

Por uma robótica institucionalista: um olhar sobre as novas metáforas da inteligência artificial

PORFÍRIO SILVA*

Introdução

Enquanto equipas de humanos jogadores de futebol disputavam em várias cidades portuguesas o Euro 2004, equipas de robôs jogadores de futebol disputaram em Lisboa, entre 27 de Junho e 5 de Julho, o RoboCup 2004 – o Campeonato Mundial de Futebol Robótico (competições e simpósio científico). Trata-se de uma iniciativa internacional, anual, que visa fomentar a investigação e educação na área da inteligência artificial (IA) e da robótica.

A robótica colectiva, na qual se integra o futebol robótico, representa uma profunda renovação da investigação em IA. Essa renovação vai de par com a adopção de novas metáforas inspiradoras, que agora se tornam mais próximas do que nunca das ciências sociais.

Este texto, que se integra numa reflexão crítica sobre as ciências do artificial, começará por dar conta de aspectos relevantes dessas mutações. Entrará, depois, no jogo da configuração do campo das metáforas que se candidatam a esse papel inspirador das novas correntes da IA. É nesse quadro que se sugere o desenvolvimento de uma abordagem institucionalista em robótica.

O que é o futebol robótico

O futebol robótico é uma das iniciativas com maior sucesso na campanha para atingir dois objectivos articulados: desenvolver modalidades de inteligência artificial; fazer aceitar pelo grande público que as máquinas podem ser inteligentes. A tentativa anterior com mais êxito nesse campo foi o xadrez computacional. É curioso que 1997 seja simultaneamente o ano em que

pela primeira vez um computador vence um campeão mundial de xadrez e o ano em que se realiza a primeira edição do RoboCup. Há, contudo, diferenças significativas entre os dois programas de investigação. A actividade de «jogar xadrez», tal como é praticada por computadores, é algo que se passa apenas «dentro da cabeça», enquanto os robôs «andam no mundo» físico, encontram obstáculos reais e outros robôs com os quais devem «colaborar» ou «competir». Desse modo, os desafios suscitados pela robótica (e em particular pela robótica colectiva) são muito mais complexos e mais ricos.

Uma das grandes vantagens do formato do RoboCup é que tem um «problema padrão»: jogar futebol. Desse modo concentram-se recursos e esforços num mesmo tipo de actividade, segundo regras comuns que todos conhecem e avaliam – e a medida do sucesso é simples: os melhores ganham os jogos. Contudo, há muita diversidade no RoboCup: há diferentes modalidades e nem sequer são todas «futebol».

A actividade mais importante do conjunto é o futebol de competição, que

* Bolseiro da Fundação para a Ciência e Tecnologia (SFRH / BD / 10127 / 2002).

inclui, além da Liga de Simulação, quatro ligas com robôs físicos. Na Liga de Pequenos Robôs, equipas de cinco robôs de 18 cm de diâmetro, com rodas, jogam futebol com uma bola de golfe cor de laranja num campo do tamanho de uma mesa de pingue-pongue. Na Liga de Robôs Médios, equipas de quatro robôs com cerca de meio metro de diâmetro jogam com uma bola de futebol cor de laranja num campo de 12x8 metros. Todos os sensores utilizados por estes robôs na sua operação vão a bordo dos mesmos. Os objectos relevantes (equipas e elementos essenciais do ambiente) distinguem-se por cores. Pode haver comunicação entre robôs, mas sem fios. Não é permitida qualquer intervenção dos humanos durante o jogo, a não ser para colocar ou retirar os robôs do campo. Na Liga de Robôs de Quatro Patas, robôs de entretenimento (cães robóticos sem fios da Sony) jogam futebol num campo de 3x5 metros. Na Liga Humanóide, robôs humanóides autónomos e bípedes competem em tarefas específicas que, de futuro, poderão ser integradas em jogos de futebol robótico (caminhar, pontapé na bola, passes entre dois jogadores, marcação de penaltis).

O RoboCup Rescue é uma das modalidades do RoboCup não relacionadas com o futebol. Uma tarefa típica desta actividade consiste num «cenário de salvamento», simulando um edifício parcialmente destruído por um tremor de terra, onde se encontram manequins que representam «vítimas» em diferentes situações: expostas numa superfície, obstruídas por destroços, encarceradas. Os manequins emitem sinais de vida simulados (calor corporal, movimentos, sons, dióxido de carbono) e os robôs devem «compreender» o significado das combinações de «sinais vitais», transmitir informações sobre elas e realizar tarefas de salvamento sem destruir o ambiente.¹ Em termos de investigação robótica, um aspecto interessante desta actividade é que ela se desenvolve em ambientes não estruturados, enquanto em aplicações mais tradicionais o ambiente é fixo ou previsível.

No âmbito do simpósio científico associado ao

RoboCup 2004, Luigia Carlucci Aiello, da Universidade de Roma La Sapienza, proferiu no domingo, 4 de Julho, uma conferência intitulada «Seven Years of RoboCup: Time to Look Ahead». Dois aspectos da sua conferência merecem aqui uma referência.

Olhando para o passado, Aiello fez um levantamento dos temas a que o movimento do RoboCup se tem dedicado. Desde o princípio que alguns tópicos estiveram muito presentes, designadamente: aprendizagem (de movimentos, de comportamentos, como construir modelos dos outros agentes); visão (reconhecimento de objectos por meio de cores); localização e navegação (como é que o agente sabe onde está no seu «mundo» e como é que se dirige para outros pontos de interesse); planeamento (de movimento) e replaneamento (quando ocorrem mudanças que a isso obrigam); cooperação e competição em sistemas multiagentes. Nos últimos anos tornaram-se importantes outros dois tópicos: robôs humanóides e comunicação entre robôs.

Olhando para o futuro, Aiello considerou que, para o RoboCup, ele está nos humanóides, porque é por eles que se poderá alcançar um progresso real na interacção entre robôs e humanos. Para que esse desenvolvimento se concretize será necessário investir nos seguintes pontos: a energia e o controlo computacional necessários terão de estar a bordo do próprio robô; o robô humanóide terá de ser capaz de se movimentar em ambientes humanos, com um «estilo» humano, manipulando os mesmos objectos que os humanos, interagindo com os humanos de forma que seja segura e seja entendida como segura, comunicando de forma intuitiva para os humanos, tendo um aspecto agradável aos humanos; a percepção do robô humanóide deverá assentar na visão. Aiello prevê que dentro de 20 anos teremos humanóides capazes de correr e com autonomia para 24 horas.

Do ponto de vista de uma crítica das ciências do artificial, o interesse do futebol robótico reside, em grande parte, no facto de ele oferecer uma

¹ Para uma apresentação mais detalhada do RoboCup consultar <http://turing-machine.weblog.com.pt/>.

espécie de «teste de Turing de segunda geração». Vejamos porquê.

O velho e o novo «teste de Turing»

Como conceito, o moderno computador digital foi «inventado» por Alan Turing num artigo de 1936 (a «máquina de Turing»). Alguns anos depois (Turing, 1950), o mesmo autor revoluciona a reflexão acerca da relação entre mente e máquina. Querendo considerar a questão «as máquinas pensam?», mas considerando-a, nessa forma, uma questão «demasiado desprovida de sentido para merecer discussão», Turing propõe-se expressá-la noutra forma. Para isso introduz o «jogo da imitação». Sejam três pessoas: A, um homem; B, uma mulher; C, um interrogador humano que permanece numa sala separada de A e B. O objectivo do jogo é: para o interrogador, determinar, com base nas perguntas que dirige a A e a B e nas respostas obtidas, qual é o homem e qual é a mulher; para a mulher, ajudar o interrogador (dizendo a verdade); para o homem, enganar o interrogador, fazendo-o crer ser ele a mulher. Para que o tom de voz de A ou de B não ajude o interrogador, as respostas ser-lhe-ão transmitidas por teletipo.

Agora, a questão «as máquinas pensam?» pode ser substituída pela questão seguinte, relativa a um computador digital C: «É verdade que, modificando esse computador para ter uma capacidade de memória adequada, aumentando satisfatoriamente a sua velocidade de trabalho e fornecendo-lhe um programa apropriado, podemos fazer com que C desempenhe satisfatoriamente o papel de A no jogo da imitação, sendo o papel de B desempenhado por um homem?» De acordo com a distribuição de papéis no jogo, o desempenho satisfatório do computador diz respeito à capacidade para evitar que o interrogador o identifique como tal.

A «aposta» de Turing é então explicitada: por volta do ano 2000 haverá computadores que jogarão tão bem o jogo da imitação que um interrogador humano médio não terá mais do que 70% de hipóteses de fazer uma identificação correcta após

cinco minutos de interrogatório, de tal modo que alguém que fale em máquinas pensantes não correrá o risco de ser contraditado.

Um aspecto central do que está em causa é que a proposta de Turing concebe uma separação radical entre inteligência e corporalidade de um sistema: «O novo problema tem a vantagem de traçar uma linha bastante nítida entre as capacidades físicas e intelectuais de um homem.» É por isso que as condições de organização do jogo têm de evitar que o interrogador veja, toque ou ouça os outros competidores. Se a questão da corporalidade não é essencial na fronteira entre humanos e máquinas, ela também não é essencial na distinção entre máquinas diferentes: apenas interessam os aspectos lógicos do seu funcionamento. Nesta senda, durante décadas o paradigma dominante na investigação em IA desprezou largamente o que não fosse processamento de informação «dentro da cabeça». Embora a IA e a robótica tenham nascido associadas, logo se separaram. Lembremo-nos (no que toca às representações dominantes) que o computador HAL (do filme *2001: Odisseia no Espaço*) se contentava em «dar ordens» e deixar a manipulação para os humanos. Antes da redescoberta da importância do corpo para a inteligência, muitos achavam aceitável que se dissesse que uma máquina podia jogar xadrez, e mesmo vencer um humano nesse jogo, sem ser sequer capaz de reconhecer um tabuleiro de xadrez. O que acontece é que o RoboCup é uma nova forma de teste de Turing, mas muito mais exigente: agora não se pode esconder nada, em particular não se pode esconder o corpo; a inteligência não é apenas raciocínio, mas também manter o corpo controlado. No sentido em que participa de uma renovação do imaginário da IA, o RoboCup é uma aventura intelectual que merece atenção.

Há vários factores que tornam o RoboCup interessante do ponto de vista da investigação em ciências do artificial lidando com múltiplos agentes, para além de encarar de forma mais directa a questão da corporeidade da inteligência. Alguns desses factores são expostos por Stone e Veloso

(2000, 370-373). Desde logo, trata-se de robótica colectiva: não temos um robô isolado, temos equipas. Temos equipas contra equipas: além de cooperação também temos competição. Temos robôs físicos autónomos (sem auxílio humano durante o jogo). Temos uma elevada exigência de coordenação interindividual (para que um passe resulte, dois robôs jogadores têm de estar no tempo e na disposição apropriada). Há uma forte conveniência em que cada robô seja capaz de estabelecer modelos dos comportamentos dos outros robôs (prever o comportamentos dos parceiros, mas também dos adversários). Dentro de uma mesma equipa, diferentes robôs com diferentes percepções do jogo podem ter «opiniões» diferentes acerca da linha de acção a adoptar, pelo que interessa o desenvolvimento de protocolos de negociação. A complexidade das tarefas exige o recurso a uma variedade de tecnologias. O requisito da operação em tempo real constitui um forte constrangimento das soluções. Há, ainda, o «factor senso comum» transportado pela visibilidade do futebol em públicos leigos face à engenharia robótica: fazer o público em geral partilhar as suas metáforas inspiradoras torna-se um desafio para as ciências do artificial.

O «jogo da imitação» (o «teste de Turing») foi uma aposta com um horizonte de 50 anos. No caso do futebol robótico há uma aposta parecida, formulada por Asada e Kitano (1999): «Até meados do século XXI, uma equipa de robôs humanóides autónomos baterá a equipa humana campeã do mundo de futebol, segundo os regulamentos oficiais da FIFA.» O RoboCup tenta ganhar essa aposta.

Cabe, no entanto, perguntar: se Asada e Kitano ganharem a sua aposta, que significado terá esse facto? Vamos regressar à questão do xadrez computacional para tornar mais clara esta pergunta.

O xadrez computacional

O *Deep Blue* foi o primeiro computador a vencer um campeão mundial de xadrez (Kasparov) num encontro em condições tradicionais de torneio.

Concebido especificamente para derrotar um humano num jogo de xadrez, recorria a meios extraordinários (por exemplo, mais de 400 processadores a trabalhar em paralelo), que lhe permitiam analisar em média 200 milhões de posições por segundo (Campbell *et al.*, 2002). Enquanto os xadrezistas humanos de topo analisam, em cada jogada, no máximo cem ou duzentas posições, o *Deep Blue* podia analisar muitos biliões.

Um jogo como o xadrez pode ser descrito por um conjunto de estados iniciais (arranjo das peças no início do jogo), um conjunto de estados finais (arranjos das peças que determinam o fim do jogo e o seu resultado) e um conjunto de regras que definem as jogadas legais. O espaço de estados de um jogo como o xadrez é o conjunto de todos os arranjos de peças que se podem obter por uma sucessão de jogadas legais a partir do estado inicial. A IA entende um jogo do tipo do xadrez como um problema de busca ou exploração do espaço de estados.

Assim sendo, a via que muitos autores consideram privilegiada para o xadrez computacional é a «força bruta», isto é, o recurso intensivo à capacidade de processamento da máquina para realizar cálculos exaustivos. Para cada posição são determinadas todas as sequências de jogadas legais que podem seguir-se, contando passo a passo com todas as respostas possíveis do adversário, e escolhe-se a linha de jogo que leva mais seguramente ao estado final desejado (vitória) ou a uma posição favorável, assumindo que o adversário fará o seu melhor. O problema é que, havendo no xadrez um número astronómico de posições legais e respectivas combinações em jogos válidos, a busca exaustiva torna-se impraticável. Calculou-se que uma análise exaustiva que antecipe uns meros cinco lances requer a análise de 30^{10} posições (que o *Deep Blue* levaria mais de 800 horas a fazer) (Amorim, 2002, 36,41-42).

Dadas as limitações da «força bruta», o xadrez computacional tem de recorrer a heurísticas. As heurísticas são critérios gerais de análise que indicam onde será provável encontrar os factores críticos na dinâmica do jogo, para que esses pos-

sam ser explorados aprofundadamente sem perder tempo a analisar todas as combinações legalmente possíveis. A implementação das heurísticas culmina na definição de uma «função de avaliação posicional» que determina como será atribuído um único valor numérico a cada posição com base em todos os factores. Tudo o que não esteja previsto ou esteja subvalorizado na função de avaliação torna-se «invisível» para o computador de xadrez.

O que acontece é que a distinção entre posições mais ou menos desejáveis é em geral pouco clara. Amorim (2002, 70-72) identifica vários tipos de posições que enganam sistematicamente os programas de xadrez. Apenas dois exemplos: a incapacidade do programa para distinguir entre uma troca de peças desfavorável e um «sacrifício»; uma peça «encravada» (obrigada a permanecer parada para proteger uma peça de valor superior), que na prática está neutralizada, continua formalmente activa. É para incorporar conhecimento deste tipo em heurísticas que se recorre a mestres xadrezistas humanos.

A forma de projectar as heurísticas mostra claramente que contém uma grande quantidade de conhecimento humano embutido. Além disso, os programas de xadrez incorporam módulos que constituem repositórios directos de conhecimento acumulado historicamente pelos humanos. As «bibliotecas de aberturas» são vastas colecções de caminhos de jogo entre o início da partida e, digamos, o vigésimo lance, permitindo que nessa fase o programa, em vez de analisar as posições, as recupere directamente da memória já avaliadas. As «bibliotecas de finais» são similares. Os «livros de partidas» são registos de partidas históricas (o do *Deep Blue* continha 700 000 partidas de grandes mestres registadas lance a lance). Um «livro de partida alargado» contém um resumo de informação sobre cada jogada dessas partidas, como: se participou numa vitória ou numa derrota, se o recurso a essa jogada é recente ou antigo (um uso recente indicia conformidade com os avanços teóricos), frequência de uso dessa jogada e categoria do jogador que a utilizou (em princípio um

jogador de mais alto nível comete menos erros) (Campbell, 1999, 67).

Vemos, então, que a força competitiva dos programas de xadrez resulta de uma combinação de força bruta com grandes quantidades de conhecimento humano embutido. Assim, compreende-se a afirmação da equipa do *Deep Thought* (um antecessor do *Deep Blue*) acerca das expectativas para o confronto com um campeão humano: «O resultado não vai revelar se a máquina pensa, mas se um esforço humano colectivo pode suplantar os maiores feitos dos seres humanos mais capazes» (Hsu *et al.*, 1990, 24). Um dos nomes importantes da IA, John McCarthy, assinala: «O facto de que este nível de jogo requeira milhões de vezes mais cálculos do que os realizados por um jogador humano é uma medida do nosso entendimento limitado dos princípios da inteligência artificial» (1997).

Esse facto é especialmente relevante se o conjurarmos com a compreensão de que há diferenças notáveis entre a forma como os humanos e os computadores «jogam xadrez». A força dos computadores é uma função da capacidade para não cometer erros, enquanto a força dos seus oponentes humanos consiste em propor problemas que não estejam previstos pela função de avaliação (da máquina ou de outro humano). Os humanos não avaliam posições por busca exaustiva, mas com base em padrões visuais e conceitos abstractos retirados da experiência. Os humanos usam uma abordagem teleológica: visualizam o tipo de posição que querem atingir e é nessa base que definem a táctica pela qual orientam os lances; não começam por analisar jogada a jogada. Como sintetiza Amorim: «A rigor, os programas de xadrez não planejam. Ao invés disso, eles procuram por soluções contingentes, como se cada posição que ocorre durante uma partida fosse um novo problema a ser resolvido» (2002, 69-81, 71 para a citação).

Isto clarifica a nossa questão anterior, que podemos recolocar assim: poderá o futebol robótico, na procura de ganhar a aposta para 2050, usar o mesmo tipo de truque que está na base do

xadrez computacional? O futebol robótico considerará que ganhou a sua aposta se, obtendo vantagem sobre os humanos, ela resultar de uma combinação de «dopingue tecnológico» (máquinas com desempenho físico desproporcionado face aos humanos, o tipo de vantagem que os atletas podem obter ingerindo substâncias químicas) com transferência directa de conhecimento humano para a máquina por via dos seus programadores? Ou, pelo contrário, o futebol robótico aceita o desafio específico de exigir dos robôs o desenvolvimento autónomo de comportamentos colectivos, de equipa, como chave do eventual sucesso da sua aposta? Neste sentido, trata-se de saber se a robótica colectiva é apenas «mais do mesmo» ou se efectivamente aceita o desafio de pensar o social como factor de inteligência.

Da inspiração biológica à «robótica de mercado»

O futebol robótico é um segmento da robótica colectiva, isto é, da investigação acerca da melhor forma de estruturar múltiplos robôs num mesmo cenário e de os controlar na implementação de uma mesma tarefa. A robótica colectiva, além de aceitar o lugar do corpo na inteligência, consagra outra ruptura fundamental com a IA clássica: ultrapassa o paradigma da inteligência como fenómeno puramente individual e começa a pensar a inteligência como inteligência da relação. Ora, uma consequência interessante desta ruptura é que a robótica colectiva deixa de poder limitar-se às velhas metáforas da IA clássica. A «metáfora do computador», que pensa a inteligência basicamente como processamento de informação dentro da cabeça, é enfraquecida. O xadrez como exemplo por excelência da inteligência tem de ser abandonado. A robótica colectiva tem de procurar outras metáforas.

Num primeiro momento, a fonte mais significativa das metáforas inspiradoras da robótica colectiva são as sociedades de animais. Isso está bem exemplificado em Arnaud (2000), que propõe uma taxonomia dos grupos para aplicação tanto a grupos

de animais como a grupos de robôs. É pela forma de deslocação que distingue os colectivos: rebanho, bando, cardume, enxame, entre outros. Nesta base, os problemas que lhe interessam são relativos ao posicionamento dos robôs no espaço: navegação (posição do próprio) e localização (posição dos outros).

A inspiração biológica continua a ser dominante neste tipo de investigação, mas começa-se a procurar inspiração em colectivos mais sofisticados do que cardumes ou rebanhos. Uma das primeiras ocorrências de uma estratégia consistente para «robotizar» conceitos das ciências sociais e humanas aparece em Mataric (1995), que se apropria da noção de comunicação para descrever a circulação de sinais no seio de sistemas robóticos, como se isso não supusesse qualquer dificuldade conceptual. Não é que não se registem resistências – de dentro da própria comunidade das ciências do artificial – a essa importação simplista. Por exemplo, Jung e Zelinsky (2000) vêm defender que é preciso distinguir entre o que parece comunicação aos nossos olhos (há transmissão de sinal entre emissor e receptor) e a efectiva existência de comunicação (o que requer uma mensagem com conteúdo e que o sinal seja interpretado pelo receptor). Também Billard e Dautenhahn (2000) defendem que só há comunicação entre dois agentes se houver ancoragem de uma linguagem simbólica no mundo e se ela for partilhada – isto é, se eles tiverem desenvolvido uma interpretação similar para um conjunto de sinais arbitrários em termos das percepções dos seus próprios sensores.

Na verdade, pensamos estar a assistir à emergência e afirmação de uma «segunda geração de metáforas inspiradoras» da robótica colectiva. A primeira geração decorreria da inspiração biológica; a segunda inspira-se em ciências da sociedade. Vejamos um exemplo.

Stentz e Dias (1999) introduzem o conceito de uma alternativa «de mercado» em sistemas de controlo para robótica colectiva, que deveria ultrapassar a dicotomia clássica entre controlo centralizado e controlo distribuído. Defendem estes autores

que uma abordagem «económica», «de mercado livre», combinaria o melhor das duas abordagens dominantes. Essa «robótica de mercado» assenta na ideia de que uma equipa de robôs tentaria, dada uma missão, cumpri-la com um esquema de operações que minimize uma função de custos e maximize uma função de lucros, distribuindo custos e lucros pelos robôs individuais conforme o seu contributo para a missão global. Num estudo comparativo de várias arquitecturas de controlo, Singh e Thayer (2001) referem-se a esta abordagem dizendo que ela deriva das «geralmente aceites teorias económicas do mercado livre», «em particular um modelo de *laissez-faire*», que substitui a inspiração biológica muito presente noutros sistemas e que é «similar à forma como as empresas competitivas operando livremente resultam numa sociedade mais produtiva» (pp. 27-28).

Gerkey e Mataric (2002) inspiram-se nesta abordagem económica para propor, como método de coordenação inteligente de grupos de robôs, o «leilão». O que pretendem é uma forma de atribuição dinâmica de recursos em grupos de robôs. Como os robôs em causa não dispõem de um modelo do mundo, não é possível seguir uma abordagem de planeamento: cada tarefa nova aparece-lhes como gerada aleatoriamente. A comunicação é anónima, os robôs nunca são «tratados pelo nome». Quando uma mensagem é difundida ela não é endereçada a determinados destinatários, mas a determinados «assuntos». Os assuntos representam recursos: uma mensagem com o assunto «câmara móvel desocupada» dirige-se àqueles recursos robóticos disponíveis que são necessários, por exemplo, para ir a algum sítio e observar alguma coisa. Daí dizer-se que a comunicação é centrada em recursos. Diríamos, nós, mais: a «comunicação» é entre recursos.

O «leilão» funciona basicamente como se segue:

- (i) O leiloeiro anuncia uma tarefa, anúncio dirigido a todos os recursos capazes de a realizar; o anúncio inclui os critérios de avaliação da capacidade de concretização (por exemplo, se a tarefa é ir até um certo ponto e apanhar

um objecto, um dos elementos de avaliação será a distância entre a posição actual do robô e o objectivo);

- (ii) Cada recurso interessado faz uma oferta, isto é, publica os seus dados relativos aos critérios de avaliação;
- (iii) O leiloeiro atribui a tarefa à melhor oferta, com um contrato com um prazo;
- (iv) Durante a execução da tarefa, o leiloeiro vai periodicamente verificando os progressos e, se ela é satisfatória, vai renovando o contrato; se a progressão é insuficiente ou, no limite, o robô «desapareceu» (avariou ou foi recarregar baterias), a tarefa é retirada e será posta a leilão novamente.

Estamos claramente no campo da exploração de novas analogias: «Os leilões, numa ou outra forma, têm sido utilizados nas sociedades através da história para distribuição de recursos escassos a indivíduos ou grupos» (p. 761). E: «resolvemos este problema (...) seguindo a divisão de trabalho usada pelos humanos» (p. 764).

Para confirmar como as metáforas se embebem umas nas outras, é curioso observar que os autores dizem que, embora possa parecer contra-intuitivo empregar um mecanismo de competição (o leilão) para gerar cooperação, a diferença entre competição e cooperação está apenas na presença ou ausência de egoísmo – e eles garantem (porque foram eles que os projectaram!) que estes robôs não são egoístas! Mas o que significa ser egoísta quando não há mais ninguém no mundo (é claramente explicitado que tanto produtores como consumidores de mensagens não estão de modo nenhum cientes uns dos outros)?

Parece-nos, pois, estarmos perante a emergência de uma nova geração de metáforas inspiradoras da IA, por influência da robótica colectiva. Contudo, o exemplo que acabámos de aflorar mostra a influência de concepções ingenuamente individualistas do social em tais experiências. Vejamos, de seguida, uma aproximação diferente a essa questão.

Organizações artificiais

Drogoul e Zucker (1998, 5-10) resumem o «método Cassiopeia» para o projecto de sistemas multiagentes com comportamentos colectivos. A ideia subjacente é a de criar colectivos artificiais que sejam mais do que um aglomerado de agentes e sejam efectivamente dotados de organização (organizações artificiais).

A ideia de partida é que um agente pode desempenhar papéis de três níveis: individual, relacional, organizacional. Os papéis individuais correspondem aos comportamentos de que cada agente é capaz isoladamente. Os papéis relacionais correspondem às formas como cada agente interage com os demais. Os papéis organizacionais correspondem às formas como os agentes gerem as interacções para produzirem organizações.

As dependências entre agentes traduzem dependências entre os papéis individuais que eles desempenham. Essas dependências são designadas por relações de influência. A estrutura da organização é distribuída pelos agentes como transmissão de sinais de influência. A estrutura de dependências define implicitamente os potenciais grupos de agentes que podem ser formados. A dinâmica da organização é definida por papéis organizacionais que permitem aos agentes dotados desses papéis gerir a formação e dissolução de grupos. Um agente que tenha o papel de «iniciador» pode criar (e dissolver) grupos de agentes que tenham o papel de «participante», sendo que um determinado grupo deixa de existir quando o agente iniciador alcançou o objectivo para que o grupo foi formado. Estes autores exemplificam a aplicação da sua metodologia com o futebol robótico em simulação.

Uma equipa de investigadores portugueses (Lima *et al.*, 2000) faz uma aplicação dessas ideias ao futebol robótico implementado fisicamente (e não apenas em simulação). Vejamos aspectos da sua abordagem.

Ao nível organizacional estabelece-se a estratégia que deverá ser seguida por toda a equipa. «Defender», «atacar» e «contra-atacar» são exemplos de estratégias dinâmicas. Cenários pré-programados em função de situações típicas de jogo («pontapé de saída» ou «grande penalidade», por exemplo) também definem estratégias, que orientam o comportamento global da equipa.

A selecção de uma estratégia pelo nível organizacional define o conjunto de táticas que a implementam e cada tática define o conjunto de comportamentos admissíveis para os robôs individuais. Ao nível relacional estabelecem-se as relações entre robôs, determinando quais unidades adoptam quais comportamentos, o que pode incluir uma negociação entre agentes acerca de qual está em melhor posição para implementar determinada tarefa (em função da proximidade à bola, por exemplo).

Ao nível individual encontram-se todos os comportamentos que os robôs podem executar². Os comportamentos («guarda-redes» ou «atacante», por exemplo) são combinações de tarefas primitivas de que os robôs individuais são capazes (como «chutar», «procurar a bola» ou «rodar para a esquerda»).

Cada um dos robôs individuais é dotado dos três níveis, embora as decisões estratégicas (nível organizacional) sejam tomadas por um único agente, o «capitão da equipa» (se esse falhar, outro assumirá a função). Aos diferentes níveis, as decisões são influenciadas quer pelo «estado do mundo» (informação disponível nos sensores acerca do estado das variáveis físicas que eles captam), quer pelo «estado da equipa» (conjunto de comportamentos que estão a ser executados).

As abordagens que muito resumidamente mencionámos neste ponto ilustram o que queremos dizer: os sistemas multiagentes e a robótica colectiva preocupam-se com o facto de os colectivos não funcionarem apenas como colecções de indi-

² A variação terminológica «papel»/«comportamento» deve ser explicitada assim: um «papel» é um «comportamento» executado por um robô quando numa determinada «posição» (o que define o «guarda-redes» é um par comportamento/posição).

víduos, mas terem dimensões próprias que os caracterizam como organizações. Que essa aproximação se esteja a desenvolver é interessante, mas queremos assinalar apenas dois exemplos de uma certa ingenuidade que ainda perpassa nessas abordagens.

Em primeiro lugar, não há «poder puro» dentro destas organizações artificiais. Isto é, só são consideradas aquelas relações de dependência entre agentes que foram estabelecidas por servirem as funções definidas para a organização. Não é deixado espaço para outras dependências, resultantes de relações sociais mais complexas. Como em geral acontece com estes seres artificiais, os projectistas concebem-nos com uma pureza de laboratório: eles só existem para aquele fim, não são contaminados por nenhum outro tipo de condicionamento, têm uma absoluta racionalidade funcional direccionada para a pertença exclusiva a um único colectivo. Como se cada um deles só existisse para um dado «bem comum» e todo o poder começasse e acabasse na melhor estratégia para o alcançar. Esse tipo de simplificação pode condenar as «organizações artificiais» a «mundos de brinquedo» (um problema que assombrou a IA clássica).

Outro aspecto interessante diz respeito a um pressuposto epistemológico implícito nas experiências consideradas. Toda a informação factual é a que está disponível ao nível individual, não há uma «percepção» própria do grupo ou da organização. Pode dizer-se: sim, também numa organização humana toda a «percepção» de um colectivo depende das percepções dos indivíduos que funcionam como «sensores» da organização. Só que aqui se revela a presença implícita de uma teoria epistemológica particular: assume-se (à maneira positivista) que «factos são factos», como se os «factos» tal como são integrados não dependessem de um ponto de vista ou de uma interpretação. Nada distingue os pontos de vista colectivo e individual. A organização enquanto tal é cega – como, aliás, se adivinhava quando os autores considerados esclareciam que as competências individuais, relacionais e organizacionais eram sempre e apenas

aspectos da competência de cada robô individual. É o regresso, pela mão das ciências do artificial, de um dos programas mais ambiciosos da epistemologia positivista: representar o mundo exterior como uma construção lógica a partir dos dados dos sentidos, como projectou Russell em *Our Knowledge of External World* e Carnap levou ao máximo esplendor em *Der logische Aufbau der Welt*. Falta aqui, parece-nos, um elemento de exteriorização da organização relativamente aos indivíduos. Tal como é importante para os humanos que sejam capazes de armazenar informação, conhecimento e procedimentos fora do seu organismo, é importante conceber colectivos que não sejam «armazenados» apenas como um somatório de dimensões dos indivíduos.

Por uma robótica colectiva com instituições

Até aqui identificámos alguns aspectos da renovação da IA que resultam da investigação em robótica colectiva. Ao mesmo tempo assinalando a infiltração de certos pressupostos epistemológicos (de eficácia discutível) nas suas metáforas inspiradoras. Esses pressupostos parecem estar intimamente relacionados com o predomínio de uma visão individualista dos grupos (das «sociedades») de robôs. É como se a robótica colectiva, quando opta por sistemas de controlo distribuído em vez de controlo centralizado, deixasse de dispor de outra alternativa conceptual que não seja esse individualismo. Desse modo, a robótica colectiva ficaria impedida de extrair todas as consequências de um dos seus traços mais inovadores: conceber a inteligência como fenómeno que não pode ser completamente compreendido sem contar com a dimensão relacional e social.

É por isso que queremos concluir, num breve apontamento, com o esboço de uma metáfora inspiradora que possa abrir outros caminhos à robótica colectiva: instituições para robôs. Vamos arrancar essa metáfora inspiradora a um debate em economia, na óptica de dois autores interessados na renovação da perspectiva institucional nessa ciência. Hodgson (1988, 53-71), numa revisão dos proble-

mas do individualismo metodológico em ciências sociais, lembra que não se trata de um ponto de vista ingénuo (nem recusa a existência e impacte das entidades sociais, nem se limita ao truísmo «a sociedade é formada por indivíduos»), mas da tese segundo a qual as explicações dos fenómenos sociais são redutíveis a explicações acerca dos indivíduos. Para o individualismo metodológico, o propósito individual é a causa suficiente de toda a acção social.

Uma forma historicamente concreta (e actual) de individualismo metodológico é o paradigma dominante em economia, a chamada teoria económica neoclássica, na qual é central a ideia do «homem económico racional»: os indivíduos agem «racionalmente» num sentido muito específico, isto é, por via de um cálculo poderoso processam toda a informação relevante com o fito de maximizar um resultado expresso num único valor («utilidade» ou lucro, por exemplo). Cada indivíduo, como agente económico, agiria então na base de uma ordenação consistente e estável das suas preferências – e esse comportamento perfeitamente racional dos indivíduos na pura prossecução dos seus interesses (económicos) próprios seria a causa do estado de equilíbrio global na actividade económica (Hodgson, 1988, 73-74).

Essa concepção tem vários pressupostos interessantes. Um deles é o do conhecimento perfeito como capacidade do agente racional: por exemplo, um consumidor que agisse segundo o modelo seria capaz de conhecer e avaliar todas as características relevantes dos 10 000 produtos diferentes que encontra num supermercado de cada vez que vai às compras. Outro pressuposto importante é o da ilimitada capacidade de cálculo do agente: é como se qualquer agente económico fosse um matemático exímio capaz dos mais complexos cálculos de probabilidades sobre todas as circunstâncias que podem afectar o resultado da sua acção. Também se pressupõe a completa consciência do processo de decisão, ignorando todos os processos mentais inconscientes e subconscientes (inacessíveis a «decisão racional») que influenciam a acção. E ainda ignora que muitas vezes agimos

contra o que os nossos juízos racionais nos ditam ser o nosso interesse (por «fraqueza de vontade» ou «incontinência»).

Castro Caldas (2001) lembra como Herbert Simon (a começar na década de 1950) mostra, com a sua teoria da «racionalidade limitada», que o agente que se pode encontrar «na realidade» não corresponde à concepção neoclássica de «homem económico racional». É interessante notar como este debate corre em paralelo nos terrenos da economia e da IA.

Contudo, Castro Caldas vai mais longe para evidenciar os limites da abordagem do próprio Simon. Partilhando uma perspectiva institucionalista, assinala o facto de que a racionalidade do indivíduo está fortemente imbricada no social. O individualismo da economia neoclássica ignora o facto de que a nossa acção (como indivíduos ou empresas, por exemplo) é em grande parte orientada por hábitos, rotinas, convenções, normas, organizações. Os comportamentos, propósitos e objectivos do indivíduo não se enraízam numa célula irredutível – instituições sociais e cultura, por exemplo, contam. Os colectivos organizam a informação que recebemos e isso influencia a nossa própria percepção do mundo e as nossas preferências. O agente limitadamente racional vive em ignorância parcial e incerteza, pelo que acontecem desfasamentos entre crenças e experiências – e assim entra em jogo a aprendizagem, a qual por sua vez é social. Não é possível explicar o indivíduo apenas como instância do grupo, mas é preciso compreender como o colectivo, o social, é mais do que um mero compósito de partes.

Na tradição institucionalista próxima de Hayek, refere ainda Castro Caldas, é reconhecida a importância das regras partilhadas, normas e convenções sociais sobre as quais assenta a ordem social. Mas é feita uma distinção entre *regras concebidas deliberadamente* (produção legislativa, por exemplo) e *regras geradas de forma espontânea*, que resultam do agregado das acções desenvolvidas por cada um dos indivíduos na prossecução do seu interesse próprio, acções essas e indivíduos esses que não tinham qualquer intenção de por

essa via produzir regras. Hayek consideraria que as regras resultantes de um crescimento espontâneo são as regras próprias de uma ordem social benéfica.

Numa série de experiências de simulação de sociedades artificiais em que recorre a uma interpretação original da técnica de evolução artificial designada por algoritmo genético, Castro Caldas vai questionar profundamente esta hipótese da ordem social espontânea: será que a mera interacção dos agentes, que prosseguem o seu interesse individual, resulta necessariamente em soluções eficientes para os problemas colectivos?

Não sendo possível entrar aqui nos detalhes das suas experiências, retemos que elas apontam para a seguinte conclusão: a pura acção dos indivíduos na prossecução dos seus interesses próprios não resulta em ordem social (necessária à própria prossecução desses interesses individuais). Ora, uma vez que uma ordem social exclusivamente dependente da coacção externa (soberano absoluto) não parece desejável e que a mera distribuição do controlo (cada um controla-se a si próprio e aos demais) também se revela ineficiente, algo mais tem de ser feito pelos indivíduos para estabelecer e garantir essa ordem social. O que será esse «algo mais»? Instituições.

Mas o que são instituições? Castro Caldas, revendo a diversidade de respostas dadas por economistas e outros cientistas sociais a essa questão, considera que o que nelas existe de comum permite concluir que o conceito de «instituições» abrange «quer os *sistemas de regras partilhadas*, quer os *sistemas de crenças* que os sustentam», com as seguintes características: precedem os indivíduos e perduram para lá da sua existência; são um médium entre a percepção e a acção («as competências cognitivas são adquiridas num contexto institucional»); limitam o espaço de decisão e a autonomia do agente, ao mesmo tempo que expandem as possibilidades de acção e podem mesmo substituir-se à decisão; reduzem a incerteza, formando expectativas sobre o curso do mundo (pp. 8-9). Além disso, nesta perspectiva institucionalista, como as instituições da sociedade

não emergem pelo puro jogo da prossecução dos interesses individuais, solicitam os indivíduos a participar na sua construção deliberada.

A sugestão, que aqui deixamos como mero esboço, é que a exploração desta concepção institucionalista na robótica colectiva pode dar suporte às suas ambições mais inovadoras, ao funcionar como antídoto ao persistente domínio das metáforas individualistas nesse campo de investigação. Uma forma concreta de tal exploração poderia passar pelo desenho de experiências com robôs físicos que traduzissem a linha de experiências que Castro Caldas realiza apenas em simulação. Tratar-se-ia, assim, de projectar experiências de robótica colectiva de inspiração institucionalista envolvendo inteligência em «corpos artificiais» agindo no mundo físico. ■

Bibliografia

- AMORIM (2002), Cláudio Alves, *A Máquina e Seus Limites: Uma Investigação sobre o Xadrez Computacional*, dissertação de mestrado, Programa de Pós-Graduação em Ensino, Filosofia e História das Ciências, Universidade Federal da Baía e Universidade Estadual de Feira de Santana, 2002 (versão electrónica fornecida pelo autor).
- ARNAUD, Pierre (2000), *Des moutons et des robots*, Lausana, Presses Polytechniques et Universitaires Romandes.
- ASADA, M., e KITANO, H. (1999), «The RoboCup Challenge», in *Robotics and Autonomous Systems*, Vol. 29, pp. 3-12.
- BILLARD, Aude, e DAUTENHAHN, Kerstin (2000), «Experiments in Learning by Imitation: Grounding and Use of Communication in Robotic Agents», in *Adaptive Behavior*, 7 (3/4), pp. 415-438.
- CAMPBELL, Murray (1999), «Knowledge Discovery in Deep Blue», in *Communications of the ACM*, 42 (11), pp. 65-67.
- CAMPBELL, Murray, HOANE Jr., Joseph A., HSU, Feng-Hsiung (2002), «Deep Blue», in *Artificial Intelligence*, 134, pp. 57-83.

- CASTRO CALDAS, José Maria (2001), *Escolha e Instituições – Análise Económica e Simulação Multi-agentes*, Oeiras, Celta Editora.
- DROGOUL, Alexis, e ZUCKER, Jean-Daniel (1998), *Methodological Issues for Designing Multi-Agent Systems with Machine Learning Techniques: Capitalizing Experiences from the RoboCup Challenge*, Technical Report LIP6 1998/041, Laboratoire d'Informatique de Paris 6.
- GERKEY, B. P., e MATARIC, M. J. (2002), «Sold! Auction Methods for Multirobot Coordination», in *IEEE Transactions on Robotics and Automation*, 18 (5), pp. 758-768.
- HODGSON, Geoffrey M. (1988), *Economics and Institutions – a Manifesto for a Modern Institutional Economics* (todas as referências remetem para a versão portuguesa, *Economia e Instituições – Manifesto por Uma Economia Institucionalista Moderna*, Oeiras, Celta Editora, 1994).
- HSU, F.-H., ANANTHARMANT, T., CAMPBELL, M., NOWATZYK, A. (1990), «A Grandmaster Chess Machine», in *Scientific American*, 263 (4), pp. 18-24.
- JUNG, D., e ZELINSKI, A. (2000), «Grounded Symbolic Communication between Heterogeneous Cooperating Robots», in *Autonomous Robots*, 8 (3).
- LIMA, Pedro, VENTURA, Rodrigo, APARÍCIO, Pedro, CUSTÓDIO, Luís (2000), «A Functional Architecture for a Team of Fully Autonomous Cooperative Robots», in *RoboCup-99: Robot Soccer World Cup III*, Berlim, Springer-Verlag.
- MATARIC, Maja J. (1995), «Issues and Approaches in the Design of Collective Autonomous Agents», in *Robotics and Autonomous Systems*, 16 (2-4), pp. 321-331 (versão em linha descarregada em Fevereiro de 2003, em <http://www-robotics.usc.edu/~maja/publications/ras-si.ps.gz>).
- MCCARTHY, John (1997), «AI as Sport», in *Science*, 276, pp. 1518-1519.
- SINGH, Surya P. N., e THAYER, Scott M. (2001), *ARMS: Autonomous Robots for Military Systems – a Survey of Collaborative Robotics Core Technologies and Their Military Applications*, Technical Report CMU-RI-TR-01-16, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pensilvânia (versão em linha descarregada em Novembro de 2002 de <http://www.ri.cmu.edu/pubfiles/pub3/singhsurya20012/singhsurya20012.pdf>).
- STENTZ, Anthony, e DIAS, M. Bernardine (1999), *A Free Market Architecture for Coordinating Multiple Robots*, Technical Report CMU-RI-TR-99-42, The Robotics Institute, Carnegie Mellon University, Pittsburgh, Pensilvânia (versão em linha descarregada em Fevereiro de 2003 de <http://www.ri.cmu.edu/pubfiles/pub2/stentzanthony19992/stentzanthony19992.pdf>).
- STONE, Peter, e VELOSO, Manuela (2000), «Multi-agent Systems: a Survey from a Machine Learning Perspective», in *Autonomous Robots*, 8 (3), pp. 345-383.
- TURING, Alan M. (1950), «Computing Machinery and Intelligence», in *Mind*, 59, pp. 433-460.

TRAJECTOS

Revista de Comunicação, Cultura e Educação

Periodicidade: semestral / n.º 5, Outono de 2004

Preço: € 12

Director

José Rebelo

Conselho de Redacção

Alexandre Melo, Idalina Conde, José Luís Garcia, José Jorge Barreiros, José Machado Pais, José Manuel Paquete de Oliveira, José Manuel Prostes da Fonseca, José Rebelo, Pierre Guibentif, Teresa Seabra

Conselho Editorial

António Firmino da Costa (ISCTE), Eduarda Gonçalves (ISCTE), Eduardo Prado Coelho (Univ. Nova de Lisboa), Fernando Luís Machado (ISCTE), Gustavo Cardoso (ISCTE), Jean-Pierre Dubois (Univ. de Paris XI), Joaquim Azevedo (Ass. Empresarial Portuguesa), Manuel Castells (Univ. Aberta da Catalunha), Maria Augusta Babo (Univ. Nova de Lisboa), Maria de Lurdes Lima dos Santos (Observatório das Actividades Culturais), Mario Diani (Univ. de Trento), Michel Wiewiorka (Escola de Altos Estudos em Ciências Sociais de Paris), Muniz Sodré (Univ. Federal do Rio de Janeiro), Teresa Ambrósio (Univ. Nova de Lisboa)

Uma publicação de



Instituto Superior de Ciências do Trabalho e da Empresa

Departamento de Sociologia, Secção de Comunicação, Cultura e Educação

Av. das Forças Armadas, Edifício ISCTE

1649-026 Lisboa

Tel: 217903016 – Fax: 217903017

E-mail: jose.rebelo@iscte.pt

Edição e distribuição

EDITORIAL NOTÍCIAS

Rua Bento de Jesus Caraça, 17

1495-686 Cruz Quebrada

Tel.: 210052350 – Fax: 210052340

E-mail: geral@editorialnoticias.pt

Internet: www.editorialnoticias.pt

REPENTINAMENTO

Projecto gráfico

Tiago Cunha

Revisão

Ayala Monteiro

Pré-impressão

VHM – Produções Gráficas, Lda.

Impressão e acabamento

Multitipo – Artes Gráficas, Lda.

ISSN 1645-5983

ISBN 972-46-1577-4

Edição: 10 04 0066

Depósito legal: 180674/02