



**O futuro da (pesquisa em)
inteligência artificial:
algumas direções**

Fabio G. Cozman

resumo

A área de inteligência artificial, ocupada com a criação de dispositivos que apresentem comportamento identificado como inteligente, teve explosivo crescimento nos últimos anos. Este artigo procura classificar as várias subáreas da inteligência artificial e sugerir algumas linhas de pesquisa que devem levar a resultados promissores nos próximos anos.

Palavras-chave: pesquisa em inteligência artificial; inteligência artificial e sociedade; aprendizado de máquina.

abstract

The field of artificial intelligence, occupied with the creation of devices that are considered intelligent, has had explosive growth in recent years. This article seeks to classify the various subareas of artificial intelligence and suggest some lines of research that should lead to promising results in the coming years.

Keywords: *research in artificial intelligence; artificial intelligence and society; machine learning.*

Embara o ser humano tenha um interesse antigo em artefatos artificiais com inteligência, consciência, sagacidade, humor – e muitos outros aspectos normalmente ligados à cognição de alto nível –, apenas na década de 50 esse interesse se tornou um programa científico concreto (Turing, 1950). Durante cerca de 50 anos a área de inteligência artificial (IA) apresentou uma evolução marcada por períodos de euforia e depressão. Por exemplo, a década de 60 viu grandes promessas sobre a iminente programação de mestres de xadrez e especialistas em matemática, enquanto a década de 70 trouxe críticas aos resultados obtidos na área, um período frequentemente referido como o “inverno da IA” (Nilsson, 2009). Os últimos dez anos, por outro lado, testemunharam um crescimento explosivo das inteligências artificiais, em muitos aspectos até mais rápido do que pesquisadores da área previam na virada do século. Hoje podemos observar artefatos

artificiais que controlam investimentos, que interagem com clientes fornecendo informações e sugerindo condutas, que detectam padrões em exames, que dirigem veículos. Esse sucesso está sendo seguido com atenção pela imprensa, por governos e pela sociedade em geral.

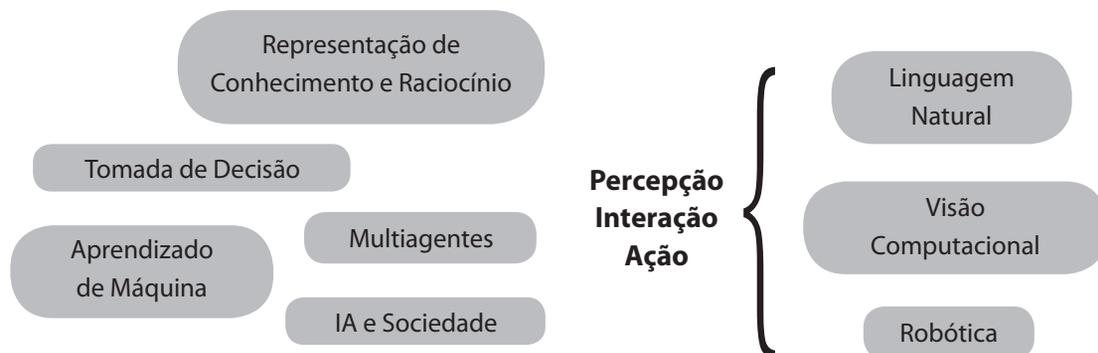
Quais são as perspectivas futuras para inteligências artificiais? Não é fácil prever como essa tecnologia irá evoluir a longo prazo; porém podemos examinar quais tendências são promissoras para a pesquisa em IA nos próximos anos. Os artefatos que emergirem dessa pesquisa contribuirão para a real evolução da área através da ação de empreendedores que conectem a fronteira de pesquisa com a prática, bem como através do debate social que levará a formas de regulação e difusão.

O objetivo deste artigo é examinar algumas direções de pesquisa que podem levar

FABIO G. COZMAN é professor titular do Departamento de Engenharia Mecatrônica da Escola Politécnica da Universidade de São Paulo.

FIGURA 1

Subáreas de pesquisa em inteligência artificial



a área de IA a um novo patamar. Com certeza outras direções levarão a grandes conquistas; aqui o objetivo não é exaurir a lista de todas as ideias com significativo potencial, nem identificar as ideias mais revolucionárias em gestação. O objetivo é listar algumas direções que já estão sendo trilhadas e que quase com certeza gerarão ganhos expressivos em desempenho. Para isso, o artigo apresenta um panorama abstrato da área de IA, classificando seus vários temas e apresentando um conjunto de linhas de pesquisa.

ENTENDENDO A ÁREA DE INTELIGÊNCIA ARTIFICIAL

Embora seja difícil traçar as fronteiras exatas do que é a pesquisa em IA, é importante começar esta discussão com algum esforço de classificação das subáreas e dos temas centrais na IA, para permitir um claro entendimento de quais técnicas são de fato essenciais¹. A Figura 1 apresenta uma classificação aproximada de subáreas, explicada a seguir.

Um dos temas mais antigos em IA é a necessidade de representar conhecimento. Uma inteligência artificial pode ser encarregada de avaliar resultados de um experimento biológico: esta inteligência precisa saber quais são os tipos de resultados possíveis e quais são os possíveis organismos envolvidos e suas características. Como estas informações são bastante especializadas, dizemos que constituem um “conhecimento de especialista”. Outras situações requerem a representação do “senso comum”: por exemplo, se uma pessoa informa a outra que uma medalha não coube em uma mala pois esta era pequena, a pessoa que recebe esta informação em geral infere que a mala era pequena (e não a medalha), levando em conta o conhecimento sobre como medalhas e malas usualmente se apresentam. Além de representar conhecimento, é preciso

¹ Baseado em listas de tópicos anunciadas por conferências de impacto na área de IA, como a International Joint Conf. on AI (IJCAI) 2019 e a AAAI Conf. on Artificial Intelligence 2019, em conjunto com o conteúdo do livro de referência mais popular da área, de autoria de Russell e Norvig (2009).

raciocinar sobre o que é sabido e sobre o que é observado – por exemplo, para aplicar regras em um jogo de damas ou para entender os parceiros em uma negociação.

O estudo do conhecimento e de sua representação é um tema antigo na filosofia; em particular, é o foco da epistemologia (Brachman & Levesque, 2004). É importante notar que o próprio termo “conhecimento” tem sido objeto de debates há séculos. O que é conhecimento? A resposta mais aceita em filosofia toma conhecimento como crença verdadeira e justificada. Na área de IA o conceito de conhecimento é bem mais elástico. Frequentemente a representação de conhecimento em IA se ocupa de representar crenças de um particular agente, sem necessariamente se preocupar com a “verdade”. E, de forma geral, representação de conhecimento em IA se ocupa não apenas de crenças que tenham caráter determinístico (por exemplo: se o rei for tomado em uma partida de xadrez, o seu jogador perdeu), mas também crenças que envolvam incerteza (por exemplo: o aumento no desemprego provavelmente diminuirá a taxa de inflação).

Ao longo do tempo muitos diferentes formalismos têm sido usados para representar conhecimento em IA. Pode-se imaginar cada um desses formalismos como uma linguagem que permite expressar alguns aspectos do mundo. A maior parte dos formalismos para representação *sem incerteza* usados em IA é baseada em linguagens formais ligadas a lógicas – por exemplo, temos a lógica proposicional, a lógica de predicados, as lógicas de descrição (Baader et al., 2003). Por outro lado, a maior parte dos formalismos para representação *com incerteza* se baseia na teoria de probabili-

dades; linguagens têm sido propostas para adaptar a vetusta teoria de probabilidades a situações que exigem recursos lógicos (Getoor & Taskar, 2007).

Uma outra subárea importante da pesquisa em IA é aquela que lida com tomada de decisão, uma atividade corriqueira com um enorme número de variantes. Tomar uma decisão envolve estabelecer metas e listar alternativas. Envolve também procurar uma solução dentre um conjunto de possíveis opções. Decisões também são tomadas para planejar uma sequência de ações com objetivo de atingir uma meta. Note-se que boa parte dessas variantes de tomada de decisão é estudada por economistas e administradores há décadas, enquanto aspectos humanos da tomada de decisão são objeto de estudo por psicólogos. Na área de IA usa-se o termo *busca* para lidar com o processo de encontrar uma alternativa vencedora em meio a um grande número de alternativas. Outro termo comum em IA é *planejamento*, em que o foco está em encontrar uma sequência de ações que leve a uma meta predefinida. Aqui se tocam aspectos de tomada de decisão e representação de conhecimento, pois para resolver um problema de planejamento é preciso primeiro representá-lo. Por exemplo, um problema clássico em IA consiste em empilhar blocos que estão em uma mesa para obter uma pilha com características desejadas. É preciso descrever os blocos: são diferenciados por cor? Por tamanho? É também preciso descrever a meta: blocos vermelhos embaixo, amarelos em cima? Ou alguma configuração especial? Após a descrição do problema, é preciso procurar uma sequência de ações no (possivelmente enorme) espaço de soluções.

Quando um número grande de agentes interage na solução de problemas, temos um *sistema multiagentes*. O estudo de tais sistemas é uma subárea importante da IA, preocupada não só com decisões, mas também com negociações e interações “sociais” entre agentes.

Uma outra subárea fundamental da IA é aquela que se ocupa de aprender com experiências – ou seja, realizar o *aprendizado de máquina*. A inteligência humana é eminentemente ligada a processos de aprendizagem: ao nascer sabemos pouco do mundo, e ao interagir com a natureza e com a sociedade ao nosso redor adquirimos nossa capacidade de agir com inteligência. Essa interação ocorre de muitas maneiras. Frequentemente a interação se dá por uma experiência direta: ao queimar a mão, uma criança aprende que não deve encostar em uma panela com água fervendo. Em outras circunstâncias a interação se dá por meio de um guia, que pode ser uma mãe ou um amigo ou um professor. Através de ensinamentos diretos aprendemos como cozinhar, como realizar operações matemáticas, como dirigir veículos. Finalmente, podemos também aprender observando dados coletados a partir de sensores ou pesquisas de opinião. Por exemplo, ao notar que a maioria das pessoas que possuem mais de um veículo também possuem mais de uma televisão, aprendemos que existe uma relação entre a posse de veículos e televisões.

Note que a expressão *machine learning* é usada na língua inglesa para se referir à subárea da IA descrita no parágrafo anterior. Essa expressão pode ser traduzida como “aprendizado de máquina” ou como “aprendizagem de máquina”. As duas tra-

duções podem ser encontradas na literatura especializada. Neste artigo usamos a primeira expressão, por uma razão simples: embora o sucesso de *machine learning* seja de fato impressionante, ainda não ocorre um processo de aprendizagem de computadores. O que temos no momento é o aprendizado de particulares tarefas, em geral a partir da análise de grandes bases de dados. Talvez cheguemos a um ponto em que discutiremos a didática para aprendizagem de máquinas adolescentes (!), mas sem dúvida não estamos perto disso.

Há cerca de 30 anos, a pesquisa em aprendizado de máquina caminhava lentamente. Havia interesse em muitos diferentes processos de aprendizado: aprender a partir de um professor explicando um tópico, aprender a partir de uma sequência de ações físicas, aprender a partir de uma tabela contendo dados coletados por um sensor. Havia também uma diversidade de técnicas, algumas com base estatística, outras inspiradas pela biologia do cérebro. Desde então houve uma grande mudança em toda a área de IA, causada pela abundância cada vez maior de dados, pelo aumento da capacidade computacional e pelo desenvolvimento de melhores algoritmos de processamento de dados: passamos a depender fundamentalmente da extração de padrões de grandes bases de dados. Considere, por exemplo, a tradução automática de textos, da língua portuguesa para a língua inglesa. Parece razoável supor que um tradutor humano leia o texto original em português e monte algum modelo mental do assunto tratado, de forma a escolher as melhores palavras no texto a ser produzido. Esse tipo de estratégia era perseguido vigorosamente até cerca de 1995:

tentava-se fazer uma análise sintática do texto de entrada, seguida de uma análise semântica sobre seu significado, e então uma tradução apropriada. Essa estratégia não funciona tão bem quanto a que emergiu na década de 90 e que tem sido cada vez mais refinada: coletamos um imenso conjunto de textos traduzidos do português para o inglês e construímos uma função que emite palavras em inglês a partir de um conjunto de palavras em português. Essa solução *não* procura “entender” o texto de entrada; procura apenas obedecer a padrões estatísticos da relação entre as línguas de interesse. A vitória desse tipo de estratégia é, de certa forma, surpreendente: quão inteligente pode-se parecer sem uma compreensão semântica das tarefas! Nas palavras de Halevy, Norvig e Pereira (2009): o processamento de grandes quantidades de dados tem uma efetividade que não parece nem razoável.

No momento existe uma considerável superposição entre aprendizado de máquina, estatística e ciência de dados, além de uma significativa superposição com a área de mineração de dados e de *big data*. Todos esses campos lidam com massas de dados a partir das quais extraem padrões, regras, fatos, possibilidades, palpites, expectativas. Em todos eles hoje predominam métodos estatísticos e técnicas inspiradas no cérebro, como as celebradas *redes neurais*, formadas por conjuntos de operações que simulam “neurônios” conectados em camadas. Cada um desses neurônios artificiais executa uma tarefa simples: por exemplo, somando suas entradas e emitindo uma função matemática da soma. Uma rede neural com muitas camadas de neurônios é chamada uma

rede neural profunda; o estudo de métodos de aprendizado de redes neurais profundas é o tema de *aprendizado profundo* (Goodfellow, Bengio & Courville, 2016). O interesse em aprendizado profundo se intensificou a partir de 2012, ano em que redes neurais profundas mostraram excelente desempenho em reconhecer objetos em uma competição internacional – a competição de reconhecimento de objetos em imagens no conjunto de imagens denominado ImageNet. Enquanto em 2011 a taxa de acerto de objetos em imagens era menor que 75%, hoje essa taxa está próxima de 97%. Esse resultado se torna mais impressionante quando sabemos que o desempenho humano está por volta de 95% de acerto; ou seja, redes neurais profundas atingem desempenho super-humano em classificação de imagens.

Há cerca de 20 anos o interesse na relação entre IA e sociedade basicamente se limitava a especulações sobre o futuro, frequentemente imersas em ficção científica, ocasionalmente fundamentada no debate social e filosófico. O sucesso pragmático atual da tecnologia de IA trouxe para o centro do debate a relação entre IA e sociedade, contrapondo visões otimistas e pessimistas (Cozman, 2018). Por um lado, existe uma visão utópica sobre a IA: a esperança que a tecnologia melhore a produtividade do ser humano, em particular tomando conta de tarefas enfadonhas e liberando-o para atividades mais recompensadoras e para o lazer. Espera-se que essa utopia permita aos seres humanos terem meios para garantir uma melhor distribuição das riquezas geradas por maior produtividade. Por outro lado, a visão distópica da IA se preocupa com a possibilidade de máquinas

violarem a privacidade de cidadãos, com perdas causadas por decisões automáticas incorretas, com os impactos no mercado de trabalho e nas relações humanas. Para além da reflexão sobre o futuro da sociedade, existe também a análise concreta sobre as relações sociais e de trabalho na presença de inteligências artificiais.

Finalmente, existem várias outras tecnologias ligadas à IA que se destinam a permitir que máquinas consigam perceber o que ocorre em seu ambiente; a interagir com esse ambiente e agir de forma a modificá-lo. Assim temos a tecnologia de processamento de linguagem natural, que se ocupa de interpretar e gerar sequências de símbolos produzidos por seres humanos em suas variadas línguas. Temos também a visão computacional, que se ocupa de compreender imagens, e a robótica, que envolve máquinas com capacidade de interagir com o mundo físico. Neste breve artigo temos menor preocupação com esses temas, não por formarem campos sem importância – ao contrário, por serem campos com tal dimensão e complexidade que requerem uma análise em separado.

PENSANDO NO FUTURO DA PESQUISA EM IA

As inteligências artificiais provavelmente se tornarão mais e mais capazes de cuidar das tarefas cotidianas e de aprender com sua experiência; prever aonde isso nos levará em um grande espaço de tempo é quase impossível. Considere um objetivo mais modesto: quais são alguns temas de pesquisa promissores na área de inteligência artificial? Vamos abordar essa questão no restante

deste artigo, apontando três caminhos que parecem particularmente auspiciosos.

A pesquisa na área certamente deve focar melhores técnicas de aprendizado de máquina. Há várias frentes em que o avanço é importante. Em primeiro lugar, é preciso aumentar a capacidade de armazenar e processar dados, já que o ritmo de produção de dados da humanidade aumenta sem cessar. Em segundo lugar, as técnicas de aprendizado de máquina devem se tornar mais robustas: pequenas variações nas entradas não podem causar desastres (como se observa às vezes com redes neurais profundas). A robustez também deve ser perseguida em outra dimensão: algoritmos devem ser de fácil entendimento e aplicação, para que sejam usados de forma correta e livre de comportamento discriminatório. É preciso também ter melhor entendimento sobre redes neurais profundas: como funcionam, como falham. Além disso, resultados obtidos via aprendizado de máquina devem ser interpretáveis: o usuário deve entender as razões que levam a particulares decisões, de forma a confiar na máquina. Todas essas frentes de trabalho devem receber atenção no futuro.

Em segundo lugar, o simples avanço da tecnologia de aprendizado de máquina não vai levar a IA a um novo nível de capacidade. Uma previsão razoável é: o próximo nível de operação de inteligências artificiais só será atingido através da integração de aprendizado de máquina, tomada de decisão e representação de conhecimento. Por exemplo, um sistema de planejamento deve ser capaz de aprender a partir de experiências e também de instruções textuais; um sistema de respostas deve ser capaz de levar em conta conhecimento acu-

mulado em séculos de evolução científica. Aprender a partir de mais e mais dados não substitui de forma eficiente todo o conhecimento acumulado pela humanidade.

Em terceiro lugar, o estudo da relação entre IA e sociedade merece atenção crescente. Preocupações relacionadas à proteção de privacidade e proteção contra decisões automáticas incorretas devem ser levadas para debate amplo, para que a legislação seja adaptada de acordo. Angústias ligadas ao mercado de trabalho também devem ser discutidas: qual a melhor forma de treinar pessoas? Como retrainar cidadãos se movimentando dentro desse novo mercado de trabalho? Que tipo de educação deve ser oferecido aos cidadãos? Em alguns casos será preciso regulamentar setores para que trabalhadores e máquinas possam prosperar em harmonia. Aqui existe um balanço delicado a ser perseguido: um controle exagerado sobre a IA pode impedir grandes avanços de qualidade de vida que esta tecnologia pode nos proporcionar. Devemos ter abertura para o potencial de ganhos em diagnósticos médicos, assistências jurídicas, processos industriais, reconhecendo que toda tecnologia nova tem riscos e benefícios – os primeiros devem ser minimizados e os últimos devem ser maximizados.

Devemos também saudar a oportunidade, como país, de dar um salto tecnológico investindo em uma área promissora e de grande impacto – o desenvolvimento da tecnologia de IA basicamente exige investimentos em formação humana, algo que precisamos fazer de qualquer maneira. Impedir a tecnologia de prosperar por receio de possíveis contratempos fará apenas com que outros países obtenham um diferencial nesta tecnologia – que precisaremos então comprar e adaptar às nossas necessidades.

É importante reconhecer que o Brasil tem significativas vantagens no panorama internacional de IA. O país tem grande população, que fala uma única língua e que compartilha sistemas governamentais amplos (por exemplo, de saúde). Essas características fazem com que grandes massas de dados estejam disponíveis para a criação de soluções para nossos problemas. Além disso, à medida que máquinas substituem tarefas repetitivas, as tarefas que envolvem comunicação e empatia se tornam mais valiosas – a chamada “economia do sentimento” (*feeling economy*) deve ter crescente impacto na riqueza das nações (Huang, Rust & Maksimov, 2019). Nada mais apropriado do que liberar a natural empatia da população brasileira para explorar e liderar essa nova economia.

BIBLIOGRAFIA

- BAADER, Franz et al. *An Introduction to Description Logic*. Cambridge, Cambridge University Press, 2017.
- BRACHMAN, Ronald J.; LEVESQUE, Hector J. *Knowledge Representation and Learning*. Burlington, Morgan Kaufmann, 2004.
- COZMAN, Fabio G. "Inteligência artificial: uma utopia, uma distopia", in *Revista Digital de Tecnologias Cognitivas*, n. 17. São Paulo, TIDD-PUC-SP, 2018, pp. 32-43.
- GETOOR, Lise; TASKAR, Ben. *Introduction to Statistical Relational Learning*. Cambridge, MIT Press, 2007.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. Cambridge, MIT Press, 2016.
- HALEVY, Alon; NORVIG, Peter; PEREIRA, Fernando. "The unreasonable effectiveness of data", in *IEEE Intelligent Systems*, IEEE, 2009.
- HUANG, Ming-Hui; RUST, Roland; MAKSIMOVIC, Vojislav. "The feeling economy: managing the next generation of artificial intelligence (AI)", in *California Management Review*, vol. 61(14), 2019, pp. 43-65.
- NILSSON, Nils J. *The Quest for Artificial Intelligence*. Cambridge, Cambridge University Press, 2009.
- RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: a modern approach*. Nova Jersey, Prentice Hall, 2009.
- TURING, Alan M. "Computing machinery and intelligence", in *Mind*, vol. 49, 1950, pp. 433-60.