

Introdução

Em que tentamos explicar por que consideramos a inteligência artificial um assunto digno de estudo, e em que procuramos definir exatamente o que é a inteligência artificial, pois essa definição é importante antes de iniciarmos nosso estudo.

INTELIGÊNCIA
ARTIFICIAL

Denominamos nossa espécie *Homo sapiens* – homem sábio – porque nossas capacidades mentais são muito importantes para nós. Durante milhares de anos, procuramos entender *como pensamos*; isto é, como um mero punhado de matéria pode perceber, compreender, prever e manipular um mundo muito maior e mais complicado que ela própria. O campo da **inteligência artificial**, ou IA, vai ainda mais além: ele tenta não apenas compreender, mas também *construir* entidades inteligentes.

A IA é umas das ciências mais recentes. O trabalho começou logo após a Segunda Guerra Mundial, e o próprio nome foi cunhado em 1956. Juntamente com a biologia molecular, a IA é citada regularmente como “o campo em que eu mais gostaria de estar” por cientistas de outras disciplinas. Um aluno de física pode argumentar com boa dose de razão que todas as boas idéias já foram desenvolvidas por Galileu, Newton, Einstein e outros. Por outro lado, a IA ainda tem espaço para o surgimento de vários Einsteins.

Atualmente, a IA abrange uma enorme variedade de subcampos, desde áreas de uso geral, como aprendizado e percepção, até tarefas específicas como jogos de xadrez, demonstração de teoremas matemáticos, criação de poesia e diagnóstico de doenças. A IA sistematiza e automatiza tarefas intelectuais e, portanto, é potencialmente relevante para qualquer esfera da atividade intelectual humana. Nesse sentido, ela é verdadeiramente um campo universal.

1.1 O que é IA?

Afirmamos que a IA é interessante, mas não dissemos o que ela é. As definições de inteligência artificial de acordo com oito livros didáticos são mostradas na Figura 1.1. Essas definições variam ao longo de duas dimensões principais. Em linhas gerais, as que estão na parte superior da tabela se relacionam a *processos de pensamento e raciocínio*, enquanto as definições da parte inferior se referem ao *comportamento*. As definições do lado esquerdo medem o sucesso em termos de fidelidade ao desempenho *humano*, enquanto as definições do lado direito medem o sucesso comparando-o a um conceito *ideal* de inteligência, que chamaremos **racionalidade**. Um sistema é racional se “faz tudo certo”, com os dados que tem.

RACIONALIDADE

Historicamente, todas as quatro estratégias para o estudo da IA têm sido seguidas. Como se poderia esperar, existe uma tensão entre abordagens centradas em torno de seres humanos e abordagens centradas em torno da racionalidade.¹ Uma abordagem centrada nos seres humanos deve ser uma ciência empírica, envolvendo hipóteses e confirmação experimental. Uma abordagem racionalista envolve uma combinação de matemática e engenharia. Cada grupo tem ao mesmo tempo desacreditado e ajudado o outro. Vamos examinar as quatro abordagens com mais detalhes.

Agindo de forma humana: a abordagem do teste de Turing

O **teste de Turing**, proposto por Alan Turing (1950), foi projetado para fornecer uma definição operacional satisfatória de inteligência. Em vez de propor uma lista longa e talvez controversa de qualificações exigidas para inteligência, ele sugeriu um teste baseado na impossibilidade de distinguir entre entidades inegavelmente inteligentes – os seres humanos. O computador passará no teste se um interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou não. O Capítulo 26 discute os detalhes do teste, e também se um computador é de fato inteligente se passar no teste. Por enquanto, observamos que programar um computador para passar no teste exige muito trabalho. O computador precisaria ter as seguintes capacidades:

TESTE DE TURING

PROCESSAMENTO DE LINGUAGEM NATURAL

REPRESENTAÇÃO DE CONHECIMENTO

RACIOCÍNIO AUTOMATIZADO

APRENDIZADO DE MÁQUINA

- ◆ **Processamento de linguagem natural** para permitir que ele se comunique com sucesso em um idioma natural.
- ◆ **Representação de conhecimento** para armazenar o que sabe ou ouve.
- ◆ **Raciocínio automatizado** para usar as informações armazenadas com a finalidade de responder a perguntas e tirar novas conclusões.
- ◆ **Aprendizado de máquina** para se adaptar a novas circunstâncias e para detectar e extrapolar padrões.

TESTE DE TURING TOTAL

O teste de Turing evitou deliberadamente a interação física direta entre o interrogador e o computador, porque a *simulação física* de uma pessoa é desnecessária para a inteligência. Entretanto, o chamado **teste de Turing total** inclui um sinal de vídeo, de forma que o interrogador possa testar as habilidades de percepção do assunto, além de oferecer ao interrogador a oportunidade de repassar objetos físicos “pela tela”. Para ser aprovado no teste de Turing total, o computador precisará:

1. Devemos salientar que, ao fazermos distinção entre comportamento *humano* e *racional*, não estamos sugerindo que os seres humanos sejam necessariamente “irracionais” no sentido de “emocionalmente instáveis” ou “insanos”. Simplesmente precisamos observar que não somos perfeitos: não somos todos grandes mestres de xadrez – nem mesmo aqueles dentre nós que conhecem todas as regras do jogo de xadrez – e, infelizmente, nem todos os seres humanos conseguem conceito A nos exames. Alguns erros sistemáticos do raciocínio humano estão catalogados em Kahneman *et al.* (1982).

Sistemas que pensam como seres humanos	Sistemas que pensam racionalmente
"O novo e interessante esforço para fazer os computadores pensarem... <i>máquinas com mentes</i> , no sentido total e literal." (Haugeland, 1985)	"O estudo das faculdades mentais pelo uso de modelos computacionais." (Charniak e McDermott, 1985)
"[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado ..." (Bellman, 1978)	"O estudo das computações que tornam possível perceber, raciocinar e agir." (Winston, 1992)
Sistemas que atuam como seres humanos	Sistemas que atuam racionalmente
"A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas." (Kurzweil, 1990)	"A Inteligência Computacional é o estudo do projeto de agentes inteligentes." (Poole <i>et al.</i> , 1998)

Figura 1.1 Algumas definições de inteligência artificial, organizadas em quatro categorias.

VISÃO DE
COMPUTADOR
ROBÓTICA

- ◆ Visão de computador para perceber objetos.
- ◆ Robótica para manipular objetos e movimentar-se.

Essas seis disciplinas compõem a maior parte da IA, e Turing merece crédito por projetar um teste que permanece relevante depois de 50 anos. Ainda assim, os pesquisadores da IA têm dedicado pouco esforço à aprovação no teste de Turing, acreditando que é mais importante estudar os princípios básicos da inteligência do que reproduzir um exemplar. O desafio do "vôo artificial" teve sucesso quando os irmãos Wright e outros pesquisadores pararam de imitar os pássaros e estudaram a aerodinâmica. Os textos de engenharia aeronáutica não definem como objetivo de seu campo criar "máquinas que voem exatamente como pombos a ponto de poderem enganar até mesmo outros pombos".

Pensando de forma humana: a estratégia de modelagem cognitiva

CIÊNCIA
COGNITIVA

Se pretendemos dizer que um dado programa pensa como um ser humano, temos de ter alguma forma de determinar como os seres humanos pensam. Precisamos *penetrar* nos componentes reais da mente humana. Existem duas maneiras de fazer isso: através da introspecção – procurando captar nossos próprios pensamentos à medida que eles se desenvolvem – e através de experimentos psicológicos. Depois que tivermos uma teoria da mente suficientemente precisa, será possível expressar a teoria como um programa de computador. Se os comportamentos de entrada/saída e sincronização do programa coincidirem com os comportamentos humanos correspondentes, isso será a evidência de que alguns dos mecanismos do programa também poderiam estar operando nos seres humanos. Por exemplo, Allen Newell e Herbert Simon, que desenvolveram o GPS, o "General Problem Solver" (Newell e Simon, 1961), não se contentaram em fazer seu programa resolver problemas de modo correto. Eles estavam mais preocupados em comparar os passos de suas etapas de raciocínio aos passos de sujeitos humanos resolvendo os mesmos problemas. O campo interdisciplinar da **ciência cognitiva** reúne modelos computacionais da IA e técnicas experimentais da psicologia para tentar construir teorias precisas e verificáveis a respeito dos processos de funcionamento da mente humana.

A ciência cognitiva é um campo fascinante, merecedora por si só de uma enciclopédia (Wilson e Keil, 1999). Não tentaremos descrever neste livro o que se conhece sobre a cognição humana. Ocasionalmente, apresentaremos comentários a respeito de semelhanças ou diferenças entre técnicas de IA e a cognição humana. Porém, a ciência cognitiva de verdade se baseia necessariamente na investiga-

ção experimental de seres humanos ou animais, e supomos que o leitor tenha acesso somente a um computador para realizar experimentação.

Nos primórdios da IA, frequentemente havia confusão entre as abordagens: um autor argumentava que um algoritmo funcionava bem em uma tarefa e que, *portanto*, era um bom modelo de desempenho humano ou vice-versa. Os autores modernos separam os dois tipos de afirmações; essa distinção permitiu que tanto a IA quanto a ciência cognitiva se desenvolvessem com maior rapidez. Os dois campos continuam a fertilizar um ao outro, em especial nas áreas de visão e linguagem natural. Em particular, a visão alcançou recentemente grandes avanços, por meio de uma abordagem integrada que considera evidências neurofisiológicas e modelos computacionais.

Pensando racionalmente: a abordagem das "leis do pensamento"

SILOGISMOS

O filósofo grego Aristóteles foi um dos primeiros a tentar codificar o "pensamento correto", isto é, processos de raciocínio irrefutáveis. Seus **silogismos** forneceram padrões para estruturas de argumentos que sempre resultavam em conclusões corretas ao receberem premissas corretas – por exemplo, "Sócrates é um homem; todos os homens são mortais; então, Sócrates é mortal". Essas leis do pensamento deveriam governar a operação da mente; seu estudo deu início ao campo chamado **lógica**.

LÓGICA

Os lógicos do século XIX desenvolveram uma notação precisa para declarações sobre todos os tipos de coisas no mundo e sobre as relações entre elas. (Compare isso com a notação aritmética básica, que foi projetada principalmente para declarações de igualdade e desigualdade a respeito de números.) Por volta de 1965, existiam programas que, em princípio, podiam resolver *qualquer* problema solucionável descrito em notação lógica.² A chamada tradição **logicista** dentro da inteligência artificial espera desenvolver tais programas para criar sistemas inteligentes.

LOGICISTA

Essa abordagem enfrenta dois obstáculos principais. Primeiro, não é fácil enunciar o conhecimento informal nos termos formais exigidos pela notação lógica, em particular quando o conhecimento é menos de 100% certo. Em segundo lugar, há uma grande diferença entre ser capaz de resolver um problema "em princípio" e resolvê-lo na prática. Até mesmo problemas com apenas algumas dezenas de fatos podem esgotar os recursos computacionais de qualquer computador, a menos que ele tenha alguma orientação sobre as etapas de raciocínio que deve tentar primeiro. Embora ambos os obstáculos se apliquem a *qualquer* tentativa de construir sistemas de raciocínio computacional, eles surgiram primeiro na tradição logicista.

Agindo racionalmente: a abordagem de agente racional

AGENTE

Um **agente** é simplesmente algo que age (a palavra *agente* vem do latino *agere*, que significa fazer). No entanto, espera-se que um agente computacional tenha outros atributos que possam distingui-lo de meros "programas", tais como operar sob controle autônomo, perceber seu ambiente, persistir por um período de tempo prolongado, adaptar-se a mudanças e ser capaz de assumir metas de outros. Um **agente racional** é aquele que age para alcançar o melhor resultado ou, quando há incerteza, o melhor resultado esperado.

AGENTE
RACIONAL

Na abordagem de "leis do pensamento" para IA, foi dada ênfase a inferências corretas. Às vezes, a realização de inferências corretas é uma *parte* daquilo que caracteriza um agente racional, porque uma das formas de agir racionalmente é raciocinar de modo lógico até a conclusão de que uma dada ação alcançará as metas pretendidas, e depois agir de acordo com essa conclusão. Por outro lado, a inferência correta não representa *toda* a racionalidade, porque com frequência ocorrem situações em que não existe nenhuma ação comprovadamente correta a realizar, mas mesmo assim algo tem de ser

2. Se não houver nenhuma solução, é possível que o programa jamais pare de procurar por uma.

feito. Também existem modos de agir racionalmente que não se pode dizer que envolvem inferências. Por exemplo, afastar-se de um fogão quente é um ato reflexo, em geral mais bem-sucedido que uma ação mais lenta executada após cuidadosa deliberação.

Todas as habilidades necessárias à realização do teste de Turing existem para permitir ações racionais. Desse modo, precisamos da habilidade de representar o conhecimento e raciocinar com ele, porque isso nos oferece a possibilidade de tomar boas decisões em uma ampla variedade de situações. Precisamos ter a capacidade de gerar sentenças compreensíveis em linguagem natural, porque enunciar essas sentenças nos ajuda a participar de uma sociedade complexa. Precisamos aprender não apenas por erudição, mas porque ter uma idéia melhor de como o mundo funciona nos permite gerar estratégias mais efetivas para lidar com ele. Precisamos da percepção visual não apenas porque ver é interessante, mas para ter uma idéia melhor do resultado de uma ação – por exemplo, ser capaz de ver uma gostosa iguaria ajuda a pessoa a se mover em direção a ela.

Por essas razões, o estudo da IA como um projeto de agente racional tem pelo menos duas vantagens. Primeiro, ele é mais geral que a abordagem de “leis do pensamento”, porque a inferência correta é apenas um dentre vários mecanismos possíveis para se alcançar a racionalidade. Em segundo lugar, ela é mais acessível ao desenvolvimento científico do que as estratégias baseadas no comportamento ou no pensamento humano, porque o padrão de racionalidade é definido com clareza e é completamente geral. Por outro lado, o comportamento humano está bem adaptado a um ambiente específico e em parte é o produto de um complicado e fundamentalmente desconhecido processo evolucionário que ainda está longe de produzir a perfeição: *Portanto, este livro se concentrará nos princípios gerais de agentes racionais e nos componentes para construí-los.* Veremos que, apesar da aparente simplicidade com que o problema pode ser enunciado, surge uma enorme variedade de questões quando tentamos resolvê-lo. O Capítulo 2 descreve algumas dessas questões com mais detalhes.

Devemos ter em mente um ponto importante: logo veremos que alcançar a racionalidade perfeita – sempre fazer tudo certo – não é algo viável em ambientes complicados. As demandas computacionais são demasiado elevadas. Porém, na maior parte do livro, adotaremos a hipótese de trabalho de que a racionalidade perfeita é um bom ponto de partida para a análise. Ela simplifica o problema e fornece a configuração apropriada para a maioria do material básico na área. Os Capítulos 6 e 17 lidam explicitamente com a questão da **racionalidade limitada** – agindo de forma apropriada quando não existe tempo suficiente para realizar todas as computações que gostaríamos de fazer.

RACIONALIDADE
LIMITADA

1.2 Os fundamentos da inteligência artificial

Nesta seção, apresentaremos um breve histórico das disciplinas que contribuíram com idéias, pontos de vista e técnicas para a IA. Como qualquer histórico, esse foi obrigado a se concentrar em um pequeno número de pessoas, eventos e idéias, e ignorar outros que também eram importantes. Organizamos o histórico em torno de uma série de perguntas. Certamente, não desejaríamos dar a impressão de que essas questões são as únicas de que as disciplinas tratam ou que todas as disciplinas estejam se encaminhando para a IA como sua realização final.

Filosofia (de 428 a.C. até a atualidade)

- Regras formais podem ser usadas para obter conclusões válidas?
- Como a mente (o intelecto) se desenvolve a partir de um cérebro físico?
- De onde vem o conhecimento?
- Como o conhecimento conduz à ação?

Aristóteles (384-322 a.C.) foi o primeiro a formular um conjunto preciso de leis que governam a parte racional da mente. Ele desenvolveu um sistema informal de silogismos para raciocínio apropriado que, em princípio, permitiram gerar conclusões mecanicamente, dadas as premissas iniciais. Muito mais tarde, Ramon Lull (1315) apresentou a idéia de que o raciocínio útil poderia na realidade ser conduzido por um artefato mecânico. Suas "rodas de conceitos" estão na capa deste livro. Thomas Hobbes (1588-1679) propôs que o raciocínio era semelhante à computação numérica, ou seja, que "efetuamos somas e subtrações em nossos pensamentos silenciosos". A automação da própria computação já estava bem próxima; por volta de 1500, Leonardo da Vinci (1452-1519) projetou, mas não construiu, uma calculadora mecânica; reconstruções recentes mostraram que o projeto era funcional. A primeira máquina de calcular conhecida foi construída em torno de 1623 pelo cientista alemão Wilhelm Schickard (1592-1635), embora a Pascaline, construída em 1642 por Blaise Pascal (1623-1662), seja mais famosa. Pascal escreveu que "a máquina aritmética produz efeitos que parecem mais próximos ao pensamento que todas as ações dos animais". Gottfried Wilhelm Leibniz (1646-1716) construiu um dispositivo mecânico destinado a efetuar operações sobre conceitos, e não sobre números, mas seu escopo era bastante limitado.

Agora que temos a idéia de um conjunto de regras que podem descrever a parte formal e racional da mente, a próxima etapa é considerar a mente um sistema físico. René Descartes (1596-1650) apresentou a primeira discussão clara da distinção entre mente e matéria, e dos problemas que surgem dessa distinção. Um dos problemas relacionados com uma concepção puramente física da mente é o fato de que ela parece deixar pouco espaço para o livre-arbítrio: se a mente é governada inteiramente por leis físicas, então ela não tem mais livre-arbítrio que uma pedra que "decide" cair em direção ao centro da Terra. Embora fosse um forte defensor da capacidade de raciocínio, Descartes também era um proponente do **dualismo**. Ele sustentava que havia uma parte da mente humana (ou alma, ou espírito) que transcende a natureza, isenta das leis físicas. Por outro lado, os animais não possuíam essa qualidade dual; eles podiam ser tratados como máquinas. Uma alternativa para o dualismo é o **materialismo**. O materialismo sustenta que a operação do cérebro de acordo com as leis da física *constituem* a mente. O livre-arbítrio é simplesmente o modo como a percepção das escolhas disponíveis se mostra para o processo de escolha.

Dada uma mente física que manipula o conhecimento, o próximo problema é estabelecer a origem do conhecimento. O movimento chamado **empirismo**, iniciado a partir da obra de Francis Bacon (1561-1626), *Novum Organum*,³ se caracterizou por uma frase de John Locke (1632-1704): "Não há nada na compreensão que não estivesse primeiro nos sentidos." A obra de David Hume (1711-1776) *A Treatise of Human Nature* (Hume, 1739) propôs aquilo que se conhece hoje como o princípio de **indução**: as regras gerais são adquiridas pela exposição a associações repetidas entre seus elementos. Com base no trabalho de Ludwig Wittgenstein (1889-1951) e Bertrand Russell (1872-1970), o famoso Círculo de Viena, liderado por Rudolf Carnap (1891-1970), desenvolveu a doutrina do **positivismo lógico**. Essa doutrina sustenta que todo conhecimento pode ser caracterizado por teorias lógicas conectadas; em última análise, a **sentenças de observação** que correspondem a entradas sensoriais.⁴ A **teoria da confirmação** de Carnap e Carl Hempel (1905-1997) tentava compreender como o conhecimento pode ser adquirido a partir da experiência. O livro de Carnap, *The Logical Structure of the World* (1928), definiu um procedimento computacional explícito para extrair conhecimento de experiências elementares. Provavelmente, foi a primeira teoria da mente como um processo computacional.

3. Uma atualização do *Organon* de Aristóteles, ou instrumento de pensamento.

4. Nesse quadro, todas as declarações significativas podem ser confirmadas ou definidas como falsas pela análise do significado das palavras ou pela execução de experimentos. Por eliminar a maior parte da metafísica, como era sua intenção, o positivismo lógico era impopular em alguns círculos.

DUALISMO

MATERIALISMO

EMPIRISMO

INDUÇÃO

POSITIVISMO LÓGICO

SENTENÇAS DE OBSERVAÇÃO
TEORIA DA CONFIRMAÇÃO

O último elemento no quadro filosófico da mente é a conexão entre conhecimento e ação. Essa questão é vital para a IA, porque a inteligência exige ação, bem como raciocínio. Além disso, apenas pela compreensão de como as ações são justificadas podemos compreender como construir um agente cujas ações sejam justificáveis (ou racionais). Aristóteles argumentava que as ações se justificam por uma conexão lógica entre metas e conhecimento do resultado da ação (a última parte deste extrato também aparece na capa deste livro):

Porém, como explicar que o pensamento às vezes esteja acompanhado pela ação e às vezes não, às vezes esteja acompanhado pelo movimento, e outras vezes não? Aparentemente, acontece quase o mesmo no caso do raciocínio e na realização de inferências sobre objetos imutáveis. Contudo, nesse caso o fim é uma proposição especulativa ... enquanto aqui a conclusão que resulta das duas premissas é uma ação. ... Preciso me cobrir; um casaco é uma coberta. Preciso de um casaco. O que eu preciso, tenho de fazer; preciso de um casaco. Tenho de fazer um casaco. E a conclusão, "tenho de fazer um casaco", é uma ação. (Nussbaum, 1978, p. 40)

Na obra *Ética a Nicômaco* (Livro III. 3, 1112b), Aristóteles desenvolve esse tópico um pouco mais, sugerindo um algoritmo:

"Não deliberamos sobre os fins, mas sobre os meios. Um médico não delibera sobre se deve ou não curar, nem um orador sobre se deve ou não persuadir, ... Eles dão a finalidade por estabelecida e procuram saber a maneira de alcançá-la; se lhes parece poder ser alcançada por vários meios, procuram saber o mais fácil e o mais eficaz; e se há apenas um meio para alcançá-la, procuram saber *como* será alcançada por esse meio, e por que outro meio alcançar *esse* primeiro, até chegar ao primeiro princípio, que é o último na ordem de descoberta. ... E o que vem em último lugar na ordem da análise parece ser o primeiro na ordem da execução. E, se chegarmos a uma impossibilidade, abandonamos a busca; por exemplo, se precisarmos de dinheiro e não for possível consegui-lo; porém, se algo parecer possível, tentaremos realizá-lo."*

O algoritmo de Aristóteles foi implementado 2.300 anos mais tarde por Newell e Simon em seu programa GPS. Agora, poderíamos denominá-lo um sistema de planejamento de regressão. (Consulte o Capítulo 11.)

A análise baseada em metas é útil, mas não nos diz o que fazer quando várias ações alcançarem a meta ou quando nenhuma ação a alcançar por completo. Antoine Arnauld (1612-1694) descreveu corretamente uma fórmula quantitativa para definir que ação executar em casos como esse (consulte o Capítulo 16). O livro de John Stuart Mill (1806-1873), *Utilitarianism* (Mill, 1863), promoveu a idéia de critérios de decisão racionais em todas as esferas da atividade humana. A teoria de decisões é mais formalmente discutida na próxima seção.

Matemática (cerca de 800 até a atualidade)

- Quais são as regras formais para obter conclusões válidas?
- O que pode ser computado?
- Como raciocinamos com informações incertas?

Os filósofos demarcaram a maioria das idéias importantes sobre a IA, mas o salto para uma ciência formal exigiu um certo nível de formalização matemática em três áreas fundamentais: lógica, computação e probabilidade.

A idéia de lógica formal pode ser traçada até os filósofos da Grécia antiga (veja o Capítulo 7), mas seu desenvolvimento matemático começou realmente com o trabalho de George Boole (1815-1864),

*Tradução direta do grego Pietro Nassetti, Editora Martin Claret, p. 63.

que definiu os detalhes da lógica proposicional ou lógica booleana (Boole, 1847). Em 1879, Gottlob Frege (1848-1925) estendeu a lógica de Boole para incluir objetos e relações, criando a lógica de primeira ordem que é utilizada hoje como o sistema mais básico de representação de conhecimento.⁵ Alfred Tarski (1902-1983) introduziu uma teoria de referência que mostra como relacionar os objetos de uma lógica a objetos do mundo real. A próxima etapa foi determinar os limites do que poderia ser feito com a lógica e a computação.

ALGORITMO

Acredita-se que o primeiro **algoritmo** não-trivial seja o algoritmo de Euclides para calcular o maior denominador comum. O estudo de algoritmos como objetos em si existe desde de Al-Khowarazmi, um matemático persa do século IX, cujos escritos também introduziram os numerais arábicos e a álgebra na Europa. Boole e outros discutiram algoritmos para dedução lógica e, no final do século XIX, foram empreendidos esforços para formalizar o raciocínio matemático geral como dedução lógica. Em 1900, David Hilbert (1862-1943) apresentou uma lista com 23 problemas que segundo sua previsão – que se mostrou correta – ocuparia os matemáticos durante a maior parte do século. O último problema pergunta se existe um algoritmo para determinar a verdade de qualquer proposição lógica envolvendo os números naturais – o famoso *Entscheidungsproblem*, ou problema de decisão. Em essência, Hilbert indagava se haveria limites fundamentais para a capacidade de procedimentos efetivos de prova. Em 1930, Kurt Gödel (1906-1978) mostrou que existe um procedimento efetivo para provar qualquer afirmação verdadeira na lógica de primeira ordem de Frege e Russell, mas essa lógica não poderia captar o princípio de indução matemática necessário para caracterizar os números naturais. Em 1931, ele mostrou que existem de fato limites reais. Seu **teorema da incompletudeza** mostrou que, em qualquer linguagem suficientemente expressiva para descrever as propriedades dos números naturais, existem afirmações verdadeiras que são indecidíveis no sentido de que sua verdade não pode ser estabelecida por qualquer algoritmo.

TEOREMA DA INCOMPLETEZA

Esse resultado fundamental também pode ser interpretado como a demonstração de que existem algumas funções sobre os inteiros que não podem ser representadas por um algoritmo – isto é, que não podem ser calculadas. Isso motivou Alan Turing (1912-1954) a tentar caracterizar exatamente que funções *são* calculáveis. Na realidade, essa noção é ligeiramente problemática, porque a noção de uma computação ou de um procedimento efetivo realmente não pode ter uma definição formal. No entanto, a tese de Church-Turing, que afirma que a máquina de Turing (Turing, 1936) é capaz de calcular qualquer função computável, em geral é aceita como uma definição suficiente. Turing também mostrou que existiam algumas funções que nenhuma máquina de Turing poderia calcular. Por exemplo, nenhuma máquina pode determinar, *de forma geral*, se um dado programa retornará uma resposta sobre uma certa entrada ou se continuará funcionando para sempre.

INTRATABILIDADE

Embora a indecidibilidade e a não-computabilidade sejam importantes para uma compreensão da computação, a noção de **intratabilidade** teve um impacto muito maior. Em termos gerais, um problema é chamado intratável se o tempo necessário para resolver instâncias do problema cresce exponencialmente com o tamanho das instâncias. A distinção entre crescimento polinomial e exponencial da complexidade foi enfatizada primeiro em meados da década de 1960 (Cobham, 1964; Edmonds, 1965). Ela é importante, porque o crescimento exponencial significa que até mesmo instâncias moderadamente grandes não podem ser resolvidas em qualquer tempo razoável. Portanto, devemos procurar dividir o problema global de geração de comportamento inteligente em subproblemas tratáveis, em vez de subproblemas intratáveis.

NP-COMPLETEZA

Como é possível reconhecer um problema intratável? A teoria da **NP-completudeza**, apresentada primeiro por Steven Cook (1971) e Richard Karp (1972), fornece um método. Cook e Karp demonstraram a existência de grandes classes de problemas canônicos de busca combinatória e de raciocínio

5. A notação proposta por Frege para a lógica de primeira ordem nunca se tornou popular, por razões que se tornaram aparentes de imediato a partir do exemplo da capa.

que são NP-completos. Qualquer classe de problemas à qual a classe de problemas NP-completos pode ser reduzida provavelmente é intratável. (Embora não tenha sido provado que problemas NP-completos são necessariamente intratáveis, a maioria dos teóricos acredita nisso.) Esses resultados contrastam com o otimismo com que a imprensa popular saudou os primeiros computadores — “Supercérebros eletrônicos” que eram “Mais rápidos que Einstein!”. Apesar da crescente velocidade dos computadores, o uso parcimonioso de recursos é que caracterizará os sistemas inteligentes. *Grosso modo*, o mundo é uma instância de um problema *extremamente* grande! Nos últimos anos, a IA ajudou a explicar por que algumas instâncias de problemas NP-completos são difíceis enquanto outras são fáceis (Cheeseman *et al.*, 1991).

Além da lógica e da computação, a terceira grande contribuição da matemática para a IA é a teoria da **probabilidade**. O italiano Gerolamo Cardano (1501-1576) foi o primeiro a conceber a idéia de probabilidade, descrevendo-a em termos dos resultados possíveis de jogos de azar. A probabilidade se transformou rapidamente em uma parte valiosa de todas as ciências quantitativas, ajudando a lidar com medidas incertas e teorias incompletas. Pierre Fermat (1601-1665), Blaise Pascal (1623-1662), James Bernoulli (1654-1705), Pierre Laplace (1749-1827) e outros pesquisadores aperfeiçoaram a teoria e introduziram novos métodos estatísticos. Thomas Bayes (1702-1761) propôs uma regra para atualizar probabilidades à luz de novas evidências. A regra de Bayes e o campo resultante chamado análise bayesiana formam a base da maioria das abordagens modernas para raciocínio incerto em sistemas de IA.

Economia (de 1776 até a atualidade)

- Como devemos tomar decisões para maximizar o lucro?
- Como devemos fazer isso quando outros não podem nos acompanhar?
- Como devemos fazer isso quando o lucro pode estar distante no futuro?

A ciência da economia teve início em 1776, quando o filósofo escocês Adam Smith (1723-1790) publicou *An Inquiry into the Nature and Causes of the Wealth of Nations*. Embora os antigos gregos e outros filósofos tenham contribuído para o pensamento econômico, Smith foi o primeiro a tratá-lo como uma ciência, usando a idéia de que podemos considerar que as economias consistem em agentes individuais que maximizam seu próprio bem-estar econômico. A maioria das pessoas pensa que a economia trata de dinheiro, mas os economistas dirão que na realidade a economia estuda como as pessoas fazem escolhas que levam a resultados preferenciais. O tratamento matemático de “resultados preferenciais” ou **utilidade** foi formalizado primeiro por Léon Walras (1834-1910) e foi aperfeiçoado por Frank Ramsey (1931) e mais tarde por John von Neumann e Oskar Morgenstern em seu livro *The Theory of Games and Economic Behavior* (1944).

TEORIA DA
DECISÃO

A **teoria da decisão**, que combina a teoria da probabilidade com a teoria da utilidade, fornece uma estrutura formal e completa para decisões (econômicas ou outras) tomadas sob a incerteza — ou seja, em casos nos quais as descrições probabilísticas captam de forma apropriada o ambiente do tomador de decisões. Isso é adequado para “grandes” economias em que cada agente não precisa levar em conta as ações de outros agentes como indivíduos. No caso das “pequenas” economias, a situação é muito mais parecida com um **jogo**: as ações de um jogador podem afetar de forma significativa a utilidade de outro (positiva ou negativamente). O desenvolvimento da **teoria dos jogos** por Von Neumann e Morgenstern (consulte também Luce e Raiffa, 1957) incluiu o surpreendente resultado de que, em alguns jogos, um agente racional deve agir de forma casual ou, pelo menos, de uma forma que pareça casual para os adversários.

TEORIA DOS
JOGOS

PESQUISA
OPERACIONAL

De modo geral, os economistas não trataram a terceira questão da listagem anterior, ou seja, como tomar decisões racionais quando os lucros das ações não são imediatos, mas sim resultam de várias ações executadas *em seqüência*. Esse tópico foi adotado no campo de **pesquisa operacional**, que emergiu na Segunda Guerra Mundial dos esforços britânicos para otimizar instalações de radar e,

mais tarde, encontrou aplicações civis em decisões complexas de administração. O trabalho de Richard Bellman (1957) formalizou uma classe de problemas de decisão seqüencial chamados **processos de decisão de Markov**, que estudaremos nos Capítulos 17 e 21.

SATISFAÇÃO

O trabalho em economia e pesquisa operacional contribuiu muito para nossa noção de agentes racionais, ainda que por muitos anos a pesquisa em IA se desenvolvesse ao longo de caminhos inteiramente separados. Uma razão para isso era a aparente **complexidade** da tomada de decisões racionais. Herbert Simon (1916- 2001), o pesquisador pioneiro da IA, ganhou o prêmio Nobel de economia em 1978 por seu trabalho inicial demonstrando que modelos baseados em **satisfação** – a tomada de decisões “boas o suficiente”, em vez de calcular laboriosamente uma decisão ótima – forneciam uma descrição melhor do comportamento humano real (Simon, 1947). Na década de 1990, ressurgiu o interesse em técnicas da teoria da decisão para sistemas de agentes (Wellman, 1995).

Neurociência (de 1861 até a atualidade)

- Como o cérebro processa informações?

NEUROCIÊNCIA

A **neurociência** é o estudo do sistema nervoso, em particular do cérebro. O modo exato como o cérebro habilita o pensamento é um dos grandes mistérios da ciência. Durante milhares de anos, observou-se que o cérebro está de alguma forma envolvido no pensamento, devido à evidência de que pancadas fortes na cabeça podem levar à incapacitação mental. Também se sabe há muito tempo que os cérebros dos seres humanos têm algumas características diferentes; em aproximadamente 335 a.C., Aristóteles escreveu: “De todos os animais, o homem é o que tem o maior cérebro em proporção a seu tamanho.”⁶ Ainda assim, apenas em meados do século XVIII o cérebro foi amplamente reconhecido como a sede da consciência. Antes disso, acreditava-se que a sede da consciência poderia estar localizada no coração, no baço e na glândula pineal.

NEURÔNIOS

O estudo da afasia (deficiência da fala) feito por Paul Broca (1824-1880) em 1861 com pacientes cujos cérebros foram danificados revigorou o campo e persuadiu a classe médica da existência de áreas localizadas do cérebro responsáveis por funções cognitivas específicas. Em particular, ele mostrou que a produção da fala estava localizada em uma parte do hemisfério cerebral esquerdo agora chamada área de Broca.⁷ Nessa época, sabia-se que o cérebro consistia em células nervosas ou **neurônios**, mas apenas em 1873 Camillo Golgi (1843-1926) desenvolveu uma técnica de coloração que permitiu a observação de neurônios individuais no cérebro (ver a Figura 1.2). Essa técnica foi usada por Santiago Ramon y Cajal (1852-1934) em seus estudos pioneiros das estruturas de neurônios do cérebro.⁸

Atualmente temos alguns dados sobre o mapeamento entre áreas do cérebro e as partes do corpo que elas controlam ou das quais recebem entrada sensorial. Tais mapeamentos podem mudar radicalmente no curso de algumas semanas, e alguns animais parecem ter vários mapas. Além disso, não compreendemos inteiramente como outras áreas do cérebro podem assumir o comando de certas funções quando uma área é danificada. Praticamente não há teoria que explique como a memória de um indivíduo é armazenada.

6. Desde então, foi descoberto que algumas espécies de golfinhos e baleias têm cérebros relativamente maiores. Hoje em dia, acredita-se que o grande tamanho do cérebro humano se deve em parte às melhorias recentes em seu sistema de resfriamento.

7. Muitos citam Alexander Hood (1824) como uma possível fonte anterior.

8. Golgi persistiu em sua convicção de que as funções do cérebro eram executadas principalmente em um meio contínuo no qual os neurônios estavam incorporados, enquanto Cajal propunha a “doutrina neuronal”. Os dois compartilharam o prêmio Nobel em 1906, mas pronunciaram discursos bastante antagônicos ao aceitarem o prêmio.

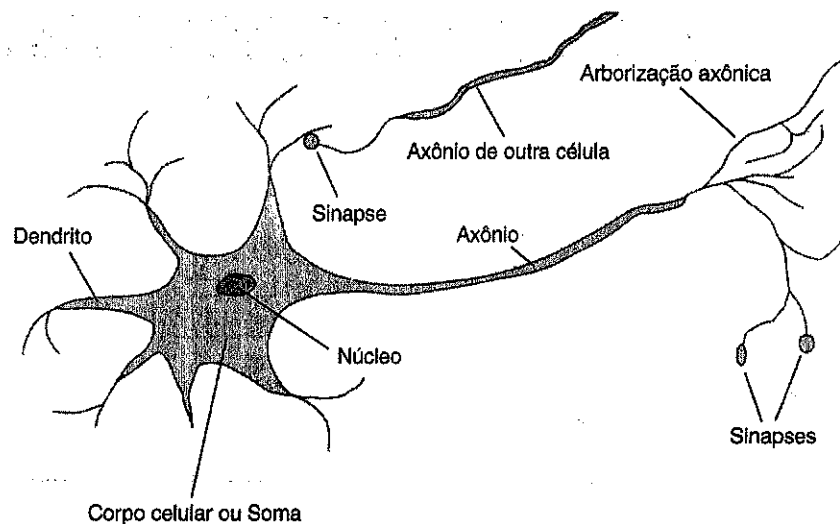


Figura 1.2 As partes de uma célula nervosa ou neurônio. Cada neurônio consiste em um corpo celular ou soma, que contém um núcleo celular. Ramificando-se a partir do corpo celular, há uma série de fibras chamadas dendritos e uma única fibra longa chamada axônio. O axônio se estende por uma longa distância, muito mais longa do que indica a escala desse diagrama. Em geral, os axônios têm 1 cm de comprimento (100 vezes o diâmetro do corpo celular), mas podem alcançar até 1 metro. Um neurônio faz conexões com 10 a 100.000 outros neurônios, em junções chamadas sinapses. Os sinais se propagam de um neurônio para outro por meio de uma complicada reação eletroquímica. Os sinais controlam a atividade cerebral em curto prazo, e também permitem mudanças a longo prazo na posição e na conectividade dos neurônios. Acredita-se que esses mecanismos formem a base para o aprendizado no cérebro. A maior parte do processamento de informações ocorre no córtex cerebral, a camada exterior do cérebro. A unidade organizacional básica parece ser uma coluna de tecido com aproximadamente 0,5 mm de diâmetro, estendendo-se por toda a profundidade do córtex, cerca de 4 mm nos seres humanos. Uma coluna contém aproximadamente 20.000 neurônios.

A medição da atividade do cérebro intacto teve início em 1929, com a invenção do eletroencefalógrafo (EEG) por Hans Berger. O desenvolvimento recente do processamento de imagens por ressonância magnética funcional (fMRI – functional Magnetic Resonance Imaging) (Ogawa *et al.*, 1990) está dando aos neurocientistas imagens sem precedentes de detalhes da atividade do cérebro, tornando possíveis medições que correspondem em aspectos interessantes a processos cognitivos contínuos. Essas medições são ampliadas por avanços na gravação da atividade dos neurônios em uma única célula. Apesar desses avanços, ainda estamos longe de compreender como realmente funciona qualquer desses processos cognitivos.



A conclusão verdadeiramente espantosa é que *uma coleção de células simples pode levar ao pensamento, à ação e à consciência* ou, em outras palavras, que *cérebros geram mentes* (Searle, 1992). A única teoria alternativa real é o misticismo: existe algum reino místico em que as mentes operam e que está além da ciência física.

Os cérebros e os computadores digitais executam tarefas bastante diferentes e têm propriedades distintas. A Figura 1.3 mostra que existem 1.000 vezes mais neurônios no cérebro humano típico do que existem portas lógicas na CPU de um computador de ponta típico. A lei de Moore⁹ prevê que, por volta de 2020, o número de portas lógicas da CPU será igual ao número de neurônios do cérebro humano. É claro que pouco se pode deduzir de tais prognósticos; além disso, a diferença na capacidade de armazenamento é secundária em comparação com a diferença em velocidade de comutação e em paralelismo. Os chips de computadores podem executar uma instrução em um nanossegundo, enquanto os neurônios são milhões de vezes mais lentos. Porém, os cérebros mais do que compensam essa diferença, porque todos os neurônios e sinapses estão ativos ao mesmo tempo, enquanto a maio-

9. A lei de Moore diz que o número de transistores por polegada quadrada dobra a cada período de 1 a 1,5 ano. A capacidade do cérebro humano dobra aproximadamente a cada período de 2 a 4 milhões de anos.

	Computador	Cérebro humano
Unidades computacionais	1 CPU, 10^8 portas lógicas	10^{11} neurônios
Unidades de armazenamento	RAM de 10^{10} bits	10^{11} neurônios
	Disco de 10^{11} bits	10^{14} sinapses
Tempo de ciclo	10^{-9} segundo	10^{-3} segundo
Largura de banda	10^{10} bits/s	10^{14} bits/s
Atualizações de memória/s	10^9	10^{14}

Figura 1.3 Uma comparação grosseira dos recursos computacionais brutos disponíveis para computadores (de 2003) e cérebros. Os números referentes ao computador aumentaram pelo menos 10 vezes desde a primeira edição deste livro e espera-se que aumentem de novo na mesma proporção ainda nesta década. Os números referentes ao cérebro não mudaram nos últimos 10.000 anos.



ria dos computadores atuais tem apenas uma ou no máximo algumas CPUs. Desse modo, *embora um computador seja um milhão de vezes mais rápido em velocidade de comutação bruta, o cérebro acaba sendo 100.000 vezes mais rápido no que faz.*

Psicologia (de 1879 até a atualidade)

- Como os seres humanos e os animais pensam e agem?

Normalmente, considera-se que as origens da psicologia científica remontam ao trabalho do físico alemão Hermann von Helmholtz (1821-1894) e de seu aluno Wilhelm Wundt (1832-1920). Helmholtz aplicou o método científico ao estudo da visão humana, e seu *Handbook of Physiological Optics* é descrito até hoje como “o mais importante tratado sobre a física e a fisiologia da visão humana” (Nalwa, 1993, p. 15). Em 1879, Wundt abriu o primeiro laboratório de psicologia experimental na Universidade de Leipzig. Wundt insistia em experimentos cuidadosamente controlados, nos quais seus trabalhadores executariam uma tarefa perceptiva ou associativa enquanto refletiam sobre seus processos de pensamento. O controle cuidadoso percorreu um longo caminho para transformar a psicologia em uma ciência, mas a natureza subjetiva dos dados tornava improvável que um pesquisador divergisse de suas próprias teorias. Por outro lado, os biólogos que estudavam o comportamento animal careciam de dados introspectivos e desenvolveram uma metodologia objetiva, como descreveu H. S. Jennings (1906) em seu influente trabalho *Behavior of the Lower Organisms*. Aplicando esse ponto de vista aos seres humanos, o movimento chamado **behaviorismo**, liderado por John Watson (1878-1958), rejeitava *qualquer* teoria que envolvesse processos mentais com base no fato de que a introspecção não poderia fornecer evidência confiável. Os behavioristas insistiam em estudar apenas medidas objetivas das percepções (ou *estímulos*) dados a um animal e suas ações resultantes (ou *respostas*). Construções mentais como conhecimento, convicções, metas e etapas do raciocínio eram abandonadas como “psicologia popular” não-científica. O behaviorismo descobriu muito sobre ratos e pombos, mas teve menos sucesso na compreensão dos seres humanos. Apesar disso, ele exerceu uma forte influência sobre a psicologia (especialmente nos Estados Unidos) de aproximadamente 1920 a 1960.

A visão do cérebro como um dispositivo de processamento de informações, uma característica importante da **psicologia cognitiva**, tem suas origens nos trabalhos de William James¹⁰ (1842-1910). Helmholtz também insistiu que a percepção envolvia uma forma de inferência lógica inconsciente. O ponto de vista cognitivo foi em grande parte eclipsado pelo behaviorismo nos Estados Unidos, mas na

10. William James era irmão do romancista Henry James. Dizem que Henry escrevia ficção como se fosse psicologia e William escrevia sobre psicologia como se fosse ficção.

BEHAVIORISMO

PSICOLOGIA
COGNITIVA

Unidade de Psicologia Aplicada de Cambridge, dirigida por Frederic Bartlett (1886-1969), a modelagem cognitiva foi capaz de florescer. *The Nature of Explanation*, de Kenneth Craik (1943), aluno e sucessor de Bartlett, restabeleceu com vigor a legitimidade de termos “mentais” como convicções e metas, argumentando que eles são tão científicos quanto, digamos, usar a pressão e a temperatura ao falar sobre gases, apesar de eles serem constituídos por moléculas que não têm nenhuma dessas duas propriedades. Craik especificou os três passos fundamentais de um agente baseado no conhecimento: (1) o estímulo deve ser traduzido em uma representação interna, (2) a representação é manipulada por processos cognitivos para derivar novas representações internas e (3) por sua vez, essas representações são de novo traduzidas em ações. Ele explicou com clareza por que esse era um bom projeto de um agente:

Se o organismo transporta um “modelo em escala reduzida” da realidade externa e de suas próprias ações possíveis dentro de sua cabeça, ele é capaz de experimentar várias alternativas, concluir qual a melhor delas, reagir a situações futuras antes que elas surjam, utilizar o conhecimento de eventos passados para lidar com o presente e o futuro e, em todos os sentidos, reagir de maneira muito mais completa, segura e competente às emergências que enfrenta. (Craik, 1943)

Após a morte de Craik em um acidente de bicicleta em 1945, seu trabalho teve continuidade com Donald Broadbent, cujo livro *Perception and Communication* (1958) incluía alguns dos primeiros modelos de processamento de informações de fenômenos psicológicos. Enquanto isso, nos Estados Unidos, o desenvolvimento da modelagem de computadores levou à criação do campo da **ciência cognitiva**. Pode-se dizer que o campo teve início em um seminário em setembro de 1956 no MIT. (Veremos que esse seminário ocorreu apenas dois meses após a conferência em que a própria IA “nasceu”.) No seminário, George Miller apresentou *The Magic Number Seven*, Noam Chomsky apresentou *Three Models of Language* e Allen Newell e Herbert Simon apresentaram *The Logic Theory Machine*. Esses três documentos influentes mostraram como modelos de computadores podiam ser usados para tratar a psicologia da memória, a linguagem e o pensamento lógico, respectivamente. Agora é comum entre os psicólogos a visão de que “uma teoria cognitiva deve ser como um programa de computador” (Anderson, 1980), isto é, ela deve descrever um mecanismo detalhado de processamento de informações por meio do qual alguma função cognitiva poderia ser implementada.

CIÊNCIA
COGNITIVA

Engenharia de computadores (de 1940 até a atualidade)

- Como podemos construir um computador eficiente?

Para a inteligência artificial ter sucesso, precisamos de inteligência e de um artefato. O computador tem sido o artefato preferido. O computador eletrônico digital moderno foi criado independentemente e quase ao mesmo tempo por cientistas de três países que participavam da Segunda Guerra Mundial. O primeiro computador *operacional* foi a máquina eletromecânica de Heath Robinson,¹¹ construída em 1940 pela equipe de Alan Turing com um único propósito: decifrar mensagens alemãs. Em 1943, o mesmo grupo desenvolveu o Colossus, uma poderosa máquina de uso geral baseada em válvulas eletrônicas.¹² O primeiro computador *programável* operacional foi o Z-3, criado por Konrad Zuse na Alemanha em 1941. Zuse também criou os números de ponto flutuante e a primeira

11. Heath Robinson foi um cartunista famoso por suas representações de aparelhos extravagantes e absurdamente complicados para realizar tarefas diárias como passar manteiga em torradas.

12. No período do pós-guerra, Turing queria usar esses computadores em pesquisas de IA – por exemplo, um dos primeiros programas de xadrez (Turing *et al.*, 1953). Seus esforços foram bloqueados pelo governo britânico.

linguagem de programação de alto nível, denominada Plankalkül. O primeiro computador *eletrônico*, o ABC, foi montado por John Atanasoff e seu aluno Clifford Berry entre 1940 e 1942 na Iowa State University, nos Estados Unidos. A pesquisa de Atanasoff recebeu pouco apoio ou reconhecimento; foi o ENIAC, desenvolvido como parte de um projeto militar secreto na University of Pennsylvania por uma equipe que incluía John Mauchly e John Eckert, dois cientistas que provaram ser os precursores mais influentes dos computadores modernos.

No meio século seguinte, cada geração de hardware de computador trouxe um aumento em velocidade e capacidade e uma redução no preço. O desempenho é duplicado a cada 18 meses aproximadamente, prevendo-se uma ou duas décadas de crescimento futuro a essa taxa. Depois disso, precisaremos da engenharia molecular ou de alguma outra tecnologia nova.

É claro que existiam dispositivos de cálculo antes do computador eletrônico. As primeiras máquinas automatizadas, datando do século XVII, foram descritas na página 8. A primeira máquina *programável* foi um tear criado em 1805 por Joseph Marie Jacquard (1752-1834) que utilizava cartões perfurados para armazenar instruções relativas ao padrão a ser tecido. Na metade do século XIX, Charles Babbage (1792-1871) projetou duas máquinas, mas não concluiu nenhuma das duas. A "Máquina Diferencial" que aparece na capa deste livro, se destinava a calcular tabelas matemáticas para projetos de engenharia e científicos. Ela foi finalmente construída e se mostrou funcional em 1991 no Science Museum em Londres (Swade, 1993). A "Máquina Analítica" de Babbage era bem mais ambiciosa: ela incluía memória endereçável, programas armazenados e saltos condicionais, e foi o primeiro artefato capaz de executar computação universal. A colega de Babbage, Ada Lovelace, filha do poeta Lord Byron, talvez tenha sido a primeira programadora do mundo. (A linguagem de programação Ada recebeu esse nome em homenagem a ela.) Ela escreveu programas para a Máquina Analítica não-concluída e até mesmo especulou que a máquina poderia jogar xadrez ou compor música.

A IA também tem uma dívida com a área de software da ciência da computação, que forneceu os sistemas operacionais, as linguagens de programação e as ferramentas necessárias para escrever programas modernos (e documentos sobre eles). Porém, essa é uma área em que a dívida tem sido paga: o trabalho em IA foi pioneiro em muitas idéias que foram aproveitadas posteriormente na ciência da computação em geral, incluindo o compartilhamento de tempo, interpretadores interativos, computadores pessoais com janelas e mouse, ambientes de desenvolvimento rápido, o tipo de dados lista encadeada, o gerenciamento automático de armazenamento e conceitos fundamentais de programação simbólica, funcional, dinâmica e orientada a objetos.

Teoria de controle e cibernética (de 1948 até a atualidade)

- Como os artefatos podem operar sob seu próprio controle?

Ctesíbio de Alexandria (cerca de 250 a.C.) construiu a primeira máquina autocontrolada: um relógio de água com um regulador que mantinha o fluxo de água em um ritmo constante e previsível. Essa invenção mudou a definição do que um artefato poderia fazer. Antes, somente os seres vivos podiam modificar seu comportamento em resposta a mudanças no ambiente. Outros exemplos de sistemas de controle realimentados auto-reguláveis incluem o regulador de máquinas a vapor, criado por James Watt (1736-1819), e o termostato, criado por Cornelis Drebbel (1572-1633), que também inventou o submarino. A teoria matemática de sistemas realimentados estáveis foi desenvolvida no século XIX.

A figura central na criação daquilo que se conhece hoje como **teoria de controle** foi Norbert Wiener (1894-1964). Wiener foi um matemático brilhante que trabalhou com Bertrand Russell, entre outros, antes de se interessar por sistemas de controle biológico e mecânico e sua conexão com a cognição. Como Craik (que também utilizou sistemas de controle como modelos psicológicos), Wiener e seus colegas Arturo Rosenblueth e Julian Bigelow desafiaram a ortodoxia behaviorista (Rosenblueth

CIBERNÉTICA

et al., 1943). Eles viram o comportamento consciente como o resultado de um mecanismo regulador tentando minimizar o “erro” – a diferença entre o estado atual e estado do objetivo. No final da década de 1940, Wiener, juntamente com Warren McCulloch, Walter Pitts e John von Neumann, organizou uma série de conferências que exploraram os novos modelos matemáticos e computacionais da cognição e influenciaram muitos outros pesquisadores nas ciências do comportamento. O livro de Wiener, *Cybernetics* (1948), tornou-se um best-seller e despertou o público para a possibilidade de máquinas dotadas de inteligência artificial.

FUNÇÃO
OBJETIVO

A moderna teoria de controle, em especial o ramo conhecido como controle estocástico ótimo, tem como objetivo o projeto de sistemas que maximizam uma **função objetivo** sobre o tempo. Isso corresponde aproximadamente à nossa visão da IA: projetar sistemas que se comportem de maneira ótima. Então, por que a IA e a teoria de controle são dois campos diferentes, especialmente quando se consideram as conexões estreitas entre seus fundadores? A resposta reside no acoplamento estrito entre as técnicas matemáticas familiares aos participantes e os conjuntos de problemas correspondentes que foram incluídos em cada visão do mundo. O cálculo e a álgebra de matrizes, as ferramentas da teoria de controle, eram adequados para sistemas que podem ser descritos por conjuntos fixos de variáveis contínuas; além disso, a análise exata em geral só é viável para sistemas *lineares*. Em parte, a IA foi criada como um meio de escapar das limitações da matemática da teoria de controle na década de 1950. As ferramentas de inferência lógica e computação permitiram que os pesquisadores da IA considerassem alguns problemas como linguagem, visão e planejamento, que ficavam completamente fora do campo de ação da teoria de controle.

Linguística (de 1957 até a atualidade)

- Como a linguagem se relaciona ao pensamento?

Em 1957, B. F. Skinner publicou *Verbal Behavior*. Essa obra foi uma descrição completa e detalhada da abordagem behaviorista para o aprendizado da linguagem, escrita pelo mais proeminente especialista no campo. Porém, curiosamente, uma crítica do livro se tornou tão conhecida quanto o próprio livro e serviu para aniquilar o interesse no behaviorismo. O autor da resenha foi Noam Chomsky, que tinha acabado de publicar um livro sobre sua própria teoria, com o *Syntactic Structures* (*Estruturas Sintáticas*). Chomsky mostrou que a teoria behaviorista não tratava a noção de criatividade na linguagem – ela não explicava como uma criança podia compreender e formar frases que nunca tinha ouvido antes. A teoria de Chomsky – baseada em modelos sintáticos criados pelo lingüista indiano Panini (cerca de 350 a.C.) – podia explicar esse fato e, diferente das teorias anteriores, era formal o bastante para poder, em princípio, ser programada.

LINGÜÍSTICA
COMPUTACIONAL

Portanto, a linguística moderna e a IA “nasceram” aproximadamente na mesma época e cresceram juntas, cruzando-se em um campo híbrido chamado **lingüística computacional** ou **processamento de linguagem natural**. O problema de compreender a linguagem logo se tornou consideravelmente mais complexo do que parecia em 1957. A compreensão da linguagem exige a compreensão do assunto e do contexto, não apenas a compreensão da estrutura das frases. Isso pode parecer óbvio, mas só foi amplamente avaliado na década de 1960. Grande parte do trabalho anterior em **representação do conhecimento** (o estudo de como colocar o conhecimento em uma forma que um computador possa utilizar) estava vinculado à linguagem e era suprido com informações da pesquisa em linguística que, por sua vez, estava conectada a décadas de pesquisa sobre a análise filológica da linguagem.

INFORIUM
BIBLIOTECA

1.3 História da inteligência artificial

Com o material que vimos até agora, estamos prontos para estudar o desenvolvimento da própria IA.

A gestação da inteligência artificial (1943-1955)

O primeiro trabalho agora reconhecido como IA foi realizado por Warren McCulloch e Walter Pitts (1943). Eles se basearam em três fontes: o conhecimento da fisiologia básica e da função dos neurônios no cérebro, uma análise formal da lógica proposicional criada por Russell e Whitehead, e a teoria da computação de Turing. Esses dois pesquisadores propuseram um modelo de neurônios artificiais, no qual cada neurônio se caracteriza por estar "ligado" ou "desligado", com a troca para "ligado" ocorrendo em resposta à estimulação por um número suficiente de neurônios vizinhos. O estado de um neurônio era considerado "equivalente em termos concretos a uma proposição que definia seu estímulo adequado". Por exemplo, eles mostraram que qualquer função computável podia ser calculada por uma certa rede de neurônios conectados, e que todos os conectivos lógicos (e, ou, não etc.) podiam ser implementados por estruturas de redes simples. McCulloch e Pitts também sugeriram que redes definidas adequadamente seriam capazes de aprender. Donald Hebb (1949) demonstrou uma regra de atualização simples para modificar as intensidades de conexão entre neurônios. Sua regra, agora chamada **aprendizagem de Hebb**, continua a ser um modelo influente até hoje.

Dois alunos do departamento de matemática de Princeton, Marvin Minsky e Dean Edmonds, construíram o primeiro computador de rede neural em 1951. O SNARC, como foi chamado, usava 3.000 válvulas eletrônicas e um mecanismo de piloto automático retirado de um bombardeiro B-24 para simular uma rede de 40 neurônios. A banca examinadora da tese de doutorado de Minsky se mostrou cética sobre esse tipo de trabalho, sem saber se ele deveria ser classificado como um trabalho de matemática. Porém, segundo contam, von Neumann teria dito: "Se não é agora, será algum dia." Mais tarde, Minsky acabou provando teoremas importantes que mostravam as limitações da pesquisa em redes neurais.

Surgiram vários exemplos de trabalhos que hoje podem ser caracterizados como IA, mas foi Alan Turing quem primeiro articulou uma visão completa da IA em seu artigo de 1950 intitulado "Computing Machinery and Intelligency". Nesse artigo, ele apresentou o teste de Turing, aprendizagem de máquina, algoritmos genéticos e aprendizagem por reforço.

O nascimento da inteligência artificial (1956)

Princeton foi o lar de outra figura influente na IA, John McCarthy. Após sua graduação, McCarthy mudou-se para o Dartmouth College, que iria se tornar o local oficial de nascimento desse campo. McCarthy convenceu Minsky, Claude Shannon e Nathaniel Rochester a ajudá-lo a reunir pesquisadores dos Estados Unidos interessados em teoria de autômatos, redes neurais e no estudo da inteligência. Eles organizaram um seminário de dois meses em Dartmouth no verão de 1956. Havia 10 participantes ao todo, incluindo Trenchard More de Princeton, Arthur Samuel da IBM, e Ray Solomonoff e Oliver Selfridge do MIT.

Dois pesquisadores da Carnegie Tech,¹³ Allen Newell e Herbert Simon, simplesmente roubaram o show. Embora os outros tivessem idéias e, em alguns casos, programas para aplicações específicas como jogos de damas, Newell e Simon já tinham um programa de raciocínio, o Logic Theorist (LT),

13. Agora Carnegie Mellon University (CMU).

sobre o qual Simon afirmou: “Criamos um programa de computador capaz de pensar não-numericamente e assim resolvemos o antigo dilema mente-corpo.”¹⁴ Logo após o seminário, o programa foi capaz de demonstrar a maioria dos teoremas do Capítulo 2 do livro *Principia Mathematica* de Russell e Whitehead. Contam que Russell ficou encantado quando Simon mostrou a ele que o programa havia criado uma prova de um teorema que era mais curta que a do livro. Os editores do *Journal of Symbolic Logic* ficaram menos impressionados; eles rejeitaram um artigo escrito em parceria por Newell, Simon e pelo Logic Theorist.

O seminário de Dartmouth não trouxe nenhuma novidade, mas apresentou uns aos outros todos os personagens importantes da história. Nos vinte anos seguintes, o campo seria dominado por essas pessoas e por seus alunos e colegas do MIT, da CMU, de Stanford e da IBM. Talvez o resultado mais duradouro do seminário tenha sido um acordo para adotar o nome sugerido por McCarthy para o campo: **inteligência artificial**. É possível que “racionalidade computacional” fosse melhor, mas “IA” foi o nome que se fixou.

Examinando a proposta do seminário de Dartmouth (McCarthy *et al.*, 1955), podemos ver por que era necessário que a IA se tornasse um campo separado. Por que todo o trabalho feito na IA não podia ficar sob o nome de teoria de controle, pesquisa operacional ou teoria da decisão que, afinal de contas, têm objetivos semelhantes aos da IA? Ou, então, por que a IA não poderia ser um ramo da matemática? Primeiro, porque a IA abraçou desde o início a idéia de reproduzir faculdades humanas como criatividade, auto-aperfeiçoamento e uso da linguagem, e nenhum dos outros campos tratava dessas questões. A segunda resposta é a metodologia. A IA é o único desses campos que claramente é um ramo da ciência da computação (embora a pesquisa operacional compartilhe uma ênfase em simulações por computador), e IA é o único campo a tentar construir máquinas que funcionarão de forma autônoma em ambientes complexos e mutáveis.

Entusiasmo inicial, grandes expectativas (1952-1969)

Os primeiros anos da IA foram repletos de sucessos – mas de uma forma limitada. Considerando-se os primitivos computadores, as ferramentas de programação da época e o fato de que apenas alguns anos antes os computadores eram vistos como objetos capazes de efetuar operações aritméticas e nada mais, causava surpresa o fato de um computador realizar qualquer atividade remotamente inteligente. Em geral, a classe intelectual preferia acreditar que “uma máquina nunca poderá realizar X”. (Veja no Capítulo 26 uma longa lista de X reunidos por Turing.) Os pesquisadores da IA respondiam naturalmente demonstrando um X após outro. John McCarthy se referiu a esse período como a era do “Olhe mamãe, sem as mãos!”.

O sucesso inicial de Newell e Simon prosseguiu com o General Problem Solver (Solucionador de problemas gerais) ou GPS. Diferente do Logic Theorist, esse programa foi projetado desde o início para imitar protocolos humanos de resolução de problemas. Dentro da classe limitada de quebra-cabeças com a qual podia lidar, verificou-se que a ordem em que o programa considerava submetas e ações possíveis era semelhante à ordem em que os seres humanos abordavam os mesmos problemas. Desse modo, o GPS talvez tenha sido o primeiro programa a incorporar a abordagem de “pensar de forma humana”. O sucesso do GPS e de programas subseqüentes como modelos de cognição

14. Newell e Simon também criaram uma linguagem de processamento de listas, a IPL, para escrever o LT. Eles não tinham nenhum compilador e fizeram a conversão para código de máquina à mão. Para evitar erros, trabalharam em paralelo, gritando números binários um para o outro à medida que escreviam cada instrução, a fim de ter certeza de que os números concordavam.

SISTEMA DE
SÍMBOLOS
FÍSICOS

levaram Newell e Simon (1976) a formular a famosa hipótese do sistema de símbolos físicos, que afirma que "um sistema de símbolos físicos tem os meios necessários e suficientes para uma ação inteligente geral". O que eles queriam dizer é que qualquer sistema (ser humano ou máquina) que exibe inteligência deve operar manipulando estruturas de dados compostas por símbolos. Veremos mais adiante que essa hipótese enfrentou desafios provenientes de muitas direções.

Na IBM, Nathaniel Rochester e seus colegas produziram alguns dos primeiros programas de IA. Herbert Gelernter (1959) construiu o Geometry Theorem Prover, que podia demonstrar teoremas que seriam considerados bastante complicados por muitos alunos de matemática. A partir de 1952, Arthur Samuel escreveu uma série de programas para jogos de damas que eventualmente aprendiam a jogar em um nível amador elevado. Ao mesmo tempo, ele contestou a idéia de que os computadores só podem realizar as atividades para as quais foram programados: seu programa aprendeu rapidamente a jogar melhor que seu criador. O programa foi demonstrado na televisão em fevereiro de 1956, causando uma impressão muito forte. Como Turing, Samuel teve dificuldades para conseguir um horário em que pudesse utilizar os computadores. Trabalhando à noite, ele usou máquinas que ainda estavam na bancada de testes na fábrica da IBM. O Capítulo 6 aborda os jogos de computador e o Capítulo 21 descreve e expande as técnicas de aprendizado usadas por Samuel.

LISP

John McCarthy saiu de Dartmouth para o MIT, e lá contribuiu com três realizações cruciais em um ano histórico: 1958. No MIT AI Lab Memo No. 1, McCarthy definiu a linguagem de alto nível Lisp, que acabou por se tornar a linguagem de programação dominante na IA. A linguagem Lisp é a segunda mais antiga entre as linguagens de alto nível importantes em uso atualmente, sendo um ano mais jovem que FORTRAN. Com Lisp, McCarthy teve a ferramenta de que precisava, mas o acesso a recursos de computação escassos e dispendiosos também eram um sério problema. Em resposta, ele e outros pesquisadores do MIT criaram o compartilhamento de tempo (*time sharing*). Também em 1958, McCarthy publicou um artigo intitulado *Programs with common sense*, em que descrevia o Advice Taker, um programa hipotético que pode ser visto como o primeiro sistema de IA completo. Como o Logic Theorist e o Geometry Theorem Prover, o programa de McCarthy foi projetado para usar o conhecimento com a finalidade de buscar soluções para problemas. Entretanto, diferente dos outros, ele procurava incorporar o conhecimento geral do mundo. Por exemplo, McCarthy mostrou que alguns axiomas simples permitiriam ao programa gerar um plano para dirigir até o aeroporto e embarcar em um avião. O programa também foi criado de forma a poder aceitar novos axiomas no curso normal de operação, permitindo assim que ele adquirisse competência em novas áreas *sem ser reprogramado*. Portanto, o Advice Taker incorporava os princípios centrais de representação de conhecimento e de raciocínio: de que é útil ter uma representação formal e explícita do mundo e do modo como as ações de um agente afetam o mundo, e ser capaz de manipular essas representações com processos dedutivos. É notável como grande parte do artigo de 1958 permanece relevante até hoje.

O ano de 1958 também marcou a época em que Marvin Minsky foi para o MIT. Porém, sua colaboração inicial com McCarthy não durou muito. McCarthy enfatizava a representação e o raciocínio em lógica formal, enquanto Minsky estava mais interessado em fazer os programas funcionarem e, eventualmente, desenvolveu uma perspectiva contrária ao estudo da lógica. Em 1963, McCarthy fundou o laboratório de IA em Stanford. Seu plano de usar a lógica para construir o Advice Taker definitivo foi antecipado pela descoberta feita por J. A. Robinson do método de resolução (um algoritmo completo para demonstração de teoremas para a lógica de primeira ordem; consulte o Capítulo 9). O trabalho em Stanford enfatizou métodos de uso geral para raciocínio lógico. As aplicações da lógica incluíam os sistemas para responder a perguntas e os sistemas de planejamento de Cordell Green (Green, 1969b) e o projeto de robótica Shakey no novo Stanford Research Institute (SRI). Esse último projeto, descrito com mais detalhes no Capítulo 25, foi o primeiro a demonstrar a integração completa do raciocínio lógico e da atividade física.

Minsky orientou vários alunos que escolheram problemas limitados cuja solução parecia exigir inteligência. Esses domínios limitados se tornaram conhecidos como **micromundos**. O programa SAINT de James Slagle (1963a) era capaz de resolver problemas de cálculo integral típicos do primeiro ano dos cursos acadêmicos. O programa ANALOGY de Tom Evans (1968) resolvia problemas de analogia geométrica que aparecem em testes de QI, como o da Figura 1.4. O programa STUDENT de Daniel Bobrow (1967) resolvia problemas clássicos de álgebra, como este:

Se o número de clientes que Tom consegue é igual ao dobro do quadrado de 20% do número de anúncios que ele publica, e se o número de anúncios publicados é 45, qual é o número de clientes que Tom consegue?

O mais famoso micromundo foi o mundo de blocos, que consiste em um conjunto de blocos sólidos colocados sobre uma mesa (ou, com maior frequência, sobre a simulação de uma mesa), como mostra a Figura 1.5. Uma tarefa típica nesse mundo é reorganizar os blocos de uma certa maneira, utilizando a mão de um robô que pode erguer um bloco de cada vez. O mundo de blocos foi a base do projeto de visão de David Huffman (1971), do trabalho em visão e propagação de restrições de David

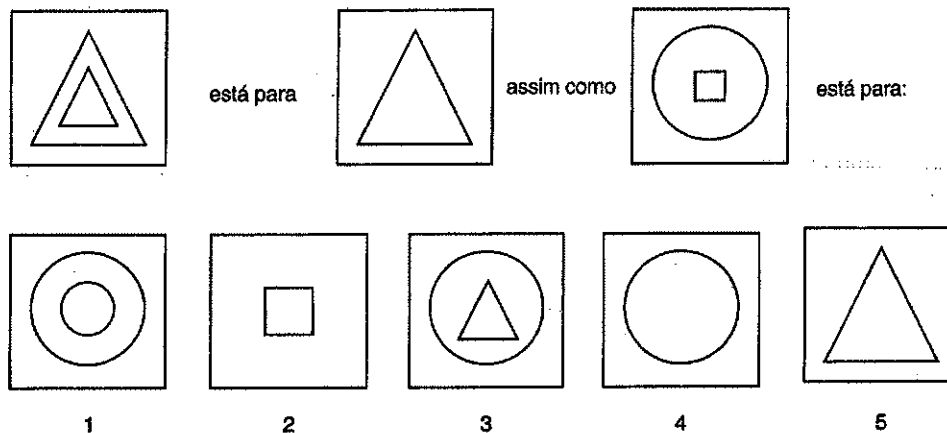


Figura 1.4 Um exemplo de problema resolvido pelo programa ANALOGY de Evans.

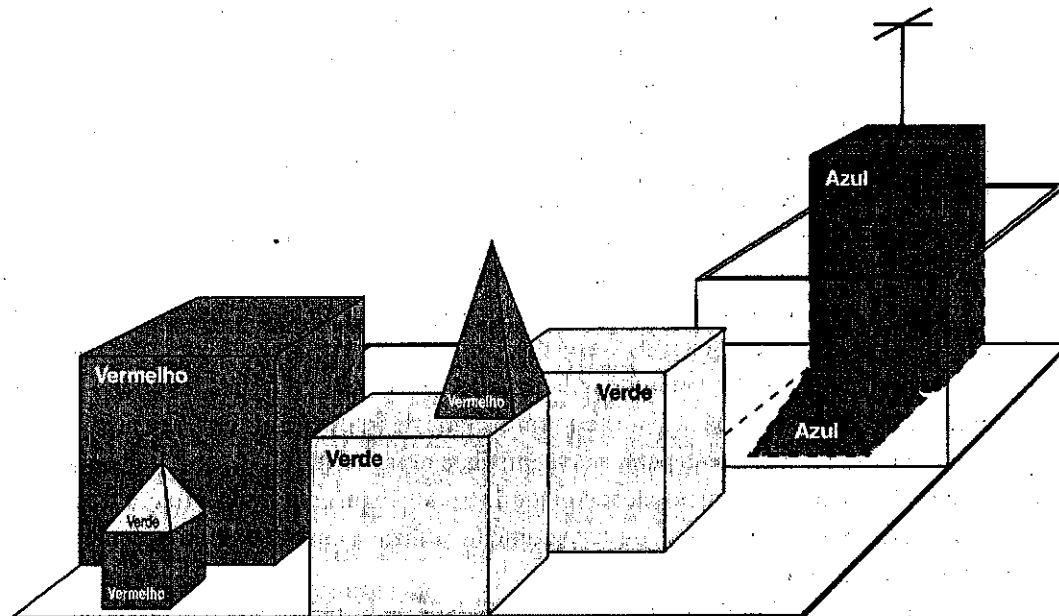


Figura 1.5 Uma cena do mundo de blocos. O programa SHRDLU (Winograd, 1972) tinha acabado de completar o comando: "Encontre um bloco mais alto que o bloco que você está segurando e coloque-o na caixa."

Waltz (1975), da teoria do aprendizado de Patrick Winston (1970), do programa de compreensão de linguagem natural de Terry Winograd (1972) e da agenda eletrônica de Scott Fahlman (1974).

O trabalho pioneiro baseado nas redes neurais de McCulloch e Pitts também prosperou. O trabalho de Winograd e Cowan (1963) mostrou que um grande número de elementos podia representar coletivamente um conceito individual, com um aumento correspondente na robustez e no paralelismo. Os métodos de aprendizado de Hebb foram aperfeiçoados por Bernie Widrow (Widrow e Hoff, 1960; Widrow, 1962), que denominou suas redes **adalines**, e por Frank Rosenblatt (1962) com seus **perceptrons**. Rosenblatt provou o **teorema de convergência do perceptron**, mostrando que seu algoritmo de aprendizado podia ajustar os pesos de conexão de um perceptron para corresponderem a quaisquer dados de entrada, desde que existisse tal correspondência. Esses tópicos são focalizados no Capítulo 20.

Uma dose de realidade (1966-1973)

Desde o início, os pesquisadores da IA eram ousados nos prognósticos de seus sucessos futuros. Esta declaração de Herbert Simon em 1957 frequentemente é citada:

Não é meu objetivo surpreendê-los ou chocá-los – mas o modo mais simples de resumir tudo isso é dizer que agora existem no mundo máquinas que pensam, aprendem e criam. Além disso, sua capacidade de realizar essas atividades está crescendo rapidamente até o ponto – em um futuro visível – no qual a variedade de problemas com que elas poderão lidar será correspondente à variedade de problemas com os quais lida a mente humana.

Termos como “futuro visível” podem ser interpretados de várias maneiras, mas Simon também fez uma previsão mais concreta: a de que dentro de dez anos um computador seria campeão de xadrez e que um teorema matemático significativo seria provado por uma máquina. Essas previsões se realizaram (ou quase) no prazo de 40 anos, em vez de 10. O excesso de confiança de Simon se devia ao desempenho promissor dos primeiros sistemas de IA em exemplos simples. Contudo, em quase todos os casos, esses primeiros sistemas acabaram falhando desastrosamente quando foram experimentados em conjuntos de problemas mais extensos ou em problemas mais difíceis.

O primeiro tipo de dificuldade surgiu porque a maioria dos primeiros programas continha pouco ou nenhum conhecimento de seu assunto; eles obtinham sucesso por meio de manipulações sintáticas simples. Uma história típica ocorreu durante os primeiros esforços de tradução por máquina, que foram generosamente subsidiados pelo National Research Council dos Estados Unidos, em uma tentativa de acelerar a tradução de documentos científicos russos após o lançamento do Sputnik em 1957. Inicialmente, imaginava-se que transformações sintáticas simples baseadas nas gramáticas russas e inglesas e a substituição de palavras com a utilização de um dicionário eletrônico seriam suficientes para preservar os significados exatos das orações. O fato é que a tradução exige conhecimento geral do assunto para solucionar ambigüidades e estabelecer o conteúdo da sentença. A famosa retradução de “o espírito está disposto mas a carne é fraca” como “a vodca é boa mas a carne é podre” ilustra as dificuldades encontradas. Em 1966, um relatório criado por um comitê consultivo descobriu que “não existe nenhum sistema de tradução automática para texto científico em geral, e não existe nenhuma perspectiva imediata nesse sentido”. Toda a subvenção do governo dos Estados Unidos para projetos acadêmicos de tradução foi cancelada. Hoje, a tradução automática é uma ferramenta imperfeita, mas amplamente utilizada em documentos técnicos, comerciais, governamentais e da Internet.

O segundo tipo de dificuldade foi a impossibilidade de tratar muitos dos problemas que a IA estava tentando resolver. A maior parte dos primeiros programas de IA resolvia problemas experi-

mentando diferentes combinações de passos até encontrar a solução. Essa estratégia funcionou inicialmente, porque os micromundos continham pouquíssimos objetos e, conseqüentemente, um número muito pequeno de ações possíveis e seqüências de soluções muito curtas. Antes do desenvolvimento da teoria de complexidade computacional, era uma crença geral que o "aumento da escala" para problemas maiores era apenas uma questão de haver hardware mais rápido e maior capacidade de memória. Por exemplo, o otimismo que acompanhou o desenvolvimento da demonstração da resolução de teoremas logo foi ofuscado quando os pesquisadores não conseguiram provar teoremas que envolviam mais que algumas dezenas de fatos. *O fato de um programa poder encontrar uma solução em princípio não significa que o programa contém quaisquer dos mecanismos necessários para encontrá-la na prática.*



EVOLUÇÃO
DA MÁQUINA

A ilusão do poder computacional ilimitado não ficou confinada aos programas de resolução de problemas. Os primeiros experimentos de **evolução da máquina** (agora chamados **algoritmos genéticos**) (Friedberg, 1958; Friedberg *et al.*, 1959) se baseavam na convicção sem dúvida correta de que, realizando-se uma série apropriada de pequenas mutações em um programa em código de máquina, seria possível gerar um programa com bom desempenho para qualquer tarefa simples. Então, a idéia era experimentar mutações aleatórias com um processo de seleção para preservar mutações que parecessem úteis. Apesar de milhares de horas de tempo de CPU, quase nenhum progresso foi demonstrado. Os algoritmos genéticos modernos utilizam representações melhores e têm mais sucesso.

A incapacidade de conviver com a "explosão combinatória" foi uma das principais críticas à IA contidas no relatório Lighthill (Lighthill, 1973), que formou a base para a decisão do governo britânico de encerrar o apoio à pesquisa da IA em todas as universidades, com exceção de duas. (A tradição oral pinta um quadro um pouco diferente e mais colorido, com ambições políticas e hostilidades pessoais, cuja descrição não nos interessa aqui.)

Uma terceira dificuldade surgiu devido a algumas limitações fundamentais nas estruturas básicas que estavam sendo utilizadas para gerar o comportamento inteligente. Por exemplo, o livro de Minsky e Papert, *Perceptrons* (1969), provou que, embora os perceptrons (uma forma simples de rede neural) pudessem aprender tudo que eram capazes de representar, eles podiam representar muito pouco. Em particular, um perceptron de duas entradas não podia ser treinado para reconhecer quando suas duas entradas eram diferentes. Embora seus resultados não se aplicassem a redes mais complexas de várias camadas, a subvenção de pesquisas relacionadas a redes neurais logo se reduziu a quase nada. Ironicamente, os novos algoritmos de aprendizado por retropropagação para redes de várias camadas que acabaram de provocar um enorme renascimento na pesquisa de redes neurais no final da década de 1980 foram, na verdade, descobertos primeiro em 1969 (Bryson e Ho, 1969).

Sistemas baseados em conhecimento: a chave para o poder? (1969-1979)

O quadro de resolução de problemas que havia surgido durante a primeira década de pesquisas em IA foi o de um mecanismo de busca de uso geral que procurava reunir passos elementares de raciocínio para encontrar soluções completas. Tais abordagens foram chamadas **métodos fracos** porque, embora gerais, não podiam ter aumento de escala para instâncias de problemas grandes ou difíceis. A alternativa para métodos fracos é usar um conhecimento mais amplo e específico de domínio que permita passos de raciocínio maiores e que possam tratar com mais facilidade casos que ocorrem tipicamente em especialidades estritas. Podemos dizer que, para resolver um problema difícil, praticamente é necessário já saber a resposta.

O programa DENDRAL (Buchanan *et al.*, 1969) foi um exemplo inicial dessa abordagem. Ele foi desenvolvido em Stanford, onde Ed Feigenbaum (um antigo aluno de Herbert Simon), Bruce Bucha-

MÉTODOS
FRACOS

nan (um filósofo transformado em cientista de computação) e Joshua Lederberg (um geneticista laureado com um prêmio Nobel) formaram uma equipe para resolver o problema de inferir a estrutura molecular a partir das informações fornecidas por um espectrômetro de massa. A entrada para o programa consiste na fórmula elementar da molécula (por exemplo, $C_6H_{13}NO_2$) e o espectro de massa que fornece as massas dos diversos fragmentos da molécula gerada quando ela é bombardeada por um feixe de elétrons. Por exemplo, o espectro de massa poderia conter um pico em $m = 15$, correspondendo à massa de um fragmento metil (CH_3).

A versão ingênua do programa gerou todas as estruturas possíveis consistentes com a fórmula, e depois previu qual seria o espectro de massa observado para cada uma, comparando esse espectro com o espectro real. Como se poderia esperar, esse é um problema intratável para moléculas de tamanho razoável. Os pesquisadores do DENDRAL consultaram especialistas em química analítica e descobriram que eles trabalhavam procurando padrões conhecidos de picos no espectro que sugerissem subestruturas comuns na molécula. Por exemplo, a regra a seguir é usada para reconhecer um subgrupo cetona ($C=O$), que pesa 28 unidades de massa:

se existem dois picos em x_1 e x_2 tais que

(a) $x_1 + x_2 = M + 28$ (M é a massa da molécula inteira);

(b) $x_1 - 28$ é um pico;

(c) $x_2 - 28$ é um pico;

(d) No mínimo um entre x_1 e x_2 é alto. :

então existe um subgrupo cetona

O reconhecimento de que a molécula contém uma subestrutura específica reduz enormemente o número de possíveis candidatos. O DENDRAL era eficiente porque:

Todo o conhecimento teórico relevante para resolver esses problemas foi mapeado de sua forma geral no [componente de previsão de espectro] ("princípios básicos") para formas especiais eficientes ("receitas de bolo"). (Feigenbaum *et al.*, 1971)

O DENDRAL foi importante porque representou o primeiro sistema bem-sucedido de *conhecimento intensivo*: sua habilidade derivava de um grande número de regras de propósito específico. Sistemas posteriores também incorporaram o tema principal da abordagem de McCarthy no Advice Taker – a separação clara entre o conhecimento (na forma de regras) e o componente de raciocínio.

Com essa lição em mente, Feigenbaum e outros pesquisadores de Stanford iniciaram o Heuristic Programming Project (HPP) para investigar até que ponto a nova metodologia de **sistemas especialistas** poderia ser aplicada a outras áreas do conhecimento humano. Em seguida, o principal esforço foi dedicado à área de diagnóstico médico. Feigenbaum, Buchanan e o Dr. Edward Shortliffe desenvolveram o MYCIN para diagnosticar infecções sanguíneas. Com cerca de 450 critérios (regras), o MYCIN foi capaz de se sair tão bem quanto alguns especialistas, e muito melhor do que médicos em início de carreira. Ele também apresentava duas diferenças importantes em relação ao DENDRAL. Primeiro, diferente das regras do DENDRAL, não havia nenhum modelo teórico geral a partir do qual as regras do MYCIN pudessem ser deduzidas. Elas tinham de ser adquiridas a partir de entrevistas extensivas com especialistas que, por sua vez, as adquiriam de livros didáticos, de outros especialistas e da experiência direta de estudos de casos. Em segundo lugar, as regras tinham de refletir a incerteza associada ao conhecimento médico. O MYCIN incorporava um cálculo de incerteza chamado **fatores de certeza** (consulte o Capítulo 13) que pareciam (na época) se adequar bem à forma como os médicos avaliavam o impacto das evidências no diagnóstico.

A importância do conhecimento de domínio também ficou aparente na área da compreensão da linguagem natural. Embora o sistema SHRDLU de Winograd para reconhecimento da lingua-

gem natural tivesse despertado bastante interesse, sua dependência da análise sintática provocou alguns problemas idênticos aos que ocorreram nos primeiros trabalhos em tradução automática (ou tradução por máquina). Ele foi capaz de superar a ambigüidade e reconhecer referências pronominais, mas isso acontecia principalmente porque o programa foi criado especificamente para uma única área – o mundo dos blocos. Diversos pesquisadores, entre eles Eugene Charniak, aluno graduado e companheiro de Winograd no MIT, sugeriram que uma compreensão robusta da linguagem exigiria conhecimentos gerais sobre o mundo e um método genérico para utilizar esses conhecimentos.

Em Yale, o lingüista transformado em pesquisador da IA Roger Schank enfatizou esse ponto, afirmando: “Não existe essa coisa de sintaxe.” Isso irritou muitos lingüistas, mas serviu para dar início a uma discussão útil. Schank e seus alunos elaboraram uma série de programas (Schank e Abelson, 1977; Wilensky, 1978; Schank e Riesbeck, 1981; Dyer, 1983), todos com a tarefa de entender a linguagem natural. Porém, a ênfase foi menos na linguagem em si e mais nos problemas de representação e raciocínio com o conhecimento exigido para compreensão da linguagem. Os problemas incluíam a representação de situações estereotípicas (Cullingford, 1981), descrição da organização da memória humana (Rieger, 1976; Kolodner, 1983) e compreensão de planos e metas (Wilensky, 1983).

FRAMES

O enorme crescimento das aplicações para resolução de problemas reais causou um aumento simultâneo na demanda por esquemas utilizáveis de representação do conhecimento. Foi desenvolvido um grande número de diferentes linguagens de representação e raciocínio. Algumas se baseavam na lógica – por exemplo, a linguagem Prolog se tornou popular na Europa, e a família PLANNER, nos Estados Unidos. Outras, seguindo a idéia de frames de Minsky (1975), adotaram uma abordagem mais estruturada, reunindo fatos sobre tipos específicos de objetos e eventos, e organizando os tipos em uma grande hierarquia taxonômica análoga a uma taxonomia biológica.

A IA se torna uma indústria (de 1980 até a atualidade)

O primeiro sistema especialista comercial bem-sucedido, o R1, iniciou sua operação na Digital Equipment Corporation (McDermott, 1982). O programa ajudou a configurar pedidos de novos sistemas de computadores; em 1986, ele estava fazendo a empresa economizar cerca de 40 milhões de dólares por ano. Em 1988, o grupo de IA da DEC tinha 40 sistemas especialistas entregues, com outros sendo produzidos. A Du Pont tinha 100 desses sistemas em uso e 500 em desenvolvimento, economizando aproximadamente 10 milhões de dólares por ano. Quase todas as corporações importantes dos Estados Unidos tinham seu próprio grupo de IA e estavam usando ou investigando sistemas especialistas.

Em 1981, os japoneses anunciaram o projeto “Fifth Generation”, um plano de 10 anos para montar computadores inteligentes que utilizassem Prolog. Em resposta, os Estados Unidos formaram a Microelectronics and Computer Technology Corporation (MCC) como um consórcio de pesquisa projetado para assegurar a competitividade nacional. Em ambos os casos, a IA fazia parte de um amplo esforço, incluindo o projeto de chips e a pesquisa da interface humana. No entanto, os componentes de IA dos projetos da MCC e Fifth Generation nunca atenderam a suas metas ambiciosas. Na Inglaterra, o relatório Alvey reabilitou o subsídio que havia cortado em consequência do relatório Lighthill.¹⁵

15. Para evitar embaraços, foi criado um novo campo chamado IKBS (Intelligent Knowledge-Based Systems), porque a IA havia sido oficialmente cancelada.

De modo geral, a indústria da IA se expandiu de alguns milhões de dólares em 1980 para bilhões de dólares em 1988. Logo depois, veio um período chamado de "inverno da IA", em que muitas empresas sofreram à medida que deixaram de cumprir promessas extravagantes.

O retorno das redes neurais (de 1986 até a atualidade)

Embora a ciência da computação tenha em grande parte abandonado o campo de redes neurais no final dos anos 70, o trabalho continuou em outros campos. Físicos como John Hopfield (1982) usaram técnicas da mecânica estatística para analisar as propriedades de armazenamento e de otimização das redes, tratando coleções de nós como coleções de átomos. Os psicólogos, incluindo David Rumelhart e Geoff Hinton, continuaram o estudo de modelos de memória de redes neurais. Conforme discutimos no Capítulo 20, o verdadeiro ímpeto ocorreu na metade da década de 1980, quando pelo menos quatro grupos diferentes recriaram o algoritmo de aprendizado por retropropagação, descoberto inicialmente em 1969 por Bryson e Ho. O algoritmo foi aplicado a muitos problemas de aprendizagem em ciência da computação e psicologia, e a ampla disseminação dos resultados na coletânea *Parallel Distributed Processing* (Rumelhart e McClelland, 1986) causou grande excitação.

CONEXIONISMO

Os chamados modelos **conexionistas** para sistemas inteligentes eram vistos por alguns como concorrentes diretos dos modelos simbólicos promovidos por Newell e Simon e da abordagem logicista de McCarthy e outros pesquisadores (Smolensky, 1988). Pode parecer óbvio que em certo nível os seres humanos manipulam símbolos – de fato, o livro de Terrence Deacon, *The Symbolic Species* (1997), sugere que essa é a *característica que define* os seres humanos, mas os conexionistas mais fervorosos questionavam se a manipulação de símbolos tinha qualquer função explicativa real em modelos detalhados de cognição. Essa pergunta permanece sem resposta, mas a visão atual é que as abordagens conexionista e simbólica são complementares, e não concorrentes.

A IA se torna uma ciência (de 1987 até a atualidade)

Nos últimos anos, houve uma revolução no trabalho em inteligência artificial, tanto no conteúdo quanto na metodologia.¹⁶ Agora, é mais comum usar as teorias existentes como bases, em vez de propor teorias inteiramente novas, fundamentar as afirmações em teoremas rigorosos ou na evidência experimental rígida, em vez de utilizar como base a intuição e destacar a relevância de aplicações reais em vez de exemplos de brinquedos.

Em parte, a IA surgiu como uma rebelião contra as limitações de áreas existentes como a teoria de controle e a estatística, mas agora ela inclui esses campos. Conforme afirmou David McAllester (1998):

No período inicial da IA, parecia plausível que novas formas de computação simbólica, como frames e redes semânticas, tornariam obsoleta grande parte da teoria clássica. Isso levou a uma forma de isolacionismo na qual a IA ficou bem separada do restante da ciência da computação. Atualmente, esse isolacionismo está sendo abandonado. Existe o reconhecimento de que o aprendizado da máquina não deve ser isolado da teoria da informação, de que o raciocínio incerto não deve ser isolado da modelagem estocástica, de que a busca não deve ser isolada da otimização clássica e do controle, e de que o raciocínio automatizado não deve ser isolado dos métodos formais e da análise estática.

16. Alguns caracterizaram essa mudança como uma vitória dos **puros** – aqueles que pensam que as teorias da IA devem se fundamentar no rigor matemático – sobre os **impuros** – aqueles que preferem experimentar muitas idéias, escrever alguns programas e depois avaliar o que parece estar funcionando. As duas abordagens são importantes. Um deslocamento em direção à pureza implica que o campo alcançou um nível de estabilidade e maturidade. Se essa estabilidade será interrompida por uma nova idéia impura é outra questão.

Em termos de metodologia, a IA finalmente adotou com firmeza o método científico. Para serem aceitas, as hipóteses devem ser submetidas a rigorosos experimentos empíricos, e os resultados devem ser analisados estatisticamente de acordo com sua importância (Cohen, 1995). Usando-se a Internet e repositórios compartilhados de dados e código de teste, agora é possível reproduzir experimentos.

O campo de reconhecimento da fala ilustra o padrão. Nos anos 70, foi experimentada uma ampla variedade de arquiteturas e abordagens diferentes. Muitas delas eram bastante *ad hoc* e frágeis, e foram demonstradas em apenas alguns exemplos especialmente selecionados. Nos últimos anos, abordagens baseadas em **modelos ocultos de Markov (MOMs)** passaram a dominar a área. Dois aspectos de MOMs são relevantes. Primeiro, eles se baseiam em uma teoria matemática rigorosa. Isso permitiu que os pesquisadores da fala utilizassem como fundamentos várias décadas de resultados matemáticos desenvolvidos em outros campos. Em segundo lugar, eles são gerados por um processo de treinamento em um grande conjunto de dados reais de fala. Isso assegura um desempenho robusto e, em testes cegos rigorosos, os MOMs têm melhorado suas pontuações de forma contínua. A tecnologia da fala e o campo inter-relacionado de reconhecimento de caracteres manuscritos já estão efetuando a transição para aplicações industriais e de consumo em larga escala.

As redes neurais também seguem essa tendência. Grande parte do trabalho em redes neurais nos anos 80 foi realizada na tentativa de definir a abrangência do que poderia ser feito e de aprender como as redes neurais diferem das técnicas "tradicionais". Utilizando uma metodologia aperfeiçoada e estruturas teóricas, o campo chegou a uma compreensão tal que, agora, as redes neurais podem ser comparadas a técnicas correspondentes da estatística, do reconhecimento de padrões e do aprendizado de máquina, podendo ser utilizada a técnica mais promissora em cada aplicação. Como resultado desse desenvolvimento, a tecnologia denominada **mineração de dados** gerou uma nova e vigorosa indústria.

MINERAÇÃO
DE DADOS

A obra de Judea Pearl, *Probabilistic Reasoning in Intelligent Systems* (1988), levou a uma nova aceitação da probabilidade e da teoria da decisão na IA, seguindo um renascimento do interesse descrito no artigo de Peter Cheeseman, "In Defense of Probability" (1985). O formalismo denominado **rede bayesiana** foi criado para permitir a representação eficiente do conhecimento incerto e o raciocínio rigoroso com a utilização desse tipo de conhecimento. Essa abordagem supera amplamente muitos problemas dos sistemas de raciocínio probabilístico das décadas de 1960 e 1970; agora ele domina a pesquisa de IA sobre raciocínio incerto e sistemas especialistas. A abordagem admite o aprendizado a partir da experiência e combina o melhor da IA clássica e das redes neurais. O trabalho de Judea Pearl (1982a) e de Eric Horvitz e David Heckerman (Horvitz e Heckerman, 1986; Horvitz *et al.*, 1986) promoveu a idéia de sistemas especialistas *normativos*: sistemas que agem racionalmente de acordo com as leis da teoria de decisão e não procuram imitar os passos do pensamento de especialistas humanos. O sistema operacional WindowsTM inclui vários sistemas especialistas de diagnóstico normativo para correção de problemas. Os Capítulos 13 a 16 examinam essa área.

Revoluções suaves semelhantes a essa ocorreram nos campos de robótica, visão computacional e representação de conhecimento. Uma compreensão melhor dos problemas e de suas propriedades de complexidade, combinada à maior sofisticação matemática, resultou em agendas de pesquisa utilizáveis e métodos robustos. Em muitos casos, a formalização e a especialização também levaram à fragmentação: tópicos como visão e robótica estão cada vez mais isolados do trabalho "principal" em IA. A visão unificada da IA como projeto de agentes racionais é uma área que pode trazer de volta a unidade para esses campos discrepantes.

O surgimento de agentes inteligentes (de 1995 até a atualidade)

Talvez encorajados pelo progresso na resolução dos subproblemas da IA, os pesquisadores também começaram a examinar mais uma vez o problema do "agente como um todo". O trabalho de Allen

Newell, John Laird e Paul Rosenbloom no SOAR (Newell, 1990; Laird *et al.*, 1987) é o exemplo mais conhecido de uma arquitetura completa de agente. O assim denominado movimento estabelecido tem como objetivo entender o funcionamento interno de agentes incorporados a ambientes reais com entradas sensoriais contínuas. Um dos ambientes mais importantes para agentes inteligentes é a Internet. Os sistemas de IA se tornaram tão comuns em aplicações da Web que o sufixo “-bot” passou a fazer parte da linguagem cotidiana. Além disso, as tecnologias da IA servem de base a muitas ferramentas da Internet, como mecanismos de pesquisa, sistemas de recomendação (*recommender systems*) e sistemas de construção de Web sites.

Além da primeira edição deste texto (Russell e Norvig, 1995), outros textos recentes também adotaram a perspectiva de agente (Poole *et al.*, 1998; Nilsson, 1998). Uma consequência da tentativa de construir agentes completos é a percepção de que os subcampos anteriormente isolados da IA talvez precisem ser reorganizados quando seus resultados tiverem de ser reunidos. Em particular, agora é amplamente aceito que os sistemas sensoriais (visão, sonar, reconhecimento da fala etc.) não podem oferecer informações perfeitamente confiáveis sobre o ambiente. Em consequência disso, os sistemas de raciocínio e planejamento devem ser capazes de lidar com a incerteza. A segunda consequência importante da perspectiva de agente é que a IA foi levada a um contato muito mais estreito com outros campos, como a teoria de controle e a economia, que também lidam com agentes.

1.4 O estado da arte

O que a IA pode fazer hoje? É difícil uma resposta concisa, porque existem muitas atividades em vários subcampos. Aqui, mostramos algumas aplicações; outras serão apresentadas ao longo do livro.

Planejamento autônomo e escalonamento: A uma centena de milhões de quilômetros da Terra, o programa Remote Agent da NASA se tornou o primeiro programa de planejamento autônomo de bordo a controlar o escalonamento de operações de uma nave espacial (Jonsson *et al.*, 2000). O Remote Agent gerou planos de metas de alto nível especificadas a partir do solo e monitorou a operação da nave espacial à medida que os planos eram executados – efetuando a detecção, o diagnóstico e a recuperação de problemas conforme eles ocorriam.

Jogos: O Deep Blue da IBM se tornou o primeiro programa de computador a derrotar o campeão mundial em uma partida de xadrez, ao vencer Garry Kasparov por um placar de 3,5 a 2,5 em uma partida de exibição (Goodman e Keene, 1997). Kasparov disse que sentiu “uma nova espécie de inteligência” do outro lado do tabuleiro. A revista *Newsweek* descreveu a partida como: “O último reduto do cérebro.” O valor das ações da IBM teve um aumento de 18 bilhões de dólares.

Controle autônomo: O sistema de visão de computador ALVINN foi treinado para dirigir um automóvel, mantendo-o na pista. Ele foi colocado na minivan controlada por computador NAVLAB da CMU e foi utilizado para percorrer os Estados Unidos – ao longo de quase 4.600km o ALVINN manteve o controle da direção do veículo durante 98% do tempo. Um ser humano assumiu o comando nos outros 2%, principalmente na saída de declives. A NAVLAB tem câmeras de vídeo que transmitem imagens da estrada para ALVINN, que então calcula a melhor forma de guiar, baseado na experiência obtida em sessões de treinamento anteriores.

Diagnóstico: Programas de diagnóstico médico baseados na análise probabilística foram capazes de executar tarefas no nível de um médico especialista em diversas áreas da medicina. Heckerman (1991) descreve um caso em que um importante especialista em patologia de gânglios linfáticos ridiculariza o diagnóstico de um programa em um caso especialmente difícil. Os criadores do programa sugeriram que ele pedisse ao computador uma explicação do diagnóstico. A máquina destacou os principais fatores que influenciaram sua decisão e explicou a interação sutil de vários sintomas nesse caso. Mais tarde, o especialista concordou com o programa.

Planejamento logístico: Durante a crise do Golfo Pérsico em 1991, as forças armadas dos Estados Unidos distribuíram uma ferramenta denominada Dynamic Analysis and Replanning Tool, ou DART (Cross e Walker, 1994), a fim de realizar o planejamento logístico automatizado e a programação de execução do transporte. Isso envolveu até 50.000 veículos, transporte de carga aérea e pessoal ao mesmo tempo, e teve de levar em conta pontos de partida, destinos, rotas e resolução de conflitos entre todos os parâmetros. As técnicas de planejamento da IA permitiram a geração em algumas horas de um plano que exigiria semanas com outros métodos. A Defense Advanced Research Project Agency (DARPA) declarou que essa única aplicação compensou com folga os 30 anos de investimento da DARPA em IA.

Robótica: Muitos cirurgiões agora utilizam robôs assistentes em microcirurgias. O HipNav (DiGioia *et al.*, 1996) é um sistema que emprega técnicas de visão computacional para criar um modelo tridimensional da anatomia interna de um paciente, e depois utiliza controle robótico para orientar a inserção de uma prótese de substituição do quadril.

Reconhecimento de linguagem e resolução de problemas: O PROVERB (Littman *et al.*, 1999) é um programa de computador que resolve quebra-cabeças de palavras cruzadas melhor que a maioria dos seres humanos, utilizando restrições sobre possíveis preenchimentos de palavras, um grande banco de dados de quebra-cabeças anteriores e uma variedade de fontes de informações que incluem dicionários e bancos de dados on-line, como uma lista de filmes e dos atores que participam deles. Por exemplo, ele descobre que a pista "Nice Story" pode ser resolvido por "ETAGE", porque seu banco de dados inclui o par pista/solução "Story in France/ETAGE" e porque reconhece que os padrões "NiceX" e "X in France" com frequência têm a mesma solução. O programa não sabe que Nice é uma cidade da França, mas consegue resolver o quebra-cabeças.

Esses são apenas alguns exemplos de sistemas de inteligência artificial que existem hoje em dia. Não é mágica ou ficção científica – mas sim ciência, engenharia e matemática, e este livro apresenta uma introdução a tudo isso.

1.5 Resumo

Este capítulo define a IA e estabelece os fundamentos culturais sobre os quais ela se desenvolveu. Alguns pontos importantes são:

- Pessoas diferentes refletem de modo distinto sobre a IA. Duas questões importantes são: você se preocupa com o pensamento ou com o comportamento? Você quer modelar seres humanos ou trabalhar a partir de um padrão ideal?
- Neste livro, adotamos a visão de que a inteligência está relacionada principalmente a uma **ação racional**. No caso ideal, um **agente inteligente** adota a melhor ação possível em uma situação. Estudaremos o problema da criação de agentes que são inteligentes nesse sentido.
- Os filósofos (desde 400 a.C.) tornaram a IA concebível, considerando as idéias de que a mente é, em alguns aspectos, semelhante a uma máquina, de que ela opera sobre o conhecimento codificado em alguma linguagem interna e que o pensamento pode ser usado para escolher as ações que deverão ser executadas.
- Os matemáticos forneceram as ferramentas para manipular declarações de certeza lógica, bem como declarações incertas e probabilísticas. Eles também definiram a base para a compreensão da computação e do raciocínio sobre algoritmos.
- Os economistas formalizaram o problema de tomar decisões que maximizam o resultado esperado para o tomador de decisões.

- Os psicólogos adotaram a idéia de que os seres humanos e os animais podem ser considerados máquinas de processamento de informações. Os lingüistas mostraram que o uso da linguagem se ajusta a esse modelo.
- Os engenheiros de computação forneceram os artefatos que tornam possíveis as aplicações de IA. Os programas de IA tendem a ser extensos e não poderiam funcionar sem os grandes avanços em velocidade e memória que a indústria de informática tem proporcionado.
- A teoria de controle lida com o projeto de dispositivos que agem de forma ótima com base no *feedback* do ambiente. Inicialmente, as ferramentas matemáticas da teoria de controle eram bem diferentes da IA, mas os campos estão se tornando mais próximos.
- A história da IA teve ciclos de sucesso, otimismo impróprio e quedas resultantes no entusiasmo e na subvenção. Também houve ciclos de introdução de novas abordagens criativas e de aprimoramento sistemático das melhores estratégias.
- A IA avançou mais rapidamente na última década, devido ao uso mais intenso do método científico nas experiências e na comparação entre as abordagens.
- O progresso recente na compreensão da base teórica da inteligência caminha lado a lado com os avanços na capacidade de sistemas reais. Os subcampos da IA se tornaram mais integrados, e a IA encontrou uma área de concordância com outras disciplinas.

Notas bibliográficas e históricas

O *status* metodológico da inteligência artificial é investigado em *The Sciences of the Artificial*, de Herb Simon (1981), que descreve áreas de pesquisas relacionadas a artefatos complexos. Ele explica como a IA pode ser visualizada ao mesmo tempo como ciência e matemática. Cohen (1995) apresenta uma visão geral da metodologia experimental dentro da IA. Ford e Hayes (1995) fornecem uma visão comentada da utilidade do teste de Turing.

Artificial Intelligence: The Very Idea de John Haugeland (1985) expõe de forma lúcida os problemas filosóficos e práticos da IA. A ciência cognitiva é bem descrita em vários textos recentes (Johnson-Laird, 1988; Stillings *et al.*, 1995; Thagard, 1996) e na *Encyclopedia of the Cognitive Sciences* (Wilson e Keil, 1999). Baker (1989) estuda a parte sintática da lingüística moderna, e Chierchia e McConnell-Ginet (1990) abordam a semântica. Jurafsky e Martin (2000) focalizam a lingüística computacional.

O início da IA é descrito no livro de Feigenbaum e Feldman, *Computers and Thought* (1963), em *Semantic Information Processing* de Minsky (1968) e na série de *Machine Intelligence*, editada por Donald Michie. Um grande número de documentos importantes foi reunido em uma antologia por Webber e Nilsson (1981), e por Luger (1995). Os primeiros documentos sobre redes neurais estão reunidos em *Neurocomputing* (Anderson e Rosenfeld, 1988). A *Encyclopedia of AI* (Shapiro, 1992) contém artigos de pesquisa sobre quase todos os tópicos relacionados à IA. Em geral, esses artigos fornecem um bom ponto de partida para o estudo da literatura de pesquisa sobre cada tópico.

O trabalho mais recente aparece nos anais das conferências sobre IA mais importantes: a bienal International Joint Conference on AI (IJCAI), a anual European Conference on AI (ECAI) e a National Conference on AI, conhecida principalmente como AAAI, que representa a organização que a patrocina. Os principais periódicos referentes à IA em geral são *Artificial Intelligence*, *Computational Intelligence*, o *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Intelligent Systems* e o jornal eletrônico *Journal of Artificial Intelligence Research*. Também existem muitas conferências e pe-

riódicos dedicados a áreas específicas, que abordaremos nos capítulos apropriados. As principais associações profissionais para a IA são a American Association for Artificial Intelligence (AAAI), o ACM Special Interest Group in Artificial Intelligence (SIGART) e a Society for Artificial Intelligence and Simulation of Behaviour (AISB). A revista da AAAI *AI Magazine* contém muitos artigos sobre tópicos variados e tutoriais, e seu Web site, aaai.org, contém notícias e informações básicas.

Exercícios

Estes exercícios foram planejados para estimular a discussão, e alguns poderiam ser definidos como projetos semestrais. Como outra alternativa, podem ser feitas tentativas preliminares para resolvê-los agora, e essas tentativas podem ser revistas após a conclusão da leitura.

1.1 Defina com suas próprias palavras: (a) inteligência, (b) inteligência artificial, (c) agente.



1.2 Leia o artigo original de Turing sobre IA (Turing, 1950). No artigo, ele discute várias objeções potenciais a suas propostas e a seu teste de inteligência. Que objeções ainda têm algum peso? Suas refutações são válidas? Você poderia imaginar novas objeções que resultem de desenvolvimentos ocorridos desde que Turing escreveu o artigo? No artigo, ele prevê que, por volta do ano 2000, um computador teria 30% de chance de passar em um teste de Turing de cinco minutos com um interrogador sem experiência. Que chance de aprovação no teste um computador teria hoje? E daqui a mais 50 anos?



1.3 Todo ano, o prêmio Loebner é entregue ao programa que chega mais perto de ser aprovado em uma versão do teste de Turing. Pesquise e faça um relatório sobre o último vencedor do prêmio Loebner. Que técnicas ele utiliza? Que avanço o programa representou para o estado da arte em IA?

1.4 Existem classes conhecidas de problemas intratavelmente difíceis para computadores, e outras classes de problemas cujo caráter indeterminado pode ser demonstrado. Isso significa que a IA é impossível?

1.5 Suponha que o programa ANALOGY de Evans seja estendido de forma a alcançar o grau 200 em um teste padrão de QI. Nesse caso, teríamos um programa mais inteligente que um ser humano? Explique.

1.6 Como a introspecção – o exame que alguém faz de seus próprios pensamentos mais íntimos – poderia ser imprecisa? Eu poderia estar errado sobre aquilo em que estou pensando? Discuta.



1.7 Examine a literatura da IA para descobrir se as tarefas a seguir podem ser resolvidas atualmente por computadores:

- a. Jogar uma partida decente de tênis de mesa (pingue-pongue).
- b. Dirigir no centro do Cairo.
- c. Comprar mantimentos para uma semana no mercado.
- d. Comprar mantimentos para uma semana na Web.
- e. Jogar uma partida decente de *bridge* em um nível competitivo.
- f. Descobrir e provar novos teoremas matemáticos.
- g. Escrever uma história intencionalmente engraçada.
- h. Dar conselhos jurídicos idôneos em uma área especializada da lei.
- i. Traduzir o inglês falado para o sueco falado em tempo real.
- j. Realizar uma operação cirúrgica complexa.

Para as tarefas inviáveis no momento, procure descobrir quais são as dificuldades e prever quando, se for o caso, elas serão superadas.

1.8 Alguns autores afirmam que a percepção e as habilidades motoras são a parte mais importante da inteligência, e que capacidades “de alto nível” são necessariamente parasitas – simples complementos dessas capacidades básicas. Certamente, a maior parte da evolução e grande parte do cérebro foram dedicadas à percepção e às habilidades motoras, enquanto a IA descobriu que, em muitos aspectos, tarefas como a participação em jogos e a inferência lógica são mais fáceis do que a percepção e a ação no mundo real. Você acredita que o foco tradicional da IA em habilidades cognitivas de alto nível é equivocado?

1.9 Por que a evolução tenderia a resultar em sistemas que agem racionalmente? Quais são os objetivos de projeto de tais sistemas?

1.10 As ações reflexas (como afastar a mão de um fogão quente) são racionais? Eles são inteligentes?

1.11 “Sem dúvida, os computadores não podem ser inteligentes – eles só podem fazer o que seus programadores determinam.” Esta última afirmação é verdadeira e implica a primeira?

1.12 “Sem dúvida, os animais não podem ser inteligentes – eles só podem fazer o que seus genes determinam.” Esta última afirmação é verdadeira e implica a primeira?

1.13 “Sem dúvida, animais, seres humanos e computadores não podem ser inteligentes – eles só podem fazer o que seus átomos constituintes determinam, de acordo com as leis da física.” Esta última afirmação é verdadeira e implica a primeira?