SZTUCZNA INTELIGENCJA JAKO PYTANIE EMPIRYCZNE

16.0 Wprowadzenie

Dla wielu osób jednym z najbardziej zaskakujących aspektów pracy w sztucznej inteligencji jest stopień, w jakim sztuczna inteligencja, a właściwie większość informatyki, okazuje się dyscypliną empiryczną. Jest to zaskakujące, ponieważ większość ludzi początkowo myśli o tych dziedzinach w kategoriach ich podstaw matematycznych lub, alternatywnie, inżynierskich. Z matematycznego punktu widzenia, czasami nazywanego „schludną" perspektywą, istnieje racjonalistyczne pragnienie wprowadzenia standardów dowodu i analizy do projektowania inteligentnych urządzeń obliczeniowych. Z perspektywy inżynierskiej lub „niechlujnej", zadanie jest często postrzegane jako po prostu tworzenie udanych artefaktów, które społeczeństwo chce nazwać „inteligentnymi". Niestety, lub na szczęście, w zależności od waszej filozofii, złożoność inteligentnego oprogramowania i niejasności związane z jego interakcjami ze światem ludzkiej działalności utrudniają analizę z czysto matematycznego lub czysto inżynierskiego punktu widzenia. Ponadto, jeśli sztuczna inteligencja ma osiągnąć poziom nauki i stać się krytycznym składnikiem nauki o inteligentnych systemach, przy projektowaniu, wykonywaniu i analizie artefaktów należy uwzględnić mieszankę metod analitycznych i empirycznych. Z tego punktu widzenia każdy program sztucznej inteligencji może być postrzegany jako eksperyment: proponuje pytanie światu przyrody, a wyniki są odpowiedzią natury. Reakcja natury na nasz projekt i zobowiązania programowe kształtuje nasze rozumienie formalizmu, mechanizmu i wreszcie natury samej inteligencji (Newell i Simon 1976). W przeciwieństwie do wielu bardziej tradycyjnych badań ludzkiego poznania, my, jako projektanci inteligentnych artefaktów komputerowych, możemy badać wewnętrzne działanie naszych „badanych". Możemy zatrzymać wykonywanie programu, zbadać stan wewnętrzny i dowolnie modyfikować strukturę. Jak zauważają Newell i Simon, struktura komputerów i ich programów wskazuje na ich potencjalne zachowanie: można je badać, a ich reprezentacje i algorytmy wyszukiwania rozumieć. Wynikiem tego jest moc komputerów jako narzędzi do zrozumienia dwoistości inteligencji. Odpowiednio zaprogramowane komputery są w stanie zarówno osiągnąć poziom złożoności semantycznej i behawioralnej, który aż prosi się o scharakteryzowanie w kategoriach psychologicznych, jak i dają możliwość wglądu w ich stany wewnętrzne, czemu w dużej mierze odmawia się naukowcom badającym większość innych form życia intelektualnego. Na szczęście dla kontynuacji pracy nad sztuczną inteligencją, a także dla stworzenia nauki o inteligentnych systemach, nowocześniejsze techniki psychologiczne, zwłaszcza te związane z fizjologią neuronalną, również rzuciły nowe światło na wiele rodzajów ludzkiej inteligencji. Wiemy teraz na przykład, że inteligentna funkcja człowieka nie jest monolityczna i jednolita. Jest raczej modułowy i rozproszony. Jego moc jest widoczna w narządach zmysłów, takich jak ludzka siatkówka, które mogą wyświetlać i wstępnie przetwarzać informacje wizualne. Podobnie ludzkie uczenie się nie jest jednolitą i jednorodną zdolnością. Uczenie się jest raczej funkcją wielu środowisk i różnych systemów, z których każdy jest przystosowany do osiągnięcia wyspecjalizowanych celów. Analiza fMRI, wraz ze skanami PET, EEG i pokrewnymi procedurami fizycznego obrazowania neuronowego, wszystko to potwierdza zróżnicowany i oparty na współpracy obraz wewnętrznego działania rzeczywistych inteligentnych systemów. Jeśli praca w AI ma osiągnąć poziom nauki, musimy też zająć się ważnymi kwestiami filozoficznymi, zwłaszcza tymi związanymi z epistemologią, czy pytaniem, jak inteligentny system „zna" swój świat. Kwestie te rozciągają się od pytania o to, co jest przedmiotem badań sztucznej inteligencji, do głębszych zagadnień, takich jak kwestionowanie ważności i użyteczności hipotezy o fizycznym systemie symboli. Dalsze pytania dotyczą tego, czym jest „symbol" w podejściu systemu symboli do sztucznej inteligencji i jak symbole mogą odnosić się do zbiorów ważonych węzłów w modelu koneksjonistycznym. Kwestionujemy również rolę racjonalizmu wyrażonego w indukcyjnym uprzedzeniu obserwowanym w większości programów nauczania i jak to się ma do nieskrępowanego braku struktury, często obserwowanego w niekontrolowanych, wzmacniających i pojawiających się

podejściach do uczenia się. Wreszcie, musimy zakwestionować rolę ucieleśnienia, usytuowania i socjologicznych uprzedzeń w rozwiązywaniu problemów. Kończymy naszą dyskusję na temat zagadnień filozoficznych, proponując konstruktywistyczną epistemologię, która pasuje do naszego zaangażowania zarówno w sztuczną inteligencję jako naukę, jak i do sztucznej inteligencji jako badań empirycznych. Dlatego w tym ostatniej części ponownie powrócimy do pytań zadanych w Części 1: Czym jest inteligencja? Czy można to sformalizować? Jak możemy zbudować mechanizmy, które to przejawiają? Jak sztuczna i ludzka inteligencja może wpasować się w szerszy kontekst nauki o inteligentnych systemach? W sekcji 16.1 zaczynamy od poprawionej definicji sztucznej inteligencji, która pokazuje, jak obecna praca w sztucznej inteligencji, choć zakorzeniona w hipotezie Newella i Simona dotyczącej fizycznego systemu symboli, rozszerzyła zarówno jej narzędzia, techniki, jak i badania w znacznie szerszym kontekście. Badamy te alternatywne podejścia do kwestii inteligencji i rozważamy ich możliwości w projektowaniu inteligentnych maszyn oraz jako składnik nauki o systemach inteligencji. W sekcji 16.2 wskazujemy, ile technik współczesnej psychologii poznawczej, neuronauki, a także epistemologii można wykorzystać do lepszego zrozumienia przedsięwzięcia związanego ze sztuczną inteligencją. Wreszcie, w sekcji 16.3 omawiamy niektóre wyzwania, które pozostają zarówno dla współczesnych praktyków sztucznej inteligencji, jak i dla epistemologów. Chociaż tradycyjne podejście do sztucznej inteligencji często było winne racjonalistycznego redukcjonizmu, nowe interdyscyplinarne spostrzeżenia i narzędzia również mają powiązane wady. Na przykład twórcy algorytmu genetycznego i projektanci badań nad życiem definiują świat inteligencji z darwinowskiego punktu widzenia: „Co jest, jest tym, co przetrwa”. Wiedza jest również postrzegana jako „wiedzieć jak”, a nie „wiedzieć co” w złożonym świecie. Dla naukowca odpowiedzi wymagają wyjaśnień, a „sukces” czy „przetrwanie” same w sobie nie wystarczą. W tym ostatnim rozdziale omówimy przyszłość sztucznej inteligencji, badając pytania filozoficzne, którymi należy się zająć, aby stworzyć obliczeniową naukę o inteligencji. Dochodzimy do wniosku, że metodologia empiryczna sztucznej inteligencji jest ważnym i być może jednym z najlepszych dostępnych narzędzi do badania natury inteligencji.

16.1 Sztuczna inteligencja: zmieniona definicja

16.1.1 Inteligencja i hipoteza systemu symboli fizycznych

Opierając się na naszym doświadczeniu z ostatnich 15 Części, oferujemy zmienioną definicję sztucznej inteligencji: sztuczna inteligencja to badanie mechanizmów leżących u podstaw inteligentnego zachowania poprzez konstrukcję i ocenę artefaktów zaprojektowanych w celu wprowadzenia tych mechanizmów.

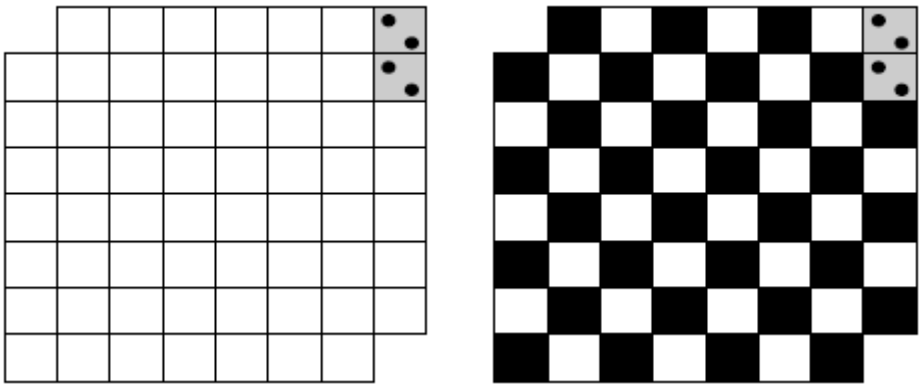
Zgodnie z tą definicją sztuczna inteligencja nie jest teorią dotyczącą mechanizmów leżących u podstaw inteligencji, a bardziej empiryczną metodologią konstruowania i testowania możliwych modeli wspierających taką teorię. Jest to zobowiązanie do naukowej metody projektowania, przeprowadzania i oceny eksperymentów w celu udoskonalenia modelu i dalszego eksperymentowania. Jednak co najważniejsze, ta definicja, podobnie jak sama dziedzina sztucznej inteligencji, bezpośrednio atakuje wieki filozoficznego obskurantyzmu dotyczącego natury umysłu. Daje ludziom, którzy mogliby zrozumieć, co jest być może naszą cechą charakterystyczną jako ludzi, alternatywę dla religii, przesądów, kartezjańskiego dualizmu, new-age placebo lub poszukiwania inteligencji w jeszcze nieodkrytych dziwactwach mechaniki kwantowej. Gdyby nauka wspierająca sztuczną inteligencję w jakikolwiek sposób przyczyniła się do ludzkiej wiedzy, polega na potwierdzeniu, że inteligencja nie jest jakimś mistycznym oparem przenikającym ludzi i anioły, ale raczej efektem zestawu zasad i mechanizmów, które można zrozumieć i stosowane w projektowaniu inteligentnych maszyn. Należy zauważyć, że nasza zmieniona definicja SI nie definiuje inteligencji; proponuje raczej spójną rolę sztucznej inteligencji w badaniu zarówno natury, jak i ekspresji inteligentnych zjawisk. Z perspektywy

historycznej dominujące podejście do sztucznej inteligencji polegało na konstruowaniu formalizmów reprezentacji i związanych z nimi mechanizmów rozumowania opartego na wyszukiwaniu. Wiodącą zasadą wczesnej metodologii sztucznej inteligencji była hipoteza dotycząca fizycznego systemu symboli, po raz pierwszy sformułowana przez Newella i Simona. Ta hipoteza stwierdza:

Warunkiem koniecznym i wystarczającym, aby system fizyczny wykazywał ogólne inteligentne działanie, jest posiadanie fizycznego systemu symboli. Wystarczający oznacza, że inteligencję można osiągnąć za pomocą dowolnego odpowiednio zorganizowanego systemu symboli fizycznych. Niezbędne oznacza, że każdy agent, który wykazuje ogólną inteligencję, musi być instancją fizycznego systemu symboli. Konieczność hipotezy dotyczącej fizycznego systemu symboli wymaga, aby każdy inteligentny agent, czy to człowiek, obcy kosmos czy komputer, osiągnął inteligencję poprzez fizyczną implementację operacji na strukturach symboli. Ogólne inteligentne działanie oznacza ten sam zakres działania, co ludzkie działanie. W granicach fizycznych system zachowuje się odpowiednio do swoich celów i dostosowuje się do wymagań otoczenia.

Newell i Simon podsumowali argumenty przemawiające za koniecznością i wystarczalnością tej hipotezy. W kolejnych latach zarówno sztuczna inteligencja, jak i kognitywistyka badały terytorium wyznaczone przez tę hipotezę. Hipoteza dotycząca fizycznego systemu symboli doprowadziła do czterech istotnych zobowiązań metodologicznych: (a) użycie symboli i systemów symboli jako medium do opisu świata; (b) projektowanie mechanizmów wyszukiwania, zwłaszcza wyszukiwania heurystycznego, w celu zbadania przestrzeni potencjalnych wniosków, które te systemy symboli mogą wspierać; oraz (c) ucieleśnienie architektury poznawczej, przez co rozumiemy, że założono, że odpowiednio zaprojektowany system symboli może zapewnić pełne wyjaśnienie przyczynowe inteligencji, niezależnie od środka jej realizacji. Wreszcie (d), z tego punktu widzenia, sztuczna inteligencja stała się empiryczna i konstruktywistyczna: próbowała zrozumieć inteligencję, budując jej działające modele. W widoku systemu symboli tokeny w języku, zwanym symbolami, były używane do oznaczania lub odniesienia do czegoś innego niż one same. Podobnie jak symbole słowne w języku naturalnym, symbole reprezentowały rzeczy w świecie inteligentnych agentów lub odnosiły się do nich. Tarski (1956, sekcja 2.3) mógłby zaproponować naukę o znaczeniu tych relacji obiekt-odniesienie. Co więcej, użycie symboli przez sztuczną inteligencję wykracza poza kwestie poruszane w semantyce Tarskiana, rozszerzając symbole tak, aby reprezentowały wszystkie formy wiedzy, umiejętności, intencji i przyczynowości. Takie konstruktywne wysiłki polegają na tym, że symbole wraz z ich semantyką, można osadzić w systemach formalnych. Definiują one język reprezentacji. Zdolność do sformalizowania modeli symbolicznych jest niezbędna do modelowania inteligencji jako działającego programu komputerowego. W poprzednich częściach szczegółowo przeanalizowaliśmy kilka reprezentacji: rachunek predykatów, sieci semantyczne, skrypty, grafy pojęciowe, ramki i obiekty. Matematyka systemów formalnych pozwala nam dyskutować o takich kwestiach, jak poprawność, kompletność i złożoność, a także omawiać organizację struktur wiedzy. Ewolucja formalizmów reprezentacyjnych pozwoliła nam ustanowić bardziej złożone (bogatsze) relacje semantyczne. Na przykład systemy dziedziczenia stanowią semantyczną teorię wiedzy taksonomicznej. Formalnie definiując dziedziczenie klas, takie języki uprościć tworzenie inteligentnych programów i dostarcz testowalne modele samej organizacji możliwych kategorii inteligencji. Pojęcie poszukiwania jest ściśle związane ze schematami reprezentacji i ich wykorzystaniem w rozumowaniu. Poszukiwanie stało się krok po kroku badaniem stanów problemowych w ramach przestrzeni stanów (zobowiązanie semantyczne a priori) w poszukiwaniu rozwiązań, celów podproblemu, symetrii problemu lub jakiegokolwiek innego aspektu problemu, który może być rozważany. Reprezentacja i wyszukiwanie są ze sobą powiązane, ponieważ zobowiązanie do określonej reprezentacji określa przestrzeń do przeszukania. Rzeczywiście, niektóre problemy mogą być trudniejsze, lub nawet niemożliwe, z powodu złego wyboru języka reprezentacji. Omówienie błędu indukcyjnego w dalszej części ilustruje ten punkt.

Dramatycznym i często przytaczanym przykładem tej wzajemnej zależności między poszukiwaniem a reprezentacją, a także trudnością w wyborze odpowiedniej reprezentacji (czy ten proces optymalizacji wyboru reprezentacji można zautomatyzować?) Jest problem umieszczania domina na ściętej szachownicy. Załóżmy, że mamy szachownicę i zestaw domina, tak aby każde domino pokryło dokładnie dwa pola szachownicy. Załóżmy również, że na planszy brakuje kilku kwadratów; na Rysunku 1 lewy górny róg i prawy dolny róg zostały usunięte.



Zadanie z szachownicą ściętą polega na pytaniu, czy istnieje sposób na umieszczenie domina na szachownicy tak, aby każde pole szachownicy było obrócone, a każde domino obejmowało dokładnie dwa pola. Możemy spróbować rozwiązać ten problem, wypróbowując wszystkie układy domina na szachownicy; jest to oczywiste podejście oparte na wyszukiwaniu i jest naturalną konsekwencją przedstawiania planszy jako prostej macierzy, pomijając takie pozornie nieistotne cechy, jak kolor kwadratów. Złożoność takiego wyszukiwania jest ogromna i wymagałaby heurystyki dla skutecznego rozwiązania. Na przykład możemy przyciąć częściowe rozwiązania, które pozostawiają pojedyncze kwadraty w izolacji. Moglibyśmy również zacząć od rozwiązania problemu dla mniejszej płytki, takiej jak 2 × 2 i 3 × 3 i spróbuj rozszerzyć rozwiązanie na sytuację 8 × 8. Bardziej wyrafinowane rozwiązanie, opierające się na bardziej złożonym schemacie reprezentacji, zwraca uwagę, że każde umieszczenie domina musi obejmować zarówno czarny, jak i biały kwadrat. Ta ścięta plansza ma 32 czarne kwadraty, ale tylko 30 białych kwadratów; w ten sposób pożądane umieszczenie nie będzie możliwe. Rodzi to poważne pytanie dotyczące wyłącznie symboli rozumując: czy mamy reprezentacje, które pozwalają rozwiązującym problemy dostęp do wiedzy z takim stopniem elastyczności i kreatywności? W jaki sposób dana reprezentacja może automatycznie zmienić swoją strukturę, gdy dowiadujemy się więcej o domenie problemowej? Heurystyki są trzecim, obok reprezentacji i wyszukiwania, składnikiem opartym na symbolach AI. Heurystyka to mechanizm organizowania wyszukiwania w ramach alternatyw oferowanych przez daną reprezentację. Heurystyki mają na celu przezwyciężenie złożoności wyczerpujących poszukiwań, stanowiących barierę dla użytecznych rozwiązań dla wielu klas interesujących problemów. W komputerach, podobnie jak w przypadku ludzi, inteligencja wymaga świadomego wyboru „co dalej robić”. W całej historii badań nad sztuczną inteligencją heurystyki przybierały różne formy. Najwcześniejsze techniki rozwiązywania problemów, takie jak wspinaczka górska w grze w szachownicę Samuela lub analiza średnich i końcowych w Newell, Shaw i Simon’s General Problem Solver , weszły do AI z innych dyscyplin, takich jak badania operacyjne, i stopniowo dojrzewały do ogólnych technik rozwiązywania problemów AI. Właściwości wyszukiwania, w tym dopuszczalność, monotoniczność i poinformowanie, są ważnymi wynikami tych wczesnych badań. Techniki te są często określane jako metody słabe. Słabe metody były ogólnymi strategiami wyszukiwania, które miały być stosowane w całych klasach dziedzin problemowych. Widzieliśmy te metody i ich właściwości w Częściach 2, 3, 4, 6 i 14. Wprowadziliśmy solidne metody rozwiązywania problemów AI z systemem eksperckim opartym na regułach, wnioskach opartych na modelach i

przypadkach oraz uczeniu się opartym na symbolach. W przeciwieństwie do słabych metod rozwiązywania problemów, silne metody koncentrują się na informacjach specyficznych dla każdego obszaru problemowego, takich jak medycyna wewnętrzna lub rachunek całkowy, a nie na projektowaniu metod heurystycznych, które uogólniają obszary problemowe. Silne metody leżą u podstaw systemów eksperckich i innych opartych na wiedzy podejść do rozwiązywania problemów. Silne metody kładą nacisk na takie kwestie, jak ilość wiedzy potrzebnej do rozwiązywania problemów, uczenia się i nabywania wiedzy, syntaktyczna reprezentacja wiedzy, zarządzanie niepewnością oraz kwestie związane z jakością wiedzy. Dlaczego nie stworzyliśmy wielu naprawdę inteligentnych systemów opartych na symbolach? Istnieje wiele zarzutów, które można zrównać z cechą inteligencji w fizycznym systemie symboli. Większość z nich można uchwycić rozważając kwestie znaczenia semantycznego i podstawy symboli inteligentnego agenta. Natura „znaczenia" oczywiście wpływa również na koncepcję inteligencji jako przeszukiwania zinterpretowanych struktur symboli i „użyteczności" ukrytej w stosowaniu heurystyk. Pojęcie znaczenia w tradycyjnej sztucznej inteligencji jest w najlepszym przypadku bardzo słabe. Co więcej, pokusa przejścia w kierunku semantyki bardziej opartej na matematyce, takiej jak podejście Tarskiana do możliwych światów, wydaje się chybiona. Wzmacnia racjonalistyczny projekt zastąpienia elastycznej i ewoluującej inteligencji wcielonego agenta światem, w którym jasne i wyraźne idee są bezpośrednio dostępne. Uziemienie znaczenia to kwestia, która od zawsze frustrowała zarówno zwolenników, jak i krytyków sztucznej inteligencji i przedsiębiorstw kognitywnych. Kwestia uziemienia pyta, jak symbole mogą mieć znaczenie. Właśnie to zwraca uwagę Searle, omawiając tak zwany pokój chiński. Searle umieszcza się w pokoju przeznaczonym do tłumaczenia chińskich zdań na angielski; tam otrzymuje zestaw chińskich symboli, wyszukuje symbole w dużym chińskim systemie katalogowania symboli, a następnie wyświetla odpowiednio połączone zestawy angielskich symboli. Searle twierdzi, że chociaż on sam absolutnie nie zna chińskiego, jego „system" może być postrzegany jako maszyna do tłumaczenia z chińskiego na angielski. Tu jest problem. Chociaż każdy, kto pracował w dziedzinie badań nad tłumaczeniem maszynowym lub rozumieniem języka naturalnego , może argumentować, że „maszyna tłumacząca" Searle'a, ślepo łącząc jeden zestaw symboli z innym zestawem symboli, dawałaby wyniki o minimalnej jakości, Faktem jest, że wiele obecnych inteligentnych systemów ma bardzo ograniczoną zdolność interpretowania zestawów symboli w „znaczący" sposób. Ten problem zbyt słabej semantyki wspierającej przenika również wiele modalności sensorycznych opartych na obliczeniach, czy to wizualnych, kinestetycznych czy werbalnych. W obszarach rozumienia języka ludzkiego Lakoff i Johnson (1999) argumentują, że zdolność tworzenia, używania, wymiany i interpretowania symboli znaczeniowych pochodzi z ludzkiego ucieleśnienia w zmieniającym się kontekście społecznym. Ten kontekst jest fizyczny, społeczny i teraz; wspiera i umożliwia człowiekowi przetrwanie, ewolucję i reprodukcję. To sprawia, że możliwy świat analogicznego rozumowania, używania i doceniania humoru oraz doświadczeń muzycznych i artystycznych. Nasza obecna generacja narzędzi i technik sztucznej inteligencji jest rzeczywiście bardzo daleka od możliwości kodowania i wykorzystywania jakiegokolwiek równoważnego systemu „znaczeń". Bezpośrednim rezultatem tego słabego kodowania semantycznego jest tradycyjne wyszukiwanie / heurystyka sztucznej inteligencji bada stany i konteksty stanów, które są wstępnie interpretowane. Oznacza to, że twórca programu AI „przypisuje" lub „nakłada" na symbole programu różne konteksty znaczenia semantycznego. Bezpośrednim rezultatem tego wstępnie zinterpretowanego kodowania jest to, że zadania bogate w inteligencję, w tym uczenie się i język, mogą wytworzyć tylko pewną obliczoną funkcję tej interpretacji. Tak więc wiele systemów AI ma bardzo ograniczone możliwości ewolucji nowych skojarzeń znaczeniowych podczas eksploracji ich otoczenia. Wreszcie, jako bezpośredni rezultat naszych obecnych ograniczonych możliwości modelowania semantycznego, te aplikacje, w których jesteśmy w stanie oderwać się od bogatego kontekstu ucieleśnionego i społecznego, a jednocześnie uchwycić podstawowe elementy rozwiązywania problemów za pomocą wstępnie zinterpretowanych systemów symboli, są naszymi

najbardziej udane przedsięwzięcia. Wiele z nich zostało omówionych w tej książce. Jednak nawet te obszary pozostają kruche, bez wielu interpretacji i tylko z ograniczoną zdolnością do automatycznego powrotu do zdrowia po awarii. W swojej krótkiej historii społeczność badawcza zajmująca się sztuczną inteligencją badała konsekwencje hipotezy dotyczącej fizycznego systemu symboli i opracowała własne wyzwania dla tego wcześniej dominującego poglądu. Jak zilustrowano w dalszych rozdziałach tej książki, jawny system symboli i wyszukiwanie nie są jedynymi możliwymi reprezentatywnymi mediami do przechwytywania inteligencji. Modele obliczeniowe oparte na architekturze mózgu zwierzęcia, a także na procesach ewolucji biologicznej zapewniają również przydatne ramy do zrozumienia inteligencji w kategoriach procesów poznanych naukowo i odtwarzalnych empirycznie. W następnych sekcjach tego ostatniego rozdziału zbadamy konsekwencje tych podejść. Istotną alternatywą dla hipotezy dotyczącej fizycznego systemu symboli są badania sieci neuronowych i innych biologicznie inspirowanych modeli obliczeń. Na przykład sieci neuronowe są obliczeniowymi i fizycznie utworzonymi modelami poznania, które nie są całkowicie zależne od wyraźnie wymienionych i zinterpretowanych symboli, które charakteryzują świat. Ponieważ „wiedza" w sieci neuronowej jest rozproszona w strukturach tej sieci, często trudno jest, jeśli nie jest to niemożliwe, wyodrębnić poszczególne pojęcia do określonych węzłów i wag sieci. W rzeczywistości każda część sieci może odgrywać zasadniczą rolę w przedstawianiu różnych koncepcji. W konsekwencji sieci neuronowe oferują kontrprzykład, przynajmniej do niezbędnej klauzuli hipotezy fizycznego systemu symboli.

Sieci neuronowe i architektury genetyczne przenoszą nacisk na sztuczną inteligencję z problemów reprezentacji symbolicznej i strategii wnioskowania dźwiękowego na kwestie uczenia się i adaptacji. Sieci neuronowe, podobnie jak ludzie i inne zwierzęta, są mechanizmami dostosowywania się do świata: struktura wyszkolonej sieci neuronowej jest kształtowana zarówno przez uczenie się, jak i przez projekt. Inteligencja sieci neuronowej nie wymaga przekształcenia świata w wyraźny model symboliczny. Sieć jest raczej kształtowana przez jej interakcje ze światem, odzwierciedlone przez ukryte ślady doświadczenia. Podejście to wniosło szereg przyczyn do naszego zrozumienia inteligencji, dając nam wiarygodny model mechanizmów leżących u podstaw fizycznego ucieleśnienia procesów umysłowych, bardziej realistyczny opis uczenia się i rozwoju, demonstrację zdolności do prostych i lokalnych adaptacji kształtować złożony system w odpowiedzi na rzeczywiste zjawiska. Wreszcie, oferują potężne narzędzie badawcze dla neuronauki poznawczej. Właśnie dlatego, że są tak różne, sieci neuronowe mogą odpowiedzieć na wiele pytań, które mogą wykraczać poza możliwości ekspresji sztucznej inteligencji opartej na symbolach. Ważna klasa takich pytań dotyczy percepcji. Natura nie jest tak hojna, aby przekazywać nasze spostrzeżenia do systemu przetwarzania jako zgrabne zestawy wyrażeń rachunku predykatów. Sieci neuronowe oferują model tego, jak możemy rozpoznać „znaczące" wzorce w chaosie bodźców zmysłowych. Ze względu na ich rozproszoną reprezentację sieci neuronowe są często bardziej niezawodne niż ich jawnie symboliczne odpowiedniki. Właściwie wyszkolona sieć neuronowa może skutecznie kategoryzować nowe instancje, wykazując raczej ludzkie postrzeganie podobieństwa niż ścisłą logiczną konieczność. Podobnie utrata kilku neuronów nie musi poważnie wpływać na wydajność dużej sieci neuronowej. Wynika to z często rozległej redundancji właściwej dla modeli sieciowych. Być może najbardziej pociągającym aspektem sieci koneksjonistycznych jest ich zdolność uczenia się. Zamiast próbować zbudować szczegółowy symboliczny model świata, sieci neuronowe polegają na plastyczności własnej struktury, aby dostosować się bezpośrednio do doświadczeń zewnętrznych. Nie tyle konstruują model świata, ile kształtują ich doświadczenia w świecie. Uczenie się jest jednym z najważniejszych aspektów inteligencji. Jest to również problem uczenia się, który rodzi jedne z najtrudniejszych pytań do pracy w komputerach neuronowych.

Dlaczego nie zbudowaliśmy mózgu?

W rzeczywistości obecna generacja skonstruowanych systemów koneksjonistycznych ma bardzo niewielkie podobieństwo do ludzkiego układu neuronowego! Ponieważ temat wiarygodności neuronowej jest krytycznym zagadnieniem badawczym, zaczynamy od tego pytania, a następnie rozważamy rozwój i uczenie się. Badania neuronauki poznawczej (Squire i Kosslyn 1998, Gazzaniga 2000, Hugdahl i Davidson 2003) wnoszą nowy wgląd w zrozumienie ludzkiej architektury poznawczej. Opisujemy pokrótce niektóre ustalenia i komentujemy ich związek z przedsiębiorstwem AI. Rozpatrujemy zagadnienia z trzech poziomów: pierwszy, neuron, drugi, poziom architektury neuronowej, a na końcu omawiamy reprezentację poznawczą lub problem kodowania. Po pierwsze, na poziomie pojedynczego neuronu identyfikują Shephard i Carlson wiele różnych typów architektur neuronowych dla komórek, z których każda jest wyspecjalizowana pod względem funkcji i roli w większym systemie neuronowym. Te typy obejmują komórki receptorów czuciowych, które zwykle znajdują się w skórze i przekazują informacje wejściowe do innych struktur komórkowych, interneurony, których głównym zadaniem jest komunikowanie się w klastrach komórek, podstawowe neurony, których zadaniem jest komunikacja między skupiskami komórek, oraz neurony ruchowe, których zadaniem jest wyjście systemu . Aktywność neuronowa jest elektryczna. Wzory jonów wpływających do i wychodzących z neuronu określają, czy neuron jest aktywny, czy spoczywa. Typowy neuron ma ładunek spoczynkowy -70mV. Kiedy komórka jest aktywna, z zakończenia aksonu uwalniane są określone substancje chemiczne. Te chemikalia, zwane neuroprzekaźnikami, wpływają na błonę postsynaptyczną, zwykle dopasowując się do określonych miejsc receptora, jak klucz do zamka, inicjując dalsze przepływy jonów. Przepływy jonów, gdy osiągną poziom krytyczny, około -50mV, wytwarzają potencjał czynnościowy, całkowicie lub żaden mechanizm wyzwalający wskazujący, że ogniwo zostało uruchomione. W ten sposób neurony komunikują się poprzez sekwencje kodów binarnych. Zmiany postsynaptyczne w zakresie potencjału czynnościowego są dwojakiego rodzaju: hamujące, występujące głównie w strukturach komórek międzyneuronowych lub pobudzające. Te dodatnie i ujemne potencjały są stale generowane w synapsach w układzie dendrytycznym. Ilekroć efektem netto wszystkich tych zdarzeń jest zmiana potencjałów błonowych powiązanych neuronów z -70 mV do około -50 mV, próg zostaje przekroczony i masowe przepływy jonów są ponownie inicjowane do aksonów tych komórek. Po drugie, na poziomie architektury neuronalnej w korze mózgowej znajduje się łącznie około 1010 neuronów, cienka pofałdowana warstwa pokrywająca całą półkulę mózgową. Duża część kory jest zwinięta, co zwiększa całkowitą powierzchnię. Z obliczeniowego punktu widzenia musimy znać nie tylko całkowitą liczbę synaps, ale także parametry wlotu i wylotu. Shephard (1998) szacuje obie te liczby na około 105. Wreszcie, poza różnicami w komórkach i architekturach systemów neuronowych i komputerowych, istnieje głęboki problem reprezentacji poznawczej. Nie wiemy na przykład, jak nawet proste wspomnienia są zakodowane w korze. Na przykład, jak rozpoznaje się twarz i jak rozpoznanie twarzy może łączyć agenta z uczuciem radości lub smutku. Wiemy bardzo dużo o fizycznych / chemicznych aspektach mózgu, ale stosunkowo niewiele o tym, jak system nerwowy koduje i wykorzystuje „wzorce” w swoim kontekście. Jednym z trudniejszych pytań, przed którymi stają badacze, zarówno w społecznościach neuronowych, jak i komputerowych, jest rola wiedzy wrodzonej w uczeniu się: czy efektywne uczenie się może kiedykolwiek nastąpić na tabula rasa lub pustej karcie, zaczynając bez wstępnej wiedzy i ucząc się całkowicie na podstawie doświadczenia ? A może uczenie się musi zaczynać się od uprzedniego uprzedzenia indukcyjnego? Doświadczenie w projektowaniu programów uczenia maszynowego sugeruje, że do uczenia się w złożonych środowiskach niezbędna jest jakaś wcześniejsza wiedza, zwykle wyrażana jako błąd indukcyjny. Zdolność sieci koneksjonistów do konwergencji w sensownym uogólnieniu z zestawu danych szkoleniowych okazała się wrażliwa na liczbę sztucznych neuronów, topologię sieci i określone algorytmy uczenia się. Razem czynniki te stanowią tak silne obciążenie indukcyjne, jakie można znaleźć w dowolnej reprezentacji symbolicznej. Badania nad rozwojem człowieka potwierdzają ten wniosek. Jest na przykład coraz więcej dowodów na to, że ludzkie niemowlęta dziedziczą szereg

„zakodowanych” uprzedzeń poznawczych, które umożliwiają uczenie się takich dziedzin pojęć, jak język i fizyka zdrowego rozsądku. Charakterystyka wrodzonych uprzedzeń w sieciach neuronowych jest aktywnym obszarem badań. Kwestia wrodzonych uprzedzeń staje się jeszcze bardziej zagmatwana, gdy rozważymy bardziej złożone problemy w nauce. Załóżmy na przykład, że opracowujemy obliczenia model odkrycia naukowego i chcemy modelować przejście Kopernika z geocentrycznego do heliocentrycznego widzenia wszechświata. Wymaga to przedstawienia w programie komputerowym zarówno poglądów kopernikańskich, jak i ptolemejskich. Chociaż moglibyśmy przedstawić te poglądy jako wzorce aktywacji w sieci neuronowej, nasze sieci nie powie nam nic o ich zachowaniu jako teoriach. Zamiast tego wolimy wyjaśnienia, takie jak „Kopernik był zaniepokojony złożonością systemu ptolemejskiego i woleliśmy prostszy model pozwalający planetom obracać się wokół Słońca”. Takie wyjaśnienia wymagają symboli. Oczywiście sieci koneksjonistyczne muszą być w stanie wspierać rozumowanie symboliczne; w końcu istoty ludzkie są sieciami neuronowymi i wydają się dość dobrze manipulować symbolami. Jednak neuronowe podstawy rozumowania symbolicznego są ważnym i otwartym problemem badawczym. Kolejnym problemem jest rola rozwoju w nauce. Ludzkie dzieci nie mogą po prostu uczyć się na podstawie dostępnych danych. Ich zdolność uczenia się w określonych dziedzinach pojawia się na dobrze zdefiniowanych etapach rozwojowych. Interesującym pytaniem jest, czy ten postęp rozwojowy jest wyłącznie wynikiem ludzkiej biologii i ucieleśnienia, czy też odzwierciedla pewne logicznie konieczne ograniczenia zdolności inteligencji do uczenia się niezmienności w jej świecie. Czy etapy rozwojowe mogą funkcjonować jako mechanizm dekompozycji problemu poznawania świata na łatwiejsze do opanowania podproblemy? Czy seria sztucznie narzuconych ograniczeń rozwojowych może zapewnić sztucznym sieciom ramy niezbędne do uczenia się o złożonym świecie? Zastosowanie sieci neuronowych do praktycznych problemów rodzi szereg dodatkowych problemów badawczych. Same właściwości sieci neuronowych, które czynią je tak atrakcyjnymi, takie jak zdolność adaptacji i odporność w świetle brakujących lub niejednoznacznych danych, również stwarzają problemy w ich praktycznym zastosowaniu. Ponieważ sieci są uczone, a nie programowane, trudno jest przewidzieć zachowanie. Istnieje kilka wskazówek dotyczących projektowania sieci, które będą odpowiednio zbieżne w danej dziedzinie problemowej. Wreszcie wyjaśnienia, dlaczego sieć doszła do określonego wniosku, są często trudne do skonstruowania i mogą przybrać formę argumentu statystycznego. To są wszystkie obszary obecnych badań. Można więc zapytać, czy sieci koneksjonistyczne i bardziej symboliczna sztuczna inteligencja są tak różne, jak modele inteligencji. Obaj mają wiele wspólnych cech, zwłaszcza że inteligencja jest ostatecznie zakodowana jako obliczenia i ma fundamentalne i formalne ograniczenia, takie jak hipoteza Church / Turinga. Oba podejścia oferują również modele umysłu ukształtowane przez zastosowanie do praktycznych problemów. Co jednak najważniejsze, oba podejścia zaprzeczają filozoficznemu dualizmowi i umieszczają podstawy inteligencji w strukturze i działaniu fizycznie zrealizowanych urządzeń. Uważamy, że pełne pogodzenie tych dwóch bardzo różnych podejść do zdobywania informacji jest nieuniknione. Kiedy to zostanie osiągnięte, teoria, w jaki sposób symbole mogą zredukować się do wzorców w sieci, a co za tym idzie, wpłynąć na przyszłą adaptację tej sieci, będzie niezwykłym wkładem. Wesprze to szereg zmian, takich jak integracja opartych na sieci percepcji i opartych na wiedzy narzędzi wnioskowania w jedną inteligencję. W międzyczasie jednak obie społeczności badawcze mają dużo pracy do wykonania i nie widzimy powodu, dla którego nie miałyby dalej współistnieć. Dla tych, którzy czują się nieswojo z dwoma pozornie niewspółmiernymi modelami inteligencji, nawet fizyka dobrze funkcjonuje z intuicyjnie sprzecznym poglądem, że światło jest czasem najlepiej rozumiane jako fala, a czasami jako cząstka, chociaż oba punkty widzenia można z powodzeniem podciągnąć do teorii strun .

16.1.3 Agenci, pojawienie się i inteligencja

Obliczenia agentowe i modularne teorie poznania rodzą kolejny zestaw interesujących problemów dla badaczy budujących sztuczną inteligencję. Jedna z ważnych szkół kognitywnych głosi, że umysł jest zorganizowany w zestawy wyspecjalizowanych jednostek funkcjonalnych. Moduły te są specjalistami i wykorzystują szereg wrodzonych struktur i funkcji, od „sztywnego" rozwiązywania problemów po indukcyjne uprzedzenia, które odpowiadają za różnorodność problemów, które jako praktyczni agenci muszą rozwiązać. Ma to sens: w jaki sposób pojedyncza sieć neuronowa lub inny system może być wyszkolony do obsługi tak różnych funkcji, jak percepcja, kontrola motoryczna, pamięć i rozumowanie wyższego poziomu? Modułowe teorie inteligencji zapewniają zarówno ramy dla odpowiedzi na te pytania, jak i kierunek dalszych badań nad takimi kwestiami, jak natura wrodzonych uprzedzeń w poszczególnych modułach, a także mechanizmy interakcji modułów. Genetyczne i powstające modele obliczeń oferują jedno z najnowszych i najbardziej ekscytujących podejść do zrozumienia zarówno ludzkiej, jak i sztucznej inteligencji. Wykazując, że globalnie inteligentne zachowanie może wynikać ze współpracy dużej liczby ograniczonych, niezależnych i indywidualnych agentów, teorie genetyczne i emergentne postrzegają złożone wyniki poprzez wzajemne powiązania stosunkowo prostych struktur. Na przykładzie Holandii mechanizmy utrzymujące duże miasto, takie jak Nowy Jork, zaopatrywane w chleb, pokazują podstawowe procesy leżące u podstaw pojawienia się inteligencji w systemie agentowym. Jest mało prawdopodobne, abyśmy mogli napisać scentralizowany planista, który z powodzeniem zaopatrywałby nowojorczyków w bogatą gamę pieczywa codziennego, do którego są przyzwyczajeni. Rzeczywiście, niefortunny eksperyment komunistycznego świata z centralnym planowaniem ujawnił ograniczenia takiego podejścia! Jednak pomimo praktycznych trudności związanych z napisaniem algorytmu scentralizowanego planowania, który zapewni zaopatrzenie Nowego Jorku w chleb, luźno skoordynowane wysiłki wielu piekarzy, kierowców ciężarówek, dostawców surowców, a także sprzedawców detalicznych w mieście rozwiązują problem całkowicie. ładnie. Podobnie jak we wszystkich nowych systemach opartych na agentach, nie ma centralnego planu. Żaden piekarz nie ma bardzo ograniczonej wiedzy na temat zapotrzebowania miasta na chleb; każdy piekarz stara się po prostu zoptymalizować swoje własne możliwości biznesowe. Rozwiązanie problemu globalnego wyłania się ze zbiorowych działań tych niezależnych i lokalnych agentów. Pokazując, jak bardzo ukierunkowane na cel, solidne, prawie optymalne zachowania mogą wynikać z interakcji lokalnych, indywidualnych agentów, modele te dostarczają kolejnej odpowiedzi na stare filozoficzne pytania dotyczące pochodzenia umysłu.

Podejście do inteligencji jest takie, że pełna inteligencja może i rzeczywiście powstaje w wyniku interakcji wielu prostych, indywidualnych, lokalnych i wcielonych inteligencji agentowych. Drugą główną cechą wyłaniających się modeli jest ich poleganie na doborze darwinowskim jako podstawowym mechanizmie kształtującym zachowanie poszczególnych czynników. Na przykładzie piekarni wydaje się, że każdy piekarz nie zachowuje się w sposób, który jest w pewnym sensie optymalny globalnie. Raczej źródłem ich optymalności nie jest centralny projekt; to prosty fakt, że piekarze, którzy słabo radzą sobie z zaspokajaniem potrzeb lokalnych klientów, generalnie zawodzą. To dzięki niestrudzonym, wytrwałym działaniom tych wybiórczych nacisków poszczególni piekarze dochodzą do zachowań, które prowadzą do ich indywidualnego przetrwania, a także do pożytecznych, wyłaniających się zbiorowych zachowań. Połączenie rozproszonej architektury opartej na agentach i adaptacyjnych presji doboru naturalnego to potężny model pochodzenia i działania umysłu. Psychologowie ewolucyjni dostarczyli modelu sposobu, w jaki dobór naturalny ukształtował ewolucję wrodzonej struktury i uprzedzeń w ludzkim umyśle. Podstawą psychologii ewolucyjnej jest pogląd na umysł jako wysoce modułowy, jako system oddziałujących, wysoce wyspecjalizowanych czynników. Rzeczywiście, dyskusje o psychologii ewolucyjnej często porównują umysł do szwajcarskiego scyzoryka, zbioru specjalistycznych narzędzi, które można zastosować do rozwiązywania różnych problemów. Istnieje coraz więcej dowodów na to, że ludzkie umysły są rzeczywiście wysoce modułowe. Fodor

przedstawia filozoficzny argument na rzecz modułowej struktury umysłu. Minsky bada konsekwencje modularnych teorii sztucznej inteligencji. Ta architektura jest ważna dla teorii ewolucji umysłu. Trudno byłoby sobie wyobrazić, jak ewolucja mogłaby ukształtować pojedynczy system tak złożony jak umysł. Jest jednak prawdopodobne, że ewolucja, działająca przez miliony lat, może sukcesywnie kształtować indywidualne, wyspecjalizowane zdolności poznawcze. Wraz z postępem ewolucji mózgu może on również pracować nad kombinacjami modułów, tworząc mechanizmy umożliwiające modułom interakcję, wymianę informacji i współpracę przy wykonywaniu coraz bardziej złożonych zadań poznawczych . Teorie selekcji neuronalnej pokazują, w jaki sposób te same procesy mogą odpowiadać za adaptację indywidualnego układu nerwowego. Darwinizm neuronowy modeluje adaptację systemów neuronowych w kategoriach darwinowskich: wzmacnianie określonych obwodów w mózgu i osłabianie innych jest procesem selekcji w odpowiedzi na świat. W przeciwieństwie do symbolicznych metod uczenia się, które próbują wydobyć informacje z danych treningowych i wykorzystać te informacje do budowy modeli świata, teorie selekcji neuronalnej badają wpływ presji selektywnych na populacje neuronów i ich interakcje. Edelman stwierdza: Rozważając naukę o mózgu jako naukę o rozpoznawaniu, sugeruję, że rozpoznawanie nie jest pouczającym procesem. Nie zachodzi żaden bezpośredni transfer informacji, podobnie jak żaden w procesach ewolucyjnych lub odpornościowych. Zamiast tego rozpoznawanie jest selektywne. Technologie agentowe oferują również modele współpracy społecznej. Korzystając z podejścia opartego na agentach, ekonomiści zbudowali informacyjne (jeśli nie całkowicie predykcyjne) modele rynków ekonomicznych. Technologie agentowe wywierają coraz większy wpływ na projektowanie rozproszonych systemów obliczeniowych, budowę narzędzi wyszukiwania internetowego oraz wdrażanie kooperacyjnych środowisk pracy. Wreszcie modele oparte na agentach wywarły wpływ na teorie świadomości. Na przykład Daniel Dennett oparł opis funkcji i struktury świadomości na agentowej architekturze umysłu. Zaczyna od argumentacji, że niewłaściwe jest pytanie, gdzie w umyśle / mózgu znajduje się świadomość. Zamiast tego, jego wielokrotna szkicowa teoria świadomości skupia się na roli świadomości w interakcjach agentów w rozproszonej architekturze mentalnej. W toku percepcji, kontroli motorycznej, rozwiązywania problemów, uczenia się i innych czynności umysłowych tworzymy koalicje oddziałujących agentów. Koalicje te są bardzo dynamiczne, zmieniające się w odpowiedzi na potrzeby różnych sytuacji. Świadomość, dla Dennetta, służy jako wiążący mechanizm dla tych koalicji, wspierając interakcję agentów i podnosząc krytyczne koalicje oddziałujących agentów na pierwszy plan przetwarzania poznawczego.

Jakie kwestie ograniczają przybliżenie inteligencji oparte na agentach? Podejścia oparte na agentach i „emergentne" otworzyły szereg problemów, które muszą zostać rozwiązane, jeśli ich obietnica ma zostać zrealizowana. Na przykład jeszcze nie wykonaliśmy wszystkich kroków, które umożliwiły ewolucję zdolności poznawczych wyższego poziomu, takich jak język. Podobnie jak wysiłki paleontologów mające na celu odtworzenie ewolucji gatunków, śledzenie rozwoju tych problemów wyższego poziomu będzie wymagało wielu dodatkowych, szczegółowych prac. Musimy oboje wyliczyć czynniki leżące u podstaw architektury umysłu i prześledzić ich ewolucję w czasie.

Innym ważnym problemem związanym z teoriami opartymi na agentach jest wyjaśnienie interakcji między modułami. Chociaż model umysłu „szwajcarskiego scyzoryka" jest użytecznym konstruktorem intuicji, moduły składające się na umysł nie są tak niezależne jak ostrza scyzoryka. Umysły wykazują rozległe, bardzo płynne interakcje między domenami poznawczymi: możemy mówić o rzeczach, które widzimy, wskazując na interakcję między modułami wizualnymi i językowymi. Możemy budować budynki, które umożliwiają określony cel społeczny, wskazując na interakcję między inteligencją techniczną a społeczną. Poeci mogą konstruować dotykowe metafory scen wizualnych, wskazujące na płynną interakcję między wizualnymi i dotykowymi zapachami. Definiowanie reprezentacji i procesów, które umożliwiają takie interakcje międzymodułowe, jest aktywnym obszarem badań. Coraz większe

znaczenie zyskują również praktyczne zastosowania technologii agentowych. Korzystając z symulacji komputerowych opartych na agentach, możliwe jest modelowanie złożonych systemów, które nie mają opisu matematycznego w postaci zamkniętej i których dotychczas nie można było zbadać tak szczegółowo. Techniki oparte na symulacji zastosowano do szeregu zjawisk, takich jak adaptacja ludzkiego układu odpornościowego i kontrola złożonych procesów, w tym akceleratorów cząstek, zachowanie globalnych rynków walutowych oraz badanie systemów pogodowych. Problemy reprezentacyjne i obliczeniowe, które należy rozwiązać, aby zaimplementować takie symulacje, nadal napędzają badania w zakresie reprezentacji wiedzy, algorytmów, a nawet projektowania sprzętu komputerowego. Dalsze praktyczne problemy, z którymi muszą sobie radzić architektury agentów, obejmują protokoły komunikacji między agentami, zwłaszcza gdy lokalni agenci często mają ograniczoną wiedzę o ogólnym problemie lub o wiedzy, którą mogą już posiadać inni agenci. Co więcej, istnieje niewiele algorytmów do dekompozycji większych problemów na spójne podproblemy zorientowane na agentów, czy też faktycznie, jak ograniczone zasoby mogą być rozdzielone między agentów. Te i inne kwestie związane z agentami zostały przedstawione w sekcji 7.4. Być może najbardziej ekscytującym aspektem wyłaniających się teorii umysłu jest ich potencjał do umieszczania czynności umysłowych w jednolitym modelu wyłaniania się porządku z chaosu. Nawet w krótkim przeglądzie przedstawionym w tej sekcji zacytowano prace wykorzystujące wyłaniające się teorie do modelowania szeregu procesów, od ewolucji mózgu w czasie, po siły, które umożliwiają jednostkom uczenie się, budowanie ekonomicznych i społecznych modeli zachowań. Jest coś niezwykle pociągającego w koncepcji, że te same procesy wyłaniającego się porządku, kształtowane przez procesy darwinowskie, mogą wyjaśniać inteligentne zachowanie przy różnych rozdzielczościach, od interakcji poszczególnych neuronów, przez kształtowanie modułowej struktury mózgu, po funkcjonowanie rynków gospodarczych i systemów społecznych. Może się zdarzyć, że inteligencja ma geometrię fraktalną, w której te same wyłaniające się procesy pojawiają się na dowolnym poziomie rozdzielczości, w jakim postrzegamy system jako całość.

16.1.4 Modele probabilistyczne i technologia stochastyczna

Już w latach pięćdziesiątych XX wieku do zrozumienia i generowania wyrażeń w języku naturalnym stosowano techniki stochastyczne. Claude Shannon zastosował modele probabilistyczne, w tym dyskretne łańcuchy Markowa, do zadania przetwarzania języka. Shannon również zapożyczył pojęcie entropii z termodynamiki jako sposób pomiaru pojemności informacyjnej wiadomości. Mniej więcej w tym czasie firma Bell Labs stworzyła pierwszy system statystyczny zdolny do rozpoznawania dziesięciu cyfr, 0, ..., 9, używanych przez jednego mówcę. Działał z dokładnością 97-99%. W latach sześćdziesiątych i siedemdziesiątych bayesowskie podejście do rozumowania było nadal obecne w kontekście działalności badawczej w dziedzinie sztucznej inteligencji. Chociaż wiele systemów eksperckich, na przykład MYCIN, stworzyło własne „algebry czynników pewności", jak widać, kilka, w tym PROSPECTOR, przyjęło podejście bayesowskie. Jednak złożoność takich systemów szybko staje się nie do rozwiązania. Jak wskazaliśmy, pełne wykorzystanie reguły Bayesa do realistycznego programu diagnostyki medycznej obejmującego 200 chorób i 2000 objawów wymagałoby zebrania i zintegrowania ośmiuset milionów informacji. Pod koniec lat osiemdziesiątych Judea Pearl zaproponował obliczeniowy model wnioskowania diagnostycznego w kontekście związków przyczynowych w dziedzinie problemowej: sieci przekonań bayesowskich. BBN rozluźniają dwa ograniczenia pełnego modelu bayesowskiego. Po pierwsze, w wnioskowaniu stochastycznym zakłada się niejawną „przyczynowość"; to znaczy, rozumowanie przechodzi od przyczyny do skutku i nie jest koliste, tj. skutek nie może zawrócić, by spowodować sam siebie. To wspiera reprezentowanie BBN jako skierowanego acyklicznego grafu. Po drugie, BBN zakładają, że bezpośredni rodzic węzła wspiera pełny wpływ przyczynowy na ten węzeł. Zakłada się, że wszystkie inne węzły są warunkowo niezależne lub mają wpływ na tyle mały, że można je zignorować. Badania Pearl odnowiły zainteresowanie

stochastycznymi podejściami do modelowania świata. Jak widzieliśmy , BBN stanowiły bardzo potężne narzędzie reprezentacyjne do wnioskowania diagnostycznego (abdukcyjnego). Jest to prawdziwe zwłaszcza w przypadku dynamicznej natury zarówno systemów ludzkich, jak i stochastycznych: gdy świat zmienia się w czasie, nasze rozumienie jest wzbogacane: niektóre przyczyny okazują się bardziej wyjaśniać to, co widzimy, podczas gdy inne potencjalne przyczyny są „wyjaśniane". Badania nad projektowaniem systemów stochastycznych, a także wspomagające je schematy wnioskowania są dopiero w powijakach. Pod koniec lat 80. nowa energia badawcza została również zastosowana do zagadnień przetwarzania języka naturalnego. Jak wspomniano w sekcji 15.4, te stochastyczne podejścia obejmowały nowe analizowanie, tagowanie i wiele innych technik ujednoznaczniania wyrażeń językowych. Pełen zakres tych podejść można znaleźć w książkach o rozpoznawaniu mowy i zadaniach w przetwarzaniu języka. Wraz z odnowionym zainteresowaniem i sukcesami w stochastycznym podejściu do charakteryzowania inteligentnego zachowania, osoba może w naturalny sposób zastanawiać się, jakie mogą być jego ograniczenia. Czy inteligencja jest zasadniczo stochastyczna?

Istnieje ogromna atrakcyjność w kierunku stochastycznego punktu widzenia interakcji agentów w zmieniającym się świecie. Wielu może argumentować, że „system reprezentacji" człowieka jest zasadniczo stochastyczny, tj. Uwarunkowany światem percepcji i przyczyn, w którym jest pogrążony. Z pewnością behawiorystyczno-empiryczny punkt widzenia uznałby to przypuszczenie za atrakcyjne. Teoretycy usytuowanego i osadzonego działania mogą pójść jeszcze dalej i postawić hipotezę, że uwarunkowane relacje agenta z jego fizycznym i społecznym otoczeniem stanowią wystarczające wyjaśnienie dostosowania się agenta do tego świata. Niektórzy współcześni badacze twierdzą nawet, że aspekty funkcji nerwowych są zasadniczo stochastyczne. Rozważmy język. Jedną z mocnych stron wypowiedzi ustnej i pisemnej, jak wskazali Chomsky i inni, jest jej generatywny charakter. Oznacza to, że w zestawie dostępnych form słownictwa i języka w naturalny sposób pojawiają się nowe i wcześniej niedoświadczone wyrażenia. Dzieje się tak zarówno na poziomie tworzenia nowych zdań, jak i pojedynczych słów, werbalizując np. „Google it!". Stochastyczny opis języka musi wykazać tę generującą moc. Obecnie ograniczenia gromadzonych informacji językowych, czy to banki drzew, czy inne kopory danych, mogą radykalnie ograniczyć wykorzystanie technologii stochastycznej. Dzieje się tak, ponieważ zebrane informacje muszą oferować odpowiednie warunki (lub wcześniejsze) do interpretacji obecnej nowej sytuacji. Stochastyczne modele dziedzin zastosowań, powiedzmy, dla silnika lotniczego lub systemu przekładniowego mogą mieć podobne ograniczenia. Tam z konieczności zawsze będzie istniał zamknięty świat lub minimalne założenia modelu, dla realistycznie złożonego systemu. Jednak dynamiczne sieci bayesowskie dają obecnie nadzieję na interpretację stanu systemu . Jednak nadal istnieją ograniczone możliwości przewidywania nowych sytuacji w czasie. Na wyższym poziomie wyjaśniającym modelowi probabilistycznemu może być trudno wyjaśnić odejście od własnego systemu wyjaśniającego lub paradygmatu. W jakim sensie, model stochastyczny może ewentualnie wyjaśniać teorie lub przeorganizowanie poglądów pojęciowych wyższego poziomu, to znaczy ponownie oceniać kwestie związane z adekwatnością samego modelu, być może z potrzebą przejść do różnych punktów widzenia? Tematy te pozostają ważnymi kwestiami badawczymi i ograniczeniami stochastycznego podejścia do zrozumienia niepewności. Faktem jest jednak, że często bez wyraźnych instrukcji agenci „tworzą" całkiem udane modele. Jak widzimy z konstruktywistycznego punktu widzenia, w sekcji 16.2, pewien model jest warunkiem sine qua non dla podmiotu, aby zrozumieć swój świat, tj. Jeśli nie ma a priori zaangażowania w to, o czym świat „chodzi", zjawiska nie są ani postrzegane, ani rozumiane. ! Oprócz przedstawionych właśnie argumentów filozoficznych, istnieje szereg praktycznych ograniczeń stosowania systemów stochastycznych. Obecna generacja BBN ma charakter propozycyjny. Dopiero niedawno było możliwe tworzenie ogólnych praw lub relacji, takich jak ∀X mężczyzna (X) → inteligentny (X) z dołączoną dystrybucją. Ponadto pożądane jest, aby

BBN zawierały relacje rekurencyjne, zwłaszcza w przypadku analizy szeregów czasowych. Badania mające na celu opracowanie stochastycznych systemów reprezentacji pierwszego rzędu, odnoszących się do tych ogólnych problemów, są ważne dla kontynuowania badań. Jak zauważono w sekcji 13.3, dostępne są teraz narzędzia do modelowania stochastycznego pierwszego rzędu i pełnego Turinga. Ważne jest również zbadanie zastosowania tych ogólnych modeli stochastycznych do danych neuro / psychologicznych, tam gdzie zostały one ostatnio zastosowane. Następnie omówimy te psychologiczne i filozoficzne aspekty ludzkiej inteligencji, które mają wpływ na tworzenie, wdrażanie i ocenę sztucznej inteligencji

16.2 Nauka o inteligentnych systemach

To nie przypadek, że duża podgrupa społeczności zajmującej się sztuczną inteligencją skupiła się w swoich badaniach na zrozumieniu ludzkiej inteligencji. Ludzie dostarczają prototypowych przykładów inteligentnej aktywności, a inżynierowie sztucznej inteligencji, chociaż zwykle tak nie jest zaangażowani w „tworzenie programów, które zachowują się jak ludzie", rzadko ignorują ludzkie rozwiązania. Niektóre zastosowania, takie jak rozumowanie diagnostyczne, są często celowo wzorowane na procesach rozwiązywania problemów ekspertów ludzkich pracujących w tej dziedzinie. Co ważniejsze, zrozumienie ludzkiej inteligencji jest samo w sobie fascynującym i otwartym wyzwaniem naukowym. Współczesna kognitywistyka, czyli nauka o systemach inteligentnych (Luger 1994), rozpoczęła się wraz z pojawieniem się komputera cyfrowego, mimo że, jak widzieliśmy w rozdziale 1, było wielu intelektualnych przodków tej dyscypliny, od Arystotelesa po Kartezjusza i Boole'a, bardziej współczesnym teoretykom, takim jak Turing, McCulloch i Pitts, twórcy modelu sieci neuronowej oraz John von Neumann, wczesny zwolennik a-life. Badanie stało się jednak nauką ze zdolnością do projektowania i przeprowadzania eksperymentów w oparciu o te teoretyczne pojęcia, i stało się to w istotnym stopniu wraz z pojawieniem się komputera. Na koniec musimy zapytać: „Czy istnieje wszechstronna nauka o inteligencji?" Możemy dalej zapytać: „Czy nauka o inteligentnych systemach może wspierać konstrukcję sztucznych inteligencji?" W kolejnych sekcjach omówimy pokrótce, w jaki sposób nauki psychologiczne, epistemologiczne i socjologiczne wspierają badania i rozwój w dziedzinie sztucznej inteligencji.

16.2.1 Ograniczenia psychologiczne

Wczesne badania kognitywne badały ludzkie rozwiązania problemów logicznych, prostych gier, planowania i uczenia się koncepcji. Zbiegając się z ich pracą nad teoretykiem logiki, Newell i Simon zaczęli porównywać swoje podejścia obliczeniowe z wyszukiwaniem strategie stosowane przez ludzi. Ich dane składały się z protokołów „myśl na głos", opisów ich myśli przez ludzi podczas procesu opracowywania rozwiązania problemu, na przykład dowodu logicznego. Newell i Simon następnie porównali te protokoły z zachowaniem programu komputerowego rozwiązującego ten sam problem. Naukowcy odkryli niezwykłe podobieństwa i interesujące różnice zarówno między problemami, jak i przedmiotami. Te wczesne projekty ustanowiły metodologię, którą dyscyplina kognitywna miałaby stosować w następnych dziesięcioleciach:

1. Na podstawie danych od ludzi rozwiązujących poszczególne klasy problemów, zaprojektuj schemat reprezentacyjny i powiązaną strategię wyszukiwania w celu rozwiązania problemu.

2. Uruchom model komputerowy, aby prześledzić zachowanie rozwiązania.

3. Obserwuj ludzi pracujących nad tymi samymi problemami i śledź mierzalne parametry ich procesu rozwiązywania, takie jak te znalezione w protokołach myślenia na głos, ruchy oczu i pisemne wyniki cząstkowe.

4. Analizować i porównywać rozwiązania ludzkie i komputerowe.

5. Popraw model komputerowy dla następnej rundy testów i porównań z ludźmi.

Ta empiryczna metodologia jest opisana w wykładzie Newell and Simon's Turing Award, cytowanym na początku tego rozdziału. Ważnym aspektem kognitywistyki jest wykorzystywanie eksperymentów do walidacji architektury rozwiązywania problemów, niezależnie od tego, czy będzie to system produkcyjny, łącznikowy, emergentny, czy też architektura oparta na interakcji rozproszonych agentów. W ostatnich latach do tego paradygmatu nadano zupełnie nowy wymiar. Teraz nie tylko programy można dekonstruować i obserwować podczas rozwiązywania problemów, ale także ludzi i inne formy życia. Szereg nowych technik obrazowania zostało włączonych do narzędzi dostępnych do obserwacji aktywności kory. Należą do nich magnetoencefalografia (MEG), która wykrywa pola magnetyczne generowane przez populacje neuronów. W przeciwieństwie do potencjałów elektrycznych generowanych przez te populacje, pole magnetyczne nie jest rozmazane przez czaszkę i skórę głowy, dzięki czemu możliwa jest znacznie większa rozdzielczość. Drugą technologią obrazowania jest pozytonowa tomografia emisyjna, czyli PET. Substancja radioaktywna, zwykle O15, jest wstrzykiwana do krwiobiegu. Kiedy dany obszar mózgu jest aktywny, więcej tego czynnika przechodzi przez czułe detektory niż w stanie spoczynku. Porównanie obrazów spoczynkowych i aktywnych może potencjalnie ujawnić funkcjonalną lokalizację przy rozdzielczości około 1 cm (patrz Stytz i Frieder 1990). Inną techniką analizy neuronowej jest funkcjonalne obrazowanie metodą rezonansu magnetycznego lub fMRI. Podejście to wyłoniło się z bardziej standardowego obrazowania strukturalnego opartego na magnetycznym rezonansie jądrowym (NMR). Podobnie jak w przypadku PET, podejście to porównuje odpoczynek z aktywnymi stanami neuronalnymi w celu ujawnienia lokalizacji funkcjonalnej. Dalszym wkładem w lokalizację funkcji mózgu, z ważnym powiązaniem z właśnie wspomnianymi technikami obrazowania, są algorytmy oprogramowania opracowane przez Baraka Pearlmuttera i jego współpracowników. Badacze ci są w stanie przyjąć złożone wzorce szumu, często postrzegane jako wynik różnych technik obrazowania neuronowego, i podzielić je na oddzielne komponenty. Jest to niezbędny krok w analizie, ponieważ wzorce normalnego, ustalonego istnienia, takie jak ruchy oczu, oddychanie i bicie serca, przeplatają się z innymi wzorcami odpalania neuronów, które chcemy zrozumieć. Wyniki badań neuronauki poznawczej znacznie poszerzyły naszą wiedzę na temat komponentów neuronalnych zaangażowanych w inteligentną aktywność. Chociaż analiza i krytyka tych wyników wykracza poza zakres tej książki, wymieniamy i odnosimy się do kilku ważnych kwestii: W obszarze percepcji i uwagi istnieje wiążący problem. Badacze tacy jak Anne Triesman i Jeff Hawkins zauważają, że reprezentacje percepcyjne zależą od rozproszonych kodów neuronowych, aby powiązać ze sobą części i właściwości obiektów, i pytają, jaki mechanizm jest potrzebny do „wiązania" informacji dotyczących do każdego przedmiotu i odróżnić go od innych. Jakie mechanizmy neuronowe wspomagają percepcję obiektów osadzonych w dużych, złożonych scenach w obszarze poszukiwań wizualnych? Niektóre eksperymenty pokazują, że tłumienie informacji z nieistotnych obiektów odgrywa ważną rolę w wyborze celów wyszukiwania. Co więcej, jak „uczymy się" widzieć? W obszarze plastyczności w percepcji Gilbert twierdzi, że to, co widzimy, nie jest ściśle odzwierciedleniem fizycznych cech sceny, ale raczej jest wysoce zależne od procesów, za pomocą których nasz mózg próbuje zinterpretować tę scenę. W jaki sposób układ korowy przedstawia i indeksuje informacje związane w czasie, w tym interpretację spostrzeżeń i produkcję aktywności motorycznej? W badaniach nad pamięcią hormony stresu uwalniane w sytuacjach pobudzenia emocjonalnego modulują procesy pamięci. Odnosi się to do problemu uziemienia: w jaki sposób myśli, słowa, spostrzeżenia mają znaczenie dla agenta? W jakim sensie może istnieć „smutek (lub łzy) w rzeczach", lacremae rerum Wergiliusza? Akustyczno-fonetyczne aspekty mowy dostarczają ważnych zasad organizujących łączenie badań neuronauki z teoriami poznawczymi i językowymi. W jaki sposób integruje się składniowe i semantyczne składniki kory?

W jaki sposób jednostka nabywa określony język i jakie etapy neurofizjologiczne wspierają ten rozwój? Jak rozumiany jest rozwój, czym jest plastyczność okresu krytycznego i reorganizacje dorosłych obserwowane w układach somatosensorycznych ssaków? Czy etapy rozwoju są krytyczne dla „budowania" inteligencji? Dalsza dyskusja - patrz Karmiloff-Smith i Gazzaniga. Praktyka sztucznej inteligencji z pewnością nie wymaga rozległej wiedzy na temat tych i pokrewnych dziedzin neuro / psychologicznych. Jednak ten rodzaj wiedzy może wspierać inżynierię inteligentnych artefaktów, a także pomóc zlokalizować badania i rozwój w dziedzinie sztucznej inteligencji w kontekście szerszej nauki o systemach inteligencji. Wreszcie tworzenie psychologicznej, neurofizjologicznej i obliczeniowej syntezy jest naprawdę ekscytujące. Ale to wymaga dojrzałej epistemologii, którą omówimy dalej.

16.2.2 Zagadnienia epistemologiczne

Rozwój sztucznej inteligencji został ukształtowany przez szereg ważnych wyzwań i pytań. Rozumienie języka naturalnego, planowanie, rozumowanie w niepewnych sytuacjach i uczenie maszynowe są typowe dla tych typów problemów, które obejmują pewien istotny aspekt inteligentnego zachowania. Co ważniejsze, inteligentne systemy działające w każdej z tych dziedzin wymagają znajomości celu, praktyki i działania w kontekstach usytuowanych i osadzonych społecznie. Aby lepiej zrozumieć te kwestie, przeanalizujemy epistemologiczne zaangażowanie programu, który ma być „inteligentny". Zaangażowanie epistemologiczne odzwierciedla zarówno semantykę wspierającą użycie symbolu, jak i strukturę użytych symboli. Zadaniem w takich sytuacjach jest odkrywanie i wykorzystanie niezmienności istniejące w domenie problemowej. „Niezmienny" to termin używany do opisania prawidłowości lub znaczących aspektów złożonych środowisk, które można manipulować. W niniejszej dyskusji terminy symbole i systemy symboli są używane ogólnie, od wyraźnych symboli tradycji Newella i Simona, węzły i architektura sieci systemu łącznikowego, wyłaniające się symbole genetycznego i sztucznego życia. Chociaż punkty, które omówimy w następnej kolejności, dotyczą większości sztucznej inteligencji, skupimy się na naszej dyskusji na temat zagadnień związanych z uczeniem maszynowym, ponieważ w tej książce stworzyliśmy wiele przykładów i algorytmów uczenia się. Mimo postępów w uczeniu maszynowym pozostaje jednym z najtrudniejszych problemów w obliczu sztucznej inteligencji. Istnieją trzy kwestie ograniczające nasze obecne rozumienie i postęp w badaniach: po pierwsze, problem uogólnienia i przeuczenia, po drugie, rola indukcyjnego uprzedzenia w uczeniu się, a po trzecie, dylemat empirysta lub zajęcie się ideą nauka bez ograniczeń. Ostatnie dwa problemy są ze sobą powiązane: ukryte indukcyjne nastawienie wielu algorytmów uczenia się jest wyrazem problemu racjonalistów polegającego na tym, że jesteśmy uprzedzeni przez oczekiwania, to znaczy to, czego się uczymy, często wydaje się być bezpośrednią funkcją tego, czego się spodziewamy się nauczyć. Z przeciwnego punktu widzenia, jak widzieliśmy w badaniach a-life, gdzie istnieje bardzo niewiele oczekiwań a priori co do tego, czego należy się nauczyć, czy naprawdę wystarczy powiedzieć: „Zbuduj to, a to się stanie"? Zgodnie z przypisem Yogi Berra na początku tej sekcji, prawdopodobnie nie będzie! W następnych sekcjach pokrótce omówiono te kwestie.

Problem uogólnienia

Przykłady, których użyliśmy do wprowadzenia różnych modeli uczenia się - opartych na symbolach, koneksjonistycznych i emergentnych - były zwykle bardzo ograniczone. Na przykład architektury koneksjonistyczne często zawierały tylko kilka węzłów lub jedną częściowo ukrytą warstwę. Jest to właściwe, ponieważ główne prawa uczenia się można odpowiednio wyjaśnić w kontekście kilku neuronów i częściowych warstw. Może to być bardzo mylące, ponieważ aplikacje sieci neuronowych są zwykle znacznie większe, a problem skali JEST ważny. Na przykład, w przypadku uczenia się z propagacją wsteczną, do rozwiązywania problemów o istotnym znaczeniu praktycznym wymagana jest duża liczba przykładów szkoleniowych i większe sieci. Wielu badaczy obszernie komentuje kwestię doboru odpowiedniej liczby wartości wejściowych, stosunku między parametrami wejściowymi a

ukrytymi węzłami oraz prób treningowych niezbędnych do osiągnięcia konwergencji. W rzeczywistości jakość i ilość danych szkoleniowych są ważnymi kwestiami dla każdego algorytmu uczenia się. Bez odpowiedniej stronniczości lub obszernej wbudowanej wiedzy algorytm uczenia się może zostać całkowicie wprowadzony w błąd, próbując znaleźć wzorce w hałaśliwych, rzadkich lub nawet złych danych.

Podobnym problemem jest kwestia „wystarczalności" w nauce. Kiedy możemy powiedzieć, że nasze algorytmy są wystarczające do wychwycenia ważnych ograniczeń lub niezmienników domeny problemowej? Czy rezerwujemy część naszych oryginalnych danych do testowania naszych algorytmów uczenia się? Czy ilość danych, które posiadamy, ma związek z jakością uczenia się? Być może ocena wystarczalności jest heurystyczna lub estetyczna: my, ludzie, często postrzegamy nasze algorytmy jako „wystarczająco dobre". Zilustrujmy ten problem uogólnienia na przykładzie, używając formy propagacji wstecznej w celu wywołania funkcji ogólnej ze zbioru punktów danych. Rysunek 16.2 może przedstawiać punkty danych, o których uogólnienie prosimy nasz algorytm. Linie w tym zestawie punktów reprezentują funkcje wywołane przez algorytm uczący się. Pamiętaj, że po przeszkoleniu algorytmu będziemy chcieli zaoferować mu nowe punkty danych i poprosić algorytm o wygenerowanie dobrego uogólnienia również dla tych danych. Indukowana funkcja f1 może reprezentować dość dokładne dopasowanie najmniejszych średnich kwadratów. Przy dalszym szkoleniu system może wytworzyć f2, co wydaje się dość „dobre" dopasować do zbioru punktów danych; ale nadal f2 nie przechwytuje dokładnie punktów danych. Dalsze szkolenie może wytworzyć funkcje, które dokładnie pasują do danych, ale mogą dać straszne uogólnienia dla dalszych danych wejściowych. Zjawisko to określa się jako przetrenowanie sieci. Jedną z mocnych stron uczenia się z propagacją wsteczną jest to, że w wielu dziedzinach aplikacji znane jest tworzenie skutecznych uogólnień, to jest przybliżeń funkcjonalnych, które dobrze pasują do danych uczących, a także odpowiednio obsługują nowe dane. Jednak określenie punktu, w którym sieć przechodzi ze stanu niedoświadczonego do przetrenowanego, jest nietrywialne. Naiwnością jest myślenie, że można przedstawić sieć neuronową lub jakiekolwiek inne narzędzie edukacyjne z surowymi danymi, a następnie po prostu odejść na bok i obserwować, jak generuje najbardziej efektywne i użyteczne uogólnienia dotyczące rozwiązywania nowych podobnych problemów. Kończymy, umieszczając tę kwestię uogólnienia z powrotem w jej epistemologicznym kontekście.

Kiedy osoby rozwiązujące problemy tworzą i wykorzystują reprezentacje (symbole, węzły sieci lub cokolwiek innego) w procesie rozwiązywania, tworzą niezmienniki i najprawdopodobniej systemy niezmienników do badania domeny problemu / rozwiązania i tworzenia powiązanych uogólnień. Ten właśnie punkt widzenia wniesiony do procesu rozwiązywania problemów wpływa na ostateczny sukces przedsięwzięcia. W następnym podrozdziale zajmiemy się dalej tą kwestią. Indukcyjne nastawienie, a priori racjonalistów Zautomatyzowane uczenie się w rozdziałach od 10 do 13, a także w większości technik sztucznej inteligencji, odzwierciedlało uprzedzenia a priori ich twórców. Problem błędu indukcyjnego polega na tym, że wynikowe reprezentacje i strategie wyszukiwania oferują medium do kodowania już zinterpretowanego świata. Rzadko oferują mechanizmy kwestionowania naszych interpretacji, generowania nowych punktów widzenia lub cofania się i zmieniania perspektyw, gdy są nieproduktywne. Ta ukryta stronniczość prowadzi do racjonalistycznej epistemologicznej pułapki widzenia świata dokładnie i tylko tego, czego oczekujemy lub jesteśmy przygotowani, aby zobaczyć. W każdym paradygmacie uczenia się należy jasno określić rolę błędu indukcyjnego. Co więcej, tylko dlatego, że nie uznaje się błędu indukcyjnego, nie oznacza to, że nie istnieje i ma krytyczny wpływ na parametry uczenia się. W uczeniu się opartym na symbolach odchylenie indukcyjne jest zwykle oczywiste, na przykład użycie sieci semantycznej do uczenia się koncepcji. W algorytmach uczenia Winstona, uprzedzenia obejmują reprezentację związku koniunkcyjnego oraz znaczenie używania „bliskich trafień" dla udoskonalenia ograniczeń. Widzimy podobne błędy w używaniu określonych

predykatów do przeszukiwania przestrzeni wersji, drzew decyzyjnych w ID3 .Na przykład ograniczenia sieci perceptronowych doprowadziło do wprowadzenia ukrytych węzłów. Możemy zapytać, jaki wkład mają ukryte węzły w generowanie rozwiązania. Jednym ze sposobów zrozumienia roli ukrytych węzłów jest to, że dodają one wymiary do przestrzeni reprezentacji. Jako prosty przykład widzieliśmy w sekcji 11.3.3, że punkty danych dla problemu wykluczającego lub problemu nie można było rozdzielić liniowo w dwóch wymiarach. Wyuczona waga ukrytego węzła nadaje reprezentacji inny wymiar. W trzech wymiarach punkty można rozdzielić za pomocą dwuwymiarowej płaszczyzny. Biorąc pod uwagę dwa wymiary przestrzeni wejściowej i ukrytego węzła, warstwę wyjściową tej sieci można następnie postrzegać jako zwykły perceptron, który znajduje płaszczyznę oddzielającą punkty w trójwymiarowej przestrzeni. Uzupełniająca się perspektywa polega na tym, że wiele „różnych" paradygmatów uczenia się ma wspólne (czasami nieoczywiste) wspólne uprzedzenia indukcyjne. Wskazaliśmy na wiele z nich: związek między grupowaniem z CLUSTER / 2 w sekcji 10.5, perceptronem w sekcji 11.2 i sieciami prototypowymi w sekcji 11.3. Zauważyliśmy, że kontrpropagacja, sprzężona sieć, która wykorzystuje nienadzorowane konkurencyjne uczenie się w warstwie Kohonena wraz z nadzorowanym uczeniem się języka Hebba w warstwie Grossberga, jest pod wieloma względami podobna do uczenia się wstecznego. W kontrpropagacji skupione dane na warstwie Kohonena odgrywają rolę podobną do uogólnień poznanych przez ukryte węzły, które używają propagacji wstecznej. Pod wieloma ważnymi względami prezentowane przez nas narzędzia są podobne. W rzeczywistości nawet odkrycie prototypów reprezentujących klastry danych stanowi uzupełniający przypadek przybliżenia funkcji. W pierwszej sytuacji próbujemy sklasyfikować zbiory danych; w drugiej generujemy funkcje, które jawnie dzielą od siebie skupiska danych. Widzieliśmy to, gdy algorytm klasyfikacji minimalnej odległości użyty przez perceptron podał również parametry określające liniową separację danych. Nawet uogólnienia, które tworzą funkcje, można postrzegać z wielu różnych punktów widzenia. Na przykład techniki statystyczne od dawna umożliwiają odkrywanie korelacji danych. Do przybliżenia większości funkcji można użyć iteracyjnego rozwinięcia szeregu Taylora. Algorytmy aproksymacji wielomianów są używane od ponad wieku do aproksymacji funkcji dopasowywania punktów danych. Podsumowując, zobowiązania podjęte w ramach schematu uczenia się, niezależnie od tego, czy są oparte na symbolach, koneksjonistyczne, emergentne czy stochastyczne, w bardzo dużym stopniu wpływają na wyniki, można oczekiwać po wysiłku rozwiązywania problemów. Kiedy doceniamy ten efekt synergii w całym procesie projektowania komputerowych rozwiązań do rozwiązywania problemów, często możemy zwiększyć nasze szanse na sukces, a także dokładniej zinterpretować nasze wyniki.

Dylemat empirysty

Jeśli obecne podejścia do uczenia maszynowego, zwłaszcza uczenia się nadzorowanego, mają dominującą tendencję indukcyjną, uczenie się nienadzorowane, w tym wiele podejść genetycznych i ewolucyjnych, musi zmierzyć się z przeciwnym problemem, czasami nazywanym dylematem empirysta. Tematy tych obszarów badawczych obejmują: pojawią się rozwiązania, ewoluują alternatywy, populacje odzwierciedlają przetrwanie najlepiej przystosowanych. Jest to potężna rzecz, szczególnie umieszczona w kontekście równoległych i rozproszonych możliwości wyszukiwania. Ale jest problem: skąd możemy wiedzieć, że jesteśmy gdzieś, skoro nie jesteśmy pewni, dokąd zmierzamy? Platon, ponad 2000 lat temu, postawił ten problem słowami niewolnika Menona: A jak możesz dociekać, Sokratesie, o to, czego jeszcze nie wiesz? Co podasz jako przedmiot zapytania? A jeśli dowiesz się, czego chcesz, skąd będziesz wiedzieć, że tego nie wiedziałeś? Kilku badaczy wykazało, że Meno miał rację, i twierdzenia Wolperta i Macready'eg o zakazie obiadu. W rzeczywistości empirysta potrzebuje resztek a priori racjonalistów, aby ocalić naukę! Niemniej jednak istnieje wielkie podekscytowanie nienadzorowanymi i ewolucyjnymi modelami uczenia się; na przykład w tworzeniu sieci opartych na wzorach lub minimalizacji energii, które można postrzegać jako atraktory punktów stałych lub baseny dla złożonych niezmienności relacyjnych. Obserwujemy, jak punkty danych

„osiedlają się” w kierunku atraktorów i mamy pokusę, by postrzegać te nowe architektury jako narzędzia do modelowania zjawisk dynamicznych. Jakie, możemy zapytać, są granice obliczeń w tych paradygmatach?

W rzeczywistości badacze wykazali, że sieci rekurencyjne są obliczeniowo kompletne, to znaczy równoważne klasie maszyn Turinga. Ta równoważność Turinga rozszerza wcześniejsze wyniki: Kołmogorow wykazał, że dla każdej funkcji ciągłej istnieje sieć neuronowa, która oblicza tę funkcję. Wykazano również, że sieć propagacji wstecznej z jedną warstwą ukrytą może przybliżyć dowolną z bardziej ograniczonej klasy funkcji ciągłych. Podobnie, widzieliśmy w sekcji 11.3, że von Neumann stworzył automaty skończone, które były kompletne według Turinga. Tak więc sieci łącznikowe i automaty skończone wydają się być tylko dwiema innymi klasami algorytmów zdolnych do obliczania praktycznie dowolnej funkcji obliczalnej. Co więcej, biasy indukcyjne DOTYCZY mają zastosowanie do nienadzorowanych, a także genetycznych i nowych modeli uczenia się; reprezentacyjne odchylenia dotyczą projektowania węzłów, sieci i genomów, a algorytmiczne błędy dotyczą operatorów wyszukiwania, nagrody i selekcji. Co zatem mogą zaoferować uczący się bez nadzoru, czy to koneksjoniści, genetycy, czy też ewoluujące maszyny skończone w różnych formach?

1. Jedną z najbardziej atrakcyjnych cech uczenia się koneksjonistów jest to, że większość modeli opiera się na danych lub przykładach. Oznacza to, że nawet jeśli ich architektury są wyraźnie zaprojektowane, uczą się na przykładzie, generalizując dane z określonej domeny problemowej. Jednak wciąż pojawia się pytanie, czy dane są wystarczające lub wystarczająco czyste, aby nie zakłócać procesu rozwiązywania. A skąd projektant może wiedzieć?

2. Algorytmy genetyczne wspierają również potężne i elastyczne przeszukiwanie przestrzeni problemowej. Wyszukiwanie genetyczne jest napędzane zarówno przez różnorodność wymuszoną przez mutację, jak i przez operatory, takie jak krzyżowanie i inwersja, które zachowują ważne aspekty informacji rodziców dla przyszłych pokoleń. W jaki sposób projektant programu może zachować i pielęgnować ten kompromis między różnorodnością a ochroną?

3. Algorytmy genetyczne i architektury koneksjonistyczne można postrzegać jako przykłady przetwarzania równoległego i asynchronicznego. Czy rzeczywiście zapewniają wyniki poprzez równoległe asynchroniczne wysiłki, które nie są możliwe w przypadku jawnego programowania sekwencyjnego?

4. Chociaż inspiracja neuronalna i socjologiczna nie jest ważna dla wielu współczesnych praktyków uczenia się koneksjonizmu i genetyki, techniki te odzwierciedlają wiele ważnych aspektów naturalnej ewolucji i selekcji. Widzieliśmy modele uczenia się z redukcją błędów z perceptronem, wsteczną propagacją i modelami Hebbian. W sekcji 11.3.4 widzieliśmy również autosocjatywne sieci Hopfielda. Różne modele ewolucji znalazły odzwierciedlenie w paradygmatach Części 12.

5. Wreszcie, wszystkie paradygmaty uczenia się są narzędziami do badania empirycznego. Gdy wychwytujemy niezmienniki naszego świata, czy nasze narzędzia są wystarczająco potężne i ekspresyjne, aby zadawać dalsze pytania dotyczące natury percepcji, uczenia się i rozumienia?

W następnej części proponujemy, aby konstruktywistyczna epistemologia, w połączeniu z eksperymentalnymi metodami współczesnej sztucznej inteligencji.

Zbliżenie konstruktywisty

Konstruktywiści stawiają hipotezę, że wszelkie rozumienie jest wynikiem interakcji między wzorcami energii w świecie a kategoriami umysłowymi narzuconymi światu przez inteligentnego agenta. Korzystając z opisów Piageta, asymilujemy zjawiska zewnętrzne zgodnie z naszym obecnym

rozumieniem i dostosowujemy nasze rozumienie do „wymagań" zjawisk. Konstruktywiści często używają terminu schemata do opisania struktury a priori używanej do organizowania doświadczenia świata zewnętrznego. Termin ten pochodzi od brytyjskiego psychologa Bartletta, a jego filozoficzne korzenie sięgają Kanta. Z tego punktu widzenia obserwacja nie jest bierna i neutralna, ale aktywna i interpretująca. Postrzegana informacja, wiedza Kanta a posteriori, nigdy nie pasuje dokładnie do naszych z góry przyjętych, a priori, schematów. W wyniku tego napięcia uprzedzenia oparte na schemacie, których podmiot używa do organizowania doświadczenia, są modyfikowane lub zastępowane. Potrzeba akomodacji w obliczu nieudanych interakcji z otoczeniem napędza proces równowagi poznawczej. Zatem konstruktywistyczna epistemologia jest zasadniczo ewolucją i udoskonalaniem poznawczym. Ważną konsekwencją konstruktywizmu jest to, że interpretacja każdej sytuacji wiąże się z narzuceniem pojęć i kategorii obserwatora rzeczywistości (uprzedzenie indukcyjne). Kiedy Piaget zaproponował konstruktywistyczne podejście do rozumienia, nazwał je epistemologią genetyczną. Brak wygodnego dopasowania aktualnych schematów do świata „takiego, jakim jest", stwarza napięcie poznawcze. To napięcie napędza proces rewizji schematu. Rewizja schematu, akomodacja Piageta, jest ciągłą ewolucją rozumienia agenta w kierunku równowagi. Rewizja schematu i ciągły ruch w kierunku równowagi jest genetyczną predyspozycją czynnika dostosowującego się do struktur społeczeństwa i świata. Łączy obie te siły i reprezentuje ucieleśnioną predyspozycję do przetrwania. Modyfikacja schematu jest zarówno a priori wynikiem naszej genetyki, jak i funkcją a posteriori społeczeństwo i świat. Odzwierciedla ucieleśnienie agenta nastawionego na przetrwanie, istoty w czasie i przestrzeni. Występuje tu mieszanka tradycji empirystycznej i racjonalistycznej, w której pośredniczy cel przetrwania agentów. Wcielone podmioty nie mogą pojąć niczego poza tym, co najpierw przechodzi przez ich zmysły. Jako przychylni agenci przetrwają dzięki poznaniu ogólnych wzorców świata zewnętrznego. To, co jest postrzegane, jest zapośredniczane przez to, czego się oczekuje; to, czego się oczekuje, zależy od tego, co jest postrzegane: te dwie funkcje można zrozumieć tylko w kategoriach siebie. W tym sensie modele stochastyczne - zarówno bayesowskie, jak i markowskie - są właściwe, ponieważ wcześniejsze doświadczenie warunkuje obecne interpretacje. Wreszcie, my, jako agenci, rzadko zdajemy sobie sprawę ze schematów, które wspierają nasze interakcje ze światem. Jako źródła uprzedzeń i uprzedzeń zarówno w nauce, jak iw społeczeństwie, często jesteśmy świadomi a priori schemata. Są one konstytutywne dla naszej równowagi ze światem, a nie (zwykle) dostrzegalnym elementem świadomego życia psychicznego. Wreszcie, dlaczego konstruktywistyczna epistemologia jest szczególnie użyteczna w rozwiązywaniu problemów związanych ze zrozumieniem inteligencji? W jaki sposób agent w środowisku może zrozumieć swoje własne rozumienie tej sytuacji? Uważamy, że konstruktywizm odnosi się również do problemu epistemologicznego dostępu zarówno w filozofii, jak i psychologii. Od ponad wieku toczy się walka w obu tych dyscyplinach między dwiema frakcjami, pozytywistą, który proponuje wnioskowanie o zjawiskach psychicznych na podstawie obserwowalnych zachowań fizycznych, oraz podejściem bardziej fenomenologicznym, które pozwala na wykorzystanie raportów z pierwszej osoby w celu uzyskania dostępu do zjawisk poznawczych. Ten frakcyjność istnieje, ponieważ oba sposoby dostępu do zjawisk psychologicznych wymagają jakiejś formy konstrukcji modelu i wnioskowania. W porównaniu z przedmiotami fizycznymi, takimi jak krzesła i drzwi, które często naiwnie wydają się być bezpośrednio dostępne, stany psychiczne i dyspozycje agenta wydają się szczególnie trudne do scharakteryzowania. W istocie twierdzimy, że ta dychotomia między bezpośrednim dostępem do zjawisk fizycznych a pośrednim dostępem do mentalności jest iluzoryczna. Analiza konstruktywistyczna sugeruje, że żadne doświadczenie rzeczy nie jest możliwe bez użycia jakiegoś modelu lub schematu organizacji tego doświadczenia. W badaniach naukowych, jak również w naszych normalnych ludzkich doświadczeniach, sugeruje to, że wszelki dostęp do zjawisk odbywa się poprzez eksplorację, przybliżanie i ciągłe udoskonalanie modelu.ferowała narzędzia i techniki do kontynuowania eksploracji nauki o inteligentnych systemach.

Jaki jest więc projekt praktyka AI?

Jako praktycy AI jesteśmy konstruktywistami. Budujemy, testujemy i udoskonalamy modele. Ale co przybliżamy w naszym budowaniu modeli? Omówimy tę kwestię w następnych akapitach, ale najpierw dokonamy obserwacji epistemologicznej: Zamiast próbować uchwycić istotę „rzeczy poza nami", rozwiązaniu problemów AI najlepiej służy próba naśladowania budowy modelu, udoskonalania i heurystyki równoważenia samego inteligentnego agenta. Tylko skrajny solipsysta (lub umysłowo upośledzony) zaprzeczyłby „rzeczywistości" świata pozaprzedmiotowego. Ale co to jest tak zwany „prawdziwy świat"? Oprócz tego, że jest złożoną kombinacją „rzeczy twardych" i „rzeczy miękkich", jest to także układ atomów, cząsteczek, kwarków, grawitacji, teorii względności, komórek, DNA i (być może nawet) superstrun. Wszystkie te koncepcje są jedynie modelami eksploracyjnymi napędzanymi przez wyjaśniające wymagania czynników opartych na równościach. Ponownie, te modele eksploracyjne nie dotyczą świata zewnętrznego. Raczej chwytają dynamiczne, równoważące napięcia inteligentnego i społecznego sprawcy, materialnej inteligencji ewoluującej i nieustannie kalibrującej się w przestrzeni i czasie. Ale dostęp do i tworzenie „rzeczywistości" jest również osiągane poprzez zaangażowanie agentów. Wcielony podmiot tworzy rzeczywistość poprzez egzystencjalną afirmację, że postrzegany model jego oczekiwań jest wystarczająco dobry, aby zaspokoić niektóre z jego praktycznych potrzeb i celów. Ten akt zaangażowania ugruntowuje symbole i systemy symboli, których agent używa w swoim materialnym i społecznym kontekście. Konstrukty te są ugruntowane, ponieważ są uznane za wystarczająco dobre, aby osiągnąć pewne aspekty swojego celu. To uziemienie jest również widoczne w używaniu języka agentów. Searle ma rację w swoim pojmowaniu zjawisk mowy jako aktów. Ta kwestia uziemienia jest jednym z powodów, dla których komputery mają fundamentalne problemy z przejawami inteligencji, w tym z demonstracjami języka i uczenia się. Jaką dyspozycję można dać komputerowi, który zapewni mu odpowiednie cele i cele? Chociaż Dennett przypisywał uziemienie komputerowi rozwiązującemu problemy wymagające i używające „inteligencji", brak wystarczającego uziemienia jest łatwo dostrzegalny w uproszczeniach komputera, kruchości i często ograniczonej ocenie kontekstu. Użycie i uziemienie symboli przez ożywionych agentów oznacza jeszcze więcej. Szczegóły dotyczące ucieleśnienia i kontekstów społecznych agenta ludzkiego pośredniczą w jego interakcjach ze światem. Systemy słuchowe i wizualne wrażliwe na określone pasmo; postrzeganie świata jako wyprostowanego dwunożnego, mającego ręce, nogi, dłonie; przebywanie w świecie z pogodą, porami roku, słońcem i ciemnością; część społeczeństwa o zmieniających się celach i celach; osoba, która rodzi się, rozmnaża i umiera: są to krytyczne elementy wspierające metafory rozumienia, uczenia się i języka; one pośredniczą w naszym pojmowaniu sztuki, życia i miłości.

Czy mam cię porównać do letniego dnia?

Jesteś piękniejszy i bardziej umiarkowany:

Ostry wiatr wstrząsa ukochanymi pąkami maja,

A letnia dzierżawa ma zbyt krótką datę ...

Szekspirowski Sonet XVIII Kończymy podsumowaniem krytycznych zagadnień, które wspierają i ograniczają nasze wysiłki w tworzeniu nauki o inteligentnych systemach.

16.3 AI: obecne wyzwania i przyszłe kierunki

Chociaż wykorzystanie technik sztucznej inteligencji do rozwiązywania problemów praktycznych dowiodło swojej użyteczności, wykorzystanie tych technik do stworzenia ogólnej nauki o inteligencji jest trudnym i ciągłym problemem. W tej ostatniej części wracamy do pytań, które doprowadziły nas

do wejścia w dziedzinę sztucznej inteligencji i do napisania tej książki: czy możliwe jest formalne, obliczeniowe ujęcie procesów, które umożliwiają inteligencję? Obliczeniowa charakterystyka inteligencji rozpoczyna się od abstrakcyjnej specyfikacji urządzeń obliczeniowych. Badania w latach trzydziestych, czterdziestych i pięćdziesiątych XX wieku zapoczątkowały to zadanie, a Turing, Post, Markov i Church wnieśli wkład w formalizm opisujący obliczenia. Celem tych badań było nie tylko sprecyzowanie, co oznacza obliczenie, ale raczej określenie granic tego, co można obliczyć. Universal Turing Machine jest najczęściej badaną specyfikacją, chociaż reguły przepisywania Posta, podstawa obliczania systemu produkcyjnego (po 1943), są również ważnym wkładem. Model Churcha, oparty na funkcjach częściowo rekurencyjnych, oferuje obsługę nowoczesnych języków funkcjonalnych wysokiego poziomu, takich jak Scheme, Ocaml i Standard ML. Teoretycy udowodnili, że wszystkie te formalizmy mają równoważną moc obliczeniową w tym sensie, że każda funkcja obliczalna przez jednego jest obliczalna przez inne. W rzeczywistości można wykazać, że uniwersalna maszyna Turinga jest odpowiednikiem każdej nowoczesnej maszyny obliczeniowej. Opierając się na tych wynikach, hipoteza Church-Turinga wysuwa jeszcze silniejszy argument: nie można zdefiniować żadnego modelu obliczeń, który byłby silniejszy niż te znane modele. Po ustaleniu równoważności specyfikacji obliczeniowych uwolniliśmy się od środka mechanizacji tych specyfikacji: możemy zaimplementować nasze algorytmy za pomocą lamp próżniowych, krzemu, protoplazmy lub zabawek druciarza. Zautomatyzowany projekt w jednym medium może być postrzegany jako odpowiednik mechanizmów w innym. To sprawia, że metoda badania empirycznego jest jeszcze bardziej krytyczna, ponieważ eksperymentujemy na jednym medium aby sprawdzić nasze zrozumienie mechanizmów zaimplementowanych w innym. Jedną z możliwości jest to, że uniwersalna maszyna Turinga i Posta może być zbyt ogólna. Paradoksalnie, inteligencja może wymagać mniej wydajnego mechanizmu obliczeniowego z bardziej skoncentrowaną kontrolą. Levesque i Brachman zasugerowali ,że inteligencja może wymagać bardziej wydajnych obliczeniowo (chociaż mniej wyrazistych) reprezentacji, takich jak klauzule Horn dla rozumowania, ograniczenie wiedzy faktycznej do literałów naziemnych oraz wykorzystanie obliczeniowo wykonalnych systemów utrzymywania prawdy. Oparte na agentach i wyłaniające się modele inteligencji również wydają się wspierać tę filozofię. Kolejną kwestią, do której odnosi się formalna równoważność naszych modeli mechanizmów, jest kwestia dwoistości i problem ciała i umysłu. Przynajmniej od czasów Kartezjusza , filozofowie postawili pytanie o interakcję i integrację umysłu, świadomości i ciała fizycznego. Filozofowie zaproponowali każdą możliwą odpowiedź, od całkowitego materializmu po zaprzeczenie egzystencji materialnej, a nawet wspierającą interwencję łagodnego boga! Badania nad sztuczną inteligencją i kognitywistyką odrzucają kartezjański dualizm na rzecz materialnego modelu umysłu opartego na fizycznej implementacji lub konkretyzacji symboli, formalnej specyfikacji mechanizmów obliczeniowych służących do manipulacji tymi symbolami, równoważności paradygmatów reprezentacji oraz mechanizacji wiedzy i umiejętności w modelach ucieleśnionych. Sukces tych badań wskazuje na słuszność tego modelu . Wiele dalszych pytań pozostaje jednak w epistemologicznych podstawach inteligencji w systemie fizycznym. Podsumowując: po raz ostatni kilka z tych krytycznych kwestii.

1. Problem reprezentacji. Newell i Simon postawili hipotezę, że fizyczny system symboli i wyszukiwanie są niezbędnymi i wystarczającymi cechami inteligencji (patrz Rozdział 16.1). Czy sukcesy modeli neuronowych lub sub-symbolicznych oraz genetycznych i wyłaniających się podejść do inteligencji obalają hipotezę dotyczącą fizycznego symbolu, czy są to po prostu inne jej przykłady? Nawet słaba interpretacja tej hipotezy - że symbol fizyczny system jest wystarczającym modelem dla inteligencji - przyniósł wiele potężnych i użytecznych wyników w nowoczesnej dziedzinie kognitywistyki. To dowodzi, że możemy zaimplementować fizyczne systemy symboli, które zademonstrują inteligentne zachowanie. Wystarczalność pozwala na tworzenie i testowanie modeli opartych na symbolach dla

wielu aspektów ludzkiej wydajności (Pylyshyn 1984, Posner 1989). Ale silna interpretacja - że fizyczny system symboli i poszukiwania są niezbędne dla inteligentnej aktywności - pozostaje otwarta

2. Rola ucieleśnienia w poznaniu. Jednym z głównych założeń hipotezy systemu symboli fizycznych jest to, że konkretna instancja fizycznego systemu symboli nie ma znaczenia dla jego działania; liczy się tylko jego struktura formalna. Zostało to zakwestionowane przez wielu myślicieli, którzy zasadniczo argumentują, że wymagania inteligentnego działania na świecie wymagają fizycznego ucieleśnienia, które pozwala agentowi na pełną integrację z tym światem. Architektura współczesnych komputerów nie wspiera tego stopnia usytuowania, wymagając od sztucznej inteligencji interakcji ze swoim światem przez niezwykle ograniczone okno współczesnych urządzeń wejścia / wyjścia. Jeśli to wyzwanie jest słuszne, to chociaż jakaś forma inteligencji maszynowej może być możliwa, będzie wymagała zupełnie innego interfejsu niż ten oferowany przez współczesne komputery.

3. Kultura i inteligencja. Tradycyjnie sztuczna inteligencja skupiała się na indywidualnym umyśle jako jedynym źródle inteligencji; zachowywaliśmy się tak, jakby wyjaśnienie sposobu, w jaki mózg koduje wiedzę i manipuluje nią, byłoby całkowitym wyjaśnieniem pochodzenia inteligencji. Moglibyśmy jednak również argumentować, że wiedzę najlepiej postrzegać jako konstrukcję społeczną, a nie indywidualną. W opartej na memach teorii inteligencji (Edelman 1992) samo społeczeństwo posiada podstawowe składniki inteligencji. Możliwe, że zrozumienie społecznego kontekstu wiedzy i ludzkich zachowań jest tak samo ważne dla teorii inteligencji, jak zrozumienie dynamiki indywidualnego umysłu / mózgu.

4. Charakterystyka charakteru interpretacji. Większość modeli obliczeniowych w tradycji reprezentacji działa z już zinterpretowaną domeną: to znaczy, istnieje ukryte i a priori zaangażowanie projektantów systemu w kontekst interpretacyjny. W ramach tego zobowiązania istnieje niewielka możliwość zmiany kontekstów, celów, lub reprezentacje w miarę ewolucji rozwiązywania problemu. Obecnie niewiele jest wysiłku, aby naświetlić proces, za pomocą którego ludzie konstruują interpretacje. Tarskianowski pogląd na semantykę jako odwzorowanie między symbolami i przedmiotami w dziedzinie dyskursu jest z pewnością zbyt słaby i nie wyjaśnia na przykład faktu, że jedna dziedzina może mieć różne interpretacje w świetle różnych celów praktycznych. Lingwiści próbowali zaradzić ograniczeniom semantyki Tarskiana, dodając teorię pragmatyki (Austin 1962). Analiza dyskursu, z jej fundamentalną zależnością od użycia symboli w kontekście, zajmowała się tymi kwestiami w ostatnich latach. Problem jest jednak szerszy, ponieważ dotyczy ogólnej awarii narzędzi referencyjnych. Tradycja semiotyczna zapoczątkowana przez C. S. Peirce'a i kontynuowana przez Eco, Seboek i innych przyjmuje bardziej radykalne podejście do języka. Umieszcza symboliczne wyrażenia w szerszym kontekście znaków i interpretacja znaków. Sugeruje to, że znaczenie symbolu można zrozumieć jedynie w kontekście jego roli interpretatora, czyli w kontekście interpretacji i interakcji z otoczeniem

5. Nieokreśloność reprezentacyjna. Przypuszczenie Andersona o nieokreśloności reprezentacji sugeruje, że w zasadzie niemożliwe może być ustalenie, który schemat reprezentacji najlepiej przybliża człowieka rozwiązującego problemy w kontekście konkretnego aktu umiejętnego wykonania. To przypuszczenie opiera się na fakcie, że każdy schemat reprezentacji jest nierozerwalnie związany z większą architekturą obliczeniową, a także strategiami wyszukiwania. W szczegółowej analizie umiejętności ludzkich może być niemożliwe kontrolowanie procesu na tyle, abyśmy mogli określić reprezentację; lub ustal reprezentację do punktu, w którym proces może być określony w sposób unikalny. Podobnie jak w przypadku zasady nieoznaczoności w fizyce, gdzie zjawiska mogą być zmieniane przez sam proces ich pomiaru, jest to ważna kwestia przy konstruowaniu modeli inteligencji, ale nie musi ograniczać ich użyteczności. Ale co ważniejsze, tę samą krytykę można skierować na sam model obliczeniowy, w którym indukcyjne uprzedzenia symboli i poszukiwań w kontekście hipotezy Churcha-Turinga wciąż ograniczają system. Postrzegana potrzeba jakiegoś optymalnego schematu

reprezentacji może równie dobrze być pozostałością marzenia racjonalisty, podczas gdy naukowiec po prostu potrzebuje modeli dostatecznie solidnych, aby ograniczyć pytania empiryczne. Dowodem jakości modelu jest jego zdolność do interpretacji, przewidywania i rewizji.

6. Konieczność projektowania modeli obliczeniowych, które są falsyfikowalne. Popper i inni argumentowali, że teorie naukowe muszą być falsyfikowalne. Oznacza to, że muszą zaistnieć okoliczności, w których model nie będzie skutecznym przybliżeniem zjawiska. Oczywistym powodem jest to, że jakakolwiek liczba potwierdzających instancji eksperymentalnych nie jest wystarczająca do potwierdzenia modelu. Ponadto wiele nowych badań jest podejmowanych w bezpośredniej odpowiedzi na niepowodzenie istniejących teorii. Ogólny charakter hipotezy fizycznego systemu symboli, a także umiejscowienia a pojawiające się modele inteligencji mogą uniemożliwić ich sfałszowanie, a tym samym ich ograniczone wykorzystanie jako modeli. Ta sama krytyka może dotyczyć przypuszczeń tradycji fenomenologicznej (patrz punkt 7). Niektóre struktury danych AI, takie jak sieć semantyczna, są tak ogólne, że mogą modelować prawie wszystko, co można opisać lub, jak w przypadku uniwersalnej maszyny Turinga, dowolną obliczalną funkcję. Tak więc, gdy badacz AI lub kognitywisty zostanie zapytany, w jakich warunkach jego model inteligencji nie zadziała, odpowiedź może być trudna.

7. Ograniczenia metody naukowej. Wielu badaczy twierdzi, że najważniejsze aspekty inteligencji nie są i w zasadzie nie mogą być modelowane, a w szczególności nie mają reprezentacji symbolicznej. Obszary te obejmują uczenie się, rozumienie języka naturalnego i tworzenie aktów mowy. Kwestie te mają głębokie korzenie w naszej tradycji filozoficznej. Na przykład krytyka Winograda i Floresa opiera się na zagadnieniach fenomenologii .

Większość założeń współczesnej sztucznej inteligencji ma swoje korzenie od Carnapa, Frege i Leibniza, poprzez Hobbesa, Locke'a i Hume'a do Arystotelesa. Ta tradycja twierdzi, że inteligentne procesy są zgodne z uniwersalnymi prawami i są w zasadzie zrozumiałe. Heidegger i jego zwolennicy reprezentują alternatywne podejście do rozumienia inteligencji. Heidegger uważa, że refleksyjna świadomość opiera się na świecie ucieleśnionego doświadczenia (świat życia). To stanowisko, podzielane przez Winograda i Floresa, Dreyfusa i innych, dowodzi, że ludzkie rozumienie rzeczy jest zakorzenione w praktycznej działalności polegającej na „używaniu" ich w radzeniu sobie z codziennym światem. Ten świat jest zasadniczo kontekstem społecznie zorganizowanych ról i celów. Ten kontekst i ludzkie w nim funkcjonowanie nie jest czymś wyjaśnianym przez zdania i rozumianym przez twierdzenia. Jest to raczej strumień, który kształtuje i sam jest nieustannie tworzony. W podstawowym sensie, ludzka wiedza nie polega na wiedzy, ale raczej w świecie ewoluujących norm społecznych i ukrytych celów, wiedząc, jak to zrobić. Z natury nie jesteśmy w stanie umieścić naszej wiedzy i większości naszych inteligentnych zachowań w języku, ani formalnym, ani naturalnym. Rozważmy ten punkt widzenia. Po pierwsze, jako krytyka czystej tradycji racjonalistycznej jest słuszna. Racjonalizm twierdzi, że wszelka ludzka działalność, inteligencja i odpowiedzialność mogą, przynajmniej w zasadzie, być reprezentowane, sformalizowane i zrozumiane. Większość refleksyjnych ludzi nie wierzy, że tak jest, rezerwując ważne role dla emocji, autoafirmacji i odpowiedzialnego zaangażowania (przynajmniej!). Sam Arystoteles powiedział w swoim eseju o racjonalnym działaniu: „Dlaczego nie czuję się zmuszony do wykonania tego, co się z tym wiąże?" Istnieje wiele działań człowieka poza dziedziną nauki, które odgrywają zasadniczą rolę w odpowiedzialnych interakcjach międzyludzkich; nie można ich powielać ani przenosić na maszyny. Mając to jednak na uwadze, naukowa tradycja badania danych, konstruowania modeli, przeprowadzanie eksperymentów i badanie wyników wraz z udoskonalaniem modeli do dalszych eksperymentów przyniosło społeczności ludzkiej istotny poziom zrozumienia, wyjaśnienia i zdolności przewidywania. Metoda naukowa to potężne narzędzie zwiększające ludzkie zrozumienie. Niemniej jednak istnieje wiele zastrzeżeń dotyczących tego podejścia, które naukowcy

muszą zrozumieć. Po pierwsze, naukowcom nie wolno mylić modelu z modelowanym zjawiskiem. Model pozwala nam stopniowo przybliżać zjawisko: z konieczności zawsze będzie istnieć „pozostałość", której nie da się empirycznie wyjaśnić. W tym sensie również nieokreśloność reprezentacji nie jest problemem. Model służy do eksploracji, wyjaśniania i przewidywania; a jeśli pozwoli to naukowcom to osiągnąć, to się powiedzie. Rzeczywiście, różne modele mogą z powodzeniem wyjaśniać różne aspekty zjawiska, takie jak teorie fal i cząstek światła. Co więcej, kiedy badacze twierdzą, że aspekty inteligentnych zjawisk są poza zakresem i metodami tradycji naukowej, samo to stwierdzenie można zweryfikować tylko przy użyciu tej właśnie tradycji. Metoda naukowa jest jedynym narzędziem, jakim dysponujemy, aby wyjaśnić, w jakim sensie kwestie mogą nadal pozostawać poza naszym obecnym zrozumieniem. Każdy punkt widzenia, nawet ten z tradycji fenomenologicznej, jeśli ma mieć jakiekolwiek znaczenie, musi odnosić się do naszych aktualnych pojęć wyjaśniania - a nawet być spójny co do zakresu, w jakim zjawiska nie mogą być wyjaśnione. Najbardziej ekscytującym aspektem pracy nad sztuczną inteligencją jest to, że aby zachować spójność i wnieść wkład w przedsięwzięcie, musimy zająć się tymi kwestiami. Aby zrozumieć rozwiązywanie problemów, uczenie się i język, musimy zrozumieć filozoficzny poziom reprezentacji i wiedzy. W pokorny sposób jesteśmy proszeni o rozwiązanie napięcia Arystotelesa między teoriami i praktyki, aby stworzyć związek zrozumienia i praktyki, teorii i praktyki, by żyć między nauką a sztuką. Praktycy sztucznej inteligencji są twórcami narzędzi. Nasze reprezentacje, algorytmy i języki są narzędziami do projektowania i budowania mechanizmów wykazujących inteligentne zachowanie. Poprzez eksperyment sprawdzamy zarówno ich obliczeniową adekwatność do rozwiązywania problemów, jak i nasze własne rozumienie inteligentnych zjawisk.