Inteligência Artificial

Aspectos Filosóficos

Prof. Paulo Martins Engel

Fronteiras da IA

- As discussões filosóficas pretendem investigar as fronteiras entre o natural e o artificial.
- Discutem aspectos mais especulativos.
  - Quais os limites da IA?
  - O que um computador não poderá fazer?
  - Por quê?
- Essas discussões são úteis para guiar novos projetos de investigação.
- Entretanto, as posições filosóficas muitas vezes são antagônicas.
- Dependem da própria experiência e posicionamento individual dos filósofos.

Motivação

- Fronteiras entre o natural e o artificial
- O que seria inteligência? Mente? Livre arbítrio?
- Uma máquina de fato inteligente pode não ser humana?
- Pode não ter processos mentais conscientes? E inconscientes?
- O que distingue a ciência da computação convencional da IA?
- É possível simular processos mentais com algoritmos?
- A mente de um cérebro é apenas um programa?
- Quais são os modelos básicos do comportamento e como construir máquinas para simulá-los?
- Até que ponto a inteligência pode ser descrita recorrendo à avaliação de regras, inferências, à dedução e à computação de padrões?
- Qual é o desempenho das máquinas que simulam tais comportamentos através destes métodos?

Habilidade essenciais do comportamento inteligente

- Responder a situações de modo flexível
- Tirar proveito de circunstâncias fortuitas
- Perceber o sentido de mensagens contraditórias ou ambíguas
- Reconhecer a importância relativa de elementos diferentes de uma dada situação
- Encontrar similaridades entre situações apesar de diferenças que existam entre elas
- Encontrar diferenças entre situações apesar das similaridades existentes
- Sintetizar novos conceitos a partir de conceitos já conhecidos relacionando estes de outras formas
- Imaginar novas idéias

IA forte e IA fraca

- Tese da IA fraca:
  - As máquinas podem *simular* um comportamento inteligente, agir como se fossem inteligentes
- Tese da IA forte:
  - As máquinas podem *realmente* pensar e não apenas simular o pensamento

IA forte e IA fraca

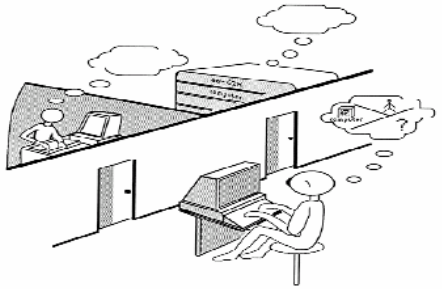
- Tese da IA forte:
  - A inteligência humana pode ser replicada
  - As explicações das funções cognitivas são apoiadas por processos computacionais
- Tese da IA fraca:
  - A inteligência humana pode, no melhor dos casos, apenas ser simulada
  - A meta para se atingir a inteligência artificial consiste em projetar máquinas que sejam capazes de exibir um comportamento inteligente
- As questões fundamentais podem ser resumidas na seguinte forma (Hofstadter, 1979): um sistema cognitivo (sistema lógico formal, teoria matemática, cérebro como "base" da mente) pode ocupar-se de si mesmo?
- Esta pergunta remonta aos problemas propostos por Hilbert (1900), Russel e Whitehead (*Principia Mathematica* tenta derivar toda matemática a partir da lógica) e outras questões relativas a provas matemáticas.

As máquinas podem pensar?

- Os filósofos estão interessados em comparar as duas arquiteturas – a humana e a da máquina.
- Máquinas podem pensar?
- Máquinas podem voar?
- Máquinas podem nadar?

Alternativa

- Em vez de perguntar se máquinas podem pensar, Turing sugeriu um teste comportamental de inteligência (teste de Turing).



A objeção da inaptidão

- Uma máquina nunca poderá:
  - ser amável, bonita, amigável, ter iniciativa, ter senso de humor, cometer enganos, apaixonar-se, gostar de morangos com creme, aprender a partir da experiência, usar palavras corretamente, ser o sujeito de seu próprio pensamento, fazer algo realmente novo, fazer alguém se apaixonar por ela, ...

A objeção da inaptidão

- Entretanto, os computadores hoje em dia fazem muita coisa:
  - jogam xadrez, damas, ...
  - inspecionam peças em linhas de montagem
  - verificam a grafia de documentos
  - pilotam aviões e helicópteros
  - fazem diagnóstico de doenças
  - .....

A objeção matemática

- Certas sentenças matemáticas são em princípio insolúveis para sistemas formais específicos.
- Exemplo: teorema da incompletude de Gödel.
- Filósofos como J. R. Lucas (1961) afirmaram que esse teorema mostra que as máquinas são mentalmente inferiores aos seres humanos.
- Roger Penrose também escreveu dois livros com argumentos nessa linha:
  - The emperor's new mind (1989)
  - Shadows of the mind (1994)

IA forte e IA fraca

- Existe um método (definitivo), ou seja um processo mecânico, que possa ser aplicado a uma declaração matemática, o qual possa dizer se esta pode ser provada?
- Não: Hardy (1928): afirma mas não prova formalmente
- Turing (1936)
  - superioridade da mente frente a dispositivos mecânicos
  - determinismo: modelagem de problemas complexos a partir de elementos (máquina de estados, autômatos, etc.)

IA forte e IA fraca

- A matemática é completa *e* consistente?
- Não (Gödel, 1931) e o Teorema da Incompletude: se S for consistente (não contém um teorema que seja a negação de um outro), então existem sentenças verdadeiras que são teoremas em S e que não podem ser demonstradas:
- É possível escrever declarações/sentenças matemáticas que se referem a elas mesmas (do tipo "Eu estou mentindo", "Esta sentença é falsa" ou "Esta declaração não pode ser provada") tais declarações não podem ser demonstradas ou seja provadas como sendo verdadeiras (levariam à uma contradição) ou como sendo falsas (idem)

IA forte e IA fraca

- O que isto tem a ver com a IA e a questão da mente?
- Um sistema formal S é um aparato simbólico (regras bem determinadas) o que implica em combinações simbólicas (teoremas de S) advindas das regras de S a máquina de Turing é um sistema formal (Teorema de Gödel se aplica)
- Se o cérebro funcionar como uma máquina de Turing, permanecer consistente implica em uma limitação (por exemplo, não conseguirá demonstrar todas as sentenças verdadeiras da aritmética elementar) mas, como o cérebro/mente (sistema S) consegue demonstrar (fora de S) que uma sentença (não demostrável no sentido de Gödel) é de fato verdadeira, então o cérebro/mente não é uma máquina de Turing, e um corolário seria: cérebro/mente não "pensa" algoritmicamente (Lucas, 1961; Penrose, 1989)

O argumento da informalidade do comportamento

- O comportamento humano é complexo demais para ser capturado por qualquer conjunto simples de regras, e como os computadores não podem fazer nada mais além de seguir um conjunto de regras, eles não podem gerar um comportamento tão inteligente quanto o dos seres humanos.
- O principal defensor desta visão foi o filósofo Hubert Dreyfus: What computers can't do (1972). What computers still can't do (1992).
- Mas as críticas se referem a sistemas programados logicamente a partir de fatos e regras que descrevem o domínio.

IA Forte: as máquinas podem realmente pensar?

- Somente quando uma máquina conseguir escrever um soneto ou compor um concerto em conseqüência de ter pensado e ter sentido emoções, e não pela disposição aleatória de símbolos, poderemos concordar que a máquina irá se equiparar ao cérebro – isto é, se ela não apenas escrever, mas souber o que escreveu (Geoffrey Jefferson, 1949)
- Argumento da consciência e da intencionalidade.

Cérebro, Mente e Consciência

- Definição de mente:
  - Intelecto, pensamento, entendimento; alma, espírito (parte distinta do corpo)
  - Capacidade mental ou psíquica; capacidade cognitiva e intelectual (em contraposição às emoções e intenções)
  - Memória, recordações
  - Concepção, imaginação
  - Local do sujeito da consciência; sistema relacionado aos fenômenos cognitivos e emocionais, experiências conscientes de um indivíduo
  - Intenção, intuito, desígnio, disposição, desejo
- Minsky, Pinker: mente como sistema (conjunto) de órgãos
- Máquinas podem possuir uma mente? Animais? Há correntes que afirmam que apenas seres humanos possuem mente (mas: mesmo sem ter um definição do que é pensar, se pensa)

Cérebro, Mente e Consciência

- Definições de consciência:
  - Ter um sentimento ou conhecimento das sensações ou sentidos de si mesmo ou de coisas externas
  - Ter um sentimento ou conhecimento do que uma coisa é ou de que uma coisa existe ou acontece
  - Percepção de si como um ser pensante; saber o que se está fazendo e porquê
  - O conhecimento do que acontece ao redor de uma pessoa; totalidade dos pensamentos de uma pessoa; sentimentos e impressões; mente; lembrança do que se fez recentemente (logo, a discussão sobre consciência em máquina é inútil)
- Visões de alguns pesquisadores sobre o tema mente e consciência (2001):
- Minsky: consciência não existe (em entrevista à ZDNET)
- Dennett: máquina (robô) consciente é possível; dificuldade é apenas financeira
- Davis: dificuldade é reproduzir o funcionamento do cérebro; progressos foram feitos apenas em relação a sensores e próteses artificiais
- Kurzweil: máquinas serão conscientes ainda na primeira metade do sec. XXI

O Problema Mente-Corpo

- Teorias
  - Mente é tudo (visão mentalista)
  - Mente nada mais é que um processo físico (visão materialista)
  - O mental e o material coexistem (visão dualista)
- Histórico
- Descartes (sec. XVII)
  - paralelo entre autômatos e corpo humano: estímulo externo ativa receptores do sistema nervoso → cérebro → nervos atuadores (ex.: joelho)
- Kant (sec. XVIII)
  - fundamentos metodológicos da ciência cognitiva (estrutura lógica do método de inferência de processos mentais)
  - Pré Kant: visão empirista do modelo da mente (Hume por exemplo mostrava que o unicórnio não existe mostrando que as duas impressões que podem ser adquiridas por um dos nossos sentidos no caso a visão ou seja a do cavalo e a do chifre, não ocorrem juntas

O Problema Mente-Corpo

- Mente e cérebro
- Senso estrito:
  - Mente é diferente do cérebro; a mente é o subconjunto das ações do cérebro que estão relacionadas com processamento de informações (em última análise, computação); a mente é o conjunto de processos que levam o cérebro de um estado a outro (Minsky)
- Teoria computacional da mente:
  - A informação permanece a mesma independente do meio que a transmite;
  - Desejos e crenças (conceitos tipicamente associados com a mente) são informações representadas por símbolos
  - Símbolos são armazenados em neurônios que são disparados por sensações
  - Novas crenças e desejos são formados
  - Um dado comportamento ocorre
- Debate atual:
  - A teoria computacional da mente resolve o problema mente-corpo?
  - Críticos: Searle (1980) e Penrose (1994)

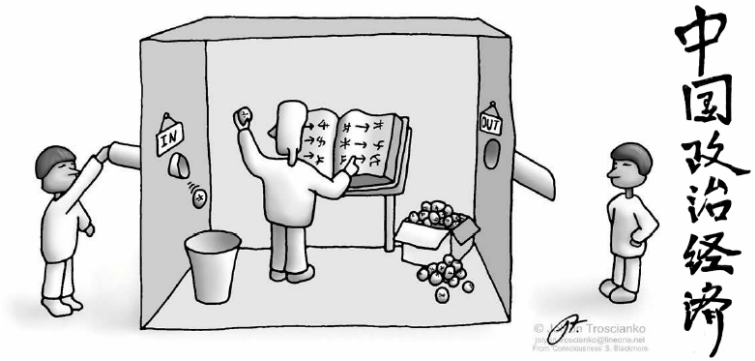
Cérebro e Mente

- "A mente de um cérebro é como um programa de computador?
- Não. Um programa simplesmente manipula símbolos, enquanto que um cérebro atribui significado a eles". (Searle)

O experimento do quarto chinês

- Searle descreve um sistema hipotético que passa no teste de Turing mas que (segundo Searle) o programa não entende nada de suas entradas e saídas.
- O sistema: um ser humano, que compreende apenas o português, equipado com um livro de regras escrito em português e diversas pilhas de papel, sendo algumas em branco e outras com inscrições indecifráveis (o ser humano é a CPU, o livro de regras o programa e o papel o dispositivo de armazenamento).
- O sistema está num quarto com uma pequena abertura para o exterior.
- Por essa abertura passam papéis com símbolos indecifráveis.
- O ser humano encontra símbolos correspondentes no livro de regras e segue as instruções que podem incluir escrever símbolos em novas folhas de papel, encontrar símbolos nas pilhas, reorganizar as pilhas, etc.
- Eventualmente, as instruções farão com que um ou mais símbolos sejam transcritos em uma folha de papel que será repassada ao exterior do quarto.

O quarto chinês



O experimento do quarto chinês

- Do exterior percebemos um sistema que está recebendo a entrada na forma de instruções em chinês e está gerando respostas em chinês, que são sem dúvida "inteligentes".
- Searle argumenta que a pessoa no quarto não entende o chinês (dado inicial).
- O livro de regras e o papel não entendem chinês.
- Então, não está acontecendo nenhuma compreensão do chinês.
- Por conseguinte, de acordo com Searle, a execução do programa correto não gera necessariamente compreensão.

O experimento da prótese cerebral
(Glymour 1970, Searle 1980, Moravec 1988)

- Suponha que a neurofisiologia tenha se desenvolvido a ponto de que o comportamento de entrada/saída e a conectividade de todos os neurônios do cérebro humano estejam perfeitamente compreendidos.
- Suponha que possamos construir dispositivos que imitem este comportamento e possam ter uma interface uniforme com o tecido neural.
- Suponha que uma técnica cirúrgica especial possa substituir gradualmente os neurônios da cabeça de alguém pelos dispositivos artificiais sem interromper o funcionamento do cérebro.

O experimento da prótese cerebral

- O que aconteceria com o comportamento externo da pessoa?
- E como ficaria a sua consciência?

IA: Discursos e Conceitos

- Propósito: Estudar a mente x Fazer programas?
- Observação: Comportamento x Estrutura?
- Natureza: Simular a Inteligência x Construir?
- Referência: Inteligência Humana x Geral?
- Modelo: Simbólico x Conexionista?

IA: Discursos e Conceitos

- Corporificação?
- Racionalidade?
- Intencionalidade?
- Consciência?
- Representação?
- Emoções?