



ON THE NATURE OF LEARNING: PROBLEMS AND
RESEARCH DIRECTIONS

PUBLISHED IN POLISH UNDER THE TITLE:
O NATURZE UCZENIA SIE - PROBLEMY I KIERUNKI
BADAWCZE

by

R. S. Michalski

Informatyka Part 1, No. 2 and Informatyka Part 3, No. 3, (Translators: E. Pierzchala and
P. Zielczynski), 1988.

2

1988

informatyka

Prof. Ryszard S. Michalski
o naturze uczenia się
System dydaktyczny do nauki języka ADA/SM
Analizator-tester protokołów

O naturze uczenia się — problemy i kierunki badawcze (I)

Artykuł zawiera przegląd zadań i kierunków badawczych w dziedzinie maszynowego uczenia się, mając w zamierzeniu autora służyć jako przewodnik w tej dziedzinie. Omówiono w nim zasadnicze aspekty procesu uczenia się, sklasyfikowano główne kierunki badawcze oraz przedstawiono poglądy autora na temat związków pomiędzy paradygmatami, strategiami i orientacjami uczenia się.

CZY POTRZEBNE SĄ MASZYNY UCZĄCE SIĘ?

Sztuczna inteligencja przeżywa obecnie niebywały rozkwit. Jej idee i wypracowane przez nią metody znajdują zastosowanie w wielu innych dziedzinach. Rozwój systemów ekspertowych, praktyczne implementacje systemów rozumienia języka naturalnego, znaczące postępy w dziedzinie komputerowych systemów wizyjnych i systemów rozumienia mowy, nowe spojrzenie na budowę skutecznych systemów wnioskujących i systemów rozumowania jakościowego (ang. qualitative reasoning) stanowią najbardziej widoczne i najważniejsze osiągnięcia. Szybkie rozszerzanie się sfery wpływów sztucznej inteligencji pozwala oczekiwać w najbliższej przyszłości jej nowych sukcesów.

W tym kontekście ważne staje się pytanie, jakie są ograniczenia obecnych metod sztucznej inteligencji i w jakim kierunku rozwiną się badania w tej dziedzinie. Jedno z oczywistych ograniczeń, wyznaczające zarazem kierunek przyszłych badań, dotyczy maszynowego uczenia się. Zagadnienie to jest związane z rozwojem obliczeniowej teorii uczenia się i budową systemów uczących się.

Obecne systemy sztucznej inteligencji, za wyjątkiem tych, których budowa ma na celu badanie maszynowego uczenia się, przejawiają niewielkie lub nie przejawiają żadnych umiejętności uczenia się. Cała wiedza, jaką dysponują, musi być uprzednio przygotowana i zaprogramowana. Jeśli w trakcie ich pracy wystąpi błąd, to nie potrafią samodzielnie go poprawić. Jest on powtarzany tyle razy, ile razy dana procedura jest wykonywana. Systemy te nie potrafią też doskonalić się w miarę zbierania doświadczeń ani gromadzić wiedzy ze swej dziedziny drogą eksperymentowania. Nie są w stanie automatycznie generować dla własnych potrzeb nowych algorytmów, formułować nowych abstrakcyjnych pojęć ani proponować nowych rozwiązań przez analogię do dotychczasowych czy też drogą swobodnego procesu odkrywania (ang. discovery). Ogólnie rzecz biorąc, systemom tym brakuje umiejętności indukcyjnego wyciągania wniosków z podanej im informacji. Można powiedzieć, że niemal wszystkie istniejące systemy są dedukcyjne, gdyż wyciągają wnioski opierając się na dostępczej wiedzy i nie potrafią samodzielnie przyswajać ani generować nowej wiedzy.

Dla porównania, jedną z najbardziej uderzających cech w inteligentnym zachowaniu się człowieka jest zdolność przyswajania nowej wiedzy, nabierania wprawy i doskonalenia przez praktykę. Wykorzystanie możliwości uczenia się czyni z młodego, niedoświadczonego człowieka wykwalifikowanego inżyniera, wychowawcę, artystę lub lekarza. W powszechnym rozumieniu osoba, która powtarza wciąż te same błędy, raczej nie zasługuje na miano inteligentnej. Umiejętność uczenia się na błędach jest uważana za podstawową cechą zarówno człowieka, jak i całego społeczeństwa [1, 6, 13, 14, 28, 29].

Skoro zdolność uczenia się wydaje się być tak ściśle związana z inteligentnym zachowaniem, obecny stan rozwoju sztucznej inteligencji skłonił niektórych badaczy do

wysunięcia tezy, że jednym z najważniejszych celów badań w tej dziedzinie powinno być zrozumienie istoty procesu uczenia się oraz wyposażenie maszyn w zdolność uczenia się [19, 33]. Celowość przewyżczenia wspomnianych ograniczeń wyznacza więc zadania badawcze na przyszłość.

W tym miejscu rodzi się pytanie: czy taki cel jest osiągalny, a jeśli tak, to czy jest pożądany. Próba odpowiedzi na pytanie o możliwość osiągnięcia tego celu zmusza do sprecyzowania niektórych pojęć. Czy można wskazać ogólne kryteria, których spełnienie zapewnia, że maszyna ma zdolność uczenia się?

Jak tego dowiodły badania w dziedzinie maszynowego uczenia się, zdolność uczenia się nie jest właściwością podlegającą dwuwartościowej ocenie: „tak lub nie”. Jest z nią związana cała gama czynności polegających na przetwarzaniu informacji: od bezpośredniego zapamiętywania faktów i prostego organizowania informacji, do bardzo skomplikowanych procesów wnioskowania, prowadzących do tworzenia nowych pojęć i odkrywania nowej wiedzy. Zawsze pociąga to za sobą zmiany w systemie — człowieku lub maszynie — które ulepszają go w określonym sensie.

Odkładając do następnego punktu odpowiedź na pytanie o definicję uczenia się, należy stwierdzić, że maszynowe uczenie się przeżywa obecnie swoje odrodzenie po okresie stabilizacji i powolnego rozwoju. Wysiłki zmierzające do budowy programów przejawiających umiejętności uczenia się zostały w ostatnich latach znacznie zintensyfikowane. Ta młoda dyscyplina ma już na swym koncie pewne sukcesy. Rezultaty niektórych spośród tych wysiłków opisano w [21]. Kontynuacją tej pracy jest książka [22], prezentująca kluczowe zagadnienia i współczesny stan wiedzy w dziedzinie maszynowego uczenia się.

Opierając się na dotychczasowych rezultatach, można dojść do wniosku, że już dziś możliwe jest wyposażenie maszyn w elementarne zdolności uczenia się. Istnieją już programy zdolne formułować nowe pojęcia i wykrywać nieznane dotąd regularności w danych, tworzyć reguły decyzyjne lepsze od reguł formułowanych przez człowieka, dostarczać interesujące analogie, automatycznie uczyć się heurystyk pomagających w rozwiązywaniu problemów lub budować uogólnione plany akcji prowadzących do osiągnięcia określonego celu. Wiele z tych programów omówiono w [21]. Mniej jasne jest, jaki postęp w dziedzinie maszynowego uczenia się można osiągnąć przy użyciu konwencjonalnego sprzętu komputerowego i obecnych metod programowania. Jak zawsze w nauce, na takie pytania można odpowiedzieć, kontynuując badania i budując eksperymentalne systemy uczące się.

Nowe kierunki badań w dziedzinie maszynowego uczenia się otwierają się wraz z rozwojem tzw. maszyn neuronowych (ang. connection machines), systemów komputerowych piątej generacji i innych nowatorskich architektur komputerowych [8, 12]. Na przykład, Hinton, Sejnowski i Ackley [9] opisują uczenie się w tzw. maszynach Boltzmann. Wiedza gromadzona przez te maszyny jest reprezentowana jako trwałość połączeń między prostymi elementami podobnymi do neuronów. Badania prowadzone w tym kierunku powinny umożliwić pokonanie ograniczeń wczesnych systemów tego rodzaju, takich jak perceptrony [26]. Nowy potencjał badawczy w dziedzinie maszynowego uczenia się wyłania się w związku z rozwojem nowych systemów programowania, a w szczególności programowania

logicznego i jego pierwszego wcielenia w postaci Prologu [31].

Dlaczego konieczna jest budowa maszyn uczących się? Wydaje się, że ich rozwój jest warunkiem dalszego postępu w badaniach nad sztuczną inteligencją i w dyscyplinach pokrewnych. Jest to szczególnie istotne w takich dziedzinach, jak: systemy ekspertowe i systemy z bazami wiedzy, komputerowe systemy wizyjne, rozumienie mowy, rozumienie języka naturalnego, inteligentne systemy edukacyjne oraz prawdziwe przyjazne systemy współpracy człowiek-maszyna. Im bardziej skomplikowane zadanie stawia się przed systemami sztucznej inteligencji, tym więcej wiedzy trzeba w nich gromadzić. Wiedza taka musi obejmować fakty i reguły specyficzne dla danej dziedziny, zdroworozsądkowe heurystyki i ograniczenia, jak również ogólne pojęcia i teorie dotyczące otaczającego świata. Zakres wiedzy każdego systemu musi być tak poszerzony, aby uniknąć powszechnie występującego dziś zjawiska, nazywanego „przepaścią wiedzy” (ang. *konwledge cliff*) [5] albo „kruchością wiedzy” (ang. *brittleness*) [24] (rozdział 20), [15]. Polega ono na tym, że system pracuje poprawnie w wyznaczonej mu dziedzinie, natomiast nieznaczone wyjście poza nią drastycznie pogarsza jego właściwości. J. H. Holland w rozdziale 20 [24] omawia uniwersalne algorytmy uczące się, oparte na równoległej architekturze regułowej. Twierdzi on, że rozumowanie indukcyjne w takich systemach pomoże przezwyciężyć „kruchość” dzisiejszych systemów sztucznej inteligencji, biorąc się ze zbyt wąskiego ukierunkowania ich wiedzy.

Wypełnianie bazy wiedzy nowego systemu jest procesem bardzo złożonym, czasochłonnym i podatnym na błędy; wymagającym dużego doświadczenia. Na przykład, budowa systemu ekspertowego wymaga wspólnego wysiłku wysoko wykwalifikowanych specjalistów — przynajmniej jednego eksperta z danej dziedziny oraz inżyniera wiedzy [2, 4, 7]. Zadanie to można uprościć stosując techniki uczenia się, które umożliwiają automatyczne konstruowanie reguł decyzyjnych na podstawie przykładów decyzji eksperta oraz drogą automatycznej analizy zawartość bazy faktów.

Szybki wzrost liczby danych i wiedzy gromadzonej przez społeczeństwo wymaga nie tylko gromadzenia, organizowania i udostępniania informacji, lecz również wykorzystania jej w nowy, twórczy sposób. Ponieważ wiedzę można rozpatrywać jako skondensowaną informację [30], potrzebujemy maszyn potrafiących dokonywać kompresji baz danych i systemów informacyjnych na bazy wiedzy drogą automatycznej analizy pojęciowej ich treści. Jak podkreśla Michie [25], „najbardziej frapującym technicznie, nawet jeśli obecnie nie najważniejszym ekonomicznie, wyzwaniem jest to, w jaki sposób zastosować komputery nie tylko do bezpośredniej obserwacji badań naukowych, prowadzonych za pomocą teleskopów, mikroskopów, komór iskrowych i innych urządzeń, lecz również w procesach rozpoznawania i wnioskowania, dzięki którym chaos danych ostatecznie przeradza się w odkrycie naukowe”.

Autor tego artykułu sądzi, że oprócz komputerów pełniących rolę inteligentnych asystentów naukowca i technika, będziemy potrzebowali inteligentnych asystentów osobistych. Obywatele rozwijającego się społeczeństwa informacyjnego będą potrzebowali ich po to, aby poradzić sobie z przytłaczającą masą informacji i złożonością codziennych procesów podejmowania decyzji. Aby asystenci mogli spełniać wyznaczoną im rolę, ich funkcjonowanie i wiedza powinny być dynamiczne; powinni oni łatwo przystosowywać się do zmniejszających się wymagań i mieć umiejętność automodyfikacji. Inaczej mówiąc, powinni dysponować zdolnością uczenia się.

Umiejętność uczenia się jest również niezbędna w dziedzinie wizji komputerowej i rozumienia mowy. Aby zbudować komputerowy system wizyjny, należy wprowadzić do niego wiele transformacji charakterystycznych dla wizji, pojęć geometrycznych oraz opisów fizycznych i funkcjonalnych obiektów wizualnych, które system ma rozpoznawać [38, 39]. Ręczne wprowadzenie tej informacji jest trudne. Znacznie łatwiejsze byłoby nauczenie systemu konkretnych pojęć przez pokazanie mu przykładów tak, aby mógł przyswoić sobie odpowiednie uogólnienia i opisy w sposób, w jaki czynią to ludzie.

System zdolny do rozumienia języka naturalnego i komunikowania się, za jego pomocą z człowiekiem powinien mieć wiedzę o składniowych właściwościach języka [17], jak również uwzględniać wiele pojęć i ich struktur (takich jak ramy, scenariusze (ang. *scripts*) i schematy), opisują-

cych semantyczne i pragmatyczne aspekty języka [24], rozdział 19 i 21, [32, 37]. Szacuje się, że w zaawansowanym programie rozumiejącym język naturalny, liczba takich pojęć i struktur pojęciowych może osiągnąć dziesiątki tysięcy, a nawet więcej. Zaprogramowanie całej tej wiedzy jest ogromnym zadaniem. Dobrze byłoby więc uprościć je, wykorzystując techniki uczenia się. Gdyby nawet w pewnym momencie maszyna dysponowała pełną wiedzą, system rozumiejący język naturalny nie mógłby i tak długo pracować poprawnie, nie mając zdolności uczenia się. Znaczenie pojęć używanych przez człowieka jest bowiem dynamiczne. Zmienia się ono z upływem czasu, aby dostosować się do nowych okoliczności. Nowe pojęcia powstają i rozwijają się, inne natomiast wychodzą z użycia. Dlatego też, tak jak w podanych wypadkach potrzebny jest system uczący się, zdolny do przyswajania nowych pojęć i struktur pojęciowych drogą uogólniania przykładów lub przez analogię do poprzedniej wiedzy. System taki powinien być zdolny do elastycznego modyfikowania, ukonkretniania i uogólniania starych pojęć.

Inteligentne systemy edukacyjne powinny operować materiałem dydaktycznym na poziomie dostosowanym do poziomu wiedzy ucznia. Aby sprostać temu zadaniu, system musi kontrolować wiedzę ucznia i nadać za jej zmianami. Pożądane jest uzyskiwanie tej informacji nie drogą cyklicznego i bezpośredniego sprawdzania, lecz drogą uczenia się w trakcie interakcyjnych sesji na podstawie wskazówek, przyjętego modelu ucznia oraz jego zachowania. Tak więc, zdolność uczenia się jest wymagana nie tylko od ucznia, ale także od systemu edukacyjnego [35, 36].

Dzięki umiejętności uczenia się, komputery przyszłości powinny być zdolne do zdobywania wiedzy bezpośrednio z dokumentów i książek drogą konwersacji z ludźmi i uogólniania obserwacji dokonywanych sztucznymi zmysłami. Powinny być zdolne do doskonalenia się przez praktykę i zdobywanie doświadczenia. Przyszłe systemy uczące się nie będą podlegały ograniczeniom, jakim podlegają ludzie (krótka pamięć, trudności w skoncentrowaniu uwagi, niska efektywność, trudności w przekazywaniu nabytej wiedzy). Jeśli powstanie jeden egzemplarz jakiegoś systemu uczącego się, to możliwe będzie zbudowanie teoretycznie nieograniczonej liczby jego kopii, które będą mogły zdobywać wiedzę w różnorodnych dziedzinach. W dodatku każdą nową porcję wiedzy zdobytej przez system uczący się będzie można skopiować do innego systemu szybko i tanio (inaczej niż w wypadku wiedzy zdobywanej przez ludzi, która musi być starannie przekazywana każdemu nowemu uczniowi).

Oczywiście daleko jest jeszcze do tej idealnej wizji, można sobie jednak wyobrazić, że takie systemy uczące się będą mogły być budowane w przyszłości. Warto przy tym rozważyć nie tylko spodziewane korzyści, lecz także niepożądane konsekwencje. Ten drugi problem można zresztą ominąć stwierdzając, że każda nowa technologia daje możliwość użycia jej w niewłaściwym celu, a to jak dotąd nie powstrzymywało człowieka przed ich rozwijaniem. Na ogół uważa się, że zagadnienia te leżą poza dziedziną badań naukowych czy technicznych. Mimo to, należy dokładnie je zbadać, ponieważ stworzenie maszyn zdolnych do samodzielnego zdobywania wiedzy wprowadza dodatkową złożoność i wpływa na wybór drogi dalszego rozwoju maszynowego uczenia się.

Podstawowym problemem jest „nieprzezroczystość” systemów samomodyfikujących się. Przewidywanie zachowania maszyn umiejących uczyć się indukcyjnie jest znacznie trudniejsze niż przewidywanie zachowania maszyn nie mających tej umiejętności. Najważniejszą właściwością maszyn uczących się jest zdolność tworzenia wiedzy mogącej zaskoczyć samych twórców. Może to zresztą prowadzić do niespodziewanych trudności, a nawet niebezpieczeństw, jeśli użyje się takiego systemu do rozwiązania poważnych zadań bez zrozumienia jego ograniczeń. Ponadto zwiększenie nieprzewidywalności działania maszyn uczących się zwiększa niebezpieczeństwo ich nieprawidłowego użycia.

Niektórzy eksperci twierdzą, że przewidywanie zachowania złożonych systemów komputerowych jest już teraz wystarczająco trudne. Wyposażenie komputerów w zdolność uczenia się uważają oni za powiększenie tych trudności, nie zaś za jakościowy krok naprzód. Niezależnie od punktu widzenia należy jednak spodziewać się, że potencjalne korzyści tej technologii sownie wynagrodzą nam jej niepożądane konsekwencje. Co zaś się tyczy możliwości ich

niewłaściwego użycia, to dlaczego nie można wykorzystać tych sprytnych maszyn uczących się do nadzorowania innych, aby zapobiec ich nadużyciu?

Oprócz trudności w przewidywaniu zachowania maszyn uczących się, warto rozważyć jeszcze jedno zagadnienie, tkwiące korzeniami w samej naturze każdego rodzaju wiedzy różnej od wiedzy zdobytej w wyniku prostej obserwacji faktów. Jak zauważył Hume, a później Popper [11, 28] i inni, wiedza taka ze swej natury ma charakter domyślny. Znacząco to, że w zasadzie nie można udowodnić poprawności żadnej wiedzy tworzonej drogą uogólniania określonych obserwacji lub przez analogię do znanych faktów, choć można dowiedzieć jej błędności.

Jest tak dlatego, że rozumowanie przez indukcję nie zapewnia zachowania „prawdy” (ang. truth-preserving), lecz zachowuje jedynie „fałsz” (ang. falsity-preserving), tzn. z prawdziwych informacji nie zawsze wyprowadza się prawdziwe informacje, natomiast z fałszywych informacji zawsze wyprowadza się fałszywe [21, 22, 23]. Aby to zilustrować, rozważmy zdanie: „Wszyscy naukowcy w Laboratorium Sztucznej Inteligencji w Massachusetts Institute of Technology MIT są błyskotliwi”. Drogą dedukcji można wyprowadzić z tego zdania konkluzję, że Roger Light, pracujący w Laboratorium Sztucznej Inteligencji, jest błyskotliwy. Jeśli pierwotna przesłanka jest prawdziwa, to ten wniosek dedukcyjny musi także być prawdziwy. Natomiast przykładem wniosku uzyskanego z pierwotnej przesłanki drogą indukcji jest następujący: „Wszyscy naukowcy w MIT są błyskotliwi”. W tym wypadku, prawdziwość pierwotnej przesłanki nie gwarantuje prawdziwości wniosku uzyskanego drogą indukcji. Jeśli jednak pierwotna przesłanka jest fałszywa, to wniosek indukcyjny również jest fałszywy. Tak więc, inaczej niż w systemie dedukcyjnym, poprawne dane wejściowe do systemu indukcyjnego nie gwarantują otrzymania poprawnych wyników. Ponadto każdemu zestawowi danych wejściowych odpowiada nieskończona liczba dających się z niego wyciągnąć wniosków indukcyjnych. Te, które są akurat wybierane, odzwierciedlają preferencje, założenia i ograniczenia, stosowane przy formułowaniu uogólnień [20, 24] rozdział 5.

Z tych powodów, aby można generować wiedzę przydatną dla człowieka, maszyny uczące się powinny być wyposażone w wiedzę dotyczącą wszystkich istotnych ograniczeń i założeń przyjmowanych przez ludzi. Ponieważ jest mało prawdopodobne, aby wszelkie subtelne ograniczenia i preferencje jednostek i społeczeństw były kiedykolwiek poznane przez maszyny, jest możliwe, że wiedza generowana przez nie będzie naruszać niektóre z tych ograniczeń. Warto przytoczyć pogląd Hofstadtera [10]: „jeżeli program nie będzie dysponował wierną kopią ludzkiego ciała..., to prawdopodobnie będzie miał zupełnie odmienny punkt widzenia na to, co jest ważne, interesujące itd.”. Ponieważ postrzeganie tego, co jest istotne i interesujące, stanowi nieodzowny element tworzenia nowej wiedzy [16], odmienność ta jest znacząca. Gdy wiedza wytworzona przez maszynę zostanie wykorzystana w praktyce, będzie można otrzymać rozwiązania poprawne z technicznego, jednak nie do przyjęcia ze społecznego punktu widzenia.

Wiąże się z tym niebezpieczeństwo zbyt dużego zaufania do wiedzy tworzonej przez maszyny. Podobne zjawisko zaobserwowano, na przykład, u ludzi nadmiernie ulegających wpływowi komputerowej analizy statystycznej, gdy nie rozumieją oni dokładnie założeń leżących u jej podstaw lub u ludzi przypisujących osobowość komputerowemu systemowi konsultacyjnemu, na przykład takiemu jak ELIZA [40]. O ile naukowcy wiedzą o tym, że wiedza uzyskana drogą indukcji jest z natury niepewna, może to być mniej oczywiste dla laików.

Powyższe rozważania prowadzą do istotnego wniosku, że wszelka wiedza wygenerowana przez maszynę powinna być poddana ścisłej weryfikacji ze strony człowieka. Ukazuje to istotny cel badawczy w dziedzinie maszynowego uczenia się: jeżeli ludzie mają rozumieć i akceptować wiedzę generowaną przez maszynę, to systemy uczące się powinny być zdolne do udzielania wyjaśnień. Ponadto wiedza tworzona przez maszyny powinna być wyrażana w formie ściśle odpowiadającej opisom przyjętym przez człowieka i jego modelom myślowym, tzn. powinna spełniać to, co autor nazywa zasadą zrozumiałości [21, 22, 23]. Projektując mechanizmy systemu uczącego się odpowiedzialne za wyjaśnienie, należy dążyć do ułatwienia człowiekowi zrozumienia nie tylko końcowych rezultatów, ale także tkwiących u ich podstaw zasad, założeń i teorii.

Ktoś może wysunąć przypuszczenie, że istnienie zaawansowanych maszyn uczących się wyeliminuje wprawdzie obecne wąskie gardło w przyswajaniu wiedzy, ale wprowadzi za to nowe — w weryfikowaniu wiedzy. Maszyny będą bowiem generowały nową wiedzę w takiej ilości, że ludziom będzie trudno ją sprawdzać i akceptować. Jeśli tak się stanie, to przyszli badacze będą mieli do rozwiązania ciekawy problem, który pozwoli im wypełnić chwile bezczynności. Oczami wyobraźni możemy już ujrzeć badaczy konstruujących wymyślne maszyny uczące się, służące do generowania eksperymentów testujących wiedzę wytworzoną przez inne maszyny.

Mając na uwadze powyższe problemy i pamiętając o tym, jak ważne jest maszynowe uczenie się, spojrzmy nieco dokładniej na zasadnicze cechy procesu uczenia się.

CZYM JEST UCZENIE SIĘ?

Jak zauważono wcześniej, uczenie się — według powszechnego przekonania — pociąga za sobą zmiany w systemie, doskonalące go w pewien sposób. Termin „doskonalenie” wymaga jednak uściślenia. Bezspornie wino „doskonałe” się w miarę upływu czasu, nikt jednak nie nazwie tego uczeniem się (przykład ten pochodzi od Steve’a Tanimoto z University of Washington w Seattle). Simon [34] podaje dokładniejszą definicję: „Uczenie się oznacza zmiany w systemie, które mają charakter adaptacyjny w tym sensie, że pozwalają systemowi wykonać następnym razem takie samo zadanie lub zadania pochodzące z tej samej populacji bardziej efektywnie”.

Na ogół wszyscy zgadzają się z wymaganiem, aby rezultatem uczenia się była efektywniejsza praca systemu. Są jednak czynności zaliczane do uczenia się, do których trudno jest zastosować kryterium udoskonalenia. Minsky w swej wnikliwej teorii procesów myślowych [27], całkowicie odrzuca ten warunek, stwierdzając: „Uczenie się jest wprowadzaniem zmian do sposobu pracy naszych umysłów”.

W języku angielskim uczeniem się często nazywa się uzyskiwanie nowej informacji, jak w zdaniu: „Ponieważ satelita splonął w atmosferze, astronauta z laboratorium kosmicznego dowiedział się (ang. learned, dosłownie — nauczył się), że satelita miał dodatkową antenę”. W tym wypadku astronauta uzyskał po prostu pewną informację, która jest bezużyteczna, jeśli chodzi o przyszłą obsługę tego właśnie satelity. **Przyswajanie wiedzy** wydaje się być centralną czynnością w wielu aktach uczenia się. Są jednak przykłady uczenia się, gdzie przyswajanie wiedzy odgrywa raczej małą rolę, szczególnie w tzw. procesach **zdobywania wprawy** (ang. skill acquisition). Zdobywanie wprawy jest to stopniowe doskonalenie zdolności ruchowych lub poznawczych drogą ponawiania prób, z niewielkim lub żadnym udziałem świadomości [3]. Podkreśla się więc jeszcze raz aspekt przyswajania wiedzy w procesie uczenia się, na co zwrócono szczególną uwagę w książce [24].

Aby uzyskać wiedzę o czymkolwiek, trzeba ją oczywiście w jakiś sposób przedstawić, np. w postaci zdań deklaratywnych, procedur, przez połączenie obu tych reprezentacji lub też jeszcze inaczej [18]. Ten fakt i wcześniejsze rozważania prowadzą do następującego wniosku. „Uczenie się jest konstruowaniem lub modyfikowaniem reprezentacji tego, co jest doświadczane”.

Pojęcie **doświadczenia** obejmuje tu wszelkie bodźce zewnętrzne, jak również wewnętrzne doświadczenia myślowe. Te bodźce i eksperymenty myślowe służą do postrzegania rzeczywistości, którą system uczący się próbuje reprezentować. Wewnętrzne procesy myślowe mogą być same przedmiotem uczenia się. W ocenie konstruowanych reprezentacji bierze się pod uwagę przede wszystkim: wiarygodność, efektywność i poziom abstrakcji. **Wiarygodność** określa stopień, w jakim reprezentacja odpowiada rzeczywistości. **Efektywność** charakteryzuje przydatność reprezentacji do osiągnięcia danego celu. Im bardziej efektywna jest reprezentacja, tym lepiej funkcjonuje system. To kryterium jest najważniejsze w tych zastosowaniach, w których sprawnym działaniem systemu jest zagadnieniem pierwszoplanowym. **Poziom abstrakcji** odpowiada zakresowi, szczegółowości i precyzji pojęć użytych w opisie. Określa on **moc opisową** reprezentacji. Te trzy kryteria stanowią łącznie o **jakości uczenia się**.

Reprezentacje mogą mieć postać opisów symbolicznych, algorytmów, modeli symulacyjnych, procedur sterujących, planów, obrazów, jak również ogólnych teorii formalnych.

Jeśli rozszerzy się pojęcie reprezentacji, aby był możliwy opis fizycznych i fizjologicznych zmian zachodzących w układzie nerwowym w trakcie zdobywania umiejętności, to rozważane procesy uczenia się obejmują również zdobywanie umiejętności.

Z tego punktu widzenia podstawowym pytaniem we wszelkich badaniach z dziedziny maszynowego uczenia się jest pytanie o formę i metodę użytą do reprezentowania i modyfikowania zdobywanej wiedzy lub umiejętności. Gdy idzie o modyfikację wiedzy, ważne jest określenie tych składowych i tych właściwości reprezentacji, które mają być modyfikowane przez system.

W systematyce badań w dziedzinie maszynowego uczenia się [3] wyróżniono trzy kryteria, jako szczególnie pomocne w klasyfikowaniu i porównywaniu metod: **strategia uczenia się, reprezentacja wiedzy i dziedzina zastosowań**. Strategia uczenia się mówi o rodzaju wnioskowania zastosowanego przez system w trakcie uczenia się. Pewne dodatkowe idee, odzwierciedlające ostatnie osiągnięcia w tej dziedzinie, przedstawiono w drugiej części tego artykułu. Kryteria reprezentacji wiedzy i dziedziny zastosowań są wyczerpująco przeanalizowane w [3], wobec czego nie będą omawiane. Zostaną natomiast przedyskutowane nieco bardziej szczegółowo dwa inne kryteria klasyfikacji: **paradygmaty badawcze oraz orientacje uczenia się**. Kryterium paradygmatu badawczego odnosi się tu do metod wykorzystywanych w budowie systemu, natomiast orientacja uczenia się mówi o zakresie i temacie badań.

tłum. i oprac.
EDMUND PIERZCHAŁA
PIOTR ZIELCZYŃSKI

LITERATURA

- [1] Berkson W., Wettersten J.: Learning from Error. Open Court Publishing Company, La Salle (IL), 1984
- [2] Buchanan B. G., Schortliffe E. H. (eds.): Rule-based Expert Systems. Addison-Wesley, Reading (MA), 1984
- [3] Carbonell J. G., Michalski R. S., Mitchell T. M.: An Overview of Machine Learning. (21)
- [4] Davis R., Lenat D. B.: Knowledge-Based Systems in Artificial Intelligence. McGraw-Hill, New York, 1982
- [5] Feigenbaum A. E.: First U. S. — China Joint Seminar on Automation and Intelligent Systems, Beijing, 28 May — 1 June 1984
- [6] Hayes-Roth F.: Using Proofs and Refutations to Learn from Experience. (21)
- [7] Hayes-Roth F., Waterman D. A., Lenat D. B. eds.: Building Expert Systems. Addison-Wesley, Reading (MA), 1983
- [8] Hillis W. D.: The Connection Machine (Computer Architecture for the New Wave). AI Memo No. 646, MIT, Cambridge (MA), September 1981
- [9] Hinton G. E., Sejnowski T. J., Ackley D. H.: Boltzmann Machines — Constraint Satisfaction Networks that Learn. Technical Report CMU-CS-84-119, Carnegie-Mellon University, Pittsburgh (PA), 1984
- [10] Hofstadter D. R.: Gödel, Escher, Bach — Eternal Golden Braid. Vintage Books, New York, 1980
- [11] Hume D.: A Treatise of Human Nature. L. S. Selby-Bigge (ed.), Clarendon Press, Oxford, 1988
- [12] Kawanobe K.: Present Status of the Fifth Generation Computer Systems Project. ICOT Journal, No. 5, October 1984
- [13] Kuhn T. S.: The Structure of Scientific Revolutions. Sec. edition. University of Chicago Press, 1970
- [14] Lakatos I.: Falsification and the Methodology of Scientific Research Programmes. Pp. 91—196, Criticism and the Growth of Knowledge. A. Musgrave, I. Lakatos (eds.), Cambridge University Press, 1980
- [15] Larkin J., Reif F., Carbonell J., Gugliotta A.: FERMI — Flexible Expert Reasoning with Multi-Domain Inferencing, Cognitive Science, 1985
- [16] Lenat D. G.: The Role of Heuristics in Learning by Discovery — Three Case Studies. (21)
- [17] Marcus M. P.: A Theory of Syntactic Recognition for Natural Language. MIT Press, Cambridge (MA), 1980
- [18] McCarthy J.: Programs with Common Sense. Semantic Information Processing. M. Minsky (ed.), MIT Press, Cambridge (MA), 1968
- [19] McCarthy J.: President's Quarterly Message — AI Needs More Emphasis on Basic Research. AI Magazine, Vol. 4, pp. 4—5, 1983
- [20] Medin D. L., Wattenmaker W. D., Michalski R. S.: Constraints in Inductive Learning — An Experimental Study Comparing Human and Machine Performance, Cognitive Science, 1985
- [21] Michalski R. S., Carbonell J. G., Mitchell T. M. (eds.). Machine Learning — An Artificial Intelligence Approach. TIOGA, Palo Alto (CA), 1983
- [22] Michalski R. S.: Theory and Methodology of Inductive Learning. Artificial Intelligence, No. 20, 1983
- [23] Michalski R. S. (ed.): Proc. of the International Machine Learning Workshop, Allerton House, University of Illinois, 22—24 June 1983
- [24] Michalski R. S., Carbonell J. G., Mitchell T. M. (eds.): Machine Learning. Vol. 2, Morgan and Kaufmann Publishers, 1986
- [25] Machice D.: Machine Intelligence and Related Topics. Gordon and Breach Science Publishers, New York, 1982
- [26] Minsky M., Papert S.: Perceptrons. MIT Press, Cambridge (MA), 1969
- [27] Minsky M.: The Society of Mind. MIT, Cambridge (w druku)
- [28] Popper K. R.: The Logic of Scientific Discovery. Basic Books, New York, 1959
- [29] Popper K. R.: Objective Knowledge — An Evolutionary Approach, Revised edition. Oxford, 1981
- [30] Rendell L. A.: Toward a Unified Approach to Conceptual Knowledge Acquisition. AI Magazine, Vol. 4, No. 4, 1983
- [31] Robinson J. A.: Logic Programming — Past, Present and Future, New Generation Computing, Vol. 1, No. 2, 1983
- [32] Schank R. C.: Dynamic Memory — A Theory of Reminding and Learning in Computer and People. Cambridge University Press, 1982
- [33] Schank R. C.: The Current State of AI — One Man's Opinion, AI Magazine, Vol. 4, No. 1, pp. 3—8, Winter/Spring 1983
- [34] Simon H. A.: Why Should Machines Learn? (21)
- [35] Sleeman D., Brown J. S. (eds.): Intelligent Tutoring Systems. Academic Press, New York, 1982
- [36] Sleeman D. H.: Inferring Student Models for Intelligent Computer-Aided Instruction. (21)
- [37] Wingrad T.: What Does it Mean to Understand Language? Perspectives on Cognitive Science, Norman (ed.), Ablex Publishing Corporations, Norwood NJ, 1981
- [38] Winston P. H., Binfor T. O., Katz B., Lowry M. R.: Learning Physical Descriptions from Functional Definitions — Examples and Precedents. National Conference on Artificial Intelligence, Washington (DC), 1983
- [39] Weizenbaum J.: Computer Power and Human Reason. W. H. Freeman, San Francisco (CA), 1976
- [40] Zagoruiko N.: Empirical Prediction Algorithms, Computer Oriented Learning Processes, J. C. Simon (ed.), Noordhoff, Leyden, 1976

W dniach 18—20 listopada 1987 r. odbyło się w Katowicach międzynarodowe sympozjum nt.

„Zastosowanie mikrokomputerów w informacji naukowej, technicznej i ekonomicznej”

Miało ono na celu popularyzację zastosowania mikrokomputerów w pracy służb inte, wymianę poglądów i informacji.

Sympozjum, pod protektoratem Urzędu Postępu Naukowo-Technicznego i Wdrożeń, zorganizowano z okazji 35-lecia Instytutu Metali Nieżelaznych w Gliwicach. Oprócz Instytutu imprezę organizował Ośrodek Postępu Technicznego w Katowicach i Komisja ds. INTE Stowarzyszenia Inżynierów i Techników Przemysłu Hutniczego. W Sympozjum wzięło udział ok. 500 uczestników z Ośrodków inte różnego szczebla, ośrodków PAN, wyższych uczelni, bibliotek, ośrodków obliczeniowych oraz studenci kierunków INTE Uniwersytetu Śląskiego. Oprócz Polaków obecni byli specjaliści z Bułgarii, Czechosłowacji, NRD, Węgier i Związku Radzieckiego.

W trakcie trwania sympozjum odbyła się wystawa mikrokomputerów, sprzętu towarzyszącego i oprogramowania do celów inte, na której swoje wyroby prezentowało 10 firm. Na specjalnych pokazach można się było zapoznać ze stosowanymi programami do wyszukiwania inte na Uniwersytecie Śląskim.

Uczestnicy sympozjum mieli także okazję zwiedzenia ekspozycji odbywającej się równocześnie XXIII Międzynarodowej Wystawy Literatury Firmowej.

Ustalono, że za dwa lata odbędzie się kolejne sympozjum na ten temat.

3

1988

informatyka

Algorytmy kombinatoryczne

Smalltalk-80

Baza danych Holmes



Cena 200 zł

○ naturze uczenia się – problemy i kierunki badawcze (2)

W drugiej, końcowej części artykułu omówiono paradygmaty, strategie i orientacje uczenia się.

PARADYGMATY BADAWCZE

Od chwili rozpoczęcia badań nad maszynowym uczeniem się, w latach pięćdziesiątych, w różnych okresach koncentrowano się na różnych podejściach i stawiano sobie różne cele badawcze. Można wyróżnić trzy główne paradygmaty badań w tej dziedzinie:

- modelowanie sieci neuronów i techniki teoriodezyjne,
- symboliczne przyswajanie pojęć,
- uczenie się w wąskiej dziedzinie z intensywnym wykorzystaniem wiedzy.

Podejścia te różnią się przede wszystkim pod względem wiedzy posiadanej *a priori* przez system uczący się i pod względem sposobu, w jaki wiedza jest reprezentowana i modyfikowana w systemie.

W podejściu polegającym na modelowaniu sieci neuronów dąży się do zbudowania uniwersalnych systemów uczących się, zaczynających pracę od niewielkiej wiedzy początkowej. Systemy takie nazywa się sieciami neuronowymi lub systemami samoorganizującymi się. System tego rodzaju składa się z sieci wzajemnie połączonych elementów, zazwyczaj podobnych w swym działaniu do neuronów, realizujących jakąś prostą funkcję logiczną, na przykład funkcję progową. Taki system uczy się drogą przyrostowego modyfikowania trwałości połączeń między elementami, zwykle przez zmianę wag związanych z tymi połączeniami. Początkowa wiedza systemu pochodzi z wyboru elementów wejściowych, reprezentujących określone atrybuty badanych obiektów oraz ze struktury i początkowej trwałości połączeń w sieci. Może to być struktura losowa, struktura przygotowana przez projektanta lub jakies rozwiązanie pośrednie. Do systemów tego rodzaju zalicza się *Perceptron* [23], *Pandemonium* [24] i każdą maszynę uczącą się, wykorzystującą funkcje dyskryminacyjne [20]. Przykładami nowszych systemów wywodzących się z tego paradygmatu są różne adaptacyjne systemy sterowania [24]. Badania w tej dziedzinie doprowadziły do powstania podejścia opartego na zastosowaniu teorii decyzji w rozpoznawaniu obrazów. Wiąza się z tym badania nad ewolucyjnym uczeniem się [5, 6] i algorytmy genetyczne [8, 18, rozdział 20]. Jak wspomniano, obserwuje się ponowny wzrost zainteresowania tym paradygmatem uczenia się, w związku z ostatnimi próbami budowy nowych maszyn neuronowych [7].

Charakterystycznymi cechami systemów budowanych według tego paradygmatu są: niski poziom wiedzy wbudowanej *a priori* oraz wykorzystanie parametrów zmieniających się w sposób ciągły, w trakcie uczenia się. Inną cechą tych systemów jest numeryczny charakter metod i algorytmów uczenia się. Kontrastuje to z dwoma następnymi paradygmatami, w których główny nacisk kładzie się na tworzenie i manipulowanie skomplikowanymi strukturami symbolicznymi.

W metodzie symbolicznego przyswajania pojęć (ang. *symbolic concept acquisition, SCA*) system uczy się, konstruując symboliczną reprezentację danego zbioru pojęć drogą analizy przykładów i kontrprzykładów tych pojęć. Reprezentacja może mieć postać wyrażenia logicznego, drzewa decyzyjnego, reguł produkcji lub sieci semantycznej. Niektóre spośród systemów skonstruowanych według tego paradygmatu znalazły praktyczne zastosowanie w wielu

dziedzinach. Przykładami mogą być programy: *ARCH* [26], *AQVAL* [11] i *ID3* [22]. W tym paradygmacie, atrybuty oraz predykaty odnoszące się do danego pojęcia są wprowadzane do systemu przez nauczyciela (instruktora).

W paradygmacie uczenia się w wąskiej dziedzinie z intensywnym wykorzystaniem wiedzy (ang. *knowledge-intensive domain-specific learning, KDL*) system zawiera liczne (zdefiniowane z góry) pojęcia, struktury wiedzy, ograniczenia dziedzinowe, reguły heurystyczne i wbudowane na stałe transformacje odnoszące się do konkretnej dziedziny. Nie wszystkie niezbędne atrybuty i pojęcia są znane na wstępie; oczekuje się, że system wygeneruje nowe w procesie uczenia się (autor nazywa proces tego rodzaju *indukcją konstruktywną*). Tak więc, główną różnicą między paradygmatem uczenia się w wąskiej dziedzinie z intensywnym wykorzystaniem wiedzy a paradygmatem symbolicznego przyswajania pojęć jest ilość i rodzaj wiedzy podstawowej wprowadzonej do systemu. Systemy uczące się oparte na tym pierwszym podejściu są zwykle budowane dla konkretnej dziedziny i nie mogą być bezpośrednio wykorzystywane w innej dziedzinie. Badania związane z tym paradygmatem dotyczą nie tylko uczenia się na przykładach, lecz również uczenia się przez analogię i uczenia się przez obserwację i odkrywanie (p. następny punkt). Przykładami systemów opartych na tym podejściu są: *Meta-DENDRAL* [3] i *AM* [10].

Wiele budowanych ostatnio systemów wykorzystuje obydwie powyższe podejścia. Ciekawą kombinację paradygmatów *SCA* i *KDL* zastosowano w systemach opartych na idei wymiennego modułu wiedzy. System taki wiąże ogólne mechanizmy uczenia się z wbudowanymi mechanizmami definiowania i używania wiedzy specjalistycznej. Jeśli ma on być zastosowany do rozwiązania określonego problemu, to niezbędna wiedza musi być najpierw dostarczona przez nauczyciela. Dzięki oddzieleniu ogólnych umiejętności wnioskowania od wiedzy specjalistycznej, takiego systemu można używać w wielu różnych dziedzinach, zachowując przy tym zalety systemów specjalistycznych. Na tej zasadzie jest oparty m.in. system *INDUCE*, który uczy się strukturalnego opisu obiektów na podstawie przykładów [12]. Innym przykładem takiego podejścia jest program *Winston* uczący się przez analogię [27]. Również system *LEX* udoskonalający heurystyki [19] oraz program *Eurisko* odkrywający nowe heurystyki [10] mogą być zaliczone do tej kategorii. W książce [18, rozdział 14] opisano podejście tego rodzaju oparte na analogii wynikowej (ang. *derivational analogy*).

Historię badań związanych z tymi trzema paradygmatami przedstawiono w [18, rozdział 1]. Przykłady aktualnie prowadzonych badań nad systemami samoorganizującymi się można znaleźć w publikacji [4]. Langley i Carbonell dokonali przeglądu podejść do maszynowego uczenia się [9]. Książka [18] koncentruje się przede wszystkim na symbolicznym przyswajaniu wiedzy i uczeniu się w wąskiej dziedzinie, z intensywnym wykorzystaniem wiedzy.

STRATEGIE UCZENIA SIĘ

Uczeń transformuje informację dostarczoną mu przez nauczyciela (lub środowisko) na pewną nową postać, którą zapamiętuje. Sposób tego przekształcania określa typ użytej strategii uczenia się. Wyróżnia się następujące podstawowe strategie: uczenie się na pamięć, uczenie się przez instruktaż, uczenie się przez dedukcję, uczenie się przez analogię

i uczenie się przez indukcję. W tej ostatniej strategii można wyodrębnić uczenie się na przykładach oraz uczenie się przez obserwację i odkrywanie. Powyższe strategie zostały wymienione w kolejności rosnącej złożoności procesu przekształcającego początkowe informacje w docelową wiedzę. Kolejność ta odpowiada wzrostowi wysiłku ze strony ucznia i zmniejszaniu wysiłku ze strony nauczyciela. Wyróżnienie wymienionych strategii jest istotne z punktu widzenia dydaktyki, jak również przyszłego rozwoju systemów uczących się. Człowiek w procesie uczenia się stosuje zwykle łącznie kilka strategii. Pomimo że większość obecnych systemów uczących się jest oparta na jednej strategii, należy oczekiwać, że w przyszłości więcej uwagi poświęci się systemom wielostrategicznym. Poniżej podano krótką charakterystykę strategii uczenia się. Opis szczegółowy można znaleźć w [13].

W uczeniu się na pamięć w zasadzie nie następuje żadne przekształcenie informacji. Informacja jest zapamiętywana w postaci podawanej przez nauczyciela. Głównym zadaniem systemu jest odpowiednie poindeksowanie informacji w celu usprawnienia procesu wyszukiwania. W uczeniu się przez instruktora informacja od nauczyciela podlega selekcji i przeformułowaniu (głównie na poziomie syntaktycznym). W uczeniu się przez dedukcję uczeń przeprowadza dedukcję na podstawie wiedzy i zapamiętuje użyteczne wnioski (strategia ta została wyodrębniona jako osobna kategoria dopiero ostatnio [13, 17]).

Jeśli w procesie uczenia się najpierw uogólnia się informację, a następnie ocenia wyniki (czyli przeprowadza się wnioskowanie indukcyjne), to mamy do czynienia z uczeniem się przez indukcję. Uczenie się przez analogię łączy w sobie uczenie się przez dedukcję i indukcję. Porównuje się w nim opisy z różnych dziedzin w celu wyodrębnienia wspólnej podstruktury, która jest następnie odwzorowywana przez analogię. Poszukiwanie wspólnej struktury jest oparte na wnioskowaniu indukcyjnym, natomiast odwzorowywanie przez analogię jest formą dedukcji. Uczenie się przez przypominanie (ang. learning by reminding) można również uważać za pewną formę uczenia się przez analogię.

Uczenie się przez analogię opisano w trzech rozdziałach książki [18]. Burstein [18, rozdział 13] przedstawił model uczenia się z wykorzystaniem rozumowania przez analogię, opisując go w kontekście przyswajania semantyki instrukcji przypisania języka Basic. Zgodnie z tym modelem, użycie analogii do uczenia się pojęć w nowej dziedzinie zależy ściśle od przyczynowych abstrakcji utworzonych uprzednio w jakiejś dziedzinie pokrewnej. Analogie te są stopniowo rozbudowywane w celu dostosowywania ich do coraz większej liczby sytuacji. Z kolei, Carbonell [18, rozdział 14] przedstawił teorię analogii wynikowej i jej wpływ na wnioskowanie i przyswajanie ekspertyz. Wnioskowanie prowadzące do rozwiązania pewnej klasy problemów może być powtarzane i modyfikowane do rozwiązania nowych, bardziej skomplikowanych problemów. Metodę tę można wykorzystać do automatycznego przyswajania wiedzy i umiejętności w systemach ekspertowych. Dershowitz [18, rozdział 15] omówił wykorzystanie analogii w automatycznym programowaniu, pokazując jak można posłużyć się analogiami między specyfikacjami programów na etapie uruchamiania lub modyfikowania istniejącego programu w celu rozszerzenia jego możliwości. Analogie mogą być również wykorzystane do wyodrębnienia schematu abstrakcyjnego programu ze zbioru programów oraz do ukonkretniania schematu w celu otrzymania właściwego programu.

Uczenie się przez indukcję można podzielić na uczenie się na przykładach i uczenie się przez obserwację i odkrywanie. W uczeniu się na przykładach (zwanym również przyswajaniem pojęć) celem jest ustalenie uogólnienia tłumaczącego wszystkie przykłady pozytywne i wykluczającego wszystkie przykłady negatywne danego pojęcia. Źródłem przykładów może być nauczyciel, znający dane pojęcie lub środowisko, w którym uczeń wykonuje doświadczenia i z którego otrzymuje odpowiedzi. W tym drugim wypadku mamy do czynienia z uczeniem się przez eksperymentowanie, które obejmuje uczenie się przez działanie i uczenie się przez rozwiązywanie problemu. Uczenie się na zasadzie pobudzenie-odpowiedź (ang. stimulus-response learning) może być również sklasyfikowane jako uczenie się na przykładach.

Najnowsze badania wyodrębniły dwa ciekawe rodzaje uczenia się na przykładach: uogólnianie na zasadzie egzemplarz-klasa (ang. instance-to-class generalization) oraz uogólnianie na zasadzie część-całość (ang. part-to-whole generalization). W uogólnieniu na zasadzie egzemplarz-klasa

system analizuje niezależne przykłady obiektów należących do pewnej klasy celem wyindukowania ogólnego opisu klasy. Większość dotychczasowych prac związanych z uczeniem się na przykładach dotyczy uogólniania na zasadzie egzemplarz-klasa. Obiektami mogą tu być struktury złożone z bloków, figury geometryczne, opisy chorób, opowiadania, rozwiązania problemów, operatory itp. Różne aspekty zadań tego rodzaju przedyskutowano w [18].

Winston [18, rozdział 3] podjął próbę interpretacji takich zagadnień, jak uczenie się z wykorzystaniem precedensów i ćwiczeń, użycie przybliżeń (ang. near-misses) w uczeniu się, uogólnienie reguł „if-then” i zastosowanie warunków „unless” zabezpieczających przed niewłaściwym użyciem reguły. Rola warunków „if-then” polega na zablokowaniu reguły „unless” pomimo spełnienia warunków zawartych w regule. Sposób ten ułatwia stopniowe korygowanie reguł. Utgoff [18, rozdział 5] zbadał rolę wskaźnika skłonności lub preferencji w określaniu prawdopodobnych hipotez uczenia się przez indukcję. Przedstawił pełną metodologię oraz program STABB, modyfikujący wskaźnik skłonności w trakcie uczenia się na przykładach. Quinlan [18, rozdział 6] rozważył wpływ szumów (zakłóceń) zawartych w analizowanych przykładach na proces odkrywania reguł klasyfikacyjnych i ich dokładność. Podał też szereg ciekawych propozycji, jak należy formułować cel uczenia się w wypadku, gdy przykłady będą zawierały szum.

Sammot i Banerji [18, rozdział 7] badali rolę przyswojonych już pojęć w uczeniu się nowych pojęć oraz zagadnienie uczenia się przez indukcję w wypadku aktywnego ucznia. Uczeń taki nie tylko akceptuje biernie przykłady pochodzące od nauczyciela, lecz również sam je generuje, pytając nauczyciela o ich przynależność do badanej klasy. Podobny problem rozważył Lebowitz [18, rozdział 8], który badał możliwość wykorzystania przechowywanych w pamięci pojęć w procesie uogólniania złożonych opisów strukturalnych. Jego metoda pamięci opartej na uogólnianiu (ang. generalisation-based memory) pozwala wybrać pojęcia do nauczenia się oraz sformułować ich definicje. Do jej realizacji służą dwa programy — program oceniający pojęcia oraz program uogólniający złożone opisy strukturalne. Konratoff i Ganascia [18, rozdział 9] rozważyli różne teoretyczne aspekty procesu uogólniania i pokazali, jak można przeprowadzić uogólnianie tworząc związki między uczeniem się na przykładach (połączenia są reprezentowane przez wiązanie zmiennych). Przegląd innych metod uogólniania można znaleźć w [2] i [16].

W uogólnianiu na zasadzie część-całość dane są wybrane części obiektu (np. sceny, sytuacji, procesu), a celem jest wywnioskowanie opisu całego obiektu. Przykładem takiego wnioskowania może być rekonstrukcja wyglądu całego pomieszczenia na podstawie serii zdjęć jego części. Innym przykładem takiej formy uogólniania jest ustalenie reguły charakteryzującej ciąg obiektów lub proces na podstawie kilku kolejnych obiektów tego ciągu lub części procesu. Podstawy teoretyczne i metodologię uogólniania na zasadzie część-całość przedstawiono w [18, rozdział 4]. Opisano tu ogólną metodę opartą na kilku modelach regułowych, które pozwalają znaleźć regułę charakteryzującą dany ciąg obiektów i wyznaczającą dopuszczalne ciągi obiektów. Każdy obiekt ciągu jest opisany przez pewną liczbę atrybutów, z których niektóre są podane a priori, inne zaś wyprowadzane na podstawie reguł wnioskowania i transformacji ciągów. Z zagadnieniami uogólniania na zasadzie część-całość są ściśle powiązane badania nad przewidywaniem procesów jakościowych [17].

W uczeniu się przez obserwację i odkrywanie (zwanym również uogólnianiem opisowym) uczeń poszukuje bez pomocy nauczyciela reguł wyjaśniających zaobserwowane fakty. Ta forma uczenia się obejmuje grupowanie koncepcji (ang. conceptual clustering), czyli tworzenie klas obiektów dających się opisać przez proste pojęcia, klasyfikowanie, odkrywanie praw wyjaśniających zbiór obserwacji i formułowanie teorii opisujących działanie systemu. Jako warianty tej strategii można traktować algorytmy genetyczne [18, rozdział 20] i algorytmy przewidywania empirycznego [28]. Różne jej aspekty przedyskutowano w [18].

Langley i in. [18, rozdział 16] opisali cztery systemy dotyczące różnych rodzajów odkryć naukowych. Program BAKON6 formułuje prawa empiryczne charakteryzujące dowolne liczbowe dane doświadczalne, GLAUBER odkrywa jakościowe prawa reakcji chemicznych, STAHL zajmuje się określeniem składników substancji w tych reakcjach, a

DALTON koncentruje się na formułowaniu strukturalnych modeli takich reakcji. W [18, rozdział 17] omówiono prace nad grupowaniem koncepcji, tzn. klasyfikowaniem obserwacji przez identyfikowanie podklas odpowiadających prostym pojęciom. W przeciwieństwie do poprzednich prac nad niezależną od celu (ang. goal-free) klasyfikacją obiektów niestrukturalnych, nowe podejście zajmuje się ukierunkowaną (ang. goal-oriented) klasyfikacją obiektów strukturalnych. Autorzy pokazali na przykładach, w jaki sposób wykorzystuje się koncepcje uczenia i reguły wnioskowania do tworzenia takich klasyfikacji. Amarel rozważał zagadnienia formowania teorii w kontekście syntezy programów. Swą metodę zilustrował w [18, rozdział 18] za pomocą zadania „wynioskowania” programu z powiązań danych wejściowo-wyjściowych, określonych w dziedzinie częściowo uporządkowanych struktur. Metoda ta akcentuje rolę modeli algebraicznych i geometrycznych oraz znaczenie wprowadzenia reprezentacji problemu do zadania syntezy programu. Inny kierunek obrał De Jong [18, rozdział 19]. Jego metoda uczenia się przez obserwację polega na kierowaniu procesem uogólniania na podstawie pojedynczego przykładu, przy wykorzystaniu wewnętrznych ograniczeń między pojęciami zawartymi w podstawowej wiedzy systemu. Jako przykłady dobrał on opowiadania o zachowaniu się ludzi przy rozwiązywaniu zadań.

Podstawą uczenia się przez indukcję i analogię jest wnioskowanie indukcyjne. Wnioskowanie indukcyjne opiera się na zbiorze faktów (obserwacji) oraz ewentualnym zbiorze hipotez a priori dotyczących tych faktów, a w wyniku daje najprawdopodobniejsze uogólnienie wyjaśniające te fakty. Jak już wspomniano, jest to wnioskowanie nie zachowujące prawdy, które przeprowadza się za pomocą reguł wnioskowania uogólniającego [14]. Jak stwierdził Popper [21], „czysta” indukcja, tzn. bezpośrednio wywnioskowanie teorii z faktów bez pomocy pojęć wyjaśniających, jest niemożliwa. Pojęcia te są potrzebne do opisu obserwacji i stanowią część wiedzy podstawowej ucznia. Ta wiedza jest niezbędnym składnikiem każdego procesu indukcyjnego. Zawiera ona również cele uczenia się, ograniczenia specyficzne dla danej dziedziny, związki przyczynowo-skutkowe, heurystyki i skłonności ukierunkowujące proces uogólniania oraz kryteria oceny stawianych hipotez.

Można wyróżnić dwie techniki kierowania i ograniczania procesu uogólniania: technikę opartą na podobieństwach lub na ograniczeniach. **Technika oparta na podobieństwach** bada powiązania między przykładami, tzn. analizuje pozytywne i negatywne przykłady pojęcia w celu stworzenia jego opisu. Poszukuje się cech wspólnych dla faktów lub przykładów tej samej klasy, jak również związków przyczynowo-skutkowych i wyjaśnień, dzięki którym pozornie różne przykłady można zaliczyć do tej samej klasy. Uogólnianie następuje w wyniku ignorowania różnic między przykładami lub przez formułowanie pojęć określających te różnice. Kilka wczesnych metod wykorzystujących tę technikę opisano w [18, rozdział 3]

Technika oparta na ograniczeniach bada związki między przykładami, które ograniczają pojęcia wyjaśniające dane przykłady. Każde z zastosowanych uogólnień musi spełniać te ograniczenia. Na przykład, uogólniając fakt „pudełko leży na stole” trzeba uwzględnić ograniczenie, że przedmiot leżący na stole nie może być ani zbyt ciężki (żeby nie załamał stołu), ani zbyt duży (żeby zmieścić się na stole). Jeden z wariantów tej techniki opisano w [1], wprowadzając pojęcie **uzasadniania** hipotezy. Inny wariant jest nazywany przez niektórych autorów **uogólnianiem opartym na wyjaśnianiu**, w celu podkreślenia znaczenia stosowanej wiedzy wyjaśniającej (termin ten nie jest zbyt precyzyjny, gdyż każdy rodzaj uczenia się przez indukcję zawiera w definicji poszukiwanie wyjaśnień stwierdzonych faktów). W [18, rozdział 19] De Jong omówił metodę wykorzystania tej techniki do rozumienia opowiadań. Techniki oparte na podobieństwach i ograniczeniach dopełniają się wzajemnie i w systemach uczących się mogą być stosowane łącznie.

ORIENTACJE UCZENIA SIĘ

W poprzednich punktach podano dwa ważne kryteria klasyfikacji badań nad maszynowym uczeniem się: paradygmaty uczenia się i strategię uczenia się. Pierwsze kryterium dotyczy rodzaju wiedzy reprezentowanej i przetwarzanej w systemie, drugie jest związane z rodzajem wykorzystanego wnioskowania. W tym punkcie omówiono pokrótce jeszcze jedno kryterium klasyfikacji, a mianowicie

orientację badań, które określają zakres i temat prowadzonych prac. Używając obrazowego porównania, paradygmatowi odpowiada punkt startowy i teren, po którym się poruszamy, strategia określa środek lokomocji, a orientacja — kierunek jazdy.

W badaniach dotyczących maszynowego uczenia się można wyróżnić trzy wzajemnie powiązane orientacje:

- analiza teoretyczna i rozwój ogólnych algorytmów uczenia się,
- rozwój obliczeniowych modeli opisujących procesy ludzkiego uczenia się,
- badanie zastosowań wraz z tworzeniem specjalistycznych systemów uczących się (zwane również orientacją inżynierską).

Pierwsza orientacja bada teoretyczne (lub uproszczone rzeczywiste) zadania uczenia się i rozwija algorytmy, które realizują te zadania niezależnie od dziedziny zastosowań. Nie wprowadza się żadnych ograniczeń co do rodzaju konstruowanego algorytmu. Nie musi on być zgodny z metodami uczenia się stosowanymi przez ludzi. Niektórzy autorzy postulują jednak, aby przynajmniej struktury wiedzy powstającej w wyniku maszynowego uczenia się były podobne do tych, które może utworzyć człowiek; same procesy tworzenia struktur wiedzy mogą być różne [15]. Ten kierunek badań zmierza do określenia teoretycznej przestrzeni algorytmów uczenia się. Prace typowe dla tej orientacji przedstawiono w [18, rozdziały 3, 5, 7 i 9].

Druga orientacja, zwana również **modelowaniem poznawczym** (ang. cognitive modeling), koncentruje się na sposobach uczenia się przez człowieka, konstruując teorie obliczeniowe i modele eksperymentalne. Badania te mogą mieć znaczny wpływ na rozwój technik budowy systemów maszynowego uczenia się, jak też na szeroko rozumianą edukację człowieka. Rosenbloom i Newell [18, rozdział 10] opisali pewne podejście do modelowania procesów leżących u podstaw zwiększania umiejętności przez praktykę. Ich model praktyki jest oparty na pojęciu grupowania podcelów w cele. Pokazali oni, że zaproponowany model odzwierciedla znaną zasadę mówiącą o skuteczności ludzkiej praktyki. Anderson [18, rozdział 11] rozważył mechanizmy uczenia się wykorzystywane w kompilacji wiedzy, tzn. w procesie przechodzenia od deklaratywnej reprezentacji umiejętności do reprezentacji proceduralnej. Pokazał on, jak mechanizmy **kompozycji** (łączenia kilku reguł produkcji w jedną) i **proceduralizacji** (tworzenia reguł produkcji na podstawie wiedzy deklaratywnej) mogą symulować początkowe stadia zdobywania umiejętności w dziedzinie uczenia się programowania.

Trzecia orientacja badawcza koncentruje się na praktycznych zadaniach uczenia się i budowie systemów realizujących te zadania. Jej przykładem może być program uczący się rozpoznawać warunki niebezpieczne dla samolotu w czasie lotu. W tego rodzaju zadaniach pojawiają się na ogół inne problemy, nie związane bezpośrednio z uczeniem się, np. interpretacja sygnałów wejściowych i przetwarzanie danych. W orientacji tej korzysta się z osiągnięć dwóch pozostałych orientacji badawczych. Często po znalezieniu rozwiązania konkretnego problemu, podejmuje się próbę uogólnienia go w celu skonstruowania metody rozwiązywania szerszej klasy podobnych problemów. Przykłady takich badań opisano w [18, rozdział 4].

Powyższe trzy orientacje tworzą triadę wzajemnie powiązanych i uzupełniających się kierunków badań w dziedzinie maszynowego uczenia się. Taki podział przeniknął również do pozostałych działów sztucznej inteligencji.

* * *

Niniejszy artykuł był w pierwotnej wersji wprowadzeniem do książki [18]. Jej zawartość dotyczy głównie paradygmatów symbolicznego przyswajania pojęć oraz uczenia się w wąskiej dziedzinie z intensywnym wykorzystaniem wiedzy i jest poświęcona omówieniu strategii uczenia się przez indukcję i przez analogię. Reprezentowane są w niej oba główne rodzaje indukcyjnego uczenia się, tj. uczenie się na przykładach oraz uczenie się przez obserwację i odkrywanie.

Książka zawiera również bibliografię prac na temat maszynowego uczenia się, począwszy od 1980 roku, z kilkoma ważniejszymi uzupełnieniami z lat wcześniejszych (obszerną bibliografię dotyczącą poprzedniego okresu zawarto w

książce [13]). Bibliografia zawiera też skorowidz zawartości, opracowany według strategii uczenia się, dziedziny zastosowań i metody badawczej. Do książki włączono także słowniczek terminów używanych w dziedzinie maszynowego uczenia się oraz noty bibliograficzne wszystkich autorów.

Artykuł powstał podczas pobytu autora w Artificial Intelligence Laboratory, w Massachusetts Institute of Technology. Autor pragnie podziękować Patrickowi Winstonowi za zaproszenie do współpracy oraz wiele zapładniających uwag i komentarzy na temat poprzednich wersji tego artykułu. Niepowtarzalna, twórcza atmosfera AT Laboratory i dyskusje prowadzone w Learning Group były pomocne w opracowaniu przedstawionych tu idei. Częściową pomoc finansową uzyskano w ramach kontraktu N00014-80-C-0505 z Advanced Research Projects Agency.

Autor dziękuje współredaktorom książki Jaime Carbonellowi i Tomowi Michelowi za współpracę i uwagi krytyczne. Dyskusje z Randy Davisem, z MIT AI Laboratory, były bardzo pomocne w udoskonalaniu artykułu. Zaproponował on zastosowanie maszyn uczących się do zabezpieczenia się przed niewłaściwym wykorzystaniem ich samych. Richard Doyle, Michel Kaskhet, Boris Katz i David Kirsh, z MIT AI Laboratory, oraz Allan Collins, z firmy Bolt, Beranek and Newman Inc. przedstawili ważne opinie i komentarze dotyczące pierwotnej wersji artykułu. Autor wyraża wdzięczność Bobowi Steppowi, Larry'emu Rendellowi, Jeffowi Backerowi, Bruce'owi Katzowi i Brianowi Stoutowi, z AI Laboratory Uniwersytetu stanu Illinois, za wartościowe sugestie i konstruktywną krytykę. Wiele pożytecznych uwag przedstawił Ken Forbus, z Uniwersytetu stanu Illinois. Autor jest zobowiązany Gail Thornburg z Wyższej Szkoły Bibliotekarskiej za cenne uwagi i krytykę. Ważne sugestie przedstawił Jan Gorecki z Wydziału Socjologii Uniwersytetu stanu Illinois. Praca ta była finansowana przez National Science Foundation w ramach umowy DCR-8406801.

Thum. i oprac.

**EDMUND PIERZCHAŁA
PIOTR ZIELCZYŃSKI**

LITERATURA

- [1] Andreae P. M., Constraint Limited Generalization — Acquiring Procedures from Examples. Proc. AAAI-84, Austin (TX), August 1984
- [2] Cohen P. R., Feigenbaum E. A. (eds.): The Handbook of Artificial Intelligence, Vol. III. Kaufmann, Los Altos (CA), 1982
- [3] Buchanan B. G., Shortliffe E. H. (eds.): Rule-based Expert Systems. Addison-Wesley, Reading (MA), 1984
- [4] Calaniello E. R., Musso G.: Cybernetic Systems — Recognition, Learning, Self-Organization. Research Studies Press, Letchworth (Hertfordshire), Wiley, New York, 1984
- [5] Conrad M.: Adaptability. Plenum Press, New York 1983
- [6] Fogel L., Owens A., Walsh M.: Artificial Intelligence Through Simulated Evolution, Wiley, New York, 1966
- [7] Hinton G. E., Sejnowski T. J., Ackley D. H.: Machines — Constraint Satisfaction Networks that Learn. Technical Report CMU-CS-84-119, Carnegie-Mellon University, Pittsburgh (PA), 1984
- [8] Holland J.: Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, 1975
- [9] Langley P., Carbonell J. G.: Approaches to Machine Learning. Journal of the American Society for Information Science, Vol. 35, No. 5, 1984
- [10] Lenat D. G.: The Role of Heuristics in Learning by Discovery — Three Case Studies. [18]
- [11] Michalski R. S.: Variable-valued Logic and Its Applications to Pattern Recognition and Machine Learning. Pp. 506—534, Computer Science and Multiple-valued Logic — Theory and Applications, D. C. Rine (ed.), North-Holland, Amsterdam, 1975
- [12] Michalski R. S.: Pattern Recognition as Rule-Guided Inductive Inference. IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol. PAMI-2, No. 4, pp. 349—381, July 1980
- [13] Michalski R. S., Carbonell J. G., Mitchell T. M. (eds.): Machine Learning — An Artificial Intelligence Approach. TIOGA, Palo Alto (CA), 1983
- [14] Michalski R. S.: A Theory and Methodology of Inductive Learning. Artificial Intelligence, No. 20, 1983
- [15] Michalski R. S. (ed.): Proc. of the International Machine Learning Workshop, Allerton House, University of Illinois, 22—24 June 1983
- [16] Michalski R. S.: Learning Strategies and Automated Knowledge Acquisition — An Overview. Knowledge-Based Learning Systems, Bolc L. (ed.), Springer-Verlag, 1985
- [17] Michalski R. S., Ko H., Chen K.: Qualitative Process Prediction — A Method and a Program SPARC/G, Report ISG-12, Dept. of Computer Science, University of Illinois, Urbana (IL), 1985
- [18] Michalski R. S., Carbonell J. G., Mitchell T. M. (eds.): Machine Learning. Vol. 2. Morgan and Kaufmann Publishers, 1986
- [19] Mitchell T. M., Utgoff P. E., Banerji R.: Learning by Experimentation — Acquiring and Refining Problem-Solving Heuristics. [18]
- [20] Nilson N. J.: Maszyny uczące się, PWN, Warszawa, 1988
- [21] Popper K. R.: Objective Knowledge — An Evolutionary Approach, Revised edition. Oxford, 1981
- [22] Quinlan J. R.: Discovering Rules from Large Collections of Examples — A Case Study. Expert Systems in the Microelectronics Age, D. Michie (ed.), Edinburgh University Press, 1979
- [23] Rosenblatt F.: The Perceptron — A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, Vol. 65, p. 386—407, 1958
- [24] Selfridge M.: Pandemonium — A paradigm for Learning. Pp. 511—529, Proc. of the Symposium on Mechanization of Thought Processes, D. Blake, A. Uttley eds., HMSO, London, 1959
- [25] Cypkin J. Z.: Podstawy teorii układów uczących się. WNT, Warszawa, 1973
- [26] Winston P. H.: Learning Structural Descriptions from Examples. The Psychology of Computer Vision, P. H. Winston (ed.), McGraw-Hill, New York, 1975
- [27] Winston P. H.: Learning and Reasoning by Analogy. Communications of the ACM, Vol. 19, No. 3, 1982
- [28] Zagorulko N.: Empirical Prediction Algorithms, Computer Oriented Learning Processes, J. C. Simon (ed.), Noordhoff, Leyden, 1976.