo de rega deva levar em consideração a temperatura ambiente e a umidade do solo. Considerando seus conhecimentos sobre jardinagem, por trás de sua decisão sobre o tempo de rega, mesmo que de modo inconsciente, talvez você esteja executando uma inferência que pode ser representada pela seguinte regra: Se a temperatura está alta, e o solo está muito seco, então, regar por muito tempo. Repare que se trata de um tipo de regra que nós não temos a menor dificuldade de entender, mesmo que envolva vários conceitos imprecisos. Veja que a regra envolve termos como “alta”, “muito seco” e “muito tempo”. Apesar das imprecisões, nosso cérebro compreende e processa facilmente essa regra, mesmo que sua interpretação seja essencialmente qualitativa, ou seja, que não haja números que representem o que significa “alta”, “muito seco” ou “muito tempo”. A proposta de um sistema nebuloso é justamente reproduzir esse tipo de raciocínio.

Os módulos que compõem um sistema nebuloso e a maneira como se interconectam estão ilustrados na imagem a seguir:



Módulos de um sistema nebuloso

Variáveis nebulosas

De modo geral, uma variável é uma forma de armazenar um valor que está associado a determinada grandeza. No caso de uma variável nebulosa (ou linguística), seu valor é descrito qualitativamente por um termo linguístico e quantitativamente por uma função de inclusão. O termo linguístico expressa conceitos e conhecimento em linguagem natural, ao passo que a função de inclusão processa os dados numéricos de entrada. Em resumo, uma variável linguística combina, em uma mesma estrutura, um valor simbólico e um valor numérico.

Para medir determinada grandeza, uma variável linguística combina vários conjuntos nebulosos em um universo de discurso. Esses conjuntos têm algum grau de superposição, permitindo que um valor x pertença a mais de um conjunto ao mesmo tempo. Formalmente, uma variável nebulosa é definida como uma 4-upla {N, R, X, C}, em que:

N

Nome da variável.

R

Conjunto de rótulos (nomes) dos conjuntos nebulosos.

X

Universo de discurso.

C

Descrições dos conjuntos nebulosos.

Para que o conceito fique mais claro, observe o exemplo a seguir, que apresenta uma variável nebulosa que qualifica alunos de acordo com três conceitos:

X: aluno
R: {ruim, médio, bom}
U: $\{x \in \mathbb{R}/0 \leq x \leq 10\}$
M:

$$\mu_{\text{ruim}}(x) = \begin{cases} 1 & \text{se} \quad x \leq 2 \\ (-x+4)/3 & \text{se} \quad 24 \end{cases}$$

$$\mu_{\text{médio}}(x) = \begin{cases} 0 & \text{se} \quad x \leq 2 \\ (x-2)/3 & \text{se} \quad 28 \end{cases}$$

$$\mu_{\text{bom}}(x) = \begin{cases} 0 & \text{se} \quad x \leq 6 \\ (x-6)/2 & \text{se} \quad 68 \end{cases}$$

Veja a representação gráfica da variável:

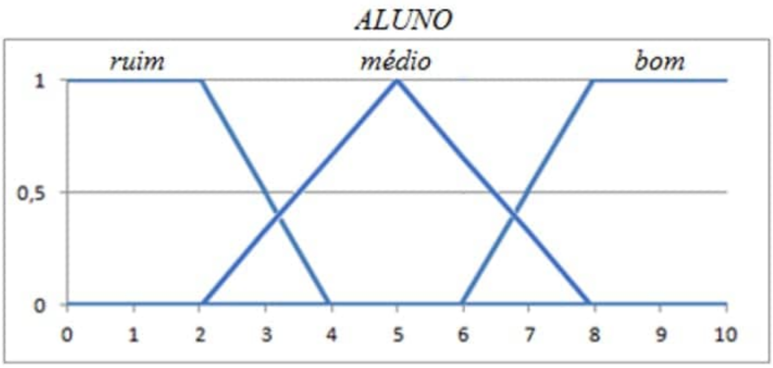


Gráfico: Variável

A partir da representação gráfica, observamos que as transições do conceito *ruim* para o conceito *médio* e do conceito *médio* para o conceito *bom* acontecem de forma suave, o que condiz com a maneira como nós interpretamos essas transições.

Podemos observar, ainda, que, nessas regiões de transição, um mesmo valor pertence a mais de um conjunto ao mesmo tempo, por exemplo:

Valor 6,5

Pertence tanto ao conjunto *médio* quanto ao conjunto *bom*, mas com uma intensidade maior no conjunto *médio* e menor no conjunto *bom*. À medida que a nota aumenta, o pertencimento ao conjunto *médio* diminui, enquanto a pertinência ao conjunto *bom* aumenta, até que, em dado momento, a situação se inverte.

Valor 7,5

A intensidade com que esse valor pertence ao conjunto *bom* é maior do que a intensidade de pertencimento ao conjunto *médio*.

Comentário

A modelagem de uma variável nebulosa é um processo que exige certa dose de criatividade. Afinal, as decisões sobre a quantidade de conjuntos a ser distribuída ao longo do universo de discurso, bem como os tipos das funções de inclusão que vão descrever esses conjuntos, dependem diretamente da grandeza a ser representada e da precisão exigida nos resultados para o problema em estudo.

Compatibilização

Apesar de um sistema nebuloso processar dados imprecisos, em geral, seus dados de entrada são precisos. No exemplo anterior, do regador de jardim, se quiséssemos que um sistema se encarregasse de controlar automaticamente o tempo de rega, os valores de temperatura ambiente e de umidade do solo a serem informados como entrada ao sistema provavelmente seriam fornecidos por meio de leituras de sensores de temperatura e de umidade. Portanto, os valores provenientes desses sensores seriam precisos, por exemplo, 18°C e 30%.

Entretanto, como a inferência é efetuada sobre valores imprecisos, a etapa de compatibilização (*fuzzification*) encarrega-se de transformar valores precisos em valores imprecisos. Esse processo de transformação compreende o mapeamento do valor de entrada na variável nebulosa correspondente, para identificar os graus de inclusão daquele valor nos conjuntos nebulosos que compõem a variável. O grau de inclusão é calculado substituindo o valor x da função de inclusão correspondente ao conjunto pelo valor de entrada.

Imaginemos que o sensor de umidade do solo tenha registrado um valor igual a 75. Considerando a variável nebulosa umidade do solo representada pelo gráfico a seguir, primeiramente, identificamos que o valor 75 mapeia para os conjuntos normal e úmido:

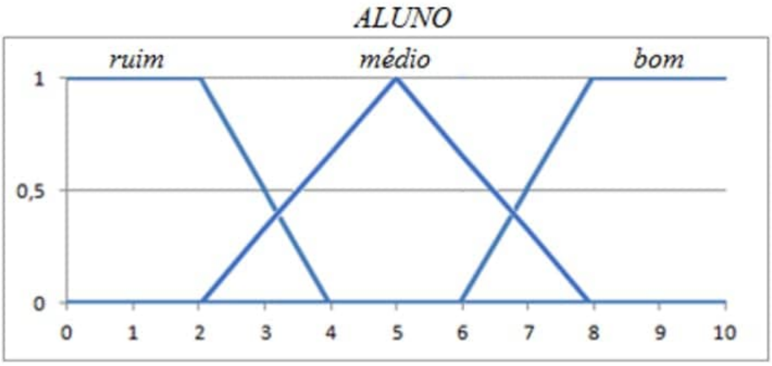


Gráfico: Representação gráfica da variável nebulosa umidade do solo

Em seguida, precisamos calcular com que intensidade o valor 75 pertence aos conjuntos *normal* e *úmido*. Para isso, precisamos substituir o valor 75 nas funções de inclusão correspondentes aos conjuntos normal e úmido. Assim, temos:

$$\mu_{\text{normal}}(75) = \frac{(80-75)}{80-50} = 0,17$$

$$\mu_{\text{imido}}(75) = \frac{(75-60)}{80-60} = 0,75$$

Agora sim, temos o valor preciso de entrada devidamente compatibilizado para que seja endereçado ao módulo de inferência.

Base de conhecimento

Um dos elementos-chave do raciocínio impreciso é a base de conhecimento. Como o próprio nome sugere, é ali que está armazenado o conhecimento sobre determinado domínio que será usado nas inferências efetuadas pelo sistema. De modo geral, esse conhecimento é fornecido por um especialista naquele domínio e é representado por um conjunto de regras que manipulam termos imprecisos.

> As regras nebulosas associam uma condição descrita por variáveis e conjuntos nebulosos a uma conclusão que, em geral, também é nebulosa. Em resumo, as regras são um esquema de captura de conhecimento impreciso fornecido por especialistas.

Para entendermos melhor como é formada uma base de conhecimento, voltemos ao exemplo do sistema nebuloso para controlar automaticamente o tempo de rega do jardim.

Considerando que o tempo de rega seja determinado de acordo com a temperatura ambiente e com a umidade do solo, isso nos leva à conclusão de que o sistema em questão tem:

Duas variáveis nebulosas de entrada

Temperatura ambiente e umidade do solo.

Uma variável nebulosa de saída

Tempo de rega.

Aqui, não vamos nos preocupar com a descrição completa das variáveis. Por ora, basta supor que os rótulos dos conjuntos nebulosos que compõem as variáveis sejam os seguintes:

temperatura ambiente: {muito frio, frio, médio, quente, muito quente} umidade do solo: {seco, normal, úmido} tempo de regra: {curto, médio, longo}

A partir dessas variáveis, considerando os respectivos rótulos de seus conjuntos nebulosos, poderíamos pensar em algumas regras:

Regra 1

Se temperatura ambiente é muito frio, e umidade do solo é seco, **então**, tempo de rega é longo.

Regra 2

Se temperatura ambiente é médio, e umidade do solo é normal, **então**, tempo de rega é médio.

Regra 3

Se temperatura ambiente é muito quente, e umidade do solo é úmido, **então**, tempo de rega é curto.

O conjunto completo de possíveis regras está representado no quadro a seguir, em que as três regras listadas anteriormente estão destacadas:

		temperatura ambiente				
		muito frio	frio	médio	quente	muito quente
umidade do solo	**seco**	longo	longo	longo	longo	longo
	normal	curto	médio	médio	Longo	Longo
	úmido	curto	curto	curto	curto	curto

Quadro: Base de regras do regador de jardim.
Elaborado por: Eugênio Silva

Atenção

As regras apresentadas são meros exemplos e estão longe de representar um conhecimento correto sobre jardinagem. Para isso, precisaríamos consultar alguém com bons conhecimentos sobre o assunto para nos guiar na elaboração das regras.

Repare que a quantidade de regras da base está diretamente relacionada à quantidade de variáveis de entrada e à quantidade de conjuntos nebulosos em cada variável. No exemplo em estudo, como são duas variáveis de entrada, uma com 3 e outra com 5 conjuntos, a base será composta por 15 regras.

Nem todas as regras têm a mesma importância para o processo de inferência. Algumas podem até mesmo ser eliminadas sem que isso prejudique os resultados retornados pelo sistema, mas esse é um assunto que está fora de nosso escopo de discussão e, por isso, não será tratado aqui. Observe, ainda, que todas as regras são do tipo condicional (**se... então**), e que tanto o antecedente (condição da regra) quanto o consequente (conclusão da regra) são nebulosos.

Saiba mais

Há outros tipos de regras, como regras incondicionais e regras cujos consequentes não são nebulosos, que são explorados em detalhes na obra de Yen e Langari (1999).

Inferência nebulosa

Agora, chegamos ao cérebro do processo de raciocínio impreciso. Uma vez que o sistema nebuloso recebe os valores de entrada e efetua a compatibilização, ou seja, mapeia esses valores em suas respectivas variáveis nebulosas de entrada, o próximo passo é acionar o módulo de **inferência**.

Esse processo é responsável tanto pela seleção das regras da base que precisam ser executadas quanto pela execução propriamente dita dessas regras, que compreende as seguintes etapas:

- Avaliação
- Agregação
- Implicação
- Combinação

Vamos entender melhor cada uma delas.

Avaliação

Esta etapa consiste em resolver as regras ativadas pelos valores de entrada: aquelas que manipulam conjuntos em que, na compatibilização, os valores de entrada têm graus de inclusão maior que zero.

Na avaliação de uma regra, levamos em consideração os conectivos eventualmente presentes no antecedente e no consequente da regra. Em regras com o conectivo e, usamos a função *min* para combinar os graus de inclusão. Já em regras com o conectivo ou, combinamos os graus de inclusão com a função *max*.

Para que isso fique mais claro, voltemos ao nosso exemplo do regador de jardim.

Imagine que, para determinado par de valores de temperatura ambiente e de umidade do solo, a compatibilização tenha resultado em um grau de inclusão igual 0,6 no conjunto *médio* da variável *temperatura ambiente* e 0,2 no conjunto *normal* da variável *umidade do solo*. Nesse caso, uma das regras ativadas da base é a seguinte:

***Se** temperatura ambiente é médio (0,6), **e** umidade do solo é normal (0,2), **então**, tempo de rega é médio.*

O resultado da avaliação dessa regra será: *min(0,6, 0,2) = 0,2*. O valor resultante, no caso o 0,2, determina a intensidade com que o conjunto *médio* da variável de saída *tempo de rega* participa da resposta do sistema para as entradas consideradas.

Agregação

Esta é uma etapa da inferência que precisa ser aplicada em situações em que são ativadas regras com antecedentes diferentes, mas que mapeiam para o mesmo consequente com intensidades distintas. Nesses casos, é preciso resolver essa espécie de conflito, em que duas regras diferentes indicam que um mesmo conjunto da variável de saída participa da resposta com intensidades distintas. Para resolver esse conflito, o método mais comum a ser utilizado é aquele que combina os resultados das avaliações por meio da função *max*.

Voltando ao nosso exemplo do regador de jardim, vamos imaginar que, para determinado par de valores de temperatura e umidade, as seguintes regras tenham sido ativadas e avaliadas:

***Se** temperatura ambiente é frio (0,70), **e** umidade do solo é úmido (0,60), **então**, tempo de rega é curto (0,60).*

***Se** temperatura ambiente é médio (0,15), **e** umidade do solo é úmido (0,60), **então**, tempo de rega é curto (0,15).*

Observando os resultados das avaliações das regras, notamos que há certo conflito sobre qual deve ser a participação do conjunto curto na resposta do sistema. Para resolver esse conflito, a agregação considera simplesmente o máximo entre os valores 0,60 e 0,15. Assim, a intensidade de participação do conjunto curto na resposta seria de 0,60.

Saiba mais

Há, também, outra forma de agregação que simplesmente soma os resultados das avaliações, respeitando um limite máximo de 1,0. Mais detalhes sobre esse método podem ser obtidos na obra de Oliveira Junior (1999).

Implicação

A etapa da agregação determina a intensidade com a qual determinado conjunto da variável de saída participa da resposta, mas não diz como essa intensidade é representada no conjunto. Esse é o papel da etapa de implicação, que define a forma (semântica) da resposta inferida. O critério mais comumente utilizado para definir esse formato da resposta é chamado de **método de corte**. Por esse método, simplesmente, cortamos o conjunto na altura do grau de inclusão (intensidade) da resposta inferida.

Recorrendo mais uma vez ao exemplo do regador de jardim, pelo método de corte, temos a resposta inferida pela regra na variável de saída *tempo de rega*:

Se temperatura ambiente **é** médio (**0,6**), **e** umidade do solo **é** normal (**0,2**), **então**, tempo de rega **é** médio (**0,2**).

A Resposta inferida pela regra a seguir na variável de saída *tempo de rega* está representada no próximo gráfico:

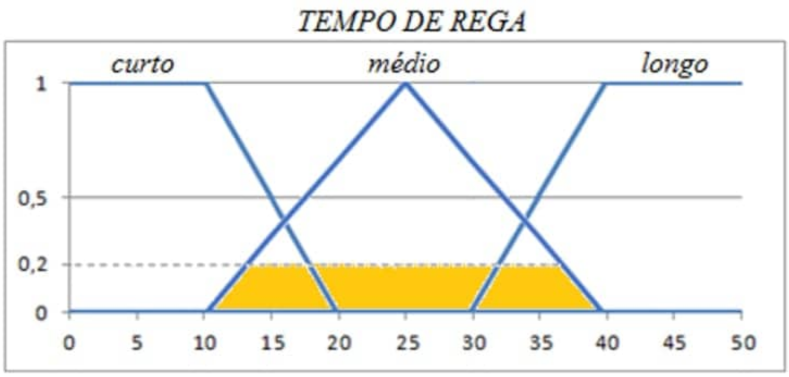


Gráfico: Resposta inferida pelo método de corte.

Repare que a resposta é formada por toda a região que se encontra abaixo do grau de inclusão 0,2 no conjunto *médio*.

Saiba mais

Além do método de corte, também há a possibilidade de se usar o método de escala. Detalhes sobre esse método estão na obra de Yen e Langari (1999).

Combinação

A combinação é o último passo do processo de inferência. É por meio dela que combinamos as regiões de reposta inferidas por todas as regras em uma região única. Em outras palavras, trata-se da superposição, utilizando a função *max* das conclusões nebulosas inferidas sobre a variável de saída.

No caso do exemplo do regador de jardim, imagine que as regras avaliadas infiram respostas tanto no conjunto *curto*, com intensidade 0,6, quanto no conjunto *médio*, com intensidade 0,2, da variável de saída *tempo de rega*. A combinação dessas respostas teria o formato apresentado no gráfico a seguir:

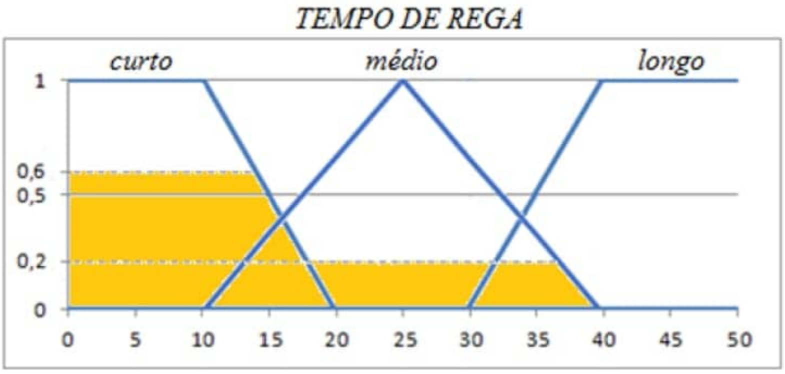


Gráfico: Combinação de respostas inferidas

Condensação

Nem sempre, o mundo físico – aquele que espera o resultado a ser emitido pelo sistema nebuloso – é capaz de interpretar a resposta qualitativa produzida pelo processo de inferência. Nesses casos, a saída nebulosa precisa ser transformada em uma saída quantitativa, ou seja, em um número que possa ser interpretado mais facilmente. Esse número corresponde a um valor no universo de discurso da variável nebulosa de saída.

Há pelo menos dois métodos para efetuar essa transformação de um resultado qualitativo em um quantitativo. São eles:

- Método do centro de gravidade (COG)
- Método da média dos máximos (MOM)

Aqui, vamos explorar apenas o método do COG.

Saiba mais

Para obter detalhes sobre o funcionamento do método MOM, consulte as obras de Cox (1995) e Yen e Langari (1999).

O método COG propõe uma abordagem bem interessante que deriva do conceito de **centro de gravidade** estudado na Física. De maneira bastante informal, podemos entender o centro de gravidade de um objeto como seu ponto de equilíbrio. Trazendo essa ideia para o contexto dos sistemas nebulosos, podemos imaginar que um número que representaria bem uma região de resposta nebulosa produzida pela inferência seria o centro de gravidade dessa região. É como se a região de resposta fosse uma placa de madeira cortada naquele formato, e que estivéssemos tentando encontrar o ponto de equilíbrio dessa placa.

É claro que, quando substituímos a região de resposta por apenas um valor, haverá perda de informação, mas esperamos que essa perda não seja significativa. Com isso, a saída numérica correspondente deve representar, com fidelidade, o resultado desejado como decisão em relação aos valores lidos como entrada.

Considerando *R* a região de resposta nebulosa produzida pela inferência, o COG(R) é dado pela equação 2.5, quando o universo de discurso da variável de saída é formado por valores contínuos:

Eq. 2.5

$$\mathrm{COG}(R) = \frac{\int \mu_R(x) \cdot x dx}{\int \mu_R(x) dx}$$

Quando o universo é formado por valores discretos, o COG(R) é dado pela equação 2.6:

Eq. 2.5

$$\mathrm{COG}(R) = \frac{\sum_x \mu_R(x) \cdot x}{\sum_x \mu_R(x)}$$

Entretanto, para nos livrarmos das dificuldades que estão por trás do cálculo das integrais da equação 2.5, podemos empregar a equação 2.6, mesmo quando estamos lidando com um universo contínuo. Para isso, a fim de que o cálculo da equação 2.6 obtenha um resultado bem próximo daquele que seria obtido a partir do cálculo da equação 2.5, basta discretizar os valores x do universo em intervalos bem pequenos.

Imagine que o gráfico a seguir represente a resposta nebulosa R inferida sobre a variável de saída *tempo de rega* do sistema de regador de jardim:

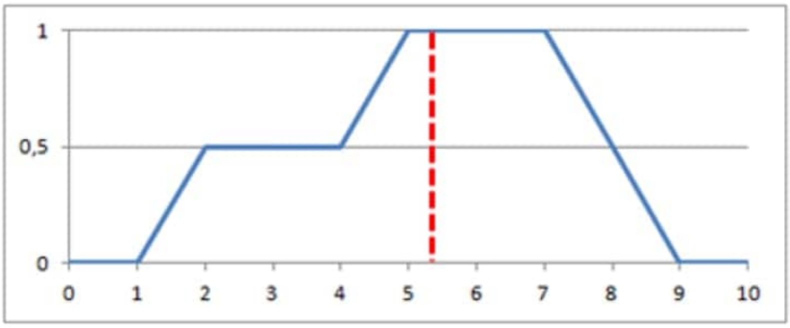


Gráfico: Resposta nebulosa inferida na variável de saída

Note que a variável tempo de rega tem o universo contínuo, ou seja, x pode assumir qualquer valor no intervalo entre 0 e 10. Entretanto, para facilitar nossos cálculos, vamos discretizar o valor de x e empregar o cálculo da equação 2.6.

Considerando que x será discretizado de 1 em 1, ou seja, que x assume os valores 0, 1, 2, ..., 10, o resultado da condensação será:

$$\mathrm{COG}(R) = \frac{2*0,5+3*0,5+4*0,5+5*1,0+6*1,0+7*1,0+8*0,5}{0,5+0,5+0,5+1,0+1,0+1,0+0,5} = 5,3$$

Então, esse será o resultado retornado pelo sistema nebuloso, ou seja, para os valores de entrada apresentados, o tempo de regra deve ser 5,3 unidades de tempo. Para um resultado mais preciso, basta discretizar x em intervalos menores, como 0,5 em 0,5. Entretanto, em termos práticos, a diferença no resultado final, em geral, não chega a ser significativa.

Utilizando lógica nebulosa

No vídeo a seguir, apresentaremos na prática a aplicação de lógica nebulosa em situações de imprecisão.

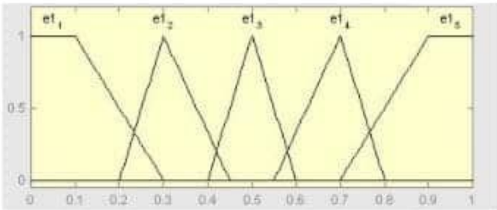
Conteúdo interativo

Acesse a versão digital para assistir ao vídeo.

Verificando o aprendizado

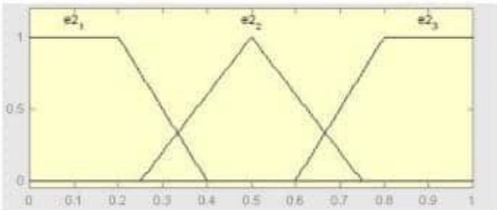
Questão 1

Seja S um sistema nebuloso cujas variáveis de entrada e de saída e o conjunto de regras estão definidos nas tabelas a seguir:



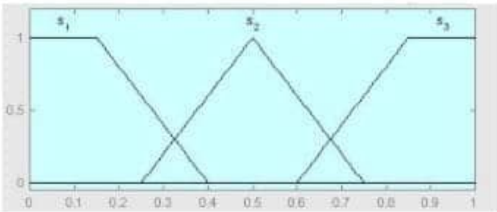
nome: *e1*
tipo: entrada

e1₁: [0,00 0,00 0,10 0,30]
e1₂: [0,20 0,30 0,45]
e1₃: [0,40 0,50 0,60]
e1₄: [0,55 0,70 0,80]
e1₅: [0,70 0,90 1,00 1,00]



nome: *e2*
tipo: entrada

e2₁: [0,00 0,00 0,20 0,40]
e2₂: [0,25 0,50 0,75]
e2₃: [0,60 0,80 1,00 1,00]



nome: *s*
tipo: saída

s₁: [0,00 0,00 0,15 0,40]
s₂: [0,25 0,50 0,75]
s₃: [0,60 0,85 1,00 1,00]

Regras

	e2 1	**e2 2**	**e2 3**
e1 1	s_1	s_2	s_3
e1 2	s_3	s_2	s_1
e1 3	s_1	s_2	s_3
e1 4	s_3	s_2	s_1
e1 5	s_1	s_2	s_3

Gráficos: Atividade

Considerando como entrada os valores e1 = 0,27 e e2 = 0,70, assinale a alternativa que lista todas as regras da base de conhecimento ativadas:

A

se $e1_4$ e $e2_1$ então s_3
se $e1_4$ e $e2_2$ então s_2
se $e1_5$ e $e2_1$ então s_1
se $e1_5$ e $e2_2$ então s_2

B

se $e1_1$ e $e2_2$ então s_3
se $e1_1$ e $e2_3$ então s_2
se $e1_2$ e $e2_2$ então s_1
se $e1_2$ e $e2_3$ então s_2

C

se $e1_1$ e $e2_2$ então s_2
se $e1_1$ e $e2_3$ então s_3
se $e1_2$ e $e2_2$ então s_2
se $e1_2$ e $e2_3$ então s_1

D

se $e1_1$ e $e2_2$ então s_2
se $e1_2$ e $e2_2$ então s_2

E

se $e1_1$ e $e2_1$ então s_1
se $e1_1$ e $e2_2$ então s_2
se $e1_1$ e $e2_3$ então s_3

A alternativa C está correta.

Para o valor **e1 = 0,27**, os conjuntos da variável e1 em que esse valor tem grau de inclusão maior que zero são **$e1_1$** e **$e1_2$**. Já o valor **e2 = 0,70** tem grau de inclusão maior que zero nos conjuntos **$e2_2$** e **$e2_3$** da variável **e2**. Assim, as regras acionadas serão aquelas que combinam os conjuntos da variável **e1** com os da variável **e2**. Há duas alternativas que mostram as combinações corretas dos conjuntos nos antecedentes das regras, mas, em apenas uma dessas alternativas, os consequentes das regras estão corretos.

Questão 2

Considerando o mesmo sistema nebuloso S do exercício anterior e as mesmas entradas, o valor retornado como saída pelo sistema é

A

0,50.

B

0,42.

C

0,65.

D

0,27.

E

0,34.

A alternativa C está correta.

Passo 1: Compatibilização

$\mu_{e11}(0,27) = \frac{(0,30-0,27)}{(0,30-0,10)} = 0,15 \quad amp;\mu_{e12}(0,27) = \frac{(0,27-0,20)}{(0,30-0,20)} = 0,70$

$u_{e22}(0,70) = \frac{(0,75-0,70)}{(0,75-0,50)} = 0,20 \quad amp;\mu_{e23}(0,70) = \frac{(0,70-0,60)}{(0,80-0,60)} = 0,50$

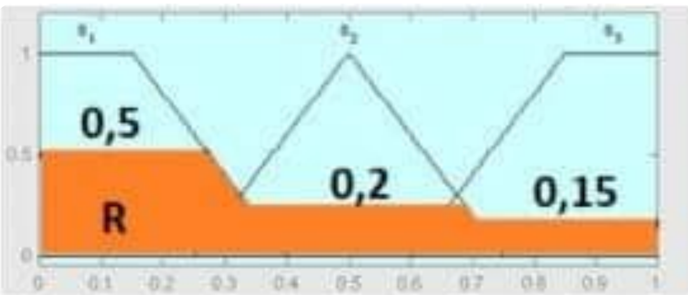
Passo 2: Inferência

Avaliação

se $e1_1$ (0,15) e $e2_2$ (0,20) então s_2 (0,15) se $e1_1$ (0,15) e $e2_3$ (0,50) então s_3 (0,15) se $e1_2$ (0,70) e $e2_2$ (0,20) então s_2 (0,20) se $e1_2$ (0,70) e $e2_3$ (0,50) então s_1 (0,50)

Agregação: s2 (0,20)

Implicação e combinação



Passo 3: Condensação

$$\begin{aligned} amp; \\ COG(R) = \\ amp; \frac{0,5*(0,1+0,2)+0,4*0,3+0,2*(0,4+0,5+0,6+0,7)+0,15*(0,8+0,9+1,0)}{0,5+0,5+0,4+0,2+0,2+0,2+0,2+0,15+0,15+0,15} \\ amp; \approx 0,42 \\ amp; \end{aligned}$$

Considerações finais

Para a maioria dos problemas reais para os quais queremos tomar algum tipo de decisão, o conhecimento prévio sobre esses problemas é incerto ou impreciso.

Assim, para a criação de modelos de inferência automática que possam ser empregados em problemas reais com essas características, os fundamentos teóricos de probabilidades e de lógica nebulosa são extremamente úteis.

Os sistemas de inferência que podem ser construídos a partir desses fundamentos são ferramentas bastante apropriadas para auxiliar especialistas de diversas áreas do conhecimento a tomar decisões mais acertadas.

Podcast

Para encerrar, ouça sobre os tipos de raciocínio automático, suas aplicações e limitações.

Conteúdo interativo

Acesse a versão digital para ouvir o áudio.

Explore +

Para conhecer uma aplicação bem interessante de redes bayesianas, busque na internet o artigo **Aplicação de Redes Bayesianas para análise de Programas Sócio Torcedor**, de Pâmela de Souza Dias, Plínio Rafael Reis Monteiro e Evandro Marcos Saidel Ribeiro.

Para conhecer uma aplicação de redes bayesianas que é de especial interesse da área de Computação, pesquise na internet o artigo **Empregando Redes Bayesianas para modelar automaticamente o conhecimento dos aprendizes em Lógica de Programação**, de Juliano Vier, João Gluz e Patrícia Augustin Jaques.

Para saber mais sobre fundamentos e aplicações de lógica nebulosa, pesquise na internet o artigo **Estado da arte dos fundamentos e ideias da lógica fuzzy aplicada às ciências e tecnologia**, de Luciana Maria da Silva, Rodrigo Mikosz Gonçalves, Leandro Mendes Ferreira, Esdras Jafet Aristides da Silva e Betânia Queiroz da Silva.

Para conhecer algumas ferramentas para a modelagem de sistemas nebulosos, pesquise na internet o artigo **Análise comparativa de ferramentas computacionais para modelagem de lógica fuzzy**, de Dayse Mourão Arruda, Gabriel Moreira Dias Abud, Felipe Arruda Pontes, Rogério Macedo Pontes e Bruno Barcellos Farias de Oliveira.

Referências

ALENCAR FILHO, E. **Iniciação à lógica matemática**. São Paulo: Nobel, 2002.

COPPIN, B. **Inteligência artificial**. Rio de Janeiro: LTC, 2010.

COX, E. **The fuzzy systems handbook** – a practitioner's guide to building, using, and maintaining fuzzy systems. New York: AP Professional, 1995.

OLIVEIRA JUNIOR, H. A. **Lógica difusa** – aspectos práticos e aplicações. Rio de Janeiro: Interciência, 1999.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2004.

SOUZA, J. N. **Lógica para Ciência da Computação e áreas afins** – uma introdução concisa. 3. ed. Rio de Janeiro: Elsevier, 2015.

YEN, J.; LANGARI, R. **Fuzzy logic** – intelligence, control and information. New Jersey: Prentice Hall, 1999.