



UNIVERSIDADE
ESTADUAL de LONDRINA

Guilherme Henrique Gonçalves Silva

SÍNTESE DO PRIMEIRO CAPÍTULO
DO LIVRO “ARTIFICIAL
INTELLIGENCE: A MODERN
APPROACH” DE STUART RUSSELL

Londrina 2023

INTRODUÇÃO

O campo da inteligência artificial (IA) explora a natureza da inteligência humana e visa criar entidades inteligentes. Surgiu após a Segunda Guerra Mundial e tornou-se uma área proeminente da ciência e da engenharia. A IA abrange vários subcampos, incluindo aprendizado, percepção, jogo de xadrez, prova de teoremas matemáticos, redação de poesia, direção e diagnóstico de doenças. É considerado um campo universal, aplicável a qualquer tarefa intelectual.

1.1 O QUE É IA?

O texto discute diferentes abordagens e definições de Inteligência Artificial (IA). As configurações variam de acordo com o foco nos processos de pensamento ou comportamento e fidelidade ao desempenho humano ou à racionalidade ideal.

A IA abrange várias disciplinas, como processamento de linguagem natural, representação do conhecimento, computador automatizado, aprendizado de máquina, visão computacional e robótica. O Teste de Turing é mencionado como um marco para avaliar a inteligência de um sistema de IA. No entanto, o texto argumenta que os investigadores têm mais interesse em entender os princípios subjacentes da inteligência do que em passar sem testar. Isso é comparado à inovação na aviação, que avançou quando se deixou de tentar imitar os pássaros e se passou a entender melhor a aerodinâmica.

O texto explora como programas de computador podem imitar o pensamento humano, utilizando métodos como introspecção, experimentos psicológicos e imageamento cerebral. Se um programa de computador imita o comportamento humano, isso pode indicar que ele usa uma negociação semelhante aos humanos para o pensamento. Um exemplo mencionado é o "General Problem Solver", um programa projetado para replicar passos de raciocínio humano.

Além disso, o texto discute a ciência cognitiva, que combina IA e psicologia para criar teorias testáveis da mente humana. No entanto, a ciência cognitiva genuína depende de experimentação com humanos ou animais reais.

Finalmente, o texto enfatiza a importância de distinguir a eficiência de um algoritmo de sua capacidade de imitar o desempenho humano, pois esses são aspectos diferentes que permitirão o rápido desenvolvimento da IA e da ciência cognitiva.

O texto discute a evolução do pensamento lógico e racional, começando com as contribuições de Aristóteles para o estudo da lógica e do pensamento irrefutável, até as tentativas modernas de implementar esses conceitos em inteligência artificial (IA). A "tradição logicista" na IA procura construir sistemas inteligentes com base em programas capazes de resolver problemas expressos em notação lógica. No entanto, existem desafios em formalizar conhecimento informal em notação lógica e em aplicar essas soluções na prática.

Outro conceito discutido é o de "agente racional", que age de forma a obter o melhor resultado possível. Embora a inferência correta seja uma parte da racionalidade, nem todas as ações racionais envolvem inferência. O texto argumenta que a abordagem do agente racional tem vantagens sobre outras abordagens da IA sendo mais geral e mais adequada para o desenvolvimento científico. No entanto, ele também reconhece que a racionalidade perfeita não é viável em ambientes complexos devido às demandas computacionais.

1.2 OS FUNDAMENTOS DA INTELIGÊNCIA ARTIFICIAL

O texto apresenta de forma concisa uma história das disciplinas que contribuíram para o desenvolvimento da Inteligência Artificial (IA), com foco na Filosofia. Aborda questões fundamentais relacionadas à aplicabilidade de regras formais para deduzir conclusões válidas, à origem do conhecimento, à relação entre mente e cérebro físico, bem como à conexão entre conhecimento e ação.

Inicia-se mencionando Aristóteles, o primeiro a formular leis racionais para o pensamento. Em seguida, o texto faz referência a pensadores como Thomas Hobbes, que comparou o raciocínio ao cálculo numérico, e Rene Descartes, que discutiu a distinção entre mente e matéria.

A seguir, são abordados o empirismo, movimento que sugere que o conhecimento é adquirido por meio dos sentidos, e o positivismo lógico, que combina racionalismo e empirismo.

Posteriormente, o texto explora a conexão entre conhecimento e ação, enfatizando a importância dessa ligação para a IA, uma vez que a inteligência requer tanto ação quanto raciocínio. Nesse sentido, são citadas as contribuições de Aristóteles e sua conexão lógica entre objetivos e resultados de ações, bem como o algoritmo baseado em objetivos para tomada de decisões.

Em seguida, o texto apresenta as bases matemáticas da IA começando pela formalização da lógica, computação e probabilidade. Destaca-se que a lógica formal teve origem com os filósofos da Grécia Antiga, sendo aprimorada por George Boole, Gottlob Frege e Alfred Tarski. Além disso, são abordados os limites da lógica e da computação, mencionando-se o teorema da incompletude de Kurt Gödel, que demonstrou a existência de funções incomputáveis, e o trabalho de Alan Turing na definição de funções computáveis.

O texto discute também a importância da divisão de problemas complexos em subproblemas tratáveis, introduzindo a teoria da NP-completude. Além disso, trata-se da teoria da probabilidade, cuja origem remonta a Gerolamo Cardano, e que é amplamente utilizada nas abordagens modernas de raciocínio incerto em sistemas de IA.

No que diz respeito à ciência econômica, o texto descreve sua relevância na maximização de resultados. Destaca-se a origem da economia com Adam Smith, bem como a formalização da teoria da utilidade por Leon Walras, Frank Ramsey e outros. Também são mencionados a teoria da decisão, que combina a teoria da probabilidade com a teoria da utilidade para decisões tomadas sob incerteza, e a teoria dos jogos de John von Neumann e Oskar Morgenstern. O texto conclui abordando a questão da decisão racional em situações sequenciais, uma área de estudo no campo da pesquisa operacional.

Em seguida, o texto explora como o cérebro processa informações, apresentando o desenvolvimento histórico e as técnicas atuais na área da neurociência. Destaca-se que, apesar do entendimento limitado sobre como o cérebro possibilita o pensamento, sabe-se que esse órgão é responsável pela consciência. São mencionados comentários históricos, como a observação de Aristóteles sobre o tamanho proporcionalmente grande do cérebro humano e o trabalho de Paul Broca na identificação de áreas cerebrais responsáveis por funções cognitivas específicas.

Outros marcos históricos incluem a técnica de coloração desenvolvida por Camillo Golgi, que permitiu a observação de neurônios individuais, e os estudos pioneiros de Santiago Ramon y Cajal sobre as estruturas neuronais do cérebro.

Apesar de alguns dados sobre o mapeamento entre áreas cerebrais e partes do corpo estarem disponíveis atualmente, ainda há muito a ser compreendido sobre a memória individual e a funcionalidade do cérebro após danos em determinadas áreas.

O texto discute a evolução da medição da atividade cerebral, desde o eletroencefalograma (EEG) de Hans Berger até a moderna imagem por ressonância magnética funcional (fMRI), além dos avanços no registro da atividade de neurônios individuais.

Por fim, o texto compara os cérebros aos computadores digitais, observando as diferenças em termos de velocidade de processamento, capacidade de armazenamento e interconexão. Embora os supercomputadores tenham capacidades comparáveis às do cérebro humano, a completa replicação da inteligência cerebral ainda é um desafio.

O texto explora também a evolução do estudo do comportamento e do pensamento humano e animal, desde suas raízes na psicologia experimental até a moderna ciência cognitiva.

Destaca-se o trabalho inicial de Hermann von Helmholtz e Wilhelm Wundt na psicologia experimental, enfatizando a importância dos experimentos controlados, embora a subjetividade dos dados gerados tenha limitado seu alcance. O movimento behaviorista, liderado por John Watson, é mencionado por seu foco em medidas objetivas de estímulos e respostas, encontrando mais sucesso no estudo de animais do que de humanos.

A psicologia cognitiva, que considera o cérebro como um dispositivo de processamento de informações, é traçada até os trabalhos de William James e Helmholtz. Embora tenha sido obscurecida pelo behaviorismo nos Estados Unidos, floresceu em Cambridge sob a direção de Frederic Bartlett.

Kenneth Craik, aluno e sucessor de Bartlett, reforçou a legitimidade de termos "mentais", como crenças e objetivos, argumentando que eles eram tão científicos quanto a pressão e a temperatura, por exemplo. Craik detalhou os três principais passos de um agente baseado em conhecimento: tradução do estímulo em uma representação interna, manipulação da representação por meio de processos cognitivos para obter novas representações internas.

O texto conclui destacando que o desenvolvimento da modelagem por computador nos Estados Unidos levou à criação do campo da ciência cognitiva, que compartilha a visão de que "uma teoria cognitiva deve ser como um programa de computador".

A história e a evolução dos computadores são exploradas, desde os primeiros dias durante a Segunda Guerra Mundial até os dias atuais. Máquinas como o Heath Robinson, o Colossus, o Z-3, o ABC e o ENIAC foram desenvolvidas inicialmente para propósitos específicos, como a decifração de mensagens e cálculos militares.

O texto observa que cada geração subsequente de hardware de computador trouxe aumentos na velocidade e capacidade, enquanto os preços diminuía. O desempenho dobrava a cada 18 meses, aproximadamente, até cerca de 2005, quando problemas de dissipação de energia levaram os fabricantes a aumentar o número de

núcleos da CPU em vez de aumentar a velocidade do relógio. Agora, espera-se que futuros aumentos de poder venham do paralelismo maciço, o que curiosamente converge com as propriedades do cérebro.

Antes do computador eletrônico, existiam dispositivos de cálculo, cujos primeiros remontam ao século XVII. A primeira máquina programável foi um tear inventado por Joseph Marie Jacquard em 1805, que usava cartões perfurados para armazenar instruções. Charles Babbage também é destacado por seu papel na história da computação.

Por fim, o texto reconhece o papel crucial da ciência da computação no lado do software para a inteligência artificial, fornecendo sistemas operacionais, linguagens de programação e ferramentas necessárias. O trabalho em IA tem fornecido muitas ideias que foram incorporadas à ciência da computação convencional.

1.3 A HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL

O texto examina os primeiros anos da Inteligência Artificial (IA), marcados por avanços significativos considerando as limitações das ferramentas da época. Os pioneiros da IA desafiaram o pensamento estabelecido de que as máquinas só poderiam realizar tarefas aritméticas. Warren McCulloch e Walter Pitts se destacaram ao propor um modelo de neurônios artificiais baseado na fisiologia dos neurônios cerebrais, lógica proposicional e teoria da computação de Turing. Em 1950, Marvin Minsky e Dean Edmonds construíram o SNARL, o primeiro computador de rede neural. Minsky também contribuiu com teoremas limitantes para a pesquisa em redes neurais. Além disso, Alan Turing é reconhecido como uma das figuras mais influentes na IA tendo apresentado conceitos como o Teste de Turing, aprendizado de máquina, algoritmos genéticos e aprendizado por reforço.

John McCarthy, outra figura influente na IA, desempenhou um papel crucial ao ajudar a organizar um workshop no Dartmouth College em 1956, considerado o marco inicial do campo. Durante esse evento, Allen Newell e Herbert Simon se destacaram com seu programa de raciocínio, o Logic Theorist (LT). McCarthy mais tarde fez importantes contribuições ao definir a linguagem de programação Lisp, ajudar a inventar o compartilhamento de tempo e descrever o Advice Taker, um programa hipotético que representava um sistema completo de IA. Em 1958, McCarthy mudou-se para o MIT, onde fundou o laboratório de IA em Stanford.

Marvin Minsky também se juntou ao MIT em 1958, colaborando inicialmente com McCarthy, mas adotando posteriormente uma abordagem antilógica. Durante sua carreira, Minsky supervisionou uma série de estudantes que trabalharam em micromundos, problemas limitados que pareciam exigir inteligência para serem resolvidos. Além disso, seu trabalho com redes neurais, baseadas nos princípios estabelecidos por McCulloch e Pitts, é mencionado como uma importante contribuição para o campo da IA.

O texto destaca as previsões otimistas feitas pelos primeiros pesquisadores de Inteligência Artificial (IA), como Herbert Simon, que, em 1957, antecipou que máquinas seriam capazes de pensar, aprender e criar, expandindo rapidamente suas capacidades. Embora suas previsões específicas, como um computador se tornando campeão de xadrez, tenham se concretizado em 40 anos, e não em 10 como ele previu, o otimismo inicial foi seguido por desafios e dificuldades.

Os primeiros programas de IA mostraram promissores em exemplos simples, mas enfrentaram dificuldades ao lidar com problemas mais complexos. Essas dificuldades decorreram principalmente da falta de conhecimento aprofundado sobre os domínios específicos envolvidos, resultando em erros em tarefas como a tradução automática.

O texto apresenta um relato sobre o êxito dos sistemas especialistas na indústria durante a década de 1980. Um exemplo notável foi a Digital Equipment Corporation, que obteve sucesso na implementação do RI, um sistema especialista comercial, resultando em significativas economias. No entanto, apesar dos esforços ambiciosos em desenvolver a inteligência artificial (IA) nos EUA, Japão e Reino Unido, como o projeto "Fifth Generation" no Japão e o consórcio de pesquisa MCC nos EUA, essas iniciativas não conseguiram alcançar seus objetivos ousados, resultando no período conhecido como "Inverno da IA".

Na metade da década de 1980, ocorreu a redescoberta do algoritmo de aprendizagem por retro propagação, que foi utilizado em vários problemas de aprendizagem nas áreas da ciência da computação e psicologia. Esse evento gerou um entusiasmo significativo e levou ao desenvolvimento de modelos conexionistas de sistemas inteligentes, que foram considerados como concorrentes dos modelos simbólicos e da abordagem logicista. Embora a questão de se a manipulação de símbolos desempenha um papel explicativo nos modelos detalhados de cognição ainda não tenha sido completamente respondida, a visão atual é que as abordagens conexionistas e simbólicas são complementares.

O texto descreve a evolução recente da inteligência artificial (IA), marcada por um maior foco na validação rigorosa de teorias por meio de evidências experimentais e uma maior relevância para aplicações no mundo real. A IA, que inicialmente se mostrava em desacordo com campos existentes como teoria de controle e estatística, agora os abraça.

A metodologia da IA está firmemente enraizada no método científico, com hipóteses sendo submetidas a experimentos empíricos rigorosos e os resultados sendo analisados estatisticamente. Alguns exemplos notáveis de avanços incluem o reconhecimento de fala, que se beneficiou de modelos ocultos de Markov (HMMs) baseados em teorias matemáticas rigorosas, e a tradução automática, que retornou a uma abordagem baseada em sequências de palavras.

As redes neurais e a mineração de dados também seguiram essa tendência, resultando no surgimento de uma indústria vigorosa. A aceitação da teoria da probabilidade e da teoria da decisão na IA, exemplificada pelo trabalho de Judea Pearl, contribuiu para um domínio da pesquisa em IA no que diz respeito ao raciocínio incerto e sistemas especialistas.

Além disso, o texto menciona avanços em robótica, visão computacional e representação do conhecimento, todos se beneficiando de uma melhor compreensão dos problemas, propriedades de complexidade e maior sofisticação matemática. A reintegração desses campos com a IA convencional está resultando em benefícios significativos.

O texto discute a recente mudança de foco na pesquisa em inteligência artificial (IA), em direção à abordagem de "agente completo", influenciada pelos avanços em subáreas da IA. Essa mudança implica a necessidade de reorganizar subcampos anteriormente isolados da

1.4 O ESTADO DA ARTE

Atualmente, a inteligência artificial (IA) abrange uma ampla variedade de atividades em diversos subcampos, demonstrando sua crescente importância e impacto. Existem diversas aplicações notáveis que ilustram esse fato. Por exemplo, temos os veículos robóticos, como o STANLEY, um carro autônomo equipado com sensores e software avançados, que conquistou o DARPA Grand Challenge em 2005. Além disso, os sistemas de reconhecimento de fala têm sido empregados para guiar conversas automatizadas, como no caso da reserva de voos com a United Airlines. A NASA desenvolveu o programa Remote Agent, capaz de planejar e programar operações de forma autônoma para espaçonaves, enquanto o DEEP BLUE, da IBM, derrotou o campeão mundial de xadrez, evidenciando a capacidade da IA em realizar tarefas complexas.

Os algoritmos de aprendizado têm desempenhado um papel crucial no combate ao spam, classificando eficientemente as mensagens indesejadas, e o planejamento logístico baseado em IA, como o DART, tem contribuído para agilizar as operações de transporte. Além disso, a robótica tem encontrado aplicações práticas em produtos como o aspirador Roomba e o PackBot, utilizado em zonas de conflito para auxiliar em operações de segurança. Adicionalmente, programas de tradução automática, construídos com base em modelos estatísticos, têm possibilitado traduções automáticas de qualidade.

Esses exemplos destacam as aplicações tangíveis da IA na atualidade, evidenciando a sua capacidade de interseção entre ciência, engenharia e matemática.