

Bruno Carvalho Castro Souza

Faculdade Anhanguera de Valparaíso

bruno.souza@unianhanguera.edu.br

DOMÍNIOS, COGNIÇÃO E INTELIGÊNCIA ARTIFICIAL

RESUMO

Os sistemas de pensamento humano baseiam-se na representação da realidade e na ação do sujeito cognoscente sobre o mundo. Este artigo levanta a possibilidade de transpor esses tipos de sistemas de pensamento para ambientes computacionais de inteligência artificial, utilizando os construtos teóricos de categorias e de redes hipertextuais de significados. Para isso, propõe uma arquitetura de pensamento computacional implementada em máquinas de estado finito distribuídas em camadas, com dependência do contexto e respostas orientadas a interagir com o meio-ambiente. Sistemas dessa natureza estariam aptos a apresentar, ao longo de um processo evolutivo, comportamento inteligente, intencional e, segundo alguns autores, consciente.

Palavras-Chave: cognição; inteligência artificial; redes hipertextuais; categorias e protótipos; modelos cognitivos.

ABSTRACT

Human thought systems are based on the representation of reality and on the action of the subject upon the world. This paper raises the possibility of transposing this sort of thought systems to artificial intelligence computational environments, using theoretical constructs of categories and semantic hypertext networks. In order to do that, it proposes computational thought architecture based on finite state machines distributed in layers, context-dependent and with environment-oriented answers. Systems of this nature should be capable of showing, after a learning process, intelligent behavior, intentionality and even, as some authors postulate, self-awareness.

Keywords: cognition; artificial intelligence; semantic hypertext networks; categories and prototypes; cognitive models.

Anhanguera Educacional

Correspondência/Contato

Alameda Maria Tereza, 2000

Valinhos, São Paulo

CEP 13.278-181

rc.ipade@unianhanguera.edu.br

Coordenação

Instituto de Pesquisas Aplicadas e

Desenvolvimento Educacional - IPADE

Artigo Original

Recebido em: 15/10/2009

Avaliado em: 14/12/2010

Publicação: 21 de dezembro de 2010

1. INTRODUÇÃO

Uma das grandes questões em inteligência artificial (IA) aborda os limites: um computador pode pensar como um humano? O cérebro humano é somente uma máquina? Essas perguntas vêm se tornando cada vez mais importantes desde que começaram os avanços nas pesquisas sobre IA. A chave para respondê-las é compreender os aspectos fundamentais das abordagens cognitivas humanas e as arquiteturas de sistemas computacionais.

Este artigo apresenta ideias para essas respostas por meio da discussão sobre os fundamentos teóricos das principais abordagens cognitivas – representacionismo, nova robótica e escola chilena –; da transposição dos conceitos de redes hipertextuais de significados, de categorias e de domínios a sistemas de IA; e da estruturação de um sistema de IA baseado nesses pressupostos, que pode evoluir para a capacidade de simulação do comportamento inteligente e da intencionalidade, podendo chegar, segundo alguns autores, à autoconsciência.

Também são abordados questionamentos para o futuro, incluindo a possibilidade de implementar um sistema computacional baseado na arquitetura sugerida, onde ficariam mais evidentes as semelhanças e as diferenças entre o funcionamento do cérebro humano e a IA.

2. AS ABORDAGENS COGNITIVAS: REPRESENTACIONISMO, NOVA ROBÓTICA E ESCOLA CHILENA

Tradicionalmente, a ciência cognitiva baseia-se no representacionismo – ou seja, no pressuposto de que a cognição humana é possível por meio da representação mental (DAMÁSIO, 2000) de objetos concretos, reais. Trata-se de uma abordagem *top-down*: partindo-se de atividades cognitivas superiores, como pensamento simbólico-abstrato e linguagem, entre outras, viabilizam-se as atividades mais básicas, como o senso comum (TEIXEIRA, 1998). O simbolismo e a representação mental de fato demonstram características fundamentais para o raciocínio abstrato, e deles deriva a capacidade de construir “mundos possíveis”, conforme proposto por Leibniz, citado por Ferrater-Mora (2001), tornando viável uma das principais características do pensamento humano – a antecipação do futuro – e o sistema mental de planos e produções. Por meio de representações do mundo podem-se criar modelos e simulações de eventos, dando origem a realidades alternativas que serão avaliadas conforme sua adequação a cada situação. Teoricamente, isso significa que o mundo como cada um o conhece somente

existiria a partir da perspectiva individual, internalizado em processos mentais que representam os estados desse mundo. Baseando-se nessa abordagem, as pessoas viveriam em função de uma realidade virtual-mental, sendo agentes onipotentes, oniscientes e onipresentes para todos os efeitos, uma vez que as representações são construtos mentais conscientes e deliberadamente produzidos. A teoria clássica da representação parte da estranheza do mundo em relação à mente individual que o concebe, o que leva ao entendimento de “mundo” e “mente” como duas entidades separadas e distintas. Nesses termos, a representação seria uma forma de recuperar algum aspecto do mundo externo, do qual a mente não faz parte. E mais: segundo a visão clássica, a representação possuiria propriedades especiais que a distinguiria dos objetos representados, nem poderia ser um objeto entre outros. Portanto, a representação seria mais do que uma relação física ou uma relação entre coisas do mundo (TEIXEIRA, 1998).

O representacionismo possui uma grande funcionalidade quando analisado sob a perspectiva de um indivíduo adulto, com um bom acervo de representações mentais. A apreensão de um mundo fica mais fácil se a maior parte de seus elementos já está devidamente representada, viabilizando a construção de modelos que o descrevam. Dessa forma, é possível antecipar o que acontecerá quando se executa alguma ação concreta, como levantar um garfo para comer, abrir uma porta, dirigir um automóvel, virar uma maçaneta, riscar um fósforo, acender a luz e praticamente quaisquer outras ações imagináveis. Já há descrições detalhadas para todas elas. A capacidade de antecipação permite que o indivíduo crie um modelo mental (ou mundo) que no qual será possível simular todas as ações para viabilizar a representação de, por exemplo, abrir a porta: é possível estimar a distância entre a mão e a maçaneta, detalhar os movimentos para alcançá-la, calcular a força necessária para girá-la, a pressão para segurar a maçaneta enquanto a porta é aberta, e até mesmo imaginar do que estará atrás da porta, invisível aos nossos olhos físicos.

No entanto, há questões que não podem ser respondidas pela teoria representacional. As principais dificuldades vêm dos estudos da nova robótica, desenvolvidos nos laboratórios do MIT por Brooks (1989) e pela escola cognitiva chilena, representada principalmente pelos trabalhos de Maturana, Varela Garcia e Acunã Llorens (1997). Para a nova robótica, a cognição retoma alguns conceitos formulados inicialmente por Minsky (1988), introduzindo-os às novas tecnologias do século XXI e revestindo-os com um enfoque voltado a IA. Minsky propõe que o comportamento inteligente ocorre, na verdade, de forma inconsciente ao indivíduo, executado por *demons*, ou homúnculos, que seriam independentes e superespecializados. Cada *daemon* ignora a existência de outros *demons*, pois está focado em executar apenas uma única tarefa simplificada ao

extremo. Por meio de mecanismos de inibição e excitação, os *demons* trabalhariam de forma sinérgica, construindo o que Minsky chamou de “sociedade mental”. É essa sociedade que viria a viabilizar o pensamento humano. Em outras palavras, a nova robótica trabalha com a ideia de “camadas” de comportamento, independentes e autômatas, capazes de tomar suas próprias decisões de acordo com as informações que recebe do ambiente e fornecer um resultado de saída. Cada camada funciona como uma máquina de estado finito ampliada (Figura 1), ignorante de outras máquinas e com operação limitada à sua tarefa específica. Um grupo dessas máquinas de estado finito formaria um comportamento, como “pegar alguma coisa”: uma camada tem a tarefa de verificar a distância, outra é responsável por alcançar a coisa, uma terceira por movimentar os membros para pegá-la, outra controla a pressão para segurá-la e assim por diante. Cada camada seria ativada ou inibida por uma circunstância (entrada) do ambiente externo ou por uma saída (resultado) de outra máquina, captada por meio de sensores de entrada, e controlada por um temporizador sensível ao contexto. Máquinas de estado finito formam a base da arquitetura de subsunção, idéia fundamental para a proposta de Brooks (1991) e que une a percepção de um mundo externo à capacidade de agir sobre esse mundo, sem a necessidade de representações prévias.

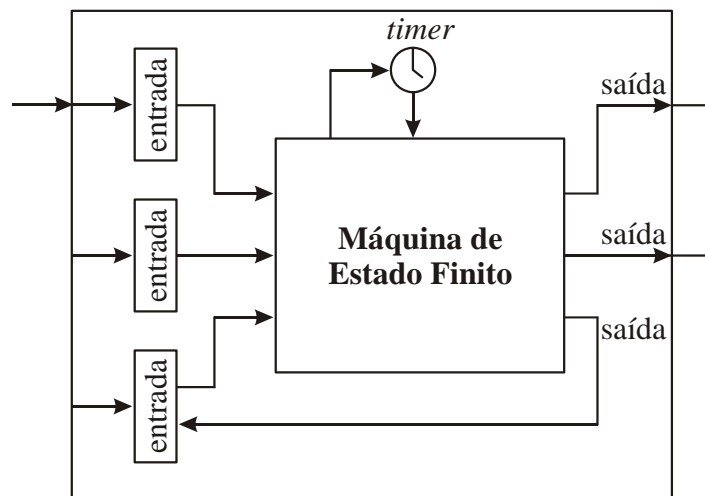


Figura 1 – Esquema da máquina de estado finito.

Fonte: Adaptado de Teixeira (1998).

Ou seja, diferentemente do representacionismo, a nova robótica encara a cognição de maneira *bottom-up*: “a simulação do comportamento inteligente deve ter como ponto de partida os comportamentos simples, mundanos, que não requerem a existência prévia de representações” (TEIXEIRA, 1998, p. 134).

Já a escola chilena reforça a crítica da nova robótica à teoria representacional e ainda acrescenta outro aspecto: a limitação imposta à abordagem cognitiva tradicional pelos paradigmas herdados da filosofia clássica, que postulavam a imaterialidade mental

para dar conta da distinção entre “mente” e “mundo”, criando uma inescrutabilidade dos processos mentais. Essa limitação originou a expressão *ghost in the machine* como sinônimo de um substrato invisível e intangível, que seria a base dos processos mentais e, conseqüentemente, do próprio conceito de significado, que é a fundação semântica de todas as representações.

Esse problema torna-se fundamental ao se examinar a intencionalidade sob a perspectiva cognitiva e consciente, conforme demonstrado por Searle (1997), citado por Teixeira (1998), quando este argumentou sobre o quarto do chinês. A questão é que a intencionalidade não pode ser expressa simbolicamente e, portanto, torna-se impossível representá-la cognitivamente. Essa limitação começou a trazer estagnação teórica e tecnológica, especialmente no campo da IA, a partir dos anos 60, devido ao fracasso em construir máquinas de tradução que fossem capazes de compreender o contexto e a intencionalidade autoral, uma vez que essas máquinas baseavam-se em conceitos representacionais, ignorando a semântica contextual. Esse problema foi muito bem expresso por Teixeira (1998, p.146),

Em suma, a idéia de conhecimento como representação parece estar na raiz das dificuldades tecnológicas aparentes envolvidas na construção desses sistemas: explosão combinatorial, comportamento rígido e assim por diante.

Já a escola chilena, representada por Maturana, Varela Garcia e Acunã Llorens (1997), citados por Teixeira (1998, p. 147), definem a cognição como “ação efetiva: história do acoplamento estrutural que faz emergir um mundo (...) através de uma rede de elementos interconectados capazes de mudanças estruturais ao longo de uma história ininterrupta”. Tal definição pressupõe a inteligência como uma interação entre sujeito e ambiente ao longo de um período de tempo contínuo. Nesse aspecto, a cognição de Maturana e Varela aproxima-se à de Piaget (MONTANGERO; NAVILLE, 1999), que explica que o aprendizado se dá como consequência de processos de assimilação de um ambiente externo e da adaptação a esse ambiente por intermédio de um mecanismo de equilíbrio contínuo e ininterrupto. Nessa perspectiva, cognição envolveria agir sobre o mundo e produzir mudanças estruturais que serviriam como base para a construção dos modelos mentais referentes para futuras ações em contextos semelhantes.

Tanto as perspectivas da nova robótica quanto as da escola chilena levantam questões pertinentes à teoria cognitiva representacional, principalmente sob os aspectos da intencionalidade e da necessidade da existência de representações prévias para a cognição. Em uma análise mais funcional, entretanto, as representações ainda possuem méritos, pois oferecem um ferramental válido e viável para a solução de problemas concretos, desde que esses problemas possam ser devidamente simbolizados e

contextualizados. É necessário se considerar as bases do cálculo matemático para as soluções desses problemas, principalmente quanto à precisão dos resultados: a arquitetura representacional pode, a partir de um modelo construído, prever estatisticamente os resultados matemáticos prováveis. Esses cálculos, no entanto, não têm como levar em conta os inúmeros fatores reais presentes no mundo no momento da aplicação da solução, permitindo apenas uma aproximação.

[...] o sistema real vê-se submetido a uma multidão de perturbações que vão desde a respiração do experimentador presente na sala até o movimento de um núcleo de hidrogênio em Alfa do Centauro. Ora, essas perturbações tornam-se rapidamente expressivas, a um ponto tal que o resultado do cálculo, embora exato, ainda assim estará muito distante do resultado observado. [...] Os engenheiros da NASA sabem perfeitamente que os computadores à sua disposição, embora extremamente potentes, não podem calcular o trajeto exato das naves espaciais a partir das condições astronômicas e astronáuticas iniciais (LÉVY, 1998, p. 105).

Isso significa que as representações podem ser usadas para testar modelos e simular resultados, mas, na vida real, sua aplicação limita-se à precisão do modelo usado para descrever o mundo. Uma vez que modelos são construídos com base na simplificação da realidade, conclui-se que é extremamente complicada a construção de um modelo que consiga abranger todos os aspectos a serem simulados, como coloca Lévy de forma contundente:

No caso de uma simulação, todas as decisões sobre o que poderia ser pertinente para a evolução do sistema estudado foram tomadas de uma vez só, no momento da formalização do modelo. A experiência real, por sua vez, sempre pode deixar aparecer a importância de um fator no qual não se pensara no momento de sua elaboração (LÉVY, 1998, p. 106).

Assim, o mesmo argumento que argui a derrocada do representacionismo pode ser desenvolvido como um de seus sustentáculos: uma vez que é inviável construir um modelo perfeito – ou uma representação perfeita – da realidade, adota-se a melhor descrição possível do real como ponto de partida, com o objetivo de interagir com o mundo. As novas técnicas de IA de fato utilizam justamente esse princípio para desenvolver sistemas inteligentes: o raciocínio baseado em casos permite que o sistema “aprenda” por meio da memória (ou, em termos cognitivos, da representação) dos casos passados, incrementando o modelo através da ação executada em cada novo episódio, ao longo de uma história contínua. Já os algoritmos genéticos usam uma ideia semelhante: “reprogramam-se” conforme as pressões do ambiente e da memória genética, acrescentando-se fatores aleatórios (que simulam as mutações genéticas da biologia) a cada nova geração ao longo de seu processo evolutivo.

3. REDES, CATEGORIAS E DOMÍNIOS: AS METÁFORAS DO PENSAMENTO

O comportamento inteligente pode existir sem qualquer representação mental prévia, conforme demonstrado pelos estudos da nova robótica. Entretanto, os mesmos estudos também deixam claro que representações podem ser construídas a partir da ação do sujeito cognoscente sobre o meio, por meio da formação de uma história contínua. Essa história viabiliza-se a partir da fisiologia da memória, seja essa humana ou simulada, como uma teia para arquivar e recuperar ações executadas e suas consequências para futuro uso em situações semelhantes. No caso da memória humana, essa organização é episódica, formando uma estrutura narrativa que cria uma rede de conhecimentos intimamente relacionados, que por sua vez se torna relevante para o cérebro devido ao seu aspecto emocional (DAMÁSIO, 1996).

A construção contínua dessa rede de conhecimentos pode ser concebida conforme as sugestões de Lévy (1993), por meio de uma rede hipertextual de significados. As relações entre os nós dessa rede formam a sua base funcional: a ativação de um nó estimula a ativação de nós interligados, que por sua vez estimulam outros nós e assim sucessivamente, até que todos os nós relacionados àquele conhecimento estejam ativados. Isso acontece de forma automática, gerando uma renegociação das relações entre os nós para destruir ou reforçar conceitos e significados prévios em função de novas experiências. Uma das principais vantagens em se conceber o pensamento como uma rede hipertextual é a facilidade para acomodar as diversas formas de perceber o mundo: imagens, palavras, sons, roteiros, sentimentos – todo tipo de coisa pode se tornar um nó. As conexões entre os nós, por sua vez, são construídas por intermédio da observação de relações de causa-efeito na realidade objetiva. Quando o sujeito cognoscente age sobre a realidade e provoca uma mudança, essa mudança é percebida pelo seu sistema cognitivo, que gera uma nova ponte entre nós e cria uma regra lógica: se *isto*, então *aquilo*. A validade dessa regra será testada inúmeras vezes ao longo da história do indivíduo, levando-o a reforçá-la ou a descartá-la como inválida.

No entanto, a criação efetiva de redes hipertextuais está condicionada à sua relevância para o sujeito. A sua consolidação na memória será efetivada apenas se o fator emocional estiver presente – ou seja, se a *história* em que se criou tal situação tiver *significado*, seja através da analogia com experiências prévias, por meio da identificação de uma instância futura ou devido a uma reação emocional aprendida. Quanto mais fortes estes fatores, mais chances terá a rede criada de se instalar na memória. E quanto mais se recorrer a ela, por intermédio do acesso a seus nós, mais ativa será a sua participação nos processos cognitivos como um todo.

As redes hipertextuais fornecem um bom pano de fundo para a compreensão do pensamento humano como algo corrente, abrangente, flexível e variado. Isso é possível devido à fisiologia humana, que dispõe de um sistema nervoso central e altamente plástico, com capacidade de conectar simultaneamente diversas redes de neurônios por meio de reações eletroquímicas com o objetivo de provocar comportamentos específicos. Em um sistema de IA, a atual arquitetura de *hardware* não permite a mesma plasticidade do cérebro humano. Essa plasticidade precisa ser simulada por meio de *softwares* e pela interconexão de diversos *hardwares* externos. Apesar de conseguir simular redes hipertextuais com relativo sucesso, essas técnicas deixam muito a desejar devido às limitações tecnológicas e conceituais embutidas em seu próprio desenvolvimento. Questões teóricas não respondidas, como os aspectos relativos à intencionalidade da rede, limitam suas aplicações práticas, enquanto a pouca capacidade de processamento paralelo e a arquitetura digital, diferentemente da capacidade analógica do cérebro humano, tornam quase impossível o relacionamento inter e intra-redes, inibindo o comportamento criativo do sistema.

Porém, quando a intenção é a simulação da própria memória humana, como estruturação de um sistema de armazenamento e recuperação de informações históricas através do relacionamento de causa-efeito, a topologia das redes hipertextuais apresenta grande funcionalidade em sistemas de IA. Aplicações como *datawarehouses* e *datamining* são beneficiadas com as possibilidades de acesso a informações de diversas naturezas (palavras, sons, imagens etc.) e da pesquisa por relações entre essas informações ao invés da informação em si. Nesse ponto, há enorme proximidade do sistema digital com a inteligência humana. O ser humano, ao buscar lembrar-se de uma situação específica, busca suas referências por meio de relações entre eventos, pessoas, imagens, sons e outras situações, ao invés de procurar pela situação em si. O sistema de indexação mental humano é muito mais relacional do que diretivo.

Os conceitos de protótipos e categorias introduzidos por Berlin e Kay em 1969 e desenvolvidos por Rosch em 1978, conforme apresentado por Ungerer e Schmid (1996) e Sperber e Wilson (1996), constituem outra metáfora para a organização do pensamento humano. Os princípios fundamentais da teoria dos protótipos e categorias cognitivas baseiam-se no conceito que a percepção do mundo acontece por meio da analogia entre objetos desse mundo, numa busca de semelhanças entre eles. Quanto maior o grau de coincidência de aspectos percebidos em objetos distintos, mais próximos os mesmos estarão sob o ponto de vista cognitivo. Um grupo de objetos com características semelhantes forma uma categoria. O grau de refinamento descritivo das categorias varia conforme a experiência que o indivíduo possui com as entidades das categorias. Rosch

introduz, com base nessa observação, a noção de fronteiras difusas entre as categorias: até que ponto uma entidade pertence à categoria A e não à categoria B é uma questão de experiência individual, acontecendo continuamente em um ambiente cultural que alimenta expectativas e demanda respostas adequadas do sujeito (UNGERER; SCHMID, 1996). As fronteiras difusas explicam a sobreposição de categorias: elementos que não são claramente identificados com pertencentes à categoria A ou à B podem ser relacionados a ambas as categorias.

É interessante observar que não são as entidades físicas que se misturam umas às outras, mas as categorias das entidades, que são produtos de classificação cognitiva. Ou seja, não são os limites das entidades que são vagas, mas os das categorias destes processos cognitivos. Em outros termos, as margens difusas são áreas de atrito entre categorias adjacentes, muitas das quais são dependentes do contexto.

Uma vez que as categorias possuem membros diferentes que podem representar melhores ou piores exemplos destas categorias, torna-se imperativo investigar o que diferencia estes membros. Dessa necessidade surge a noção de atributo, definido como um conjunto de características comuns aos membros da categoria. Vale ressaltar que estes atributos podem não existir da mesma forma em todos os membros, mas apenas em alguns exemplos. Os atributos que definem um pássaro, por exemplo, podem ser: “coloca ovos”, “possui um bico”, “tem duas asas e duas pernas”, “tem penas”, “pode voar”, “é pequeno e leve”, “pia/canta”, “tem pernas pequenas e finas”, “possui uma calda curta”. Um avestruz não é pequeno, não voa e não tem pernas pequenas e finas. Um pinguim possui ainda menos atributos do que um pássaro, mas ainda assim pode-se identificá-lo como tal. O que se observa é que os atributos são uma rede de características sobrepostas, funcionando por proximidade de exemplos. Essa é a idéia de semelhança de família, definida na forma de uma série de itens da natureza de AB, BC, CD, DE e assim por diante. Isto é, cada membro possui pelo menos um elemento em comum com um ou mais membros, mas nenhum ou muito poucos elementos são comuns a todos os membros.

A estrutura de atributos de uma categoria baseada em protótipos pode ser resumida da seguinte maneira:

- a) Membros prototípicos de categorias cognitivas (bons exemplos) possuem o maior número comum de atributos com outros membros da categoria e o menor número de atributos compartilhados com membros de categorias vizinhas - ou seja, protótipos de categorias apresentam a maior distinção possível de outros protótipos;
- b) Membros marginais de categorias (maus exemplos) compartilham apenas um pequeno número de atributos com outros membros da sua categoria, mas possuem vários atributos em comum com outras categorias - ou seja,

os limites das categorias realmente são difusos.

Pode-se sintetizar os principais conceitos sobre categorias cognitivas da seguinte forma:

- a) Categorias não representam divisões arbitrárias do mundo, mas são baseadas na capacidade cognitiva da mente humana;
- b) Categorias cognitivas de cores, formas, organismos e objetos concretos são baseadas em protótipos, que têm um importante papel na formação das próprias categorias;
- c) Os limites das categorias são difusos;
- d) Variando entre os protótipos e os limites, categorias cognitivas possuem membros que podem ser avaliados conforme sejam “bons” ou “maus” exemplos da categoria.

A teoria dos protótipos dá conta de outra maneira de representar o mundo: por intermédio da analogia direta com a realidade. Os sistemas computacionais ainda não desenvolveram estruturas perceptivas que os possibilitem apreenderem a realidade como um conjunto de entidades independentes que possuem características comuns. As características compartilhadas pelas entidades do mundo podem ser físicas ou contextuais. Um membro de uma categoria pode se aproximar de outro tanto por meio de um comportamento semelhante quanto através de suas características físicas. As tecnologias de identificação de formas, por exemplo, apesar de grandes avanços nos últimos anos, ainda apresenta dificuldades em perceber até mesmo os padrões mais simples, facilmente identificáveis por crianças humanas. Quando colocadas em um ambiente social, com a infinidade de relacionamentos entre as pessoas, e levando-se em conta os diversos aspectos sociais, políticos, econômicos e culturais – dentre outros – os sistemas digitais carecem da capacidade para categorizar estas entidades de forma adequada, faltando-lhes as estruturas perceptivas¹ necessárias para apreender as sutilezas dessas relações – isso sem considerar experiência necessária para corretamente interpretar as atitudes e outras ações humanas, uma vez que se pressupõe que tal interpretação poderia vir a ser aprendida pelo sistema por meio da observação de relações de causa-efeito, uma vez que existam tecnologias adequadas para capturar esse tipo de relação.

Ainda assim, a transposição do conceito de categorias e protótipos para o formato digital pode ser útil, caso a mesma seja construída artificialmente *a priori* da utilização do sistema. Seria o caso da modelagem de uma arquitetura de categorias, com a definição prévia dos seus atributos e da entrada classificada de entidades em cada categoria definida por meio de programação. Criar-se-ia, assim, um modelo

¹ Entendem-se como estruturas perceptivas as capacidades advindas, no caso humano, dos sentidos que permitem a interação com o mundo: visão, audição, olfato, gustação e sômato-sensitividade, acrescidas da capacidade da linguagem.

computacional para um grupo de categorias, bem como de ponteiros para o reconhecimento dos atributos das entidades que as compõem e das entidades que povoam esse mundo. Tal modelo poderia simular, por exemplo, a categoria de “máquinas que andam sobre rodas”, e nelas estariam contidas as entidades “carro”, “caminhão”, “locomotiva” etc. Também deveriam estar programados os atributos necessários para que as entidades possam ser caracterizadas como membros - “ter rodas”, “possuir motor”, “andar”, “parar” e assim por diante.

A funcionalidade de um sistema computacional baseado em categorias reside em sua possível aplicação na estruturação simulada de domínios mentais. Domínios são organizados em torno de classes de conhecimentos, podendo ser entendidos como categorias mentais de grande abrangência. Quando é preciso recuperar alguma informação técnica, por exemplo, a mente humana recorre a um domínio específico em dessa informação. Quando se depara com um problema matemático, por sua vez, geralmente a resposta é procurada em domínios associados a fórmulas, equações, conjuntos, matrizes, números e assim por diante, e não em representações mentais vinculadas a arte, diversão, história etc. O uso de categorias como metáfora para a estruturação de um sistema de inteligência artificial pode, portanto, servir como base para a simulação artificial de domínios mentais, aproximando-a a funcionalidade de uma representação do mundo mais próxima aos métodos descritivos humanos e, portanto, em uma linguagem mais natural.

4. DOMÍNIOS COGNITIVOS EM SISTEMAS DIGITAIS

A aplicação simultânea das teorias de redes hipertextuais de significados e de categorias pode gerar as bases necessárias para, com as tecnologias atualmente disponíveis, construir um sistema de inteligência artificial que apresente comportamento criativo e que, segundo Chalmers (1996), poderá eventualmente vir a desenvolver autoconsciência, intencionalidade e reações bem próximas às humanas (DENNETT, 1996). Convém ressaltar, no entanto, que a natureza exata dessas intenções e reações provavelmente permanecerá desconhecida, uma vez que a base fisiológica de funcionamento dos sistemas humano e digital é fundamentalmente distinta - seria como comparar o sistema emotivo de seres humanos e de macacos: ambos sentem emoções, porém a natureza das emoções humanas é bem diferente da dos macacos.

A construção desse sistema poderia acontecer em camadas, como sugere a nova robótica, partindo-se de um “baixo nível” para tratamento direto com o ambiente (“camada perceptiva”), que teria como entradas as informações sobre o mundo coletadas

por sensores específicos. A segunda camada seria responsável por identificar as informações, enquadrando-as em categorias pré-programadas ou, caso estas não existam, armazenando-as em uma área de trabalho temporária para processamento posterior – essa camada poderia ser identificada como “camada de identificação”. A próxima camada teria a tarefa de buscar relações entre as informações recém-categorizadas e as entidades já existentes no sistema, ativando uma ou mais redes hipertextuais de significados: seria a “camada relacional”. As relações encontradas poderiam ser então, qualificadas conforme sua força ou algum outro sistema que permita hierarquizá-las para que possam ser atribuídos valores e selecionadas novas categorias para uso no tratamento da situação. Dessa forma, a qualificação originaria a “camada qualificatória”. Uma vez identificadas as melhores relações (mais fortes ou relevantes), entraria em funcionamento uma nova ‘camada relacional’, desta vez em sentido oposto – a entrada seriam as relações, e a saída, as melhores categorias que atendem às relações estabelecidas. A partir daí, entraria em funcionamento uma “camada adequada”, com a função de selecionar as categorias pertinentes à situação encontrada inicialmente e encaminhando-as novamente à camada relacional, para identificação das novas relações entre as categorias selecionadas. Por fim, as relações seriam repassadas a uma “camada interativa”, que as transformaria em ação sobre o mundo, inclusive na forma de linguagem, caso esta seja uma das relações selecionadas.

Esse tipo de implementação é análogo ao conceito de linha de montagem, porém aplicada ao sistema de pensamentos, como se todas as operações fossem realizadas em série. Porém, com *hardwares* mais potentes e processamento distribuído, algumas operações podem ocorrer simultaneamente, como as pertinentes à camada perceptiva, que recolhe informações sobre o ambiente ao mesmo tempo em que a camada de identificação analisa os dados recolhidos previamente. Na realidade, como o sistema está inserido em um contexto histórico e contínuo, o funcionamento é ininterrupto em todas as camadas.

Justamente por estar inserido em uma história e não dedicado a um único evento, sempre deverá existir no sistema uma máquina de estado finito ampliada e combinatorial (CHALMERS, 1996), responsável pelo controle de cada operação, pela armazenagem das novas situações encontradas em categorias pré-existentes, bem como pela manutenção das redes hipertextuais do sistema (incluindo reforço ou destruição e criação de novas pontes entre nós). Essa seria a “camada de controle e armazenamento”. Chalmers (1996) propõe a construção de sistemas inteligentes que poderiam tornar-se conscientes por meio da utilização dessas máquinas de estado finito ampliadas e combinatoriais, que são como máquinas de estado finito comuns, porém possuem como

entradas vetores, e não dados simples, sendo também seu funcionamento interno e saídas orientados a vetores de informações.

O tratamento do erro também demanda a construção de um subsistema dedicado. O erro acontece quando o sistema se depara com uma situação totalmente inédita (não programada), sendo incapaz de enquadrá-la nem em categorias e nem em redes pré-definidas para adequadamente negociá-la. Nesse caso, o subsistema seria programado buscar semelhanças e diferenças em outras redes e categorias, e executaria uma sobreposição dos resultados – em suma, utilizaria a experiência formada por uma rede ou categoria inapropriada como insumo para a criação de novos domínios ou para a incorporação da nova situação em categorias e redes pré-existentes, por intermédio da execução de um modelo pré-programado. Em termos computacionais, esse subsistema funcionaria de forma semelhante a um *buffer* de informações para as quais não existe tratamento específico, enquanto o sistema procura descobrir qual a melhor estratégia para tratá-las. No sistema cognitivo humano, faria um papel semelhante ao da nossa memória de trabalho.

A Figura 2 ilustra o funcionamento desse sistema de inteligência decupado em camadas.

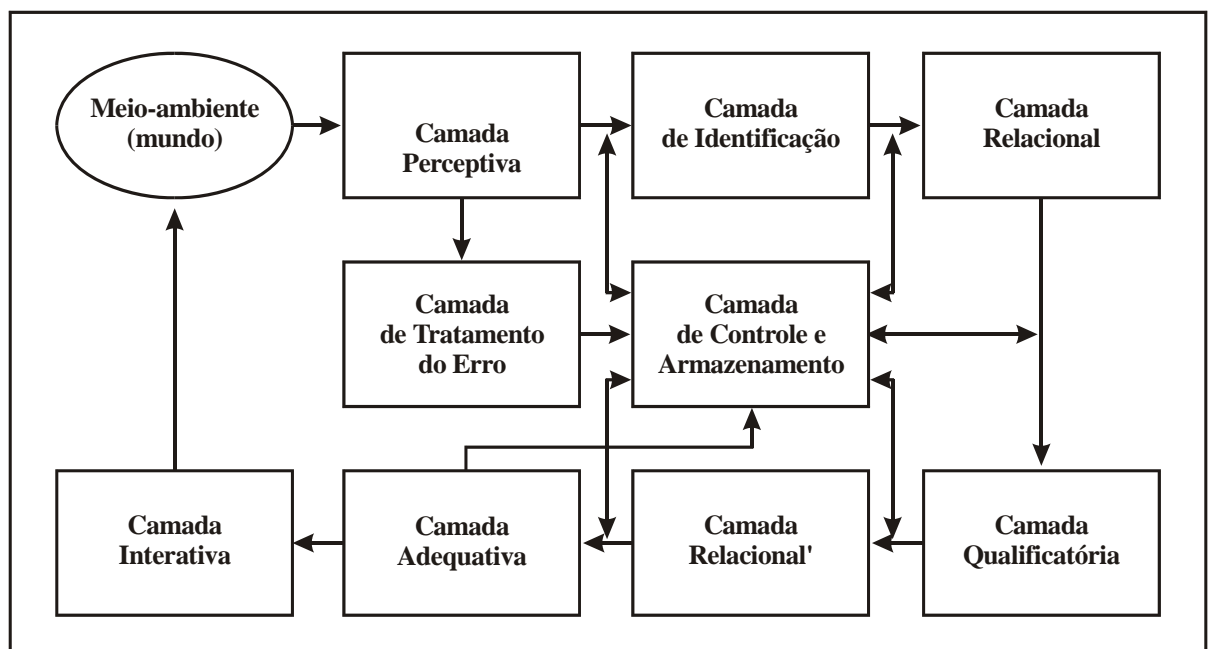


Figura 2 – Implementação de um sistema de inteligência artificial baseado em camadas.

5. CONCLUSÕES

As bases de conhecimentos e as relações entre essas bases são os componentes essenciais para a inteligência e para o comportamento inteligente. Nessa perspectiva, computadores e pessoas compartilham modelos semelhantes, porém sustentados em *hardwares* diferentes (Tabela 1), impondo a obrigatoriedade da aproximação do sistema cognitivo humano através de sua simulação.

Tabela 1 – Paralelo entre o cérebro humano e os sistemas computacionais.

| Processos de Inteligência | Cérebro Humano | Sistemas Computacionais |
|------------------------------------|---|---|
| Organização da informação | - Rede hipertextual (Lévy) - Protótipos e categorias (Rosch, Kay) | - Bases de dados (<i>datawarehouses</i>) |
| Representação do mundo | - Analógica (ícones) - Digital (linguagem) | - Digital |
| Evidência do processo | - Consciência - Varredura em paralelo, com processamento em <i>background</i> - Não-exclusivo | - Não-consciente - Varredura seriada, com processamento em primeiro plano - Exclusivo |
| Abertura a estímulos | - Aberto a estímulos externos durante o processamento | - Fechado a estímulos |
| Dependências | - Depende de conhecimentos e experiências prévias (crenças) | - Depende de bases de dados bem montadas |
| Orientação e tipo de processamento | - Por relevância - Difusa | - Coincidência de informações - Exata |

É possível, portanto, simular a cognição humana em sistemas computacionais por meio da utilização dos construtos teóricos de redes hipertextuais de significados e de categorias cognitivas, aplicadas em uma implementação de máquinas de estado finito distribuídas em camadas, tendo-se como resultado um sistema com comportamento e estruturas “mentais” bem próximas às que se acreditam existir no cérebro humano (Tabela 2).

Tabela 2 – Simulação dos processos mentais em sistemas computacionais através da arquitetura sugerida

| Processo mental | Construto para a simulação |
|------------------------|---|
| Representação do mundo | Camadas: Perceptiva Identificação Relacional |
| | Estratégias: |

| Processo mental | Construto para a simulação |
|---------------------------|--|
| | Categorização das entradas Identificação de relações (redes) |
| Processamento da situação | Camadas: Qualificatória Relacional Adequativa |
| | Estratégias: Comparação (das relações entre as entidades – redes) Categorização das relações |
| Ação sobre o mundo | Camadas: Interativa |
| | Estratégias: Seleção de categorias e relações |
| Tratamento do erro | Camadas: Tratamento do erro |
| | Estratégias: Categorização do erro Categorização da experiência Relacionamentos decorrentes da experiência (resultados integrados às redes hipertextuais) |
| Controle do processo | Camadas: Controle e armazenamento |
| | Estratégias: Categorização da experiência Manutenção das redes |

A implementação desse tipo de sistema poderia ser realizada em qualquer tipo de *hardware* computacional. A efetivação e os testes necessários, entretanto, estão além do escopo deste artigo, abrindo portas para novas pesquisas e desenvolvimento de processos e técnicas de aplicações para IA. Em especial, as ideias propostas permitem a construção de domínios cognitivos computacionais, por meio da categorização de entidades e de relações hipertextuais de significados, viabilizando o surgimento de *insights* criativos do próprio sistema através do “aprendizado”, que por sua vez iria gerar um consequente aumento da quantidade de categorias cognitivas e das relações entre elas (redes). Em aplicações voltadas à criatividade e à solução de problemas por analogia, parece adequada a utilização desse tipo de estrutura.

REFERÊNCIAS

- BROOKS, R.A. How to build complete creatures rather than isolated cognitive simulators. **Architectures for Intelligence**, K. VanLehn (ed.), Erlbaum, Hillsdale, NJ, Fall 1989, p. 225-239. Disponível em: <<http://www.ai.mit.edu/people/brooks/papers/how-to-build.pdf>>. Acesso em 28 out. 2000.
- _____. Intelligence without representation. **Artificial Intelligence Journal**, v. 47, p. 139-159, 1991. Disponível em: <<http://www.ai.mit.edu/people/brooks/papers/representation.pdf>>. Acesso em: 28 out. 2000.
- CHALMERS, David J. **The conscious mind: in search of a fundamental theory**. New York: Oxford University Press, 1996.
- DAMÁSIO, António R. **O erro de Descartes: emoção, razão e o cérebro humano**. São Paulo: Cia das Letras, 1996.
- _____. **O mistério da consciência: do corpo e das emoções ao conhecimento de si**. São Paulo: Cia das Letras, 2000.
- DENNETT, Daniel C. **Kinds of minds: towards an understanding of consciousness**. New York: Basic Books, 1996.
- FERRATER-MORA, José. **Dicionário de filosofia**. Volume 3. São Paulo: Loyola, 2001.
- LÉVY, Pierre. **A máquina universo: criação, cognição e cultura informática**. Porto Alegre: ArtMed, 1998.
- _____. **As tecnologias da inteligência**. Rio de Janeiro: Ed. 34, 1993.
- MATURANA R., Humberto; VARELA GARCIA, Francisco J; ACUÑA LLORENS, Juan. **De máquinas e seres vivos: autopoiese – a organização do vivo**. 3.ed. Porto Alegre: Artes Médicas, 1997.
- MINSKY, Marvin L. **The society of mind**. New York: Simon & Schuster, 1988.
- MONTANGERO, Jacques; NAVILLE, D. Maurice. **Piaget ou a inteligência em evolução**. Artes médicas: Porto Alegre, 1999.
- SEARLE, John R. **The mystery of consciousness**. New York: The New York Review of Books, 1997.
- SPERBER, Dan; WILSON, Deirdre. **Relevance, communication & cognition**. 2.ed. Oxford-Cambridge, Blackwell, 1996.
- TEIXEIRA, João de Fernandes. **Mentes e máquinas: uma introdução à ciência cognitiva**. Porto Alegre: Artes Médicas, 1998.
- UNGERER, Friedrich; SCHMID, Hans-Jörg. **An introduction to cognitive linguistics**. Londres, Longman, 1996.