

Trí Tuệ Nhân Tạo

Nguyễn Nhật Quang

quangnn-fit@mail.hut.edu.vn

Viện Công nghệ Thông tin và Truyền thông

Trường Đại học Bách Khoa Hà Nội

Năm học 2009-2010

Nội dung môn học:

- **Giới thiệu về Trí tuệ nhân tạo**
 - Định nghĩa
 - Các nền tảng
 - Lịch sử tóm tắt
 - Các thành tựu quan trọng
- Tác tử
- Giải quyết vấn đề: Tìm kiếm, Thỏa mãn ràng buộc
- Logic và suy diễn
- Biểu diễn tri thức
- Suy diễn với tri thức không chắc chắn
- Học máy
- Lập kế hoạch

Định nghĩa về TTNT (1)

- Các định nghĩa (quan điểm) về TTNT được chia thành 4 nhóm:
 - (1) Các hệ thống **suy nghĩ (thông minh) như con người**
 - *"The exciting new effort to make computers think ... machines with minds, in the full and literal sense."* (Haugeland, 1985)
 - *"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ..."* (Bellman, 1978)
 - (2) Các hệ thống **suy nghĩ hợp lý**
 - *"The study of mental faculties through the use of computational models."* (Charniak and McDermott, 1985)
 - *"The study of the computations that make it possible to perceive, reason, and act."* (Winston, 1992)

Định nghĩa về TTNT (2)

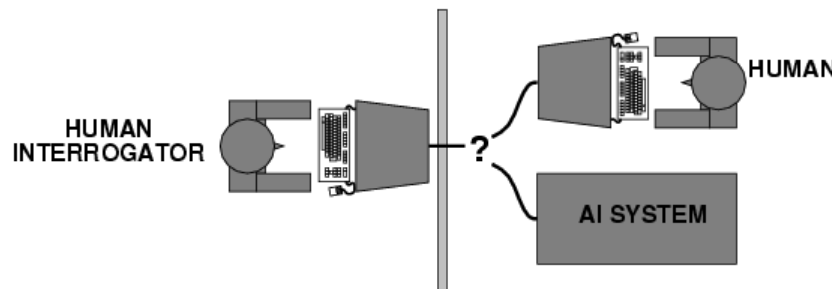
- ❑ (3) Các hệ thống **hành động (thông minh) như con người**
 - *"The art of creating machines that perform functions that require intelligence when performed by people."* (Kurzweil, 1990)
 - *"The study of how to make computers do things at which, at the moment, people are better."* (Rich and Knight, 1991)
- ❑ (4) Các hệ thống **hành động hợp lý**
 - *"Computational Intelligence is the study of the design of intelligent agents."* (Poole et al., 1998)
 - *"AI . . . is concerned with intelligent behavior in artifacts."* (Nilsson, 1998)

Định nghĩa về TTNT (3)

- Các định nghĩa (1) và (2) liên quan đến *các quá trình suy nghĩ và suy diễn*
- Các định nghĩa (3) và (4) liên quan đến *cách hành động*
- Các định nghĩa (1) và (3) đánh giá mức độ thành công (sự thông minh) theo *tiêu chuẩn của con người*
- Các định nghĩa (2) và (4) đánh giá mức độ thành công (sự thông minh) theo *tiêu chuẩn của sự hợp lý*
 - Một hệ thống hành động hợp lý, nếu nó làm các việc phù hợp đối với những gì nó (hệ thống) biết

Hành động như con người: Turing Test

- Turing (1950) “Máy tính toán và sự thông minh”:
- “Máy tính có thể suy nghĩ được không?” → “Máy tính có thể hành động một cách thông minh được không?”
- Thí nghiệm kiểm chứng hành động thông minh: Imitation Game



- Dự đoán rằng đến năm 2000, máy tính sẽ có 30% khả năng vượt qua một người không có chuyên môn đối với một bài kiểm tra (Turing test) trong 5 phút
- Turing (vào năm 1950) đã dự đoán trước các vấn đề tranh luận quan trọng trong TTNT trong vòng 50 năm sau
- Turing đã đề xuất các thành phần quan trọng của TTNT: tri thức, suy diễn, hiểu ngôn ngữ, học

Suy nghĩ như con người: Khoa học nhận thức

- Cuộc “cách mạng nhận thức” những năm 1960:
 - Xem bộ não người như một cấu trúc xử lý thông tin
 - Nghiên cứu về tâm lý nhận thức thay thế cho các nghiên cứu trước đó về hành vi ứng xử
- Cần các lý thuyết khoa học về các hoạt động bên trong của bộ não người
- Làm thế nào để xác nhận (kiểm chứng)? Yêu cầu:
 - 1) Dự đoán và kiểm chứng các hoạt động (hành vi) của chủ thể con người (hướng tiếp cận top-down), hoặc
 - 2) Nhận dạng (xác định) trực tiếp từ các dữ liệu về hệ thần kinh (hướng tiếp cận bottom-up)
- Hiện nay, cả 2 hướng tiếp cận này (Cognitive Science và Cognitive Neuroscience) được tách rời với lĩnh vực TTNT

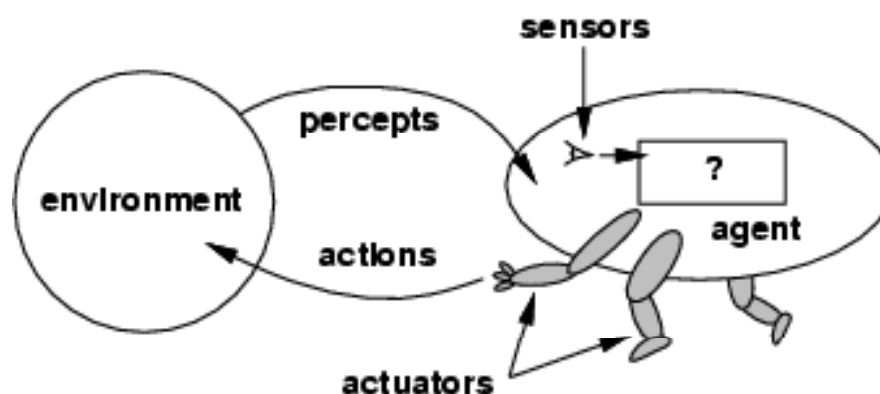
Suy nghĩ hợp lý: Các luật suy nghĩ

- Chuẩn hóa (hoặc quy tắc hóa), hơn là mô tả
- Aristotle: Thế nào là các quá trình suy nghĩ / tranh luận đúng đắn?
- Một số trường học ở Hy Lạp đã phát triển những dạng *logic*: *ký hiệu* và *các luật dẫn xuất* đối với các quá trình suy nghĩ
- Mối liên hệ trực tiếp, thông qua toán học và triết học, đối với khoa học TTNT hiện đại
- Các vấn đề:
 1. Không phải tất cả các hành vi (hành động) thông minh đều xuất phát từ các cân nhắc (suy nghĩ) logic
 2. Mục đích của sự suy nghĩ là gì? Những suy nghĩ nào mà tôi nên thực hiện, trong số các suy nghĩ mà tôi có thể có?

Hành động một cách hợp lý

- Hành động **một cách hợp lý**: thực hiện đúng việc cần làm
- *Đúng việc cần làm*: là việc (hành động) giúp cực đại hóa việc đạt được các mục tiêu, đối với các thông tin hiện có
- Không nhất thiết liên quan đến sự suy nghĩ – ví dụ, phản xạ chớp mắt
- Tuy nhiên, sự suy nghĩ nên được xem là thuộc vào hệ thống (nhóm) các hành động hợp lý
- Sự hợp lý cần phải tính đến cả độ phức tạp tính toán
 - Nếu chi phí về tài nguyên tính toán và thời gian quá cao, thì sẽ không có tính thực tế (không áp dụng được trong thực tế)

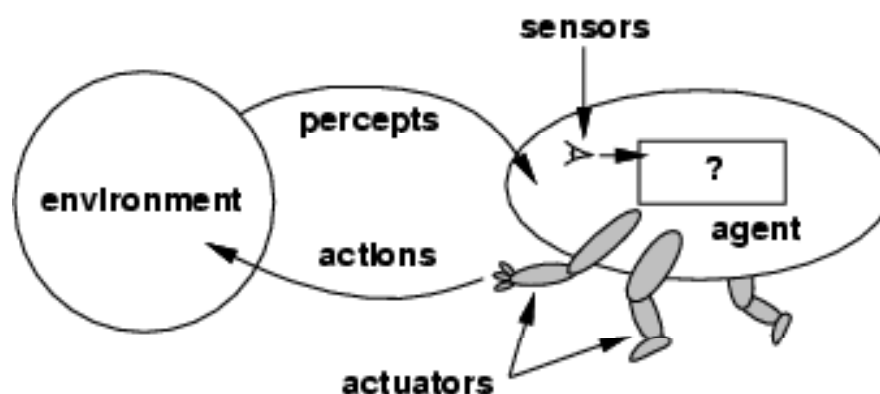
Các tác tử hợp lý (1)



- Một **tác tử** (agent) là một thực thể có khả năng nhận thức và hành động
- Một cách khái quát, một tác tử có thể được biểu diễn bằng một hàm ánh xạ: từ quá trình (lịch sử) nhận thức đến hành động:

$$f: P^* \rightarrow A$$

Các tác tử hợp lý (2)



- Đối với một tập (lớp) các môi trường và nhiệm vụ, chúng ta cần tìm ra tác tử (hoặc một lớp các tác tử) có hiệu suất tốt nhất
- Lưu ý: Các giới hạn về tính toán (của máy tính) không cho phép đạt được sự hợp lý hoàn hảo (tối ưu)
 - Mục tiêu: Thiết kế chương trình máy tính tối ưu đối với các tài nguyên máy tính hiện có

Các nền tảng của TTNT (1)

■ Triết học

- ❑ Logic
- ❑ Các phương pháp suy diễn
- ❑ Các cơ sở (nền tảng) của việc học
- ❑ Ngôn ngữ
- ❑ Sự hợp lý

■ Toán học

- ❑ Biểu diễn hình thức và các giải thuật chứng minh
- ❑ Tính toán
- ❑ Bài toán (vấn đề) giải quyết được và không giải quyết được
- ❑ Bài toán (vấn đề) áp dụng được và không áp dụng được (độ phức tạp tính toán - thời gian để giải quyết bài toán - là hàm mũ)
- ❑ Xác suất

Các nền tảng của TTNT (2)

■ Kinh tế học

- Hàm lợi ích (tiện ích)
- Lý thuyết ra quyết định

■ Khoa học thần kinh

- Nền tảng (cơ sở) tự nhiên của các hoạt động trí óc

■ Tâm lý học

- Sự thích nghi
- Các dấu hiệu của nhận thức và điều khiển vận động
- Các kỹ thuật thực nghiệm (vd: tâm sinh lý học,...)

Các nền tảng của TTNT (3)

- Công nghệ máy tính
 - Xây dựng các máy tính có tốc độ tính toán nhanh
- Lý thuyết điều khiển
 - Thiết kế các hệ thống nhằm cực đại hóa một hàm mục tiêu nào đó
- Ngôn ngữ học
 - Biểu diễn tri thức
 - Ngữ pháp (của một ngôn ngữ)

Lịch sử tóm tắt của TTNT (1)

- 1943: McCulloch & Pitts trình bày công trình nghiên cứu đầu tiên về AI, đề xuất mô hình các nơ-ron nhân tạo 2 trạng thái (on/off)
- 1950: Khái niệm về TTNT lần đầu tiên được Turing đề cập trong bài báo "Computing Machinery and Intelligence"
- 1956: Workshop đầu tiên (diễn ra trong 2 tháng) ở Dartmouth (Mỹ) bàn về lĩnh vực TTNT, khái niệm TTNT được thừa nhận
- 1952-1969: Các thành tựu ban đầu trong TTNT
- 1950s: Các chương trình TTNT đầu tiên
 - Chương trình chơi cờ của Samuel
 - Chương trình lý luận logic của Newell & Simon
 - Chương trình chứng minh các định lý hình học của Gelernter

Lịch sử tóm tắt của TTNT (2)

- 1965: Robinson đề cử giải thuật hoàn chỉnh cho việc suy diễn logic
- 1966-1973:
 - Các nhà nghiên cứu về TTNT nhận ra khó khăn về độ phức tạp tính toán
 - Gần như là không còn các nghiên cứu về các mạng nơ-ron nhân tạo
- 1969-1979: Sự hình thành và phát triển ban đầu của các hệ thống dựa trên tri thức
- 1980: TTNT trở thành một ngành công nghiệp (các hệ thống, chương trình TTNT dùng trong thương mại)
- 1980-1988: Sự xuất hiện bùng nổ của các hệ chuyên gia
- 1986: Các mạng nơ-ron nhân tạo xuất hiện trở lại, trở nên phổ biến
- 1987: TTNT trở thành một lĩnh vực khoa học
- 1995: Sự xuất hiện của các tác tử thông minh

Các thành tựu quan trọng trong TTNT (1)

- Lập kế hoạch và lập lịch tự động
 - NASA đã thiết kế được chương trình lập kế hoạch tự động (gọi là Remote Agent) để điều khiển việc xếp lịch các hoạt động của tàu vũ trụ
- Chơi cờ
 - Deep Blue (hệ thống máy tính của IBM) đã đánh bại kiện tướng cờ vua Thế giới Garry Kasparov vào năm 1997
- Điều khiển tự động
 - Một xe tải nhỏ được điều khiển tự động bởi hệ thống ALVINN (của CMU) trong suốt 98% của khoảng thời gian đi từ Pittsburgh đến San Diego (~2850 miles)
- Người máy
 - Ngày nay, rất nhiều cuộc phẫu thuật trong y tế sử dụng các trợ giúp người máy trong các thao tác vi phẫu (microsurgery)

Các thành tựu quan trọng trong TTNT (2)

■ Chuẩn đoán

- Các chương trình chuẩn đoán y tế dựa trên phân tích xác suất đã có thể thực hiện ở mức tương đương các bác sỹ chuyên môn trong một số lĩnh vực của y tế

■ Lập kế hoạch hậu cần cho quân đội

- Trong cuộc chiến tranh Vùng Vịnh năm 1991, các lực lượng của quân đội Mỹ đã triển khai sử dụng một chương trình lập kế hoạch và xếp lịch cho công tác hậu cần để di chuyển 50.000 xe cộ, hàng hóa, và quân lính

■ Hiểu ngôn ngữ và giải quyết vấn đề

- Chương trình máy tính PROVERB có thể giải được các bài toán đồ chữ (crossword puzzles) tốt hơn khả năng của nhiều người

Các tranh luận về TTNT (1)

- Khả năng của TTNT?
 - ❑ Chơi (hợp lệ) một ván bóng bàn?
 - ❑ Lái xe tự động theo một đường núi quanh co?
 - ❑ Mua trực tuyến các hàng hóa trong 1 tuần cho một hiệu tạp phẩm?
 - ❑ Phát hiện và chứng minh một lý thuyết toán học mới?
 - ❑ Hội thoại được với một người trong 1 giờ đồng hồ?
 - ❑ Thực hiện tự động một ca phẫu thuật mổ phức tạp?
 - ❑ Dịch trực tiếp (tức thời) giữa hai thứ tiếng cho một hội thoại?
 - ❑ ...
- Máy tính có thể suy nghĩ (như con người) được không?

Các tranh luận về TTNT (2)

- Nếu máy tính có thể làm thay những việc đang được làm bởi con người, thì con người sẽ càng ít việc (thất nghiệp)
- Con người sẽ có quá nhiều thời gian rảnh rỗi (so với quá ít, như hiện nay)
- Con người cảm thấy mất cảm giác sự thông minh thống trị (cao nhất) của họ
- Vì máy tính làm thay (và can thiệp) vào nhiều việc hàng ngày của con người, họ sẽ cảm thấy các quyền riêng tư bị xâm phạm
- Việc sử dụng nhiều hệ thống TTNT có thể làm giảm (mất đi) trách nhiệm giải trình trong các công việc
- Sự thành công (hoàn hảo) của TTNT có ý nghĩa như là sự kết thúc của loài người?

Tài liệu tham khảo

- R. E. Bellman. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company, San Francisco, 1978.
- E. Charniak and D. McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, Massachusetts, 1985.
- J. Haugeland. *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, Massachusetts, 1985.
- R. Kurzweil. *The Age of Intelligent Machines*. MIT Press, Cambridge, Massachusetts, 1990.
- N. J. Nilsson. *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann, San Mateo, California, 1998.
- D. Poole, A. K. Mackworth, and R. Goebel. *Computational Intelligence: A Logical Approach*. Oxford University Press, Oxford, UK, 1998.
- E. Rich and K. Knight. *Artificial Intelligence* (Second Edition). McGraw-Hill, New York, 1991.
- P. H. Winston. *Artificial Intelligence* (Third Edition). Addison-Wesley, Reading, Massachusetts, 1992.