Co wiemy o tym, na czym to polega, że coś wiemy?

*Paweł WAWRZYŃSKI**

* Instytut Informatyki, Wydział Elektroniki i Technik Informacyjnych, Politechnika Warszawska

Modelowaniem procesów poznawczych w ludzkich mózgach zajmuje się neuronauka obliczeniowa (*computational neuroscience*). Nie dostarcza ona na razie wiedzy wystarczającej do tego, żeby zbudować sztuczny mechanizm działający jak ludzki umysł. Zbliżamy się do tego dosyć powoli.

Istotą poznania czegokolwiek jest pojawienie się pewnej reprezentacji tego czegoś w naszym mózgu. Ta reprezentacja powinna być na tyle trwała, aby mózg mógł ją w jakiś sposób wykorzystać. W tym sensie nasze poznanie ma granice związane z faktem, że nie poznamy tego, czego nie zrozumiemy, bo wówczas reprezentacja tego w naszym mózgu będzie błędna. Ponadto nie poznamy tego, czego na dłużej nie zapamiętamy, ponieważ wtedy jedynie osiągniemy chwilowe poczucie, że to poznaliśmy. Odłóżmy na bok analizę procesów w mózgu prowadzących do zrozumienia złożonej rzeczywistości dookoła niego. Skupmy się na problemie prostszym, a mianowicie na czym to polega, że coś wiemy? Spróbujmy na to pytanie odpowiedzieć w sposób wystarczający do tego, aby zbudować sztuczny system, o którym można byłoby powiedzieć, że posługuje się tym samym mechanizmem.

Siedzę na ławce w gmachu mojego wydziału i czekam na kogoś. Blisko mnie stoją dwa dystrybutory z kawą. Do jednego z nich podchodzi dziewczyna, wrzuca monetę i wybiera napój. Dystrybutor wydaje różne dźwięki wskazujące na to, że coś robi, choć akurat nie kawę. Dziewczyna zdradza objawy irytacji, popycha dystrybutor, ostatecznie podchodzi do drugiego i tam już skutecznie kupuje kawę. Następnego dnia sam zaopatruję się w kawę w dystrybutorze – oczywiście tym drugim.

Gromadzenie wiedzy o otaczającym świecie i używanie jej po to, aby podejmować w nim racjonalne akcje – oto oczywista funkcja systemów nerwowych bardziej złożonych organizmów żywych. Skuteczność realizowania tej funkcji niekiedy przesądza o życiu lub śmierci. Stąd bierze się ewolucyjna presja na jej rozwój.

Historia o tym, że coś wiemy, zaczyna się od podstawowego budulca naszych aparatów poznawczych, czyli komórki nerwowej – neuronu. Jądro komórkowe neuronu akumuluje potencjał elektryczny, który z biegiem czasu rośnie. Po przekroczeniu pewnej granicy jądro komórkowe skokowo obniża swój potencjał elektryczny, jednocześnie emitując impuls. Impuls rozchodzi się wzdłuż aksonu – długiego pałąka stanowiącego część neuronu. Akson ma kilka tysięcy zakończeń, którymi poprzez synapsy jest połączony z dendrytami innych neuronów. Rozchodzące się przez akson wyładowanie, poprzez synapsy i dendryty, wpływa na tempo, w którym inne neurony (do których te dendryty należą) zwiększają swój potencjał elektryczny.

A zatem podstawowym zajęciem neuronów tworzących nasze mózgi jest strzelanie do siebie impulsami elektrycznymi. Mamy tych neuronów po około 10^{11} – każdy z nich jest połączony poprzez akson i synapsy z około 7 tysiącami innych neuronów i swoimi impulsami współokreśla chwile, w których te neurony wygenerują kolejne impulsy. Każdy neuron generuje od kilku do kilkuset impulsów na sekundę. Jak ta strzelanina prowadzi do tego, że coś wiemy?

Informacja o tym, że dziewczyna nie zdołała kupić kawy w pierwszym dystrybutorze, przez kilka minut utrzymywała się w moim mózgu w postaci cyrkulacji elektrycznej. Gdyby ktoś w tym czasie poddał mój mózg elektrowstrząsom i zresetował jego naturalną aktywność elektryczną, nie pamiętałbym tego zdarzenia. Świadomość tego, co się wokół nas dzieje, jest podtrzymywana przez strzelające do siebie neurony.

Większość tego, czego jesteśmy świadomi, zapamiętujemy na dłużej. Niekiedy na całe życie. Ja z pewnością pamiętałem następnego dnia zaobserwowaną porażkę w kupowaniu kawy i pewnie będę to pamiętał przez kolejne tygodnie czy miesiące. Ewentualny reset cyrkulacji elektrycznej w moim mózgu nie zatrze tego wspomnienia. Zatrze je natomiast czas, i za kilka lat nie pozostanie po nim ślad. Byłoby ono znacznie trwalsze,

gdyby to zdarzenie było dużo bardziej emocjonujące. Silne emocje istotnie wpływają na trwałość wspomnień.

W ciągu kilku minut cyrkulacji elektrycznej utrzymywania świadomości wydarzenia, którego doświadczamy, wydarzenie to jest równocześnie pakowane do pamięci, która ma trwalszą formę. Tą formą jest siła połączeń synaptycznych pewnych neuronów. Przez zmianę tej siły neurony postsynaptyczne (dołączone dendrytami) będą inaczej reagować na impulsy w neuronach presynaptycznych (dołączonych zakończeniami aksonu). W przyszłości fragment mózgu będzie mógł być odpytany o dane wspomnienie i w odpowiedzi przywoła je. Zarówno pytanie, jak i odpowiedź będą miały postać wyładowań elektrycznych.

W jaki sposób impulsy w neuronach są nośnikami naszej świadomości? W jaki sposób to, czego jesteśmy świadomi, trafia do trwalszej pamięci? W jaki sposób przebiega odpytywanie pamięci długotrwałej i jak ona odpowiada? Nie wiemy tego. Stan wiedzy w tej dziedzinie obejmuje modele zachowania się pojedynczego neuronu. Dzięki rozmaitym technikom obrazowania wiemy z pewną dokładnością, które fragmenty mózgu zwiększają swoją aktywność pod wpływem różnych kategorii bodźców. Jeśli chodzi o świadomość i pamięć, to zbudowaliśmy szereg opisujących je modeli o postaci prostokątów z napisami połączonymi strzałkami. Są one jednak dalece niewystarczające do tego, żeby zbudować sztuczny system o podobnej funkcjonalności.

Pewne hipotezy o funkcjonowaniu mózgu można budować, czerpiąc z obszaru dziedziny uczenia maszynowego zwanego uczeniem permanentnym (*continual learning*). Klasyczny problem rozważany w uczeniu maszynowym wygląda tak: Mamy zestaw par $\langle x_i, y_i\rangle \in \mathbb{R}^{n_x+n_y}$, $i = 1, \ldots, N$ i funkcję $m(x, \theta)$ stanowiącą model zależności istniejących w danych. Wyznacza on wektory w $\mathbb{R}^{n_y}$ z wejściami w $\mathbb{R}^{n_x}$ i parametrami w $\mathbb{R}^{n_\theta}$. Model będzie nam służył do zgadywania wektorów $y \in \mathbb{R}^{n_y}$ na podstawie znanego $x \in \mathbb{R}^{n_x}$ należącego do tej samej pary, przy założeniu, że owa para $\langle x, y\rangle$ będzie pochodziła z tego rozkładu co nasze dane. Problem sprowadza się do znalezienia minimum funkcji

$$(*) \qquad J(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|y_i - m(x_i, \theta)\|^2.$$

O uczeniu permanentnym mówimy wtedy, kiedy pary $\langle x_i, y_i\rangle$ przychodzą w pakietach i pakiety przetwarzamy sekwencyjnie – jeden po drugim. Dla przykładu, model powstaje na podstawie danych z kamer umieszczonych na samochodach. Codziennie przychodzi porcja nowych nagrań. Jest ich w sumie zbyt dużo, żeby pamiętać je wszystkie i budować model każdego dnia od nowa. Chcemy jedynie uaktualniać model na podstawie bieżącego pakietu. Aktualizacja modelu przez minimalizację funkcji J $(*)$ zdefiniowanej przez dane z bieżącego pakietu prowadzi do „katastrofalnego zapominania" poprzednich danych. Kiedy bowiem zmuszamy nasz model m do właściwego reagowania na nowe dane, zmienia się także jego odpowiedź na wcześniejsze dane. Innymi słowy, nie możemy zmienić wykresu funkcji m w niektórych punktach, pozostawiając go nietkniętym w reszcie dziedziny. Co więc zrobić?

Kluczowe podejście do uczenia permanentnego polega na tym, aby poza modelem m budować także inny model, „generatywny", pozwalający losować wcześniej widziane dane. W istocie ten drugi model stanowi pewną skompresowaną postać tych danych lub przynajmniej reprezentatywnej ich części. Wtedy aktualizacja modelu m polega na minimalizacji funkcji J $(*)$ zdefiniowanej przez dane z bieżącego pakietu i dane z modelu generatywnego. Fakt, że to podejście nie ma lepszej alternatywy, jest rozczarowujący. Chcielibyśmy aktualizować model na podstawie nowych danych bez konieczności pamiętania wszystkich. Tymczasem w jakimś sensie musimy pamiętać wszystkie dane, aby zapewnić, że model faktycznie do nich pasuje.

Taki sposób konsolidowania dawniejszej wiedzy ze świeżą jest oczywiście kosztowny obliczeniowo. Być może jednak każda jego alternatywa jest jeszcze gorsza. Argumentu wspierającego takie przypuszczenie dostarcza sen. Jedną trzecią życia spędzamy, śpiąc i śniąc. To sporo czasu spędzanego w ryzykowny sposób. W naszej przeszłości ewolucyjnej musiało to kosztować niejedno życie – odebrane przez skradające się drapieżniki (albo skonfliktowanych z nami przedstawicieli tego samego gatunku). Musi stać za tym jakaś bardzo istotna potrzeba. Być może jest to potrzeba ciągłego powtarzania wcześniejszych wspomnień, po to aby upakować je do tego samego modelu wraz z nowymi doświadczeniami.

Co faktycznie wiemy o ludzkiej pamięci długotrwałej? Ośrodkiem w mózgu, który zapewne odpowiada za tworzenie się wspomnień, jest hipokamp. Nie wygląda jednak na to, aby on sam był nośnikiem pamięci. Wiele przesłanek wskazuje raczej na to, że pamięć jest rozproszona po korze mózgowej. Ośrodki mózgu „merytorycznie" związane z przetwarzaniem informacji pewnego rodzaju biorą także udział w procesach związanych z tworzeniem się śladów pamięciowych dotyczących takich informacji, ich podtrzymywaniem i przywoływaniem.

Wiele osób intuicyjnie przypuszcza, że część mózgu jest rodzajem twardego dysku, do którego jest kompresowana informacja i potem przywoływana pewnego rodzaju zapytaniami. Jest to jednak intuicja całkowicie błędna. Naturalna pamięć jest rozproszonym, dynamicznym procesem, w którym nowa wiedza jest konsolidowana ze starą w taki sposób, aby uniknąć niekontrolowanej degradacji tej starej. Tymczasem, co takiego robią nasze neurony, aby ten proces podtrzymywać, i co takiego robią, kiedy przywołujemy naszą wiedzę? Odpowiedzi na te pytania są na razie poza granicami naszego poznania. Coraz lepiej jednak rozumiemy skalę trudności tych pytań.