

Engenharia Reversa no Cérebro

No MIT, a neurociência e a inteligência artificial começam a se misturar

Reverse-Engineering the Brain

At MIT, neuroscience and artificial intelligence are beginning to intersect

Fred Hapgood

Em http://www.technologyreview.com/printer_friendly_article.aspx?id=17111

Tradução: Pedro Lourenço Gomes

“Maggie é uma macaca muito esperta”, diz Tim Buschman, estudante de graduação do laboratório de neurociência do Professor Earl Miller. Maggie não está à vista – ela está em um reservado de segurança biológica cuja finalidade é protegê-la de germes humanos – mas os sinais de sua inteligência fluem através de dois monitores em frente a Buschman. Durante os últimos sete anos, Maggie “trabalhou” para o Department of Brain and Cognitive Sciences (BCS) do MIT. A macaca, por três horas diárias, ocupa-se no computador com jogos que (usualmente) são projetados para exigir dela a geração de representações abstratas e então usar essas abstrações como ferramentas. “Até eu tenho problemas com esse aí”, diz Buschman referindo-se a um jogo que envolve a classificação de operações lógicas. Maggie está à toda, resolvendo problemas com golpes nas teclas, levando mais ou menos meio segundo em cada um e acertando uns quatro em cada cinco deles.

A atividade de Maggie está na interseção entre inteligência artificial (AI – artificial intelligence) e neurociência. Sob a tutela de Buschman e Michelle Machon, também estudante de graduação, ela está contribuindo para a pesquisa sobre como o cérebro aprende e elabora regras lógicas, e como o desempenho dessas tarefas se compara com o de redes neurais artificiais usadas em AI.

Há quarenta anos, a idéia de que a neurociência e a AI poderiam convergir em laboratórios como o de Miller era impensável. Naquela época, as duas disciplinas operavam à distância uma da outra. Enquanto a neurociência se concentrava em descobrir e descrever os detalhes da neuroanatomia e da atividade neural, a AI tentava desenvolver uma via independente e não biológica para a inteligência. (Historicamente, a tecnologia não precisava realmente copiar a natureza tão liberalmente; os aviões não voam como pássaros e os carros não andam como cavalos). E era a AI que parecia progredir com mais rapidez. A neurociência quase nada sabia sobre o que era o cérebro, muito menos sobre como funcionava, ao passo que qualquer um com um pouco de bom senso acreditava que o dia em que os computadores seriam capazes de fazer tudo o que os humanos faziam (e fazer melhor) estava perto. Em 1962, o próprio Presidente Kennedy foi convencido disso, chamando a automação (ou como era chamada na época, “cibernação”) de principal desafio nacional da década de 1960, por causa da ameaça de que deixaria os humanos fora do mercado de trabalho.

Mas alguma coisa tirou o expresso da AI dos trilhos. Apesar dos computadores poderem ser levados a manusear objetos simples em ambiente controlado, fracassavam

vergonhosamente no reconhecimento de objetos complexos do mundo natural. Um microfone podia distinguir níveis de som, mas não resumir o que tinha sido dito; um manipulador podia pegar um objeto novo e limpo que estivesse em uma sequência ordenada, mas não um objeto sujo e velho que estivesse numa pilha desordenada. (Nem podia, num exemplo da inspiração de Marvin Minsky, inserir um traveseiro em uma fronha). Hoje em dia nos preocupamos mais com a concorrência humana de outros países do que com a concorrência das máquinas.

Enquanto o progresso da AI foi mais lento do que o esperado, a neurociência ficou muito mais sofisticada em sua compreensão do funcionamento do cérebro. Em nenhum lugar isso fica mais claro do que nos 37 laboratórios do Complexo BSC do MIT. Aqui, os grupos estão mapeando as vias neurais da maioria das funções cognitivas superiores (e seus distúrbios), incluindo aprendizagem, memória, organização de comportamentos sequenciais complexos, formação e acumulação de hábitos, formação de imagens mentais, gerenciamento e controle de números, definição e planejamento de metas, processamento de conceitos e crenças, e a habilidade de compreender o que outras pessoas pensam. O impacto potencial dessa pesquisa parece ser enorme. A descoberta de como o cérebro funciona – *exatamente* como funciona, como sabemos como um motor funciona – reescreveria quase todos os textos das bibliotecas. Só para começar, revolucionaria a justiça criminal, a educação, o marketing, as tarefas dos pais e o tratamento dos distúrbios mentais de todo tipo (Earl Miller espera que as pesquisas feitas em seu laboratório ajudem a desenvolver terapias para os distúrbios de aprendizagem).

Esse progresso é uma das razões para que a antiga e clara diferença entre neurociência e AI esteja começando a desaparecer no MIT – e não só no laboratório de Miller. As pesquisas correntes sobre visão do Instituto também ilustram como as duas disciplinas estão começando a colaborar. “Os campos cresceram separadamente”, diz James DiCarlo, professor assistente de neurociência, “mas não vão estar separados por muito mais tempo”. Hoje em dia os pesquisadores de AI seguem o avanço da neurociência com grande interesse, e a idéia de fazer engenharia reversa no cérebro já não é tão implausível como parecia antes.

Entendendo o Reconhecimento de Objetos

A maior parte do trabalho do laboratório de DiCarlo trata do reconhecimento de objetos, que é o que nos permite identificar um objeto (como uma vaca) em muitas apresentações diferentes (vacas à distância, vacas vistas de cima, vacas ao nascer do sol, uma vaca em um caminho) sem confundí-lo com objetos similares (digamos, um cavalo). DiCarlo e o aluno de graduação David Cox publicaram em agosto último na revista *Nature Neuroscience* uma pesquisa que focalizava uma das questões básicas do reconhecimento de objetos: o quanto de nosso sucesso no reconhecimento de objetos depende de circuitos inatos neurointegrados (NT – *hardwired*) e quanto de habilidades aprendidas?

DiCarlo e Cox conduziram cada um de seus experimentos com 12 pessoas, uma de cada vez. Os sujeitos se sentavam em frente de um equipamento que podia apresentar imagens de objetos e ao mesmo tempo distinguir a direção do olhar do sujeito. Os objetos eram gerados por computador e se pareciam vagamente com animais antropomorfizados, mas eram projetados para serem estranhos para os sujeitos. Um objeto aparecia em uma de três posições na tela, e o sujeito naturalmente dirigia seu olhar para ele. Certos objetos, entretanto, eram substituídos pelos pesquisadores por novos objetos enquanto os sujeitos estavam movendo os olhos. Por exemplo, digamos que um objeto de aparência atarracada, com olhos vivos, fosse apresentado à direita da tela quando o sujeito estivesse focalizando o centro. Quando a mirada do sujeito mudasse para o objeto atarracado e de olhar vivo, os

pesquisadores o substituiriam por um objeto que parecesse mais magro e com olhos baixos. Como os humanos são efetivamente cegos durante mudanças de mirada, os sujeitos não notariam a mudança. Mas seus cérebros sim.

Após um hora ou duas de exposição a diferentes objetos, alguns dos quais trocados de modo consistente quando aparecessem em posições particulares, eram apresentados aos sujeitos pares de objetos em diferentes posições na tela e era pedido a eles que os comparassem. Poderia ser esperado que os sujeitos distinguíssem os objetos sem muita dificuldade. E assim ocorreu, exceto quando os objetos tinham sido trocados – e agora reapareciam nas mesmas posições onde ocorreram as mudanças. Os sujeitos tendiam a confundir esses objetos: isto é, era mais provável que o atarracado e de olhar vivo em uma posição e o magro de olhos tristes em outra posição lhes pareciam ser o mesmo objeto. DiCarlo acha que tais erros mostram que os mecanismos cerebrais para reconhecimento do mesmo objeto em lugares diferentes dependem de uma experiência visual normal ao longo do espaço-tempo. “O achado sugere que mesmo propriedades fundamentais do reconhecimento de objetos pode se desenvolver através da experiência visual com nosso mundo”, diz ele. DiCarlo e sua equipe estão conduzindo experimentos similares com animais para examinar os padrões de atividade neuronal que são básicos no reconhecimento de objetos. (Um bom exemplo dessa pesquisa foi publicado no exemplar de 4 de novembro de 2005 da revista *Science*. DiCarlo e três colaboradores registraram e analisaram a atividade de centenas de neurônios de cérebro de macacos. Conseguiram demonstrar que informações altamente confiáveis sobre identidade e categoria de objetos estavam contidas em punhados de neurônios semelhantes).

O reconhecimento de objetos tem sido uma das principais metas, e um dos principais desapontamentos, da AI tradicional. Ao mesmo tempo em que a visão por máquinas é uma indústria de verdade, seu sucesso tem se dado em aplicações estreitamente definidas em condições altamente controladas como decodificar placas de automóveis, identificar impressões digitais, reconhecer caracteres escritos e inspecionar produtos (por exemplo, identificar uma batata frita queimada para que ela possa ser retirada da linha de produção). Cada sistema de visão por máquinas “vê” apenas um tipo específico de objeto; por exemplo, a máquina que lê placas de automóveis não seria capaz de identificar impressões digitais, e vice-versa. Ainda que a tecnologia atual possa ser boa o suficiente para nos dar máquinas que reconhecem qualquer coisa individual, a maior parte das tarefas na maioria das indústrias – montagem, manutenção, serviços de saúde, transporte, segurança – requer mais versatilidade do que isso. Os trabalhadores precisam ser capazes de reconhecer um martelo, uma chave de parafusos e uma chave inglesa, a despeito das diferenças de iluminação, da orientação do objeto e do agrupamento circundante. O fracasso em construir máquinas que possam fazer isso é especialmente frustrante dado que aves como corvos, e pequenos mamíferos como ratos, exibem rotineiramente um nível de habilidade em reconhecimento geral bem superior ao da tecnologia corrente. Existe algo na incapacidade de construir máquinas tão espertas como nós que consola nossa vaidade, mas não ser capaz de fazer uma máquina tão esperta quanto um pombo é apenas embaraçoso.

Assim, por diversos anos os pesquisadores em AI trabalharam no problema de associar padrões visuais a significados ou identidades. Esta é uma das áreas onde a AI e a neurociência têm se misturado uma à outra: a neurociência tem estudado o papel do cérebro no reconhecimento de objetos, e a AI na lógica geral do que qualquer sistema teria que fazer para resolver o mesmo problema. Depois de décadas, elas já estão ao alcance da voz uma da outra. DiCarlo imagina se já não será a hora de batizar-se uma nova disciplina que se valha dos dois campos, como “visão de máquina inspirada biologicamente”.

Nenhuma universidade se aproxima dessa interseção mais depressa do que o MIT, onde a colaboração entre engenharia e ciência é uma missão institucional. E isto, diz DiCarlo, é uma das razões porque ele veio para o MIT: ele espera que a revolução aconteça aqui.

Modelando o Reconhecimento Imediato

Uma ilustração notável do argumento de DiCarlo pode ser encontrada nos laboratórios de Tomaso Poggio. Co-diretor do Center for Biological and Computational Learning do MIT, Poggio vem estudando a visão por quatro décadas, primeiro no Max Planck Institute de Tübingen, na Alemanha, depois no laboratório de AI do MIT (que se tornou o Computer Science and Artificial Intelligence Lab), e agora no Department of Brain and Cognitive Sciences. (Poggio colaborou com DiCarlo no experimento com macacos descrito na revista *Science*). Por quase todo esse tempo, Poggio dirigiu um grupo de pesquisa em neurociência e outro em visão de máquina, e não viu qualquer razão para juntá-los. “Nós sabíamos muito pouco”, diz ele. “Sempre achei que era um erro esperar muito da neurociência”. Entretanto, resultados recentes de um projeto executado pelo aluno de pós-doutorado Thomas Serre e por Aude Oliva, professora assistente de neurociência cognitiva no BCS, o convenceram.

Atualmente o laboratório de Poggio se concentra em um tipo de reconhecimento de objetos chamado reconhecimento imediato. Esse fenômeno foi descrito pela primeira vez em 1969, num estudo da professora Mary Porter, do MIT (agora professora de psicologia no BCS), e de sua assistente de pesquisa Ellen Levy. O reconhecimento imediato é a forma mais rápida de reconhecimento que se conhece. O sujeito de um experimento clássico de reconhecimento imediato senta-se frente a uma tela e lhe pedem para apertar um de dois botões em resposta a cada imagem de um série, na dependência de conter ou não um animal. Para se certificar de que olhar uma imagem não ajuda acidentalmente os sujeitos a aprender a olhar outras, os pesquisadores escolheram imagens muito diferentes: muitas espécies, em muitas poses e perspectivas diferentes, postas contra diversos tipos de fundos. As fotos se sucedem em intervalos de alguns décimos de segundo. No começo de um estudo o sujeito pode nem perceber que lhe foi mostrada uma figura, e muito menos reconhecer o que havia nela. Entretanto, surpreendentemente, as pessoas apertavam as teclas certas com grande frequência. Ficavam melhores – e tomavam consciência do aparecimento de imagens – com a prática. Ainda assim, no início, algo no cérebro é capaz de reconhecer e categorizar objetos antes mesmo do sujeito ter consciência de estar vendo alguma coisa.

O reconhecimento imediato é importante para os pesquisadores porque é o caso mais simples possível de reconhecimento geral de objetos. Acontece rápido demais para envolver o recrutamento de muitos neurônios ou o processamento intensivo de informações ou o envio e recebimento de impulsos por mais do que uma fração de um centímetro. A informação a partir dos movimentos dos olhos, um elemento chave em outros tipos de reconhecimento (como no trabalho de DiCarlo), pode não desempenhar papel algum. Ainda assim, de algum modo as teclas certas são pressionadas (em sua maioria), o que significa que uma forma limitada de reconhecimento de objetos de propósitos gerais deve ser possível com o uso de um número relativamente pequeno de neurônios organizados de maneira relativamente simples.

Elaborando a partir de um trabalho que Poggio fez com Max Riesenhuber, PhD '00, na época um aluno de graduação do MIT e hoje professor da Georgetown University, Serre, Poggio e alguns outros do grupo de Poggio desenvolveram uma teoria sobre a parte do córtex visual primariamente responsável pelo reconhecimento imediato. Sua abordagem do processamento visual era, em muitos casos, diferente da abordagem de um engenheiro de

visão de máquinas. Por exemplo, a maioria dos programas de visão de máquinas têm um processador executando uma série de instruções em ordem consecutiva, uma arquitetura conhecida como “processamento serial”. O cérebro, por outro lado, usa o “processamento paralelo”, uma abordagem na qual um problema é subdividido em muitas partes, cada uma delas apropriada separadamente por seu próprio processador, após o que os resultados são combinados ou integrados para obter um resultado geral único – digamos, a percepção de uma vaca. Em teoria, os engenheiros poderiam usar o processamento paralelo em programas de visão de máquinas (e alguns tentaram), mas na prática raramente é óbvio como subdividir um problema de uma maneira que permita que as peças prontas sejam recombinadas com consistência perfeita.

A visão biológica resolve esse problema de muitas maneiras diferentes. Uma delas, de acordo com o grupo de Poggio, é organizar o processamento em torno de duas operações simples e então alternar essas operações de maneira ordenada através de camadas de neurônios. A camada A pode filtrar os inputs básicos do nervo óptico; a camada B integraria os resultados a partir de muitas células da camada A; a C filtraria os inputs da B; a D integraria os resultados da C; e assim por diante, talvez uma dúzia de vezes. À medida que um sinal avança através das camadas, os outputs dos processadores em paralelo combinam-se gradualmente, a identidade emerge e o ruído vai desaparecendo.

Serre e Poggio usaram essa técnica de camadas para permitir que seu modelo fizesse um processamento paralelo. Outro truque que tomaram emprestado da biologia foi aumentar o número de conexões que ligam suas unidades básicas de comutação. As unidades de comutação dos computadores convencionais têm poucas conexões, quase sempre umas três; os neurônios, as unidades básicas de comutação do cérebro, têm milhares ou mesmo dezenas de milhares. Serre e Poggio dotaram as comutações lógicas de seu modelo com um grau biologicamente plausível de conectividade. Nos casos que a ciência ainda desconhecia, eles fizeram suposições baseadas em sua experiência mais ampla com neuroanatomia.

Para testar sua teoria, Serre e Poggio desenvolveram um programa de computador de reconhecimento imediato que analisa imagens digitais. Quando arquivos de imagens digitais são introduzidos no programa, ele os processa através de múltiplas camadas alternadas de células de filtragem e integração, treinando para identificar e classificar as imagens. “A chave é construir a complexidade lentamente”, diz Serre. “Introduzir a inteligência muito depressa é um grande erro”. Esforços iniciais de AI podem ter tentado se concentrar na identidade cedo demais, deixando de lado informações que eram importantes para se obter a resposta certa.

A abordagem de Serre e Poggio foi um sucesso espetacular. De um ponto de vista neurocientífico, algumas de suas suposições acabaram predizendo características reais, como a presença de células (vamos chamá-las de células OR) que selecionam o sinal mais forte ou mais consistente em um grupo de inputs e o copiam em suas próprias fibras de output. (Imagine um grupo de três neurônios, A, B e C, todos enviando sinais para o neurônio X da OR. Se esses sinais estivessem em níveis de potência 1, 2 e 3 respectivamente, X iria suprimir A e B e copiar o sinal de C para seu output. Se as potências tivessem sido 3, 2 e 1, ele teria, ao invés, copiado o sinal de A e suprimido os de B e C).

Os resultados também foram dramáticos do ponto de vista da AI. Quando sujeitos humanos e o programa de reconhecimento imediato de Serre e Poggio passaram pelo teste de presença/ausência de animal, o computador se saiu tão bem quanto os humanos – e melhor do que o melhor programa de visão de máquinas disponível. (De fato, obteve a resposta correta 82% das vezes, enquanto que os humanos alcançaram apenas 80% em média). Esta

é quase que certamente a primeira vez que um programa de visão *geral* teve um desempenho tão bom quanto o de humanos.

Os promissores resultados fizeram com que Poggio e Serre pensassem para além do reconhecimento imediato. Poggio suspeita que o modelo poderia se aplicar muito bem na percepção auditiva. Serre acrescenta uma especulação ainda mais ousada: o reconhecimento em geral de objetos é uma das fundações da cognição. Talvez seja por isso que dizemos “Estou vendo” quando queremos indicar que entendemos alguma coisa.

Apesar da extensão de sua teoria nestas novas direções demandar mais trabalho, o modelo de Serre e Poggio já começou a se disseminar no MIT entre as comunidades de AI e de neurociência. O aluno de graduação de engenharia elétrica Stan Bileschi terminou recentemente um doutorado no qual aplicou o modelo ao reconhecimento de cenário, que é uma derivação de juízos de ordem superior – “é uma fazenda!” – a partir do reconhecimento de objetos isolados – um celeiro, uma vaca, uma cerca de troncos. Bileschi acredita que a análise do cenário geral será crítica para muitas aplicações da visão de máquinas do mundo real – vigilância, por exemplo.

O reconhecimento imediato é a base de todo reconhecimento visual, diz Poggio, mas ele não se limita a isso. Há muitos níveis de reconhecimento, e o reconhecimento imediato é um dos mais simples. Dependendo do contexto, um objeto pode ser identificado como um brinquedo, uma boneca, a Barbie, um reflexo da cultura americana, uma mulher, uma representação de uma garota com um estranho distúrbio de crescimento e assim por diante, numa extensa lista. Similarmente, quanto a problemas no jogo de xadrez, reconhecer a jogada certa pode levar segundos, minutos ou horas, dependendo da configuração das peças. Presume-se que à medida que os problemas se tornam mais difíceis sua solução exige o recrutamento de níveis superiores da função cerebral – e isso toma tempo.

Um modelo de reconhecimento imediato pode resolver os problemas de visão que impediram o desenvolvimento de robôs úteis em manutenção e construção. Ou pode-se descobrir que para serem realmente úteis esses robôs têm que ser capazes de reconhecer tanto anomalias do cenário como suas causas. Este tipo de reconhecimento, claramente, é de ordem superior.

O próximo passo é construir modelos que recrutem cada vez mais recursos, e desse modo exijam mais tempo de processamento. “Nós sabemos como o modelo poderia ser modificado para incluir o tempo”, diz Serre. “Isto poderia nos aproximar mais do pensamento – talvez”.