Teoria da Mente e Aprendizado: Desafios da Robótica Inteligente

Conrado Silva Miranda
Laboratório de Mobilidade Autônoma
Faculdade de Engenharia Mecânica
Universidade Estadual de Campinas
Email: miranda.conrado@gmail.com

***Resumo*—A robótica atual busca gerar sistemas autônomos e interativos, que são capazes de se adaptar ao ambiente dinâmico que as cerca e sociabilizar com pessoas ou outros robôs. Para interagir adequadamente neste ambiente, o robô deve ser capaz de inferir o que os outros agentes estão pensando através de seu comportamento e agir de maneira apropriada. Este trabalho visa estudar como isso é feito atualmente através da utilização da teoria da mente em robôs, levantando métodos conhecidos na literatura e mostrando como o aprendizado se apresenta como ponto crucial, e apresentar problemas existentes na aplicação de tal teoria em robôs reais juntamente com possíveis soluções, que serão exploradas em trabalhos futuros.**

***Index Terms*—Robótica Inteligente, Teoria da Mente, Aprendizado Robótico**

I. Introdução

Para desenvolvimento de robôs inteligentes, é necessário que eles possam inferir os estados mentais dos outros seres inteligentes com os quais ele interage, podendo também aprender através de um tutor. A teoria da mente fornece base para realizar tal inferência, assumindo conhecimento sobre o ser observado e o próprio ser, assim como um modelo de mundo. No entanto, as técnicas atuais para utilização de tal teoria são limitadas, funcionando adequadamente apenas em situações restritas. Este trabalho visa analisar estas falhas principais e levantar soluções para as mesmas, utilizando arquiteturas de aprendizado por curiosidade, simulação e memória

Como a mente observável a um ser é apenas sua própria, ele não é capaz de determinar os estados mentais de outros seres ou mesmo se outros seres possuem mente. Premark levanta em seu trabalho [Premack et al., 1978], através de estudos com chimpanzés, a possibilidade de seres não humanos possuírem uma teoria da mente, estudando métodos de identificá-la. Tais estudos permitem identificar como um ser com teoria da mente deve se comportar em determinadas situações, servindo como guia portanto para o desenvolvimento e teste de sistemas artificiais capazes de inferir estados mentais.

Existem duas teorias atuais que explicam como funciona a teoria da mente, ambas publicadas no livro *Theories of theories of mind* [Carruthers and Smith, 1996]: a teoria-teoria, que diz que a teoria da mente seria apenas uma teoria utilizada para raciocinar sobre a mente dos outros, sendo inata e desenvolvida automaticamente, o que impossibilitaria a construção de seres artificiais com esta capacidade, uma vez que não somos capazes de entendê-la; já a teoria da simulação diz que a teoria da mente não seria puramente teórica, sendo feita através de simulação mental da situação em questão e, portanto, passível de implementação.

No entanto, a própria teoria da simulação é dividida entre duas abordagens [Dôkic and Proust, 2002]. A primeira abordagem, que é geralmente utilizada em sistemas artificiais e será descrita em mais detalhes durante o trabalho, diz que o ser deve ser capaz de identificar seus próprios estados mentais antes de inferir sobre os estados de outras mentes. Com esta capacidade, o ser seria capaz de inferir os estados mentais de outros, interpretando as entradas e saídas do sistema observado. Outra teoria diz que uma pessoa é capaz de identificar os estados mentais próprios ou de outros seres através de rotina de ascensão, na qual perguntas sobre estados mentais são refraseadas como questões metafísicas e, através de sua resposta, infere-se sobre o estado mental atual. Esta teoria é de difícil implementação prática, pois requer que o sistema observado tenha a capacidade de elaborar tais questões e de respondê-las apropriadamente.

A ideia de desenvolvimento de robôs como sistemas inteligentes existe há muito tempo, tendo surgido com Turing [Turing and Copeland, 2004] e popularizado com Brooks [Brooks, 1990, 1993], através do seu estudo de cognição situada e incorporada, que se tornou o foco de muitas pesquisas em robótica cognitiva [Anderson, 2003]. As primeiras pesquisas com robôs utilizando teoria da mente foram realizadas para facilitar o aprendizado, utilizando um professor humano para ensinar tarefas [Lungarella and Metta, 2002]. Estes robôs eram programados com os princípios primitivos da teoria da mente, sendo capazes de interpretar o que o tutor queria dizer ao interagir com o robô, ensinando-o tarefas que não estavam previstas em sua programação inicial, possibilitando aprendizado de tarefas que o programador poderia não ter capacidade de programar diretamente.

No entanto, tais técnicas de aprendizado devem ser expandidas para robôs autônomos, uma vez que originalmente elas necessitam de um tutor e dependem da habilidade do robô reconhecer tal tutor, o que pode ser um limitante no desenvolvimento de suas habilidades. Para explorar completamente suas capacidades, o robô deve ser capaz de aprender autonomamente, além de aprender com outros seres como tutores. Ele também deve ser capaz de interagir corretamente com seres não tutores através da inferência de seus estados

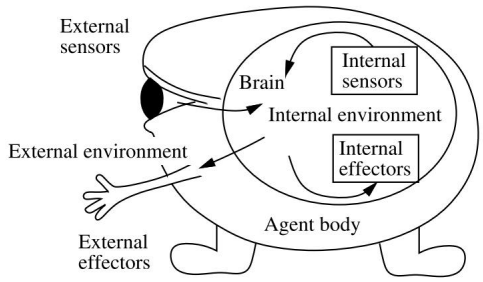

Figura 1. Agente auto-ciente e auto-efetivo. Imagem retirada de [Weng, 2002].

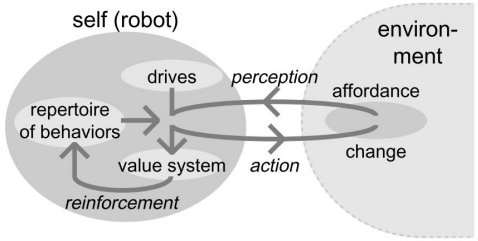

Figura 2. Intencionalidade em um robô. Imagem retirada de [Kozima, 2001].

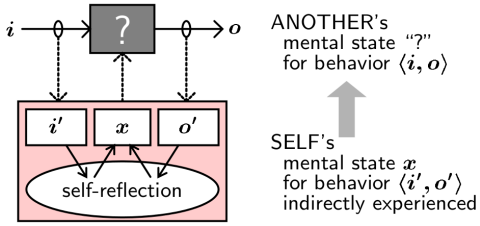

Figura 3. Inferência do estado mental do observado. Imagem retirada de [Kozima, 2001].

mentais. O restante deste trabalho aborda como isso é feito atualmente e as principais falhas existentes no sistema atual, juntamente com possíveis soluções.

Este trabalho está dividido um quatro seções: esta seção apresentou uma introdução ao problema que será abordado, sendo seguido pela seção II que explora as abordagens utilizadas atualmente para aplicação da teoria da mente. Em seguida, apresentamos na seção III uma discussão sobre os métodos apresentados e suas falhas, apresentando possíveis soluções. A seção IV finaliza o trabalho, destacando trabalhos futuros e indicando problemas possivelmente encontrados.

II. Aplicações robóticas da teoria da mente

Ao se programar um robô de tal maneira que ele tenha apenas os conhecimentos definidos na sua criação, este robô não é completo [Weng, 2002], ou seja, não é capaz de atingir o mesmo potencial que um ser humano em uma idade arbitrária, ficando limitado à capacidade fornecida em sua programação inicial. Por isso, devemos criar agentes auto-cientes e auto-efetivos, sendo estes capazes de alterar suas próprias estruturas mentais e seu conjunto de sensores e atuadores não devem perceber apenas o mundo a sua volta, mas também seu próprio ambiente interno, como mostrado na figura 1. A criação de um robô sem esse conhecimento prévio completo implica na necessidade de se estabelecer mecanismos de aprendizado posterior, enquanto o sistema estiver operando.

No intuito ultrapassar tal barreira, iniciaram-se pesquisas em robôs capazes de aprender com um professor. Tais robôs utilizam princípios da teoria da mente, como atenção compartilhada, para tentar deduzir o que a pessoa está o ensinando. Kozima [Kozima, 2001] apresenta a necessidade de: atenção compartilhada, que faz com que o robô identifique o alvo desejado pelo professor; empatia, para que seja capaz de se colocar no lugar do professor, visando compreender seus objetivos; e captura de movimento, que permite mapear entradas e saídas do professor para si mesmo e estimar seu estado mental pelo estado mental do robô sob aquelas entradas e saídas observadas.

A empatia necessária para aprendizado robótico com teoria da mente difere da empatia tradicional no ponto que o robô não deve apenas se colocar no lugar do outro, mas também tentar inferir qual a intenção das ações executadas. Desta maneira, a intencionalidade, como concebida por Dennett [Dennett, 1989], se mostra característica marcante neste desenvolvimento. Se não há intencionalidade, o sistema não será capaz de prever qual estado mental o professor possui, uma vez que ele não está relacionado com um objetivo a ser alcançado. Para Kozima, tal intencionalidade pode ser aprendida por reforço em um repertório de movimentos, conforme mostrado na figura 2. Quanto à captura de movimento, Kuniyoshi [Kuniyoshi et al., 2004] aborda o desenvolvimento de um robô que aprende a reconhecer ações de maneira simbólica, sendo capaz de executá-las depois e de notar padrões no comportamento, que podem ser imitados posteriormente.

Esta utilização da teoria da mente se baseia na primeira abordagem da teoria da simulação, que diz que o sistema inteligente se coloca no lugar de outro sistema inteligente e, através de interpretação das entradas e saídas, estima o estado mental do observado através do estado mental presente no observador, como representado na figura 3. Tal metodologia de simulação simulação foi utilizada por Buchsbaum [Buchsbaum et al., 2005] para desenvolver um ser computacional que, através de identificação dos movimentos gerados pelo observado, que eram conhecidos pelo observador, e um grafo de motivações, organizado de forma hierárquica e contendo tuplas que envolvem a motivação, uma ação e um objeto de referência, era capaz de identificar a motivação por trás dos movimentos de um ser semelhante a si. Além disso, ao observar movimento ambíguos, ele era capaz de descobrir a motivação através da análise do objeto e das possíveis tuplas motivacionais ou mesmo de descobrir uma nova utilidade, denominada *affordance*, para um dado objeto, dentre as *affordances* possíveis conhecidas.

O conhecimento das *affordances*, definidas por Gibson [Gibson, 1977], possíveis para os objetos é essencial para que o robô seja capaz de interpretar qual a intenção do professor. Sahin [Sahin et al., 2007] apresenta formalismos para

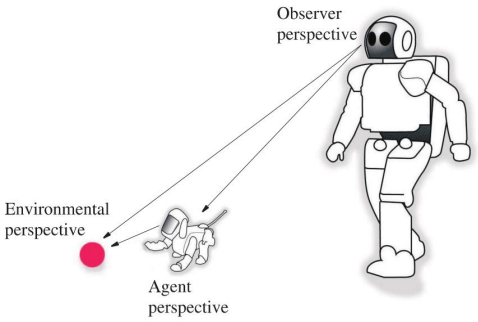

Figura 4. Perspectivas possíveis de *affordance*. Imagem retirada de [Sahin et al., 2007].

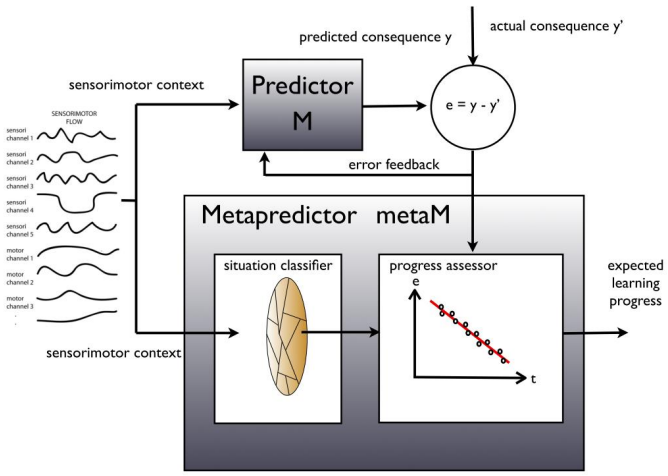

Figura 5. Preditor e metapreditor para aprendizado por curiosidade. Imagem retirada de [Kaplan and Oudeyer, 2006].

utilização de *affordances* em robôs e pesquisas relacionadas nas áreas de ciência cognitiva e distingue as *affordances* em três perspectivas possíveis, como mostrado na figura 4, sendo estas: perspectiva do agente, no qual o agente determina a *affordance* para um dado objeto através da interação com o mesmo utilizando seus próprios atuadores; perspectiva do ambiente, cujas *affordances* são natuais ao ambiente e devem ser percebidas pelo agente; e perspectiva de observador, onde a interação do agente com o ambiente é observada por um observador externo que não sabe exatamente as *affordances* que o agente possui com o dado objeto, mas é capaz de criar hipóteses sobre quais elas são. Para teoria da mente, todas estas perspectivas são importantes e devem ser levadas em conta separadamente. A perspectiva do agente juntamente com as *affordances* observadas no ambiente determinam o que o robô sabe que ele é capaz de fazer com o ambiente. No entanto, ao observar seu tutor, ele possui perspectiva de observador, devendo as *affordances* disponíveis nessa perspectiva ser utilizadas para inferência dos estados mentais. É importante ressaltar que não existe obrigatoriamente nenhuma relação entre a perspectiva de agente e observador, podendo possuir conjuntos disjuntos de *affordances* para objetos e instantes distintos.

III. Problemas existentes nas abordagens atuais

Um problema notável na utilização da teoria da simulação é que o robô deve interpretar as entradas e saídas do sistema observado, exigindo que ele possa mapear sentidos e atuadores, como feito por Buchsbaum. No entanto, tal abordagem é limitante, uma vez que esse mapeamento é pré-programado no robô, ou seja, se apresentado a um tutor diferente do esperado, ele pode não ser capaz de mapear as entradas e saídas. Para sanar este problema, a solução mais clara seria utilizar os resultados das ações para fazer essa interpretação, de tal maneira que o observador veja o resultado e interprete como ele atingiria o mesmo resultado caso estivesse na posição do observado. A teoria de *affordances* poderia então ser utilizada para descobrir como gerar determinados resultados, uma vez que ela trata das ações possíveis sobre um objeto e permite analisar seus resultados, assim como para realizar o mapeamento da entrada, através da análise de perspectivas apresentada anteriormente. Por exemplo, eu sei que um morcego possui um sonar e não possui visão, então seu conjunto de entradas ao interagir com o ambiente é diferente do meu. Uma vez que é impraticável analisar todas as entradas possíveis, posso dizer que uma *affordance* tanto do sonar como da visão é detecção de objetos, conseguindo então mapear adequadamente a entrada disponível ao observado para o observador.

No entanto, o requerimento de saber os *affordances* tanto do observado quanto do observador pode ser muito limitada, uma vez que tal conhecimento é difícil de ser adquirido. Um robô, a princípio, só é capaz de saber seus próprios *affordances* com os objetos do mundo. Como um robô que só é capaz de andar na terra é capaz de saber que um pássaro pode passar por um balão sem gerar danos a ele? Além disso, como ele pode ser capaz de assemelhar isso ao robô poder passar por um balde vazio? Ambos casos representam o mesmo *affordance* de passar por um objeto sem sofrer dano, mas os objetos e agentes são completamente diferentes entre si. Há ainda o caso de o agente não ser capaz de conceber o próprio *affordance* observador como por exemplo saber que o passáro voa e por isso ele pode alcançar lugares diferentes do robô. Para tentar dar uma solução para isso, vamos primeiro analisar as abordagens atuais para aprendizado robótico, que inclui aprendizado de *affordances*.

A principal abordagem para aprendizado não supervisionado é a utilização da curiosidade. Kaplan [Kaplan and Oudeyer, 2006] apresenta uma arquitetura, mostrada na figura 5, que utiliza um preditor para antecipar as consequências de uma determinada ação, sendo este preditor treinado para minimizar o erro entre a previsão e o resultado ocorrido. Além deste preditor, a arquitetura possui também um metapreditor, cujo objetivo é identificar qual o contexto sensorio-motor atual e estimar qual o progresso do preditor neste contexto. Assim, o robô não apenas é capaz de predizer os resultados de suas ações, mas quando um determinado contexto não for mais interessante para seu aprendizado, ou seja, aquele aprendizado possível já foi quase todo explorado, o robô automaticamente

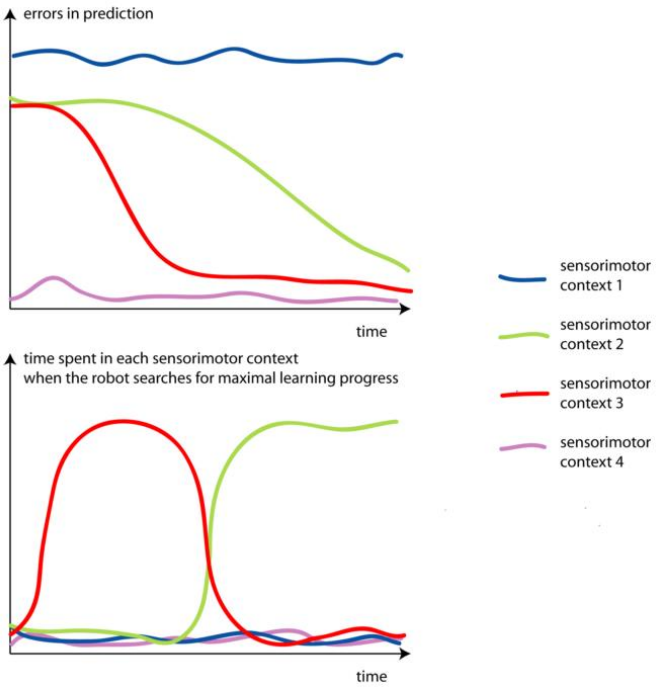

Figura 6. Contextos de aprendizagem. Imagem retirada de [Kaplan and Oudeyer, 2006].

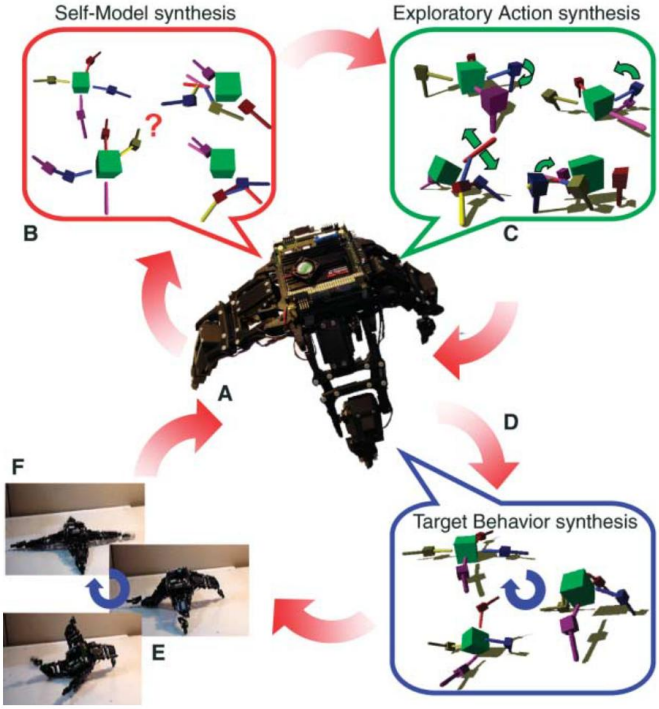

Figura 7. Ciclo de exploração e explotação de modelos internos. Imagem retirada de [Bongard et al., 2006].

busca um contexto onde haverá mais aprendizado. Dessa forma, o robô aprende sem um objetivo definido, apenas tentando aprender o que ele pode ou não fazer, e evita situações que ele é capaz de prever os resultados precisamente ou muito grosseiramente, focando o aprendizado nos contextos onde há maior oportunidade de ganho. Como o robô não possui nenhum conhecimento prévio sobre como é sua dinâmica, esse algoritmo permite que o robô aprenda a se locomover e interagir com os objetos, gerando *affordances* indiretamente, como saber que pode fechar sua boca em um objeto e ele fica preso, que pode ser representado pela *affordance* de ser mordível.

Uma arquitetura semelhante a esta pode ser utilizada para se aprender o que outros agentes em seu ambiente são capazes de fazer. Supondo que o robô saiba as entradas e saídas possíveis do observado, ele pode prever o que acontecerá baseado em observações anteriores, como por exemplo saber que um pássaro voará ao bater as asas, tendo já visto comportamento anterior. O metapreditor pode ser utilizado para escolher o agente a ser observado, visando redução do erro de predição. Assim, o robô pode se comportar como uma criança humana, que dentre um grupo de formigas, observa aquela que apresenta comportamento mais interessante. Utilizando algum algoritmo que possa estimar proximidade entre agentes, o erro de predição pode ser reduzido para vários agentes simultaneamente, uma vez que agentes parecidos provavelmente possuem comportamentos semelhantes. Tal habilidade seria semelhante à capacidade de uma pessoa prever como uma zebra vai se comportar, mesmo sem nunca ter visto uma e tendo apenas contato com cavalos. Esta observação leva a crer que essa generalização para reduzir erro de predição se baseia em semelhança morfológica entre os agentes observados, já que não é de se esperar que um escorpião se comporte como um cachorro, mesmo nunca tendo ouvido falar de escorpiões. Existem várias técnicas para extração de características de imagens [Guyon, 2006], sendo este o método mais utilizado para detecção da semelhança morfológica entre os objetos observados. Estas técnicas podem ser utilizadas em conjunto com uma alteração da arquitetura proposta por Bongard.

Bongard [Bongard et al., 2006] propõe uma arquitetura no qual um robô é capaz de estimar seu formato corporal, partindo de alguns formatos básicos como cilindros. O algoritmo se baseia no conceito de exploração e explotação, no qual se utiliza intensamente a informação previamente conhecida para gerar modelos internos e estes modelos são utilizados para descobrir qual a melhor saída a ser gerada para obter mais informações. Esta arquitetura trabalha com coevolução de modelos para o corpo e saídas interessantes, ambos evoluídos por algoritmos genéticos. Baseado em sensoriamento anterior para saídas determinadas, o robô evolui um conjunto de modelos possíveis para explicar tal comportamento, sendo estes modelos simulados internamente. Em seguida, visando reduzir ao máximo o número de modelos, o robô evolui saídas possíveis para os atuadores, que são apresentadas aos modelos para determinar qual comportamento elas trariam. Aquela saída que gerar maior discrepância entre os modelos imaginados é fornecida aos atuadores e tem seu resultado mensurado por sensores, reiniciando o ciclo. Tal ciclo é apresentado na figura 7 com o robô que utiliza este ciclo para detectar danos em seu

corpo.

Esta arquitetura de Bongard pode ser modificada para identificar quais entradas, ao invés das saídas, são mais interessantes para distinção entre os modelos. Uma vez identificada tal entrada, busca-se um agente da classe observada que possui entradas mais semelhantes à desejada, observando seu comportamento. Esta etapa se equivaleria ao metapreditor da arquitetura de Kaplan, servido de fonte de escolha da situação que reduzirá o erro de predição. Se estamos tentando determinar se um cachorro é capaz de ver, então é mais interessante observar algum que esteja em um ambiente rico com informações visuais, como dentro de uma casa, do que um que esteja em um ambiente pobre, como um campo aberto. Estas observações são utilizadas para elaborar um modelo mais próximo possível dos sensores do agente observado. Ao se observar o movimento por ele gerado, também podemos identificar quais os atuadores e a morfologia do corpo do agente. Uma vez identificada a maneira com que o agente interage com o mundo, o robô é capaz de dizer quais entradas sensoriais estão disponíveis, quais atuações ele pode gerar e quais suas consequências, podendo portanto se colocar na posição do agente observado e aplicar a teoria da simulação para prever o estado mental do agente, como explicado anteriormente. Além disso, a mesma capacidade de comparar proximidade de modelos para um dado agente observado pode ser utilizada para comparar modelos entre agentes diferentes, permitindo que o conhecimento adquirido seja expandido para agentes não observados anteriormente, mas que se assemelham a agentes conhecidos.

Bongard mostra ainda que a mesma arquitetura é capaz de estimar falhas no robô [Bongard and Lipson, 2004]. A arquitetura pode então ser modificada para aumentar sua robustez de tal maneira que falhas particulares nos agentes observados não apareçam como erro de predição para a classe como um todo, o que prejudicaria o desempenho para outras classes de agentes semelhantes também. Como exemplo, imagine que um cachorro ande batendo em paredes e objetos. Baseado em observações anteriores de cachorros, o robô é capaz de prever que o cachorro irá desviar da parede ao se aproximar da mesma. Ao observar um erro em sua predição, a arquitetura original modifica o preditor para que ele corresponda ao evento observado. No entanto, tal evento não representa um erro de predição, mas uma falha particular do agente observado. Partindo do modelo da classe do agente, neste caso cachorro, a arquitetura evolui uma série de modelos para justificar o comportamento observado do cachorro, identificando que as entradas visuais não são relevantes. Esta falta de dados sensoriais podem então ser utilizadas para interpretar as ações do cachorro corretamente, sem necessidade de alteração do preditor e, consequentemente, aumento do erro para o caso geral.

Uma terceira característica importante de levantar da arquitetura de Bongard é que a utilização de modelos não está limitada apenas ao robô ou outros agentes, podendo-se criar modelos para objetos no mundo. Estes modelos podem ser utilizados internamente em um tipo de simulador do mundo, permitindo que a criatura gere ações que maximizem o aprendizado, aumentando a velocidade apresentada pela arquitetura de Kaplan. A determinação dos contextos a serem explorados permite que o robô explore-os de maneira a otimizar a minimização do erro, mas a escolha de ações a serem realizadas em um dado contexto para explorar ao máximo essa minimização é melhor do que a escolha de ações aleatórias, como realizadas pelos robôs de Kaplan, que leva cerca de 10 minutos para o robô descobrir aleatoriamente um conjunto de ações que o permita mover um pouco, seguido de uma hora apenas sendo capaz de se locomover em uma direção e virar e três horas para movimentações completa. Este potencial é justificado no estudo de Bongard [Bongard and Lipson, 2004], que mostra que sua arquitetura necessita de, em média, apenas de três iterações com o robô físico contra 3550 iterações de outros algoritmos para recuperação de movimentos após ocorrência de falha, devido ao uso de um simulador interno baseado no modelo de corpo construído. Considerando a discrepante velocidade entre simular o modelo e atuar no mundo real, esta técnica de simulação permite que haja um ganho no desempenho do robô, além de evitar tentativas em excesso, que poderiam danificar o mesmo.

Outro problema fundamental da teoria da simulação em robôs é que este deve ter conhecimento das coisas do mundo e ser capaz de saber os conhecimentos possuídos pelo agente observado, para fazer uma inferência correta sobre o seu estado mental. Se uma pessoa vê uma criança pulando o muro de uma casa, o conhecimento de que sua bola caiu atrás do muro muda o estado mental estimado. No entanto, este conhecimento não basta, uma vez que a criança pode não saber disto e, portanto, apesar deste conhecimento estar disponível para o observador, ele não está disponível para o observado. Este problema é complexo, devido à similaridade das ações realizadas pelo agente em ambos os casos. Gentner [Gentner and Collins, 1981] realizou experimentos sofre inferência sobre a falta de conhecimento, sendo portanto uma metainferência por tratar de inferência baseada no conhecimento sobre o próprio conhecimento. Seus experimentos mostraram que a falta de conhecimento sobre uma afirmativa reduz a chance de o sujeito achar que tal afirmação é verdadeira. Além disso, quanto mais importante a informação e mais especialista for o sujeito, mais correta é a inferência sobre a falta de conhecimento. Uma medida sobre quão especialista é o sujeito é dada pela quantidade de afirmações sobre aquele assunto a pessoa é capaz de recuperar de sua memória.

Este problema ainda não possui solução ou mesmo tentativas de implementação em sistemas artificiais. Uma primeira abordagem seria semelhante à apresentada por Gentner, na qual o robô deve determinar o quão especialista ele é na área apresentada no problema, baseado em seus conhecimentos anteriores. Este grau de especialização pode ser estimado baseando-se na proporção da quantidade de conhecimento na área requisitada em relação ao conhecimento disponível em outras áreas juntamente com o conhecimento de outras vezes em que foi exposto a problemas semelhantes no passado e qual a evolução de seu erro em situações novas na área requisitada,

ou seja, o quão errado esteve anteriormente quando defrontado com situações novas na mesma área. Uma vez estimado o quão especialista o robô pode se considerar, ele deve julgar qual a probabilidade de alguma afirmação sobre o conhecimento do agente observado ser verdade, baseado no seu conhecimento disponível sobre a área e quão especialista ele é. Portanto, esta implementação seria extremamente dependente do modelo de memória utilizado, devendo este ser capaz de recuperar memórias semelhantes à situação apresentada. Para desenvolvimento de arquiteturas melhores e mais confiáveis, tanto arquiteturas de memória quanto outros experimentos sobre inferência na ausência de conhecimento devem ser mais profundamente estudados, assim como determinação de senso comum, visando criar uma arquitetura para dar cobertura a esta falha na teoria da simulação.

IV. Conclusão

Neste trabalho, discutimos a necessidade de se utilizar a teoria da mente para desenvolvimento de robôs interativos e capazes de aprender com outros seres inteligentes. A teoria da simulação, que permite implementação prática, foi analisada, apresentando usos feitos em ambiente de simulação e robôs que aprenderam a realizar movimentos sob orientação de um tutor.

Em seguida, analisamos as principais falhas da aplicação de tal teoria atualmente. A primeira diz respeito ao mapeamento de entradas e saídas do agente observado pelo observador, cuja solução seria o mapeamento através de *affordances* para identificar as características dos objetos e os resultados das ações ao invés de haver um mapeamento direto. Esta ideia porem necessita do conhecimento das *affordances* envolvidas, que devem ser aprendidas. Para tal, uma mistura de aprendizado por curiosidade com uma arquitetura evolutiva de modelos pode ser utilizada, tendo potencial não apenas de aprendizado, mas de generalização e de identificar casos que fogem à regra.

O segundo problema se refere ao conhecimento utilizado para realizar a inferência dos estados mentais, que deve ser o conhecimento do observado e não do observador para que a inferência esteja correta. Apesar de não haver estudos para construção de sistemas artificiais com esta característica, apresentamos uma primeira abordagem baseada em estudos psicológicos que analisam a realização de inferência sobre o conhecimento.

Utilizando as metodologias aqui descritas, acreditamos que será possível construir agentes mais inteligentes e com capacidades sociais ampliadas. Em trabalhos futuros, testaremos estes princípios em sistemas artificiais, validando a hipótese de que as abordagens funcionam.

Referências

M. Anderson. Embodied Cognition: A field guide. *Artificial Intelligence*, 149(1):91–130, September 2003. ISSN 00043702. doi: 10.1016/S0004-3702(03)00054-7.

J. Bongard, V. Zykov, and H. Lipson. Resilient machines through continuous self-modeling. *Science*, 314(5802): 1118, 2006.

J.C. Bongard and H. Lipson. Automated damage diagnosis and recovery for remote robotics. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 4, pages 3545–3550. IEEE, 2004.

R.A. Brooks. Elephants don't play chess. *Robotics and autonomous systems*, 6(1-2):3–15, 1990.

R.A. Brooks. Human level cognition in embodied robots. *Neural Networks*, (Berkeley 1949):1079–1084, 1993.

D. Buchsbaum, B. Blumberg, C. Breazeal, and A.N. Meltzoff. A simulation-theory inspired social learning system for interactive characters. *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 85–90, 2005. doi: 10.1109/ROMAN.2005.1513761.

P. Carruthers and P.K. Smith. *Theories of theories of mind*. Cambridge University Press, 1996. ISBN 9780521559164.

D.C. Dennett. *The intentional stance*. Bradford Books. MIT Press, 1989. ISBN 9780262540537.

J. Dôkic and J. Proust. *Simulation and knowledge of action*. Advances in consciousness research. John Benjamins Pub., 2002. ISBN 9789027251701.

D. Gentner and A. Collins. Studies of inference from lack of knowledge. *Memory & cognition*, 9(4):434–43, July 1981. ISSN 0090-502X.

J. J. Gibson. *The Theory of Affordances*. Lawrence Erlbaum, 1977.

I. Guyon. *Feature extraction: foundations and applications*. Studies in fuzziness and soft computing. Springer-Verlag, 2006. ISBN 9783540354871.

F. Kaplan and P.Y. Oudeyer. Curiosity-driven development. In *Proceedings of the International Workshop on Synergistic Intelligence Dynamics*, pages 1–8, 2006.

H. Kozima. A robot that learns to communicate with human caregivers. *Proceedings of the First International Workshop on*, 2001.

Y. Kuniyoshi, Y. Yorozu, Y. Ohmura, and K. Terada. From humanoid embodiment to theory of mind. *Embodied artificial*, pages 202–218, 2004.

M. Lungarella and G. Metta. Beyond Gazing, Pointing, and Reaching. *Citeseer*, 2002.

D. Premack, G. Woodruff, and Others. Does the chimpanzee have a theory of mind. *Behavioral and Brain sciences*, 1 (4):515–526, 1978.

E. Sahin, M. Cakmak, M. R. Dogar, E. Ugur, and G. Ucoluk. To Afford or Not to Afford: A New Formalization of Affordances Toward Affordance-Based Robot Control. *Adaptive Behavior*, 15(4):447–472, December 2007. ISSN 1059-7123. doi: 10.1177/1059712307084689.

A.M. Turing and B.J. Copeland. *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Clarendon Press, 2004. ISBN 9780198250807.

J. Weng. A theory for mentally developing robots. *Proceedings 2nd International Conference on Development and Learning. ICDL 2002*, pages 131–140, 2002. doi: 10.1109/DEVLRN.2002.1011821.