

**Data Science
Academy**

www.datascienceacademy.com.br

Introdução à Inteligência Artificial

Sistemas e Objetos Mentais

Os agentes que construímos até agora têm crenças e podem deduzir novas crenças. Ainda assim nenhum deles tem qualquer conhecimento sobre crenças ou sobre dedução. O conhecimento sobre o próprio conhecimento e sobre os processos de raciocínio é útil para controlar a inferência. Por exemplo, suponha que Alice pergunte “Qual é a raiz quadrada de 1764” e Bob responda: “Eu não sei.” Se Alice insiste, “Pense um pouco mais”, Bob deve perceber que, com um pouco mais de raciocínio, essa questão pode de fato ser respondida. Por outro lado, se a questão fosse “Sua mãe está sentada agora?”, Bob perceberia que raciocinar mais profundamente não deve ajudar. Conhecimento sobre o conhecimento de outros agentes também é importante; Bob deverá perceber que sua mãe sabe se está sentada ou não, e uma forma de descobrir isso seria lhe perguntando. O que precisamos ter é um modelo dos objetos mentais que estão na cabeça de alguém (ou na base de conhecimento de alguma coisa) e dos processos mentais que manipulam esses objetos mentais. O modelo não precisa ser detalhado. Não precisamos ter a capacidade de prever quantos milissegundos determinado agente vai demorar para fazer uma dedução. Ficaremos felizes apenas em ser capazes de concluir que a mãe sabe se está sentada ou não. Começamos com as atitudes proposicionais que um agente pode ter em direção aos objetos mentais: atitudes tais como Acredita, Sabe, Quer, Pretende e Informa. A dificuldade é que essas atitudes não se comportam como predicados “normais”. Por exemplo, suponha que nós tentamos afirmar que Lois sabe que o Super-Homem pode voar:

$$Sabe(Lois, PodeVoar(SuperHomem))$$

Uma questão menor sobre isso é que normalmente pensamos $PodeVoar(SuperHomem)$ como uma sentença, mas aqui aparece como um termo. Essa questão pode ser remendada apenas reificando $PodeVoar(SuperHomem)$, tornando-o uma proposição. Um problema mais grave é que, se for verdade que o Super-Homem é o Clark Kent, devemos concluir que Lois sabe que Clark pode voar:

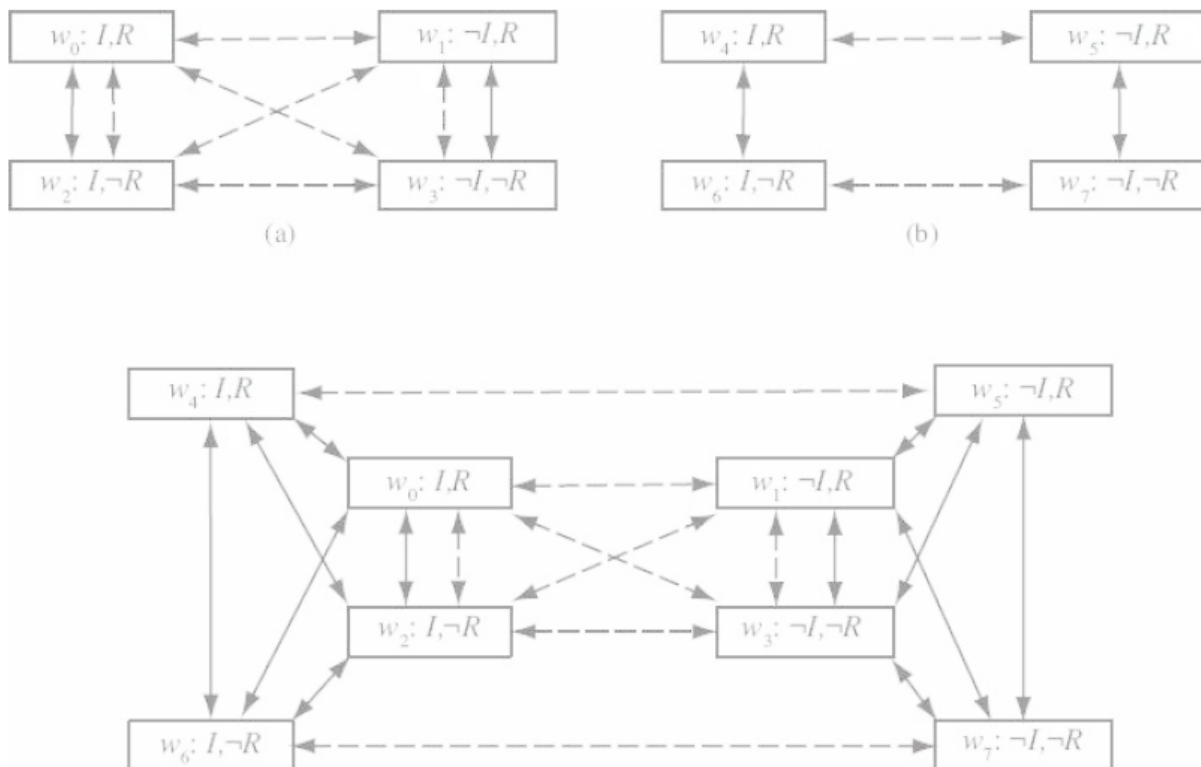
$$(SuperHomem = Clark) \wedge Sabe(Lois, PodeVoar(SuperHomem)) \\ \models Sabe(Lois, PodeVoar(Clark)).$$

Essa é uma consequência do fato de que o raciocínio com igualdade é parte da lógica. Normalmente isso é uma coisa boa, se o nosso agente souber que $2 + 2 = 4$ e $4 < 5$, então queremos que o nosso agente saiba que $2 + 2 < 5$. Essa propriedade é chamada de **transparência referencial** — não importa que termo uma lógica use para se referir a um objeto, o que importa é o objeto ao qual o termo dá nome. Mas, para atitudes proposicionais, como acredita e sabe, gostaríamos de ter opacidade referencial — os termos usados importam porque nem todos os agentes sabem que termos são correferenciais.

A lógica modal foi projetada para resolver esse problema. A lógica regular está preocupada com uma única modalidade, a da verdade, que nos permite expressar “P é verdadeiro”. A lógica modal inclui operadores modais especiais que levam sentenças (em vez de termos) como argumentos. Por exemplo, “A sabe P” é representado com a notação KAP ,

onde K é o operador modal para conhecimento. Ele recebe dois argumentos, um agente (escrito como subscrito) e uma sentença. A sintaxe da lógica modal é a mesma que a da lógica de primeira ordem, exceto que as sentenças também podem ser formadas com operadores modais.

A semântica da lógica modal é mais complicada. Na lógica de primeira ordem, um modelo contém um conjunto de objetos e uma interpretação que mapeia cada nome para o objeto apropriado, relação ou função. Na lógica modal queremos ser capazes de considerar tanto a possibilidade que a identidade secreta do Super-Homem seja Clark como que não seja. Portanto, vamos precisar de um modelo mais complicado, que consista em uma coleção de mundos possíveis em vez de apenas um mundo verdadeiro. Os mundos estão ligados em um grafo por relações de acessibilidade, uma relação para cada operador modal. Dizemos que o mundo w_1 é acessível do mundo w_0 em relação ao operador modal KA se tudo em w_1 for consistente com o que A sabe em w_0 , e escrevemos como $Acc(KA, w_0, w_1)$. Em diagramas como esse abaixo, mostramos a acessibilidade como uma seta entre os mundos possíveis. Como exemplo, no mundo real, Bucareste é a capital da Romênia, mas, para um agente que não sabia disso, outros mundos possíveis são acessíveis, inclusive onde a capital da Romênia seja Sibiu ou Sofia. Presumivelmente, um mundo onde $2 + 2 = 5$ não seria acessível a qualquer agente.



Mundos possíveis com relações de acessibilidade $K_{Super-homem}$ (setas sólidas) e K_{Lois} (setas pontilhadas). A proposição R significa que “o boletim do tempo para amanhã é de chuva” e I significa “a identidade secreta do Super-Homem é Clark Kent”. Todos os mundos são



acessíveis a eles mesmos; as setas de um mundo para si mesmo não são mostradas. Em geral, um átomo de conhecimento KAP é verdadeiro no mundo w se e somente se P for verdadeiro em todo mundo acessível de w . A verdade de sentenças mais complexas é derivada pela aplicação recursiva dessa regra e pelas regras normais da lógica de primeira ordem. Isso significa que a lógica modal pode ser usada para raciocinar sobre sentenças de conhecimento aninhadas: o que um agente sabe sobre o conhecimento do outro. Por exemplo, podemos dizer que, apesar de Lois não saber se a identidade secreta do Super-Homem é Clark Kent, ela sabe que Clark sabe:

$$K_{Lois} [K_{Clark} \text{ Identidade (SuperHomem, Clark)} \vee K_{Clark} \neg \text{ Identidade (SuperHomem, Clark)}].$$

Há um número infinito de mundos possíveis, então o truque é apresentar apenas os que você precisa para representar o que está tentando modelar. Um novo mundo possível é necessário para falar sobre os diferentes fatos possíveis (por exemplo, a chuva está prevista ou não) ou para falar sobre os diferentes estados de conhecimento (por exemplo, Lois sabe que a chuva está prevista). Isso significa que dois mundos possíveis, tais como w_4 e w_0 no diagrama acima, podem ter os mesmos fatos-base sobre o mundo, mas diferem em suas relações de acessibilidade e, portanto, em fatos sobre o conhecimento.

Podemos definir axiomas semelhantes para a crença (muitas vezes denotada por B) e outras modalidades. No entanto, um problema com a abordagem da lógica modal é que ele assume a onisciência lógica na parte dos agentes. Ou seja, se um agente sabe um conjunto de axiomas, então ele sabe todas as consequências desses axiomas. Essa é uma área instável, mesmo para a noção um tanto abstrata de conhecimento, mas parece ainda pior para a crença porque a crença tem mais conotação de se referir a coisas que são representadas fisicamente no agente, e não apenas potencialmente derivável. Houve tentativas de definir uma forma de racionalidade limitada dos agentes para dizer que os agentes acreditam nas afirmações que podem ser derivadas com a aplicação de não mais do que k etapas de raciocínio ou não mais que s segundos de computação. Essas tentativas têm sido geralmente insatisfatórias.

Referências:

Livro: Inteligência Artificial

Autor: Peter Norvig