

Architektury kognitywne, czyli jak zbudować sztuczny umysł

14.1. Wstęp

Długofalowym zadaniem stojącym przed badaniami nad sztuczną inteligencją jest stworzenie systemów, które osiągną poziom kompetencji przekraczający możliwości ludzkie w różnych dziedzinach¹. W ograniczonym zakresie jest to już oczywiście możliwe: komputery szybciej liczą, lepiej pamiętają, analizują modele zbyt trudne dla człowieka z powodu złożoności procesów, dużej ilości danych lub liczby zmiennych poddawanych optymalizacji. Komputery dokładniej rozpoznają specyficzne wzorce w sygnałach, lepiej niż ludzie radzą sobie w wielu grach planszowych, w sposób doskonalszy niż my prowadzą manewry sterując helikopterami, samolotami i pojazdami naziemnymi. A jednak pomimo ciągłego postępu w wielu dziedzinach sztuczna inteligencja jak dotąd w znacznej mierze zawiodła pokładane w niej nadzieje. Stosunkowo proste niższe funkcje poznawcze związane z percepcją, rozpoznawaniem obiektów i analizą relacji między nimi w wykonaniu komputerów nadal pozostają daleko w tyle za możliwościami nawet prymitywnych gatunków zwierząt. Znaczny postęp nastąpił jedynie na poziomie stosunkowo prostej analizy wzorców i przetwarzania sygnałów. Wyższe czynności poznawcze, związane z użyciem rozpoznanych wzorców – słów, symboli, obiektów stanowiących podstawę wnioskowania, analizy tekstów, dialogu w języku naturalnym, planowania i rozwiązywania problemów, tworzenia i wykorzystywania reprezentacji złożonych form wiedzy – okazały się znacznie trudniejsze do symulacji. Nie wystarczą tu proste inspiracje biologiczne, dotyczące działania pojedynczych neuronów lub ich niewielkich grup. Konieczne jest uwzględnienie wielkoskalowej architektury mózgu i tworzenie modeli funkcji wyspecjalizowanych obszarów.

¹ Dobrym wprowadzeniem do problematyki poruszanej w tym rozdziale byłaby lektura innych rozdziałów tej książki oraz takich pozycji, jak: Hawkins J., Blakeslee S. (2006) *Istota inteligencji*. Helion, Gliwice; Kasperski M.J. (2003) *Sztuczna Inteligencja. Droga do myślących maszyn*. Helion, Gliwice; Edelman G.M. (1999) *Przenikliwe powietrze, jasny ogień. O materii umysłu*. PIW, Warszawa.

14.2. Koncepcja informatyki neurokognitywnej

Opisane poniżej podejście można określić mianem **informatyki neurokognitywnej**. Funkcje poznawcze wspomagane są w niej przez pamięć różnego typu:

- pamięć **rozpoznawczą**, pozwalającą na identyfikację znanych obiektów lub dostrzeżenie odstępstw od oczekiwań;
- pamięć **skojarzeniową**, prowadzącą automatycznie do prostych wniosków, zapewniającą realizację procesów warunkowania klasycznego;
- pamięć **proceduralną**, czyli pamięć umiejętności manualnych i sekwencji działań;
- pamięć **semantyczną**, pozwalającą na interpretację sensu oraz na dostęp do złożonych struktur wiedzy;
- pamięć **roboczą**, pozwalającą na łączenie ze sobą w kombinatoryczny sposób różnych fragmentów informacji w większe całości.

Wszystkie te rodzaje pamięci wspierane są przez sieci neuronowe o specyficznej architekturze, dostosowanej do wymaganych funkcji (O'Reilly i Munakata 2000). Warto podkreślić, że organizacja przechowywania i dostępu do informacji przez mózg i komputer różni się na bardzo podstawowym poziomie. Matematyczne dowody uniwersalności różnych systemów obliczeniowych (maszyny Turinga, sieci neuronowych) do pewnego stopnia zaślepiły badaczy, jednak dowody te odnoszą się do tworów teoretycznych, tymczasem znane nam praktyczne architektury obliczeniowe ograniczają na wiele sposobów możliwości przetwarzania informacji. W niektórych zastosowaniach komputery przetwarzają informacje lepiej niż ludzki mózg, ale nadal w wielu ważnych zastosowaniach jest on niedoścignionym wzorem. Nie wiemy, czy architektury kognitywne, zaimplementowane na konwencjonalnych komputerach, kiedykolwiek osiągną podobne do biologicznego mózgu możliwości zarówno pod względem niższych, jak i wyższych czynności poznawczych.

14.3. Wyższe i niższe czynności poznawcze

Sztuczna inteligencja (*artificial intelligence*, AI) zajmowała się tradycyjnie wyższymi czynnościami poznawczymi: rozwiązywaniem problemów, myśleniem, reprezentacją wiedzy, analizą języka naturalnego (Russell i Norvig 2003). Próbowano tego dokonać w oderwaniu od mechanizmów biologicznych, opierając się na symbolicznych modelach złożonych form wiedzy, a więc niejako na poziomie racjonalnego umysłu, czyli reprezentacji wiedzy dającej się poznać w świadomy sposób. Cechą charakterystyczną takiego podejścia jest skupienie się na procesach sekwencyjnych, w których złożoność kombinatoryczna, wynikająca z możliwości zastosowania różnych transformacji przydatnych w rozwiązywaniu problemu, wymaga rozważenia bardzo wielu możliwości i wyboru odpowiedniej strategii poszukiwania rozwiązania. Podstawę dla działań inteligentnych stanowią w tym ujęciu algorytmy heurystycznego szukania rozwiązań i symbolicznej reprezentacji wiedzy (Newell 1990).

Niższe czynności poznawcze, realizowane przez układy biologiczne w znacznie krótszej skali czasowej w sposób całkowicie równoległy, umożliwiają percepcję, postrzeganie obiektów, kategoryzację, szybkie skojarzenia, aktywność sensomotoryczną.

W technice takie czynności są podstawą działania robotów. Inteligentne roboty mają się rozwijać podobnie jak dzieci, przyswajając sobie sens symbolicznych przekazów i poleceń poprzez działanie w realnym świecie. Tendencje rozwojowe w tej dziedzinie doprowadziły do upowszechnienia się filozofii „ucieleśnienia poznania” (*embodied cognition*) oraz enaktywizmu (Barsalou 2008), zgodnie z którą sens wszystkich symboli określać mają specyficzne sekwencje działań sensomotorycznych.

W dziedzinie tej jest wiele problemów – szczególnie ze zrozumieniem, w jaki sposób tworzą się abstrakcyjne reprezentacje mentalne oparte na reprezentacjach percepcyjno-ruchowych (Mahon i Caramazza 2008). Symboliczne aproksymacje działań sensomotorycznych mają bez wątpienia silne ograniczenia, podobnie jak wszelkie próby dyskretyzacji procesów ciągłych. Nie wiemy, jak daleko trzeba pójść w stronę „ucieleśnionych reprezentacji” podstawowych procesów, by sens, nadany pierwotnym koncepcjom poprzez działanie w realnym świecie, był wystarczający do stworzenia na tej podstawie bardziej abstrakcyjnych koncepcji (pisałem o tym już w: Duch 1994, 1996). Być może za wcześnie porzucać marzenia Newella i Simona (1963) o konstrukcji Ogólnego Rozwiązywacza Problemów (*General Problem Solver*, GPS) na poziomie manipulacji symbolami, chociaż nie da się zapewne uniknąć konieczności stworzenia rozszerzeń reprezentacji symbolicznych umożliwiających lepszą aproksymację procesów skojarzeniowych. Zastosowania w robotyce i sterowaniu czy choćby do zrozumienia zachowania organizmów biologicznych mogą wymagać dokładniejszych modeli, pozwalających na „ucieleśnienie poznania” na głębszym poziomie.

14.4. Wielkie wyzwania

14.4.1. Próby stworzenia ogólnej inteligencji

Wiele ciekawych koncepcji sprawdzających się w przypadku prostych problemów nie działa w rzeczywistych, złożonych zastosowaniach. Pojawiła się ostatnio tendencja do budowy ogólnej sztucznej inteligencji (*Artificial General Intelligence*, AGI), a więc systemów, które nie tylko potrafią rozwiązać jedno zadanie na poziomie mistrzowskim (np. grać w szachy lepiej od mistrza świata), lecz dają się zastosować do całej klasy podobnych problemów (np. nauczyć się innych gier i osiągnąć w nich wysoki poziom). Lista wyzwań, stojących przed inteligencją obliczeniową, jest długa (Duch i Mandziuk 2007). Analiza istniejących architektur poznawczych powinna pomóc w zrozumieniu ich ograniczeń i zaproponowaniu sposobów ich przezwyciężenia.

Czego należy oczekiwać od architektur poznawczych, by można je uznać za coś więcej niż wyspecjalizowane programy udające inteligencję?

AI skupiła się na praktycznych problemach tworzenia systemów doradczych, prawie całkowicie porzucając swoje pierwotne ambitne cele, takie jak stworzenie GPS² (Newell i Simon 1963). Okazało się, że nie wystarczy do tego wiedza ogólna, rozwiązywanie problemów wymaga bardzo obszernej wiedzy w wielu dziedzinach. Projekt Cyc Douglasa Lenata stworzenia systemu zawierającego encyklopedyczną wiedzę

² Czyli programu *General Problem Solver* (a nie systemu nawigacji satelitarnej).

(nazwa „Cyc” jest fragmentem angielskiego słowa enCYClopedia), konieczną do realizacji zdrowego rozsądku, jest jak dotychczas jedyną taką próbą (Panton i in. 2006), realizowaną na podstawie klasycznych metod AI.

Projekt ten rozpoczęto w 1984 roku, konstruując ogromną bazę wiedzy opartą na złożonych strukturach, nazywanych ramami. Ontologia systemu Cyc zawiera ponad 300 tysięcy pojęć i wiele milionów łączących je faktów (OpenCyc w wersji 1.0 miał ponad 3 miliony asercji dotyczących tych pojęć, częściowo ograniczających i definiujących wzajemne relacje ponad 26 tysięcy typów), a jego opracowanie kosztowało ponad 600 osobołat pracy.

Rezultaty nie są zachwycające i nadal brakuje ambitnych zastosowań, które pokazałyby, że tak złożony system istotnie daje się do czegoś użyć, chociaż zawarta w nim wiedza jest ostatnio wykorzystywana przez projekty Internetu semantycznego (przetwarzania informacji tekstowych dzięki zrozumieniu ich sensu). Lista „potencjalnych zastosowań” Cyc jest długa, ale tylko CycSecure, program do analizy zabezpieczeń sieci komputerowych, jest realnym zastosowaniem, prezentowanym na oficjalnej stronie www.cyc.com firmy Cycorp.

14.4.2. Dialog w języku naturalnym

AI zawiodła w wielu dziedzinach, ale rzeczywistość chyba najbardziej odbiega od oczekiwań w dziedzinie systemów do dialogu w języku naturalnym i w innych zagadnieniach dotyczących rozumienia języka. Napisano wiele programów do konwersacji, które nazywają się botami lub chatterbotami (od nazwy ro-bot powstały nazwy typu soft-bot, skrócone do bot). Programy te znajdują zastosowanie w firmach komercyjnych i służą do odpowiadania na pytania dotyczące działalności firmy. Test Turinga (1950), polegający na konwersacji z ludźmi i botami na niczym nieograniczony temat, jest częściowo stosowany w konkursach Loebnera (<http://www.loebner.net>). Od prawie 20 lat postęp jest jednak niewielki, w większości przypadków stosuje się bowiem techniki oparte na dopasowaniu odpowiedzi do szablonów pytań (Wallace 2003), znane od ponad pół wieku.

Programy oparte na szablonach nie mają pojęcia, jaki jest sens koncepcji pojawiających się w dialogu, nie mają też żadnego strukturalnego opisu pojęć. Dlatego można ich co najwyżej użyć do stereotypowej konwersacji w systemach odpowiadających na pytania w zaprogramowanej z góry wąskiej dziedzinie. Niektórzy sędziowie w konkursie Loebnera dają się oszukać programom przejmującym inicjatywę prowadzenia konwersacji, zadającym dużo pytań, dzięki czemu łatwiej im zawęzić oczekiwania co do intencji rozmówcy i analizować odpowiedzi dyskutantów za pomocą szablonów. Jest to jednak trik nastawiony na to, żeby bot sprawiał lepsze wrażenie, podczas gdy semantyczna analiza tekstu wypowiedzi rozmówcy (człowieka), decydująca o jego zrozumieniu, pozostaje bardzo prymitywna.

Proponowano liczne zmodyfikowane wersje testu Turinga, na przykład Carpenter i Freeman (2005) piszą o „spersonalizowanym teście”, w którym program ma się wcielić w osobę znaną rozmówcy. Jest to więc próba określenia, na ile skutecznie można zrobić model określonego umysłu, a nie jakiś ogólny, abstrakcyjny program do konwersacji. Wyniki w tym zakresie są jednak także dalekie od takich, które można by uznać za zadowalające.

Ciekawe wyzwania dla AI stwarzają gry słowne, a w szczególności znana gra w 20 pytań, stanowiąca interesujący paradygmat dla wielu innych zastosowań. Gry słowne wymagają szerokiej wiedzy o znaczeniu słów oraz o własnościach obiektów wskazywanych przez te słowa, ale nie wymagają pełnej wiedzy o złożonych relacjach między tymi obiektami. W takich zastosowaniach okazuje się, że nie ma jednej, uniwersalnej i optymalnej metody reprezentacji wiedzy. Złożone formy wiedzy, na przykład ramy Cyc, trudno szybko przeszukiwać, jeśli chcemy określić, jakie pytanie warto zadać, by zdobyć maksymalnie dużo informacji. Proste wektorowe metody reprezentacji informacji mogą w tym przypadku być bardziej przydatne (Szymański i in. 2008). Jednakże przy tych reprezentacjach nadal potrzebna jest obszerna wiedza o tysiącach różnych koncepcji, a niełatwo ją automatycznie stworzyć, mimo że dostępnych jest obecnie sporo ontologii, słowników, encyklopedii i słabiej ustrukturalizowanych źródeł wiedzy.

Systemy do odpowiadania na pytania (*Q/A systems*) są jeszcze bardziej wymagające, gdyż nie wystarczy w nich zrozumienie poszczególnych słów, lecz potrzebne jest zrozumienie sensu całego zdania. Seria konferencji TREC (*Text Retrieval Conference*)³ organizuje konkursy dla systemów wyszukiwawczych, które mają znaleźć w Internecie (lub w swoich bazach wiedzy) odpowiedzi na zbiór pytań konkursowych. Innym ważnym wyzwaniem są inteligentne systemy wspomagające nauczanie (*Intelligent Tutoring Systems, ITS*), chociaż nie ma w tej dziedzinie jednoznacznej metody na określenie postępów.

14.4.3. Systemy doradcze nowej generacji

Budowa systemów doradczych, które mogłyby służyć jako partnerzy dla ludzi, jest nadal wielkim wyzwaniem (Feigenbaum 2003). Chodzi tu nie tyle o klasyczne systemy doradcze, oparte na bazach wiedzy pozyskanej od ekspertów, co o systemy obdarzone szerszą inteligencją, pozwalające na łatwiejszą komunikację z człowiekiem, rozumujące i podpowiadające interesujące skojarzenia. Potrzebę stosowania takich systemów lokuje się w takich dziedzinach, jak matematyka, nauki o życiu czy prawo. Rezultaty łatwo ocenić: panel ekspertów może przeprowadzić egzamin, zadając pytania i domagając się szczegółowego uzasadnienia odpowiedzi, w celu oceny stopnia zrozumienia danego problemu.

Zadanie nie jest łatwe. Potrzebne są tu nie tylko moduły prowadzące ściśle wnioskowanie logiczne, ale i modele skojarzeniowe, zauważające interesujące korelacje i związki przyczynowe. Interesującym źródłem doświadczeń w omawianym tu obszarze są konkursy automatycznego dowodzenia twierdzeń matematycznych, odbywające się na konferencjach CADE (*Conference on Automated Deduction*)⁴ w wielu specjalistycznych kategoriach. Rozszerzenie możliwości takich wyspecjalizowanych modułów na metapoziom, pozwalający na komunikację z użytkownikiem i wybór odpowiedniego modułu do danego problemu, jest drogą do stworzenia interesującego „sztucznie inteligentnego” partnera dla matematyka. Inną dziedziną, w której odczuwa się zapotrzebowanie na automatyczne systemy wnioskujące, jest bioinfor-

³ <http://trec.nist.gov/>

⁴ <http://www.cs.miami.edu/~tptp/CASC/>

matyka, wykorzystująca głównie wyniki badań nad genomem człowieka. Złożoność genomicznych, metabolicznych i proteomicznych procesów jest zbyt duża, by człowiek mógł je poznać i zrozumieć, dlatego komputerowy superekspert w tej dziedzinie będzie niezbędny w bliskiej przyszłości.

Podobne zadania warto formułować w innych dziedzinach. Odległym celem jest stworzenie komputerowych doradców dla ekspertów – programów oceniających ich wnioski, wspomagających rozumowanie, podsuwających interesujące obserwacje i skojarzenia, a nawet kreatywne idee (Duch i Pilichowski 2007). Amerykańska agencja ministerstwa obrony do zaawansowanych projektów badawczych (*Defense Advanced Research Projects Agency*, DARPA)⁵ sponsoruje zakrojony na szeroką skalę program budowy spersonalizowanych asystentów osobistych (*Personalized Assistants that Learn*, PAL) i innych systemów kognitywnych⁶.

14.4.4. Sposób oceniania systemów sztucznej inteligencji

Miarą ogólnej inteligencji systemu AI może być liczba zadań lub nawet całych profesji, które dany program może wykonać lub w pełni zautomatyzować (Nilsson 2005). Już Allan Turing zauważył (1950), że taki ogólny system (nazwał go *child machine*), który mógłby nauczyć się różnych zadań, wystarczy, aby zastąpić wiele wyspecjalizowanych systemów. Liczne zawody związane są obecnie z przetwarzaniem informacji i przydatność różnych programów komputerowych w nich stosowanych można mierzyć za pomocą standardowych testów. W zawodach polegających na pracy fizycznej, wymagających koordynacji sensomotorycznej i rozwiązywania problemów związanych z percepcją (w tym analizą obrazów), potrzebne będą autonomiczne roboty. DARPA zorganizowała dwa udane konkursy dla automatycznie sterowanych pojazdów. Były to: konkurs przejazdu przez pustynię i konkurs jazdy w symulowanych warunkach miejskich (*Darpa Urban Challenge Competition* 2007). Takie zadanie wymagało przede wszystkim integracji w jednym systemie metod przetwarzania sygnałów, analizy obrazu, sterowania, planowania i rozumowania.

Jazda samochodem to zadanie znacznie prostsze niż kontrola humanoidalnego domowego robota, który miałby współpracować z ludźmi, a więc musiałby rozumieć ich intencje, rozpoznawać obiekty i relacje między nimi, kontrolować uwagę oraz uczyć się na podstawie obserwacji. Jednakże wiadomo, że nie każdy człowiek może się nauczyć jazdy samochodem. W tym zakresie automatycznie sterowane pojazdy osiągnęły więc wyższy poziom inteligencji niż niektórzy ludzie. Firmy samochodowe zapowiedziały w ciągu najbliższego dziesięciolecia zastąpienie zawodowych kierowców na dalekich trasach przez systemy automatyczne⁷.

14.4.5. Sztuczna inteligencja i ludzie

Jednym z problemów, jakie podnosi się w kontekście rozwoju sztucznej inteligencji, jest problem zagrożenia bezrobociem. Jak się wydaje, znaczenie tego czynnika bywa często

⁵ <http://www.darpa.mil/>

⁶ http://www.darpa.mil/ipto/thrust_areas/thrust_cs.asp

⁷ http://en.wikipedia.org/wiki/Driverless_car

przesadnie oceniane. Stworzenie osobistych asystentów nie musi w pełni eliminować pracy człowieka, na początek powinno wspomagać jego produktywność. Dla pełniejszej oceny brakuje uporządkowania stopnia trudności różnych zadań tak, by stawiać sobie realistyczne cele i uzyskiwać realistyczne oceny kompetencji systemów AI w tych zadaniach. Nie znaczy to jednak, że problem bezrobocia związanego z rozwojem AI jest wydumany. Wiele zawodów już zniknęło, nastąpiła znaczna redukcja liczby pracowników związanych bezpośrednio z produkcją, jak i zatrudnionych w instytucjach finansowych. W większości przypadków zmniejszenie zatrudnienia to wynik zmiany organizacji pracy, wprowadzenia nowych urządzeń o stosunkowo niskim stopniu inteligencji (automatyka przemysłowa, automaty ATM zamiast kas). Nie ulega jednak wątpliwości, że większość ambitnych projektów w dziedzinie sztucznej inteligencji wymagać będzie systemów, które powinny współdziałać w naturalny sposób i w szerokim zakresie z ludźmi, mieć podobne reakcje i funkcje. Takie systemy określane są mianem „architektur kognitywnych”. Spróbujemy je teraz bliżej przeanalizować.

14.5. Architektury kognitywne

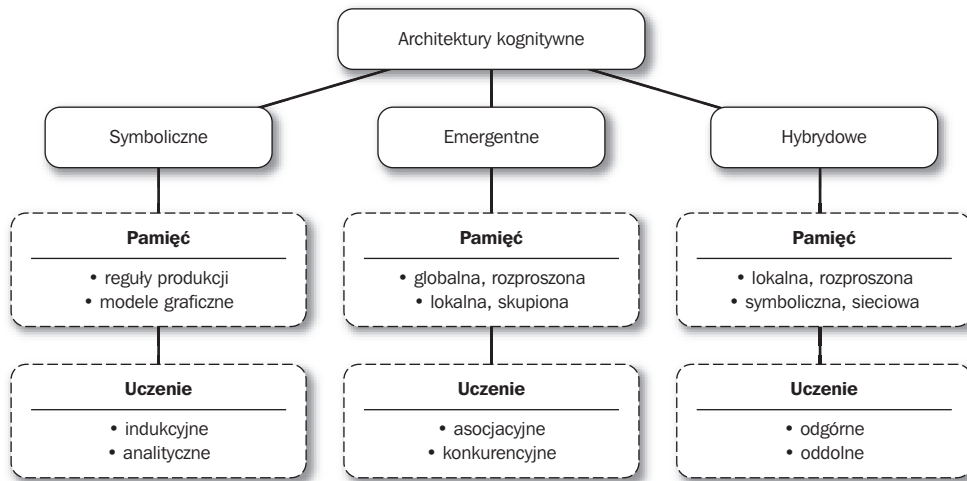
14.5.1. Potrzeba tworzenia modeli kognitywnych

Zrozumienie sposobu działania naszego aparatu poznawczego wymaga skonstruowania modelu działania człowieka, który uwzględni jego możliwości percepcyjne i poznawcze. Początkowo architektury kognitywne tworzone przede wszystkim z myślą o tym, by modelować reakcje człowieka w złożonych sytuacjach, gdy trzeba reagować w odpowiedni sposób na wiele sygnałów o różnych modalnościach (Newell 1990; Meyer i Kieras 1997). Niewłaściwie zaprojektowany system sterowania dużym samolotem może stawiać zbyt wielkie wymagania w stosunku do możliwości reakcji pilotów. Dlatego znajomość ludzkich ograniczeń, jak też wierne modele ludzkich zachowań są bardzo pożądane.

14.5.2. Kryteria oceny architektur kognitywnych

Jak oceniać architektury kognitywne? Allen Newell (1990) zaproponował aż 12 kryteriów oceny takich systemów, dotyczących ich zdolności adaptacyjnych, zachowania dynamicznego, elastyczności zachowań, możliwości stopniowych ulepszeń, możliwości całkowitej ewolucji architektury, sposobów uczenia się, integracji wiedzy, wielkości dostępnej bazy wiedzy, kompetencji językowych, działania w czasie rzeczywistym oraz możliwości implementacji w architekturach mózgowopodobnych. Anderson i Lebiere (2003) zastosowali te kryteria do analizy trzech znanych architektur (ACT-R, SOAR i klasycznej architektury konekjonistycznej⁸). Takie szczegółowe porównania trudno jednak zastosować do większej liczby systemów.

⁸ Podany termin w pierwotnej wersji tekstu występował w formie „koneksjonistycznej” a nie „konekjonistycznej”. Po dyskusji przyjęto, że w tej książce będzie używany termin „konekjonistycznej”, ponieważ taka forma występuje także w innych rozdziałach pisanych przez innych autorów, aczkolwiek warto odnotować, że ta druga forma (koneksjonistycznej) jest częstsza w Internecie, a w słowniku Kopalińskiego znajduje się taki zapis:



Rycina 14.1. Klasyfikacja architektur kognitywnych.

Zaproponowaliśmy prostszą taksonomię (Duch i in. 2008), przedstawioną powyżej, łącznie z przykładami obecnie rozwijanych architektur kognitywnych (ryc. 14.1). Większość prac w tej dziedzinie koncentruje się na konkretnych architekturach i brakuje ogólnego przeglądu, który dałby szerszą orientację. Wyjątkiem jest artykuł Vernona i in. (2007), koncentrujący się na możliwościach autonomicznego rozwoju zdolności mentalnych agentów programowych.

Dwie najważniejsze cechy wszystkich architektur kognitywnych to sposób organizacji ich pamięci i mechanizmy uczenia się. **Pamięć** stanowi repozytorium wiedzy o świecie i o sobie, a także o celach i bieżących działaniach. Rola pamięci rozumiana jest w różny sposób przez różnych autorów (Hawkins i Blakeslee 2004; Hoya 2005; Hecht-Nielsen 2007). **Uczenie** się jest procesem, który transformuje zapamiętaną wiedzę i sposób jej wykorzystania. Pamięć i uczenie się to dwa fundamenty budowy systemów poznawczych, z których rozwijać się mogą złożone wyższe czynności poznawcze. Organizacja pamięci zależy od sposobu reprezentacji wiedzy. Prosta taksonomia architektur kognitywnych oparta na tych dwóch filarach prowadzi do identyfikacji trzech głównych grup architektur: symbolicznych, emergentnych i hybrydowych.

Architektury symboliczne wykorzystują wiedzę deklaratywną, zawartą w relacjach zapisanych na poziomie symbolicznym, a także skupiają się nad użyciem tej wiedzy do rozwiązywania problemów. **Architektury emergentne** wykorzystują przepływy sygnałów przez sieć licznych, wzajemnie ze sobą oddziałujących elementów, w której pojawiają się stany emergentne, dające się zinterpretować w symboliczny

„koneksjonizm: fizj. teoria pamięci, wg której pamięć doświadczeń, faktów, rzeczy, słów itd. rozprzodowana jest po całym mózgu w licznych zespołach, a praca mózgu polega na uruchomianiu połączeń nerwowych między tymi zespołami” (przyp. red. nauk.).

sposób. **Architektury hybrydowe** są kombinacją obu tych podejść, połączonych na rozmaite sposoby.

Do rozwiązywania różnego rodzaju problemów angażowane są różne typy architektury poznawczej.

14.6. Architektury symboliczne

14.6.1. Ogólny przegląd architektur symbolicznych

Hipoteza Newella i Simona (1976) na temat fizycznych systemów symbolicznych ma swoje źródło w badaniach nad pamięcią i rozwiązywaniem problemów. Fizyczny system symboliczny może używać dowolnych wzorców fizycznych, by stworzyć odpowiednie symbole, manipulować nimi, wczytywać je, wypisywać, przechowywać i zmieniać, podejmując przy tym odpowiednie działania ukierunkowane na osiągnięcie określonych celów. Chociaż w praktyce koncepcję tę zawężono do symboli językowych, jej pierwotne sformułowanie jest dość uniwersalne i pozwala na nazywanie symbolami także specyficznych konfiguracji pobudzeń układów neuronowych. W takim przypadku pojawia się od razu problem dyskretyzacji (zignorowany przez Newella), inaczej liczba symboli oznaczających podobne stany rośnie w sposób kombinatoryczny.

Większość architektur symbolicznych używa scentralizowanej kontroli przepływu informacji od czujników sensorycznych do efektorów (kontrola ruchu lub informacji wyjściowych). Można to uznać za aproksymację funkcji wykonawczych związanych z pamięcią roboczą, wspomaganych przez pamięć semantyczną dostarczającą wiedzy. Reguły produkcji mają reprezentować cykl postrzeżenie–działanie, który służy aproksymacji procesów rozumowania, prowadzonych przez ludzi. Stosowane są również reprezentacje grafowe, zwykle w postaci grafów skierowanych, których węzły reprezentują symbole i przypisane im atrybuty, a łuki – związki między nimi. Tak jest w przypadku zarówno sieci semantycznych, jak i grafów koncepcyjnych (Sowa 1984). W graficznej postaci przedstawić też można ramy i schematy (Minsky 1975). W robotyce stosowana jest reprezentacja za pomocą zbioru reakcji w określonym kontekście (*reactive action packages*, RAPs; Firby 1989).

W systemach symbolicznych wprowadzono różne mechanizmy uczenia się, które można podzielić na **analityczne** (dedukcyjne) i **indukcyjne**.

Uczenie analityczne wykorzystuje wiedzę do wnioskowania na jej podstawie o prawdziwości nowych faktów. Przykłady tego podejścia to uczenie oparte na wyjaśnieniach (*explanation-based learning*, EBL; Mitchell i in. 1986) lub analogiach (Veloso i Carbonell 1990). Uczenie indukcyjne odkrywa nowe, ogólne reguły na podstawie obserwacji i ma na celu dokonanie aproksymacji struktury danej domeny. Przykładem jest tu oparte na wiedzy uczenie indukcyjne (*knowledge-based inductive learning*, KBIL; Larvac i Dzeroski 1994) oraz opóźnione uczenie się z krytykiem (Kaelbling i in. 1996).

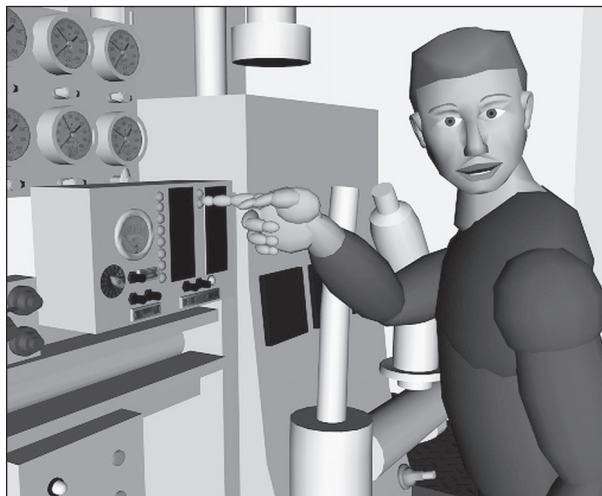
Wiele ambitnych architektur symbolicznych zarzucono po krótkim okresie entuzjazmu ich twórców; poniżej omówione zostaną najciekawsze projekty, które rozwijają się od dłuższego czasu lub mają duże szanse na dalszy rozwój.

14.6.2. Architektura SOAR

SOAR (nazwa pochodzi od haseł *State, Operator And Result*) jest klasycznym, rozwijanym od ponad 30 lat przykładem architektury kognitywnej, służącej modelowaniu ogólnej inteligencji za pomocą systemu regułowego, traktowanego jako przybliżenie systemów opartych na wiedzy (Newell 1990; Laird i in. 1987). Sytuacja rozpatrywana jest tu w przestrzeni problemu, czyli z wykorzystaniem zbioru dopuszczalnych stanów opisu problemu. Reguły produkcji mają formę: „jeśli spełnione są warunki..., to należy podjąć działania...” (IF ... THEN ...). Reguły te uporządkowane są według typu operatorów określających działania w tej przestrzeni.

Główny mechanizm uczenia oparty jest na koncepcji porcjowania (*chunking*) – technice analitycznej formułowania nowych reguł i makrooperacji na podstawie powtarzalnych kombinacji prostych operacji, które okazały się przydatne wcześniej (Laird i in. 1987). Nowe „porcje wiedzy” prowadzą do rozwiązań niejako na skróty. System dostępny jest w różnych wersjach na stronie projektu⁹.

SOAR zastosowano do replikacji wyników wielu eksperymentów psychologicznych, planowania, pracy z dużymi i złożonymi bazami wiedzy, sterowania eskadrą myśliwców, sterowania awatarami (ryc. 14.2) na przykład w grach komputerowych, rozwiązywania problemów i zrozumienia języka naturalnego (NL-SOAR) w czasie rzeczywistym¹⁰.



RYCINA 14.2. Steve, awatar sterowany za pomocą SOAR, uczy obsługi skomplikowanej maszynarii w wirtualnym świecie.

Architektura SOAR uległa w ostatnich latach znacznej ewolucji: uczenie z krytykiem zastosowano do określenia strategii nadawania preferencji operatorom, wprowadzono uczenie epizodyczne do analizy ewolucji stanu problemu oraz pamięć semantyczną do opisu abstrakcyjnej wiedzy deklaratywnej. Pojawiły się też modne

⁹ <http://sitemaker.umich.edu/soar/home>

¹⁰ Opis nowości w SOAR jest pod: <http://ai.eecs.umich.edu/people/laird/current-research.html>