Tomasz Smoleń

BAYESOWSKA TEORIA OCENY PODOBIEŃSTWA I KATEGORYZACJI

Wprowadzenie

Można zaryzykować stwierdzenie, że kategoryzacja, uogólnianie, ocena podobieństwa i tworzenie pojęć są, na pewnym poziomie ogólności, tym samym zjawiskiem. U podstawy wszystkich tych procesów leży problem określenia pewnej relacji, wyrażającej najogólniej rozumianą zgodność własności. Aby nie ograniczać możliwości formułowania wniosków odnoszących się do szerokiej klasy problemów, z jednej strony, ale i nie osłabiać siły tych wniosków przez zbyt szeroką definicję kategoryzacji – z drugiej, należy rozważnie określić jakie rodzaje procesów oceny podobieństwa będą przedmiotem analizy w niniejszym rozdziale. Wydaje się, że zbytnim uproszczeniem, połączonym z groźbą zagubienia sensu pojęcia „kategoryzacja", byłoby rozważanie dowolnego rodzaju relacji. Także relację bliskości lub pokrewieństwa, określoną wyłącznie na podstawie cech obiektów, ich odległości i rozkładu w przestrzeni własności, chociaż może ona stanowić ważny krok w procesie kategoryzacji lub użyteczną heurystykę eksploracyjną (jak na przykład analiza skupień), trudno traktować jako podstawę do tworzenia pojęć. Najbardziej interesujące są przypadki, kiedy relacja jest określana na podstawie dostępnych cech obiektów, jednak kryterium oceny podobieństwa jest pewna cecha, która jest ważna ze względu na adaptację podmiotu do środowiska i, przy tym, nie jest bezpośrednio dostępna podmiotowi określającemu to podobieństwo.

Można wskazać dwie konsekwencje przyjęcia powyższej definicji problemu oceny podobieństwa. (a) Istnieje wiele możliwych dobrych lub najlepszych kategoryzacji tego samego zbioru. (b) Kategoryzacja może wymagać wnioskowania w celu uzyskania koniecznych do jej przeprowadzenia informacji. Chodzi tu o informacje, które są kluczowe dla procesu określenia relacji między własnościami jawnymi a własnością kryterialną i które, w nietrywialnym przypadku, nie są dostępne bezpośrednio. W szczególności, kategoryzacja ze względu na dane kryterium może być niemożliwa, w przypadku braku związku między własnością kryterialną a własnościami dostępnymi poznawczo lub gdy związek ten przybiera postać funkcji nieobliczalnej efektywnie.

Dwa opisy procesu tworzenia kategorii są rozważane najczęściej: przez poszukiwanie funkcji odległości w jakiejś przestrzeni podobieństwa oraz przez tworzenie klasy abstrakcji. Oba opisy, pomimo zasadniczych różnic na poziomie formalnym, są w istocie dwoma sposobami ujęcia tego samego problemu: ciągłym i dyskretnym. Ponieważ poszukiwanie funkcji odległości jest problemem ogólniejszym, ten właśnie proces zostanie tutaj utożsamiony z tworzeniem pojęć. Powyższe sformułowanie problemu wymaga abstrahowania od nazywania pojęć, czy ogólniej, od jakichkolwiek ich własności wynikających z ich związków z językiem.

Podawanie na tym etapie wywodu przykładu sytuacji, w której organizm staje przed zadaniem kategoryzacji, może wyglądać na zabieg stojący w sprzeczności z dążeniem do maksymalnej ogólności. Ponadto, ponieważ problem jest rozważany na dość elementarnym poziomie – może się wydawać niepotrzebne. Jednak uzyskanie intuicji co do tego, jak rozumiane są poszczególne elementy prezentowanych tu teorii, może okazać się użyteczne dla śledzenia dalszych części prezentowanego tu rozumowania.

Za przykład wnioskowania o przynależności do kategorii niech posłuży sytuacja organizmu, który mając do wyboru owoce o różnych właściwościach, musi wybrać te, które są jadalne (owoce są więc dzielone na dwie kategorie – jadalne i niejadalne). Jawne cechy, na podstawie których następuje kategoryzacja to, na przykład, rozmiar owocu, jego kolor, wysokość, na jakiej rośnie lub pora dnia, o jakiej został znaleziony. Ukryta własność, będąca kryterium przynależności do kategorii, to jadalność owocu. Załóżmy, że na początku procesu kategoryzacji organizm nie ma żadnej wiedzy o relacji między obserwowalnymi cechami owoców a ich cechą kryterialną. Sytuacja zmienia się nieco po zjedzeniu pierwszego owocu i stwierdzeniu, że był jadalny. Dla organizmu, który nie dysponuje żadnymi metodami wnioskowania o podobieństwie obiektów, ta informacja jest bezużyteczna, ponieważ owoc,

o którym informację posiada organizm, został już zjedzony. Natomiast organizm, który ma możliwość oszacowania prawdopodobieństwa, że inny owoc, mniej lub bardziej do tamtego podobny, okaże się również jadalny, dysponuje narzędziem poznawczym o trudnej do przecenienia wartości adaptacyjnej. Tak więc pytanie na jakie musi odpowiedzieć organizm w opisanej wyżej sytuacji brzmi: „Jeżeli owoc o rozmiarze *n* okazał się być jadalny, to jakie jest prawdopodobieństwo, że owoc o rozmiarze *n+a* okaże się również być owocem jadalnym?", i dalej: „Jak to prawdopodobieństwo zależy od *a*?".

Rzecz jasna, analogiczne pytanie można zadać dla każdego z wymiarów, na jakich da się opisać kategoryzowane obiekty. W środowisku naturalnym można wskazać wiele, potencjalnie wręcz nieskończenie wiele takich wymiarów opisu, przy czym często tylko bardzo niewiele z nich ma związek z własnością kryterialną. Przypadek, w którym obiekty posiadają wiele wymiarów nie mających związku z cechą kryterialną, daje się oczywiście opisać za pomocą przyjętej definicji kategoryzacji – funkcja podobieństwa dla wymiarów niezwiązanych z własnością kryterialną byłaby w takim przypadku stała. Trudno jednak oprzeć się wrażeniu, że uwzględnienie wielu nieistotnych wymiarów przy tworzeniu funkcji podobieństwa w nieuzasadniony sposób zwiększa złożoność problemu. Dlatego wydaje się, że wskazane jest podzielenie procesu tworzenia kategorii na dwa kroki: (a) wyodrębnienie wymiarów istotnych ze względu na kryterium oraz (b) określenie funkcji podobieństwa wyłącznie na owych istotnych wymiarach.

Nie zostaną tu rozważone interesujące problemy dotyczące tego, w jakiej sytuacji wprowadzenie pierwszego, opisanego wyżej, kroku staje się korzystne ani tego, jaki mechanizm wyodrębniania istotnych wymiarów jest optymalny. Pewnych sugestii dotyczących rozwiązania pierwszego z tych problemów dostarczyć mogą studia nad wstępnym przetwarzaniem (*preprocessing*, zob. Maehigashi i Miwa, 2010). Rozwiązanie problemu drugiego zaproponowali Jones i Cañas (2010), opierając je na mechanizmie uczenia się ze wzmocnieniem. Inną propozycję rozwiązania problemu selekcji istotnych wymiarów prezentuje praca Gershmana, Cohena i Niv (2010). Druga z wymienionych propozycji także zawiera elementy uczenia się ze wzmocnieniem, jednak głównym mechanizmem wnioskowania o istotności wymiarów jest w niej wnioskowanie Bayesowskie.

Wielość problemów teoretycznych, związanych z tworzeniem pojęć, klas i kategorii oraz możliwości ujęcia tego zagadnienia, znalazła odzwierciedlenie w różnorodności psychologicznych prób teoretycznego opisu tych zjawisk (np. Tversky, 1977; Rosch, 1978; Murphy i Medin, 1985; Sheppard, 1987; Jackendoff, 1990). Najszerzej dyskuto-

wane ujęcia podstawowej wersji opisywanego problemu, to jest oceny podobieństwa, pochodzą od Tversky'ego (1977) i Sheparda (1987).

Teoria Amosa Tversky'ego

Teoria Tversky'ego jest jedną z pierwszych psychologicznych teorii oceny podobieństwa. Jest ona psychologiczna w tym sensie, że przewiduje efekty, które są powszechnie obserwowane w zachowaniu ludzi (Rosch i Mervis, 1975; Rothkopf, 1985; Wish, 1967), a które są niezgodne z normatywnymi teoriami metrycznymi (np. Torgerson, 1965). Spośród efektów przewidywanych przez teorię Tversky'ego, a nieobecnych w teoriach normatywnych na największą uwagę zasługują: brak symetryczności podobieństwa, brak komplementarności podobieństwa i różnicy oraz zależność podobieństwa od kontekstu.

Teorie normatywne zakładają, że obiekt *a* jest podobny do obiektu *b* w tym samym stopniu, w jakim *b* jest podobny do *a*. Dalej, w teoriach normatywnych spełniony jest warunek komplementarności podobieństwa i różnicy, a więc im bardziej dwa obiekty są do siebie podobne tym mniej są od siebie różne. I w końcu podobieństwo dwóch obiektów w teoriach normatywnych nie zależy od tego, na tle jakich obiektów jest ono oceniane. Powyższe własności wydają się być zasadne, przynajmniej w odniesieniu do abstrakcyjnego pojęcia podobieństwa, które jest skrojone na potrzeby formalnego opisu relacji między obiektami; jak jednak zostało wspomniane wyżej badania empiryczne pokazują, że ocena podobieństwa dokonywana przez ludzi rządzi się trochę innymi prawami. Dalej zostanie zaprezentowane wyjaśnienie nowych efektów pojawiających się w teorii Tversky'ego, jednak zanim to nastąpi, konieczne będzie krótkie zaprezentowanie samej teorii.

W modelu Tversky'ego podobieństwo między obiektami zależy od wartości funkcji *f* operującej na cechach wspólnych obu obiektom oraz na cechach specyficznych dla każdego z nich. Wzór (1) obrazuje relację między podobieństwem *S* obiektów *a* i *b* a zbiorami ich cech (*A* i *B*),

$$S(a, b) = \theta f\,(A \cap B) - \alpha f\,(\mathrm{A}\backslash\mathrm{B}) - \beta f\,(\mathrm{B}\backslash\mathrm{A}). \qquad (1)$$

Teoria Tversky'ego często bywa sprowadzana do powyższego wzoru, wyrażającego model kontrastu. W rzeczywistości jest to najczęściej stosowany model oceny podobieństwa, oparty na teorii Tversky'ego, jednak nie jedyny. Co więcej, teoria Tversky'ego nie tylko nie sprowa-

dza się do jednego modelu, ale nie zostaje wyczerpana nawet przez żaden skończony zbiór modeli. Teoria ta sformułowana została na poziomie ogólniejszym niż konkretny model. Jest ona systemem opartym na zbiorze założeń, pozwalającym na wywodzenie z niego modeli oceny podobieństwa, spełniających określone warunki.

Założenia, na których opiera się teoria Tversky'ego, to: (a) dopasowanie, (b) monotoniczność, (c) niezależność, (d) rozwiązywalność i (e) niezmienność. Założenie dopasowania wymaga, aby podobieństwo dwóch obiektów $s(a, b)$ było wyrażone przez funkcję trzech argumentów: cech wspólnych obiektów a i b, $(A \cap B)$, cech należących do a i nie należących do b, $(A \setminus B)$ oraz cech należących do b i nie należących do a, $(B \setminus A)$. Dalej, model spełnia założenie monotoniczności, jeżeli $s(a, b) \geq s(a, c)$ zawsze wtedy, kiedy $A \cap B \supset A \cap C$, $A \setminus B \subset A \setminus C$ oraz $B \setminus A \subset C \setminus A$. Oznacza to, że w każdym modelu zgodnym z teorią Tversky'ego podobieństwo między obiektami zwiększa się, jeżeli zwiększa się liczba ich cech wspólnych, a zmniejsza się, jeżeli zwiększa się liczba cech specyficznych dla któregoś z obiektów. Założenie niezależności wymaga, aby uszeregowanie wspólnego efektu którychkolwiek dwóch cech ocenianych obiektów było niezależne od trzeciej cechy. Rozwiązywalność wymaga, aby rozważana przestrzeń cech była wystarczająco bogata, żeby możliwe było znalezienie rozwiązań równań podobieństwa. Można zauważyć, że ostatnie założenie nie odnosi się do modeli oceny podobieństwa, ale do problemów, do rozwiązania których używa się tych modeli. W końcu, założenie niezmienności wymaga aby zachowana była równość interwałów między cechami obiektów.

Innym niż model kontrastu, powszechnie stosowanym modelem, który spełnia warunki teorii Tversky'ego jest model proporcji. Wzór (2) obrazuje zależność podobieństwa między obiektami od liczby ich cech wspólnych i dystynktywnych według tego modelu,

$$S(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A \setminus B) + \beta f(B \setminus A)}. \tag{2}$$

Podobieństwo zdefiniowane za pomocą opisanego wyżej systemu aksjomatów cechuje się wcześniej wspomnianymi własnościami. Mianowicie: brakiem symetryczności, brakiem komplementarności z różnicą oraz zależnością od kontekstu. Łatwo pokazać, że podobieństwo nie musi być relacją symetryczną. Zarówno z modelu kontrastu, jak i z modeli proporcji wynika, że:

$$S(a, b) = S(b, a) \text{ wtw } \alpha f(A \setminus B) + \beta f(B \setminus A) = \alpha f(B \setminus A) + \beta f(A \setminus B)$$
$$\text{wtw } (\alpha - \beta) f(A \setminus B) = (\alpha - \beta) f(B \setminus A).$$

Tak więc *a* jest podobne do *b* w takim samym stopniu, jak *b* jest podobne do *a* tylko jeżeli $\alpha = \beta$ lub $f(A \setminus B) = f(B \setminus A)$, co z pewnością nie musi być prawdą w ogólnym przypadku.

Komplementarność podobieństwa i różnicy polega na tym, że ocena różnicy jest liniową funkcją oceny podobieństwa o nachyleniu –1. Oznacza to, że wzrost podobieństwa między dwoma obiektami oznacza spadek różnicy między nimi. Jednak według teorii Tversky'ego nie musi tak być zawsze. Jeżeli cechy wspólne mają większą wagę w ocenie podobieństwa niż w ocenie różnicy, to dwa obiekty, które mają więcej cech zarówno wspólnych, jak i specyficznych, mogą być ocenione zarówno jako bardziej podobne, jak i jako bardziej różne niż obiekty, które mają mniej cech wspólnych i specyficznych.

Według teorii Tversky'ego podobieństwo między obiektami zależy od kontekstu, w jakim jest oceniane. W jednym z badań Tversky (1977) poprosił osoby badane o podzielenie czterech podanych państw na dwie grupy państw najbardziej do siebie podobnych. Austria zestawiona ze Szwecją, Polską i Węgrami była najczęściej (49%) grupowana ze Szwecją, podczas gdy zestawiona ze Szwecją, Norwegią i Węgrami, najczęściej (60%) trafiała do grupy razem z Węgrami. Działo się tak dlatego, że na wyrazistość cechy, a więc pośrednio na ocenę podobieństwa, ma wpływ dystynktywność tej cechy, czyli to, jak często różnicuje ona między obiektami w zbiorze, w odniesieniu do którego następuje ocena.

Teoria Tversky'ego, będąc dużym krokiem na drodze do zrozumienia procesu oceny podobieństwa przez ludzi, ma jednak pewne wady. Jedną z nich jest duża złożoność modelu. Model Tversky'ego posiada wolne parametry, których psychologiczna interpretacja nie jest łatwa. Dobór różnych wartości tych parametrów prowadzi do otrzymania różnych rodzajów funkcji podobieństwa. Również funkcja operująca na zbiorach cech może zostać dobrana na różne sposoby, pozwalając na uzyskanie wielu różnych hipotez. Należy pamiętać, że spojrzenie z jeszcze wyższego poziomu ogólności pozwala zauważyć, że przedstawione powyżej modele proporcji i kontrastu są tylko dwoma z wielu możliwych modeli zgodnych z aksjomatyczną teorią Tversky'ego.

Teoria Rogera Sheparda

Drugi ze wspomnianych modeli, model Sheparda, również cechuje się kilkoma interesującymi własnościami. Być może na największą uwagę zasługuje sam sposób ujęcia problemu w tej teorii. Przy tym

zastosowanie metody użytej przez Sheparda nie jest ograniczone do problemu oceny podobieństwa, ale może zostać rozszerzone na wiele zagadnień z dziedziny psychologii i kognitywistyki. Shepard za punkt wyjścia przyjął nie problem wyjaśnienia danych obserwacyjnych dotyczących zachowania badanych podmiotów w sytuacji generalizacji reakcji, ale zagadnienie, jakiego problemu rozwiązaniem jest owo zachowanie. Przyjmując pewne założenia dotyczące natury tego problemu i natury środowiska, w jakich jego rozwiązanie zachodzi, Shepard doszedł (dedukcyjnie) do hipotez dotyczących zachowania ludzi (czy zwierząt w ogóle) w rozważanej sytuacji. Dzięki sformułowaniu teorii w abstrakcyjnej przestrzeni psychologicznej, Shepard uzyskał niezmienniczość predykcji ze względu na warunki eksperymentalne i cechy badanych organizmów. Jako uzasadnienie dla owej niezmienniczości Shepard przywoływał twierdzenie, że prawo generalizacji jest skutkiem adaptacji do uniwersalnych własności świata.

Shepard rozważał problem generalizacji reakcji, to jest szacował prawdopodobieństwo, że zachowanie, będące wyuczoną reakcją na bodziec wzorcowy, zostanie zaobserwowane także jako reakcja na bodziec testowy – podobny do wzorcowego. Przyjął, że reakcja na bodziec testowy o tyle jest adaptacyjna, o ile oba bodźce dzielą własność, która czyni adaptacyjną reakcję na bodziec wzorcowy. W tej sytuacji odpowiedź na pytanie o to, jak prawdopodobna jest reakcja na bodziec podobny do uczącego, sprowadza się do pytania o to, jaka jest szansa, że obydwa bodźce, wzorcowy i testowy, dzielą pewną, kluczową własność. Shepard założył, że rozważaną własność posiadają bodźce, dla których wartość pewnej innej – ciągłej – własności leży w pewnym przedziale. Tak więc prawdopodobieństwo, że jakiś nowy bodziec będzie posiadał tę samą własność, co bodziec zaobserwowany, jest równe prawdopodobieństwu, że nowy bodziec leży w owym nieznanym przedziale, przy założeniu, że leży w nim bodziec zaobserwowany. Warto zwrócić uwagę na to rozumowanie, ponieważ okaże się ono być szczególnym przypadkiem wnioskowania Bayesowskiego, które jest podstawą modelu Tenenbauma i Griffithsa.

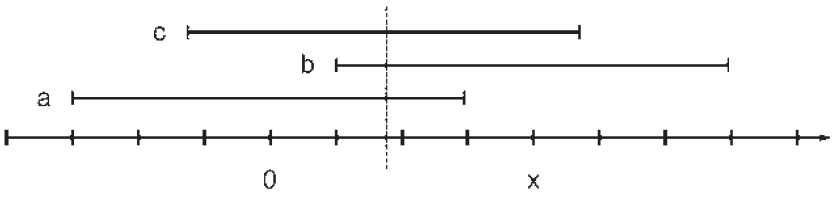
Szczegóły rozumowania Sheparda wyglądały następująco. Jeżeli przedział[1], zawierający wszystkie bodźce mające interesującą nas własność (przedział znaczący), ma określony rozmiar, to zbiór przedziałów, które zawierają zarówno zaobserwowany bodziec odniesienia

[1] Shepard rozważa nie tylko jednowymiarowe przedziały, ale również regiony dowolnej przestrzeni wielowymiarowej, jednak mówienie o przedziałach jednowymiarowych pozwoli na uproszczenie streszczenia wywodu autora bez zagubienia istoty jego rozumowania.

(przypiszmy mu wartość 0), jak i bodziec, o którym wnioskujemy (x) ogranicza się do przedziałów, których środek zawiera się w przecięciu się przedziału o środku w punkcie 0 i o środku w punkcie x (zob. rycinę 1). Ponieważ wszystkie lokalizacje przedziału znaczącego są równie prawdopodobne, warunkowe prawdopodobieństwo, że x zawiera się w przedziale znaczącym (jeżeli wiemy, że 0 się w nim zawiera) wynosi $(s-|x|)/s$, gdzie s jest rozmiarem przedziału. Ponieważ nie wiadomo, jaki jest rozmiar s przedziału znaczącego, należy rozważyć wszystkie możliwe rozmiary, biorąc pod uwagę prawdopodobieństwo $p(s)$ każdego rozmiaru. W konsekwencji tego wzór (3) na prawdopodobieństwo generalizacji ($g(x)$) własności bodźca 0 na bodziec x ma postać:

$$g(x) = \int_{|x|}^{\infty} p(s)\frac{s-|x|}{s}ds. \qquad (3)$$

Jeżeli funkcja $g(x)$ przyjmuje wartość 0 dla wszystkich wartości $s<0$ oraz oczekiwana wartość s jest skończona ($\int_0^{\infty} sp(s)ds < \infty$), to można pokazać, że funkcja prawdopodobieństwa generalizacji ($g(x)$) jest malejąca z dodatnim przyspieszeniem, a więc przypomina kształtem ujemną część funkcji wykładniczej. Faktycznie, niezależnie od rozkładu prawdopodobieństwa $p(s)$, kształt funkcji g odbiega od kształtu funkcji wykładniczej bardzo nieznacznie, a kiedy funkcja $p(s)$ przyjmuje postać rozkładu Erlanga (Angus, 2001), g staje się tożsama z rozkładem wykładniczym. Predykcje modelu, dotyczące wykładniczego kształtu zależności prawdopodobieństwa generalizacji reakcji od wielkości różnicy między bodźcem testowanym i wzorcowym, są zgodne z obserwacjami poczynionymi w wielu badaniach (m.in. Blough, 1961; McGuire, 1961; Shepard i Cermak, 1973).



Ryc. 1. Przedział o środku w punkcie 0 (a), o środku w punkcie x (b), i przykładowy przedział o środku w punkcie leżącym w przecięciu tych przedziałów (c), obejmujący zarówno 0 jaki i x

Zaproponowany przez Sheparda model przewiduje stosunkowo prostą relację pomiędzy odległością między dwoma porównywanymi bodźcami a oceną ich podobieństwa. Jak to możliwe, że taka relacja

nie została zaobserwowana, wziąwszy pod uwagę, że zjawiska oceny podobieństwa i generalizacji reakcji były wcześniej wielokrotnie badane? Można to wytłumaczyć tym, że ważną cechą modelu Sheparda jest przyjęcie jako przestrzeni odniesienia nie konkretnych wartości bodźców, ale pewnej abstrakcyjnej przestrzeni psychologicznej. Przed zastosowaniem przez Sheparda przekształcenia zwanego skalowaniem wielowymiarowym, analizowano zależności między wartościami bodźców wyrażonymi wprost (na przykład barwą wyrażoną jako długość fali świetlnej). Okazuje się, że w takim wypadku relacja między odległością między bodźcami a ich podobieństwem różni się w zależności od modalności, badanego organizmu czy typu bodźca (zob. przegląd wyników w: Shepard, 1987). Shepard przez analogię z procesem abstrakcji zastosowanej przez Newtona podczas tworzenia teorii grawitacji zdecydował się opisać bodźce jako wartości w pewnej abstrakcyjnej, wielowymiarowej przestrzeni psychologicznej, co sprawiło, że reguła wykładniczego spadku podobieństwa zostaje zachowana. Na przykład, aby móc zastosować regułę podobieństwa do oceny barw, należy opisywać je nie jako długość fali świetlnej, ale jako punkt na dwuwymiarowym Newtonowskim (1704/1979) kole barw. Podobnie porównywane tony, aby ocena ich podobieństwa przez ludzi spełniła zasadę opisaną przez Sheparda, muszą być wyrażone nie jako wysokości dźwięku, ale jako punkty na trójwymiarowej helisie (Shepard, 1958).

Za główną zaletę modelu Sheparda należy, jak się wydaje, uznać to, że wyraża on problem oceny podobieństwa wprost i przewiduje, że zachowanie będzie próbą optymalnego rozwiązania tego problemu. Pod tym względem model Sheparda różni się od modelu Tversky'ego, który odtwarza zachowanie w sytuacji oceny podobieństwa za pomocą opisu, który pomimo godnej zauważenia trafności predykcji, nie sprawia wrażenia, że udaje mu się ująć, czym *w istocie* jest ocena podobieństwa. Główną wadą modelu Sheparda jest natomiast brak ogólności. Model odnosi się tylko do kategorii, które są powiązane z obserwowaną cechą w ten sposób, że cechę kryterialną posiadają obiekty, których wartość (tj. wartość ich cechy obserwowanej) zawiera się w spójnym i ciągłym przedziale. Ponadto model Sheparda nie jest w stanie uwzględnić tendencyjności w klasyfikowaniu obiektów (na przykład tendencji do uznawania obiektu za przynależnego do kategorii pomimo braku wystarczającego uzasadnienia) lub wiedzy o tym, że poszukiwana kategoria ogranicza się do pewnych przedziałów z większym prawdopodobieństwem niż do innych. Rozwiązania zarówno problemu ogólności struktury dopuszczalnych kategorii, jak i niemożliwości uwzględnienia wiedzy apriorycznej dostarcza Bayesowski model kategoryzacji.

Model Bayesowski

Teoria Tenenbauma i Griffithsa (2001) jest oparta na prostym założeniu, że wnioskowanie o podobieństwie jest szczególnym przypadkiem wnioskowania Bayesowskiego (Bayes, 1763). Aby uczynić przedstawienie modelu Tenenbauma i Griffithsa łatwiejszym, być może wskazane będzie w tym miejscu krótkie przypomnienie na czym polega wnioskowanie Bayesowskie.

Twierdzenie Bayesa odnosi się do wnioskowania o prawdopodobieństwie warunkowym ($p(h|d)$) zdarzenia h (prawdopodobieństwo, że wystąpi zdarzenie h, jeżeli wystąpiło zdarzenie d) na podstawie prawdopodobieństwa warunkowego ($p(d|h)$) zdarzenia d oraz prawdopodobieństw bezwarunkowych zdarzeń p i h. Regułę wnioskowania obrazuje wzór (4):

$$p(h|d) = \frac{p(h|d)p(h)}{p(d)}. \tag{4}$$

Powyższa, na pierwszy rzut oka dość skomplikowana, relacja okazuje się być niesłychanie pożytecznym narzędziem, jeżeli uświadomimy sobie, że h może oznaczać prawdziwość określonej hipotezy, a d – obserwację pewnego zbioru danych. W ten sposób, znając wiarygodność danych (*likelihood*, $p(d|h)$), którą zwykle można łatwo obliczyć, prawdopodobieństwo aprioryczne hipotezy (*prior probability*, $p(h)$) oraz prawdopodobieństwo brzegowe danych (*marginal probability* $p(d)$), możemy określić prawdopodobieństwo, że rozważana hipoteza jest prawdziwa (*posterior probability*, $p(h|d)$). Określenie prawdopodobieństwa hipotezy na podstawie zaobserwowanych danych można bez zbytniej przesady uznać za główny cel empirycznego poznania świata. Chodzi tu nie tylko o poznanie za pomocą narzędzi dostarczanych przez nauki empiryczne, ale przede wszystkim o poznanie będące udziałem organizmów żyjących w świecie. Czym bowiem jest, na przykład, ocena odległości obserwowanego przedmiotu, jeżeli nie oszacowaniem rozkładu prawdopodobieństwa po hipotezach (odległościach) na podstawie danych (dwuwymiarowego obrazu na siatkówce)? Problem wnioskowania o prawdopodobieństwie hipotez na podstawie danych jest, oczywiście, także centralnym zagadnieniem wnioskowania statystycznego w naukach przyrodniczych. Jednakże, chociaż twierdzenie Bayesa dostarcza narzędzia do rozwiązania tego problemu wprost, to wciąż bardzo powszechna jest pośrednia ocena prawdopodobieństwa

hipotez na podstawie wiarygodności danych (osławiony graniczny poziom istotności *p*).

Spośród trzech wyżej wymienionych wartości, wchodzących w skład prawej strony wzoru Bayesa, uzyskanie prawdopodobieństwa brzegowego danych może nastręczyć największych trudności. Sytuacja jest jednak stosunkowo prosta, jeżeli możemy ograniczyć uniwersum rozważanych hipotez do pewnego zbioru. W tej sytuacji prawdopodobieństwo brzegowe danych $p(d)$ jest równe sumie wiarygodności brzegowych danych po wszystkich hipotezach (h') należących do tego zbioru, (H, wzór 5),

$$p(d) = \sum_{h' \in H} p(d|h')p(h'). \tag{5}$$

Podobnie, jeżeli hipoteza jest wyrażona przez wartość parametru (θ) należącą do pewnego ciągłego przedziału, prawdopodobieństwo brzegowe danych jest równe całce po wszystkich możliwych wartościach tego parametru (wzór 6),

$$p(d) = \int_{-\infty}^{\infty} p(d|\theta)p(\theta)d\theta. \tag{6}$$

Rozważmy następujący przykład. Naszym celem jest weryfikacja twierdzenia, że pewna osoba cechuje się zdolnością do prekognicji. Testem niech będzie odgadnięcie wyniku siedmiu kolejnych rzutów monetą. Załóżmy, że rzeczona osoba przechodzi test pomyślnie. Zastosowanie statystycznej analizy częstościowej zobowiązuje nas do odrzucenia hipotezy zerowej (h_0), która mówi, że badana osoba nie posiada zdolności do prekognicji i uznania, że owa zdolność jest w jej posiadaniu, ponieważ prawdopodobieństwo zaobserwowania otrzymanego wyniku jest mniejsze niż 5% (p = 0,0078), jeżeli hipoteza zerowa jest prawdziwa. Rozważmy teraz ten sam problem, stosując wnioskowanie Bayesowskie. Najpierw określmy wartości wchodzące w skład wzoru (4). Wiarygodność danych przy założeniu hipotezy negatywnej (badana osoba nie posiada zdolności do prekognicji, $p(d|h_n)$) jest równa $0{,}5^7 = 0{,}0078$. Wiarygodność danych przy założeniu hipotezy pozytywnej ($p(d|h_p)$) jest równa 1. Przyjmijmy takie aprioryczne prawdopodobieństwo hipotez, które wyraża umiarkowanie silne niedowierzanie w istnienie badanego fenomenu, niech $p(h_n)$ = 0,999, a $p(h_p)$ = 0,001. Stosując teraz wzór Bayesa, otrzymujemy, że:

$$p(h_p|d) = \frac{1 \cdot 0{,}001}{1 \cdot 0{,}001 + 0{,}0078 \cdot 0{,}999} = 0{,}11.$$

Widzimy więc, że aposterioryczne prawdopodobieństwo hipotezy pozytywnej jest znacznie mniejsze niż prawdopodobieństwo hipotezy negatywnej ($p(h_n|d) = 1\text{-}p(h_p|d) = 0{,}89$), co powinno skłonić nas do odrzucenia tej pierwszej.

Powyższy przykład został dobrany tak, aby uwypuklić pewien problem wiążący się z wnioskowaniem Bayesowskim. Dobór prawdopodobieństwa apriorycznego hipotez jest w tym przypadku nieco arbitralny. Wyraża on przekonanie analizującego, że istnienie prekognicji jest mało prawdopodobne. Można przypuszczać, że trudno byłoby uzyskać takie priory w wyniku konsensusu osób reprezentujących wszystkie możliwe stanowiska w tej kwestii. Na szczęście dobór priorów często nie stanowi problemu, ponieważ w wielu przypadkach wnioskowania mamy do czynienia z jedną z trzech następujących możliwości.

1. Wartości priorów dane są obiektywnie. Jeżeli badane zjawisko jest mniej tajemnicze niż przywołane powyżej, na przykład mamy do czynienia z testem wykrywającym pewną chorobę, której częstość występowania w danej populacji jest nam znana, możemy użyć informacji o prawdopodobieństwie apriorycznym hipotezy, zamiast je szacować. Z podobną sytuacją mamy do czynienia, kiedy replikujemy badania, których wyniki są nam znane. Wyniki poprzednich badań stanowią tutaj doskonały prior dla badań kolejnych.

2. Wartości priorów nie wpływają na wyniki analizy. Cechą wnioskowania Bayesowskiego jest to, że im większego zbioru danych używamy do analizy, tym większy wpływ na prawdopodobieństwo aposterioryczne mają dane, i tym mniejszy wpływ mają priory. W praktyce analiz statystycznych wyników badań empirycznych często zdarza się, że analizowane zbory danych są tak duże, że analiza daje te same wyniki niezależnie od wartości priorów, o ile te ostatnie nie przyjmują postaci rozkładów o skrajnych wartościach parametrów.

3. Nie chcemy przekazywać żadnej informacji za pomocą priorów. Może się zdarzyć, że nie mamy żadnych oczekiwań, co do prawdziwości hipotez. W tej sytuacji można użyć priorów nieinformacyjnych, zwanych też płaskimi. Takie priory nie faworyzują żadnej z rozważanych hipotez. W powyższym przykładzie priory płaskie miałyby wartość $p(h_n) = p(h_p) = 0{,}5$. Wnioskowanie Bayesowskie, oparte o priory nieinformacyjne zwykle daje wyniki zbliżone do wnioskowania częstościowego. Należy jednak dodać, że różnice w teoretycznych podstawach obu rodzajów analiz pozwalają interpretować takie wyniki w nieco inny sposób.

Zastosowanie wnioskowania Bayesowskiego do oceny podobieństwa wymaga odpowiedniej interpretacji pojęć „hipoteza" i „dane". W modelu Tenenbauma i Griffithsa, podobnie jak w modelu Sheparda,

pojedyncza obserwacja obiektu *a*, o którym wiemy, że należy do kategorii lub który chcemy dopiero poddać kategoryzacji, jest równoważna z pewną wartością na rozważanym wymiarze. Z kolei hipoteza *h* została w tym modelu zdefiniowana jako zbiór wartości na tym wymiarze. Hipoteza jest prawdziwa, jeżeli $a \in C$ (*a* posiada własność kryterialną) wtedy i tylko wtedy, gdy $a \in h$, gdzie *C* oznacza kategorię. Jako dane traktuje się obserwację, że określony zbiór obiektów należy do kategorii. Dla uproszczenia, w dalszej części tekstu będą rozważane przypadki, w których obiekty opisywane są na jednym wymiarze, ale model równie dobrze może zostać odniesiony do sytuacji, w której zarówno hipoteza, jak i obiekt są zdefiniowane jako, odpowiednio, wektor podzbiorów $\vec{H}$ i wektor wartości $\vec{a}$. Opis za pomocą wyżej zdefiniowanego słownika, przedstawionego wcześniej przykładu, mógłby wyglądać następująco Niech rozważanym wymiarem będzie waga owocu, a kryterium przynależności do kategorii *C*, jak poprzednio – jego jadalność. Hipotezą *h* może być przedział od 200 do 500 gramów, co oznacza, że hipoteza *h* jest prawdziwa wtedy i tylko wtedy, gdy wszystkie owoce jadalne ważą nie mniej niż 200 i nie więcej niż 500 gramów. Możemy też przyjąć, że mamy do dyspozycji dane w postaci dwóch obserwacji jadalnych owoców a_1 = 240 gramów i a_2 = 520 gramów. Nietrudno zauważyć, że podany zbiór danych falsyfikuje hipotezę *h*.

Przedstawiona wyżej definicja danych pozwala, opierając się na twierdzeniu Bayesa, obliczyć prawdopodobieństwo, że dana hipoteza jest prawdziwa, jeżeli został zaobserwowany określony zbiór danych. Prawdziwość pojedynczej hipotezy nie wystarczy jednak, żeby ocenić prawdopodobieństwo, że inny obiekt, co do którego nie znamy wartości cechy kryterialnej, należy do danej kategorii. Żeby obliczyć to prawdopodobieństwo, musimy rozważyć wszystkie możliwe hipotezy. Tak więc twierdzenie Bayesa, dopiero przy uwzględnieniu całego zbioru możliwych hipotez, pozwala na obliczenie prawdopodobieństwa generalizacji, czyli prawdopodobieństwa, że obiekt *b* o określonych cechach posiada własność kryterialną, jeżeli posiada ją obiekt *a*, czyli że rozważany obiekt należy do tej samej kategorii *C*, co pewien dany obiekt. Prawdopodobieństwo generalizacji otrzymuje się, sumując (lub w przypadku ciągłej przestrzeni hipotez – całkując) prawdopodobieństwo wszystkich hipotez obejmujących przynależność danego obiektu do kategorii, jak to przedstawia wzór (7):

$$p(b \in (C|a) = \sum_{h:b \in h} p(h|a). \tag{7}$$

Dwa ostatnie elementy modelu, konieczne do określenia funkcji podobieństwa to: wiarygodności danych ($p(a|h)$) oraz sposób określania prawdopodobieństwa apriorycznego hipotez ($p(h)$). Wiarygodność danych jest w modelu Tenenbauma i Griffithsa określona wzorem (8):

$$p(\mathrm{a}|h) = \begin{cases} \frac{1}{|h|} & \text{jeżeli } a \in h \\ 0 & \text{w innym wypadku.} \end{cases} \tag{8}$$

A więc im większy zakres hipotezy h, tym mniejsza szansa, że obiekt o konkretnej wartości leżącej w tym zakresie zostanie zaobserwowany. Takie założenie wydaje się być intuicyjnie trafne. Prawdopodobieństwo, że losowo wybrany dzień będzie pierwszym dniem tygodnia jest równe 1/7, ponieważ tydzień ma siedem dni, tymczasem prawdopodobieństwo, że losowo wybrany dzień będzie pierwszym dniem miesiąca jest tyle razy mniejsze, o ile razy liczba dni miesiąca jest większa niż liczba dni tygodnia. Należy jednak nadmienić, że nie jest to jedyny sposób obliczania wiarygodności, rozważany przez autorów. Opisany sposób został wybrany przez nich ze względu na własności opartej na nim funkcji podobieństwa.

Podobnie jak istnieje wiele możliwych sposobów obliczania prawdopodobieństwa zaobserwowania danych, istnieje także wiele sposobów określenia apriorycznego prawdopodobieństwa hipotez. Kryterium wyboru powinna stanowić zgodność z modelowanym środowiskiem. Znaczenie tego stwierdzenia może uczynić bardziej zrozumiałym następujący przykład. W prezentowanych w dalszej części tekstu analizach zbiorem wartości, które mogą być przyjmowane przez obiekty, będzie skończony przedział liczb całkowitych. Jeżeli przedział zawiera n liczb, to istnieje $(n^2+n)/2$ hipotez mieszczących się w tym przedziale, również będących przedziałami liczb całkowitych. W przypadku istnienia skończonego zbioru hipotez, narzucającym się sposobem wyrażenia początkowej ignorancji, jest przypisanie wszystkim hipotezom takiego samego prawdopodobieństwa a priori. Należy jednak zauważyć, że w opisywanym przypadku wartości leżące bliżej środka rozważanego przedziału są objęte przez większą liczbę hipotez, niż wartości leżące bliżej krańców przedziału. Jeżeli to stwierdzenie budzi wątpliwości, można rozważyć następujące hipotezy należące do przedziału (1, 3): {1}, {2}, {3}, {1, 2}, {2, 3}, {1, 2, 3}. Każda z liczb 1, 2 i 3 jest objęta przez jedną hipotezę jednoelementową i jedną hipotezę trójelementową, jednak liczba 2 jest objęta przez obie hipotezy dwuelementowe, podczas gdy liczby 1 oraz 3 – tylko przez jedną z nich każda.

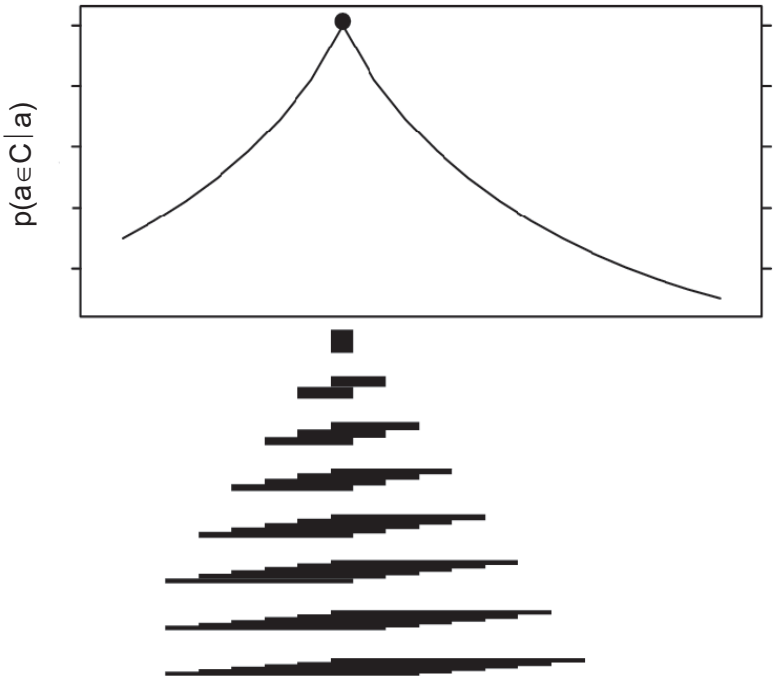
W tej sytuacji funkcja opisująca aprioryczne prawdopodobieństwo obserwacji wartości przyjmie postać odwróconej litery „U". Taka postać funkcji prawdopodobieństwa obserwacji wartości jest niepożądana, o ile nie mamy powodów przypuszczać, że w danym środowisku prawdopodobieństwo (a priori) obserwacji różni się w zależności od wartości. Sposobem na rozwiązanie powyższego problemu jest przyjęcie priorów Jeffreysa (1946), które są dobrane tak, aby prawdopodobieństwo obserwacji każdej wartości było równe (a więc też są one priorami płaskimi, tyle że nie ze względu na hipotezy, ale ze względu na możliwe wyniki). Oznacza to, że aprioryczne prawdopodobieństwo hipotez leżących bliżej krańców rozważanego przedziału musi być większe niż hipotez leżących bliżej jego środka (w powyższym przykładzie $p(h_1) = p(h_3) \approx 0{,}19$, podczas gdy $p(h_2) \approx 0{,}11$).

Określenie apriorycznego prawdopodobieństwa jest tylko jednym z problemów związanych z zagadnieniem doboru hipotez. Innym problemem jest sam wybór zbioru hipotez. Zależnie od tego, jaki zbiór hipotez zostanie rozważony, funkcja podobieństwa może przyjmować różne postaci. Ten problem może jednak zostać rozwiązany na gruncie teorii Bayesowskiej. Struktura zbioru hipotez również może być przedmiotem wnioskowania Bayesowskiego. W takim przypadku mielibyśmy do czynienia niejako z dwoma piętrami wnioskowania. Struktury hipotez byłyby ewaluowane ze względu na wynik wnioskowania o podobieństwie, dla którego owe struktury stanowiłyby dane wejściowe. Autorzy teorii poddali opisywany sposób rozumowania analizie, która odsłania duże możliwości tej metody (Tenenbaum, Griffiths i Kemp, 2006).

Działanie modelu Bayesowskiego zostało zobrazowane na ryc. 2. Krzywa przedstawia funkcję podobieństwa przyporządkowującą obiektowi (definiowanemu jako wartość na danym wymiarze) prawdopodobieństwo przynależności do kategorii. Kółko oznacza zaobserwowany obiekt *a* należący do kategorii *C*. Prostokąty oznaczają hipotezy, przy czym szerokość prostokąta to zakres hipotezy, a jego wysokość – jej aposterioryczne prawdopodobieństwo. Funkcja podobieństwa $p(b \in C|a)$ powstaje przez zsumowanie i znormalizowanie wartości prawdopodobieństwa hipotez dla każdego punktu.

Rozważmy działanie tego modelu na przykładzie podanym wyżej. Niech rozważanym zbiorem wartości będzie przedział [1, 3]. Niech zaobserwowany obiekt $a = 1$ należy do kategorii *C*. Jakie jest prawdopodobieństwo, że obiekt $b = 3$ również należy do *C*? Spośród sześciu wyliczonych wyżej hipotez musimy rozważyć te trzy, które zawierają obiekt *b*: $h_3 = \{3\}$, $h_{2,3} = \{2, 3\}$ oraz $h_{1,2,3} = \{1, 2, 3\}$. Tak więc na mocy wzoru (7) $p(b \in C|a) = p(h_3|a) + p(h_{2,3}|a) + p(h_{1,2,3}|a)$. Zarówno $p(h_3|a)$ jak i $p(h_{2,3}|a)$ na mocy wzoru (8) jest równe 0. Z kolei $p(h_{1,2,3}|a)$, jak

wynika ze wzoru Bayesa (4), jest równe 0,33 (ponieważ $|h_{1,2,3}| = 3$), pomnożone przez 0,17 (prawdopodobieństwo aprioryczne hipotezy $h_{1,2,3}$), podzielne przez 0,33 (suma wiarygodności brzegowej po wszystkich hipotezach), co jest równe 0,17. Tak więc z faktu, że obiekt o wartości 1 posiada interesującą nas cechę, możemy wyciągnąć wniosek, że z prawdopodobieństwem 17% będzie ją posiadał również obiekt o wartości 3.



Ryc. 2. Tworzenie funkcji podobieństwa na podstawie hipotez zgodnych z obserwacją w modelu Bayesowskim

Tak zdefiniowany model kategoryzacji jest bardzo ogólny. Dwie opisane wyżej teorie (teoria Tversky'ego oraz teoria Sheparda) okazują się być jego szczególnymi przypadkami. Nie trzeba wykazywać, że jest to prawdą w odniesieniu do teorii Sheparda, widać bowiem wyraźnie, w jaki sposób jest ona zagnieżdżona w teorii Tenenbauma i Griffithsa. Jeżeli na gruncie teorii Bayesowskiej jako przestrzeń hipotez przyjmiemy kontinuum (o dowolnej liczbie wymiarów) i wszystkim hipotezom przypiszemy równe prawdopodobieństwo a priori (ewentualnie uzależniając prawdopodobieństwo hipotez od ich rozmiaru), a ponadto

ograniczymy zbiór zaobserwowanych obiektów należących do kategorii do jednego, to model, który otrzymamy będzie tożsamy z modelem Sheparda.

Trochę trudniejszy jest dowód (podany za Paulewiczem, 2010) twierdzenia, że model Tversky'ego jest szczególnym przypadkiem modelu Bayesowskiego. Rozważmy alternatywną, stosowaną niekiedy, wersję modelu kontrastu (wzór 9):

$$S(a,b) = 1/\left(1 + \frac{f(A \cap B) + \alpha f(\mathrm{A}\backslash\mathrm{B}) - \beta f(\mathrm{B}\backslash\mathrm{A})}{f(A \cap B)}\right). \tag{9}$$

Przyjmijmy teraz, że hipotezy w modelu Bayesowskim będą odpowiadać posiadaniu przez obiekt określonych cech. Tak więc wiarygodność danych $p(a|h)$ jest równa jeden, kiedy obiekt a posiada cechę h, czyli $h \in Q(a)$, gdzie $Q(a)$ jest zbiorem cech obiektu a, a analogicznie jest równa zero, kiedy obiekt a nie posiada tej cechy ($p(a|h) = 0$ jeżeli $h \notin Q(a)$). Jeżeli przyjmiemy, że aprioryczne prawdopodobieństwo wszystkich hipotez jest równe, to prawdopodobieństwo generalizacji $p(b|a)$ będzie stosunkiem liczby cech wspólnych dla obu obiektów do liczby cech obiektu a. Tak więc:

$$\begin{aligned} p(b|a) &= \frac{|Q(a) \cap Q(b)|}{|Q(a)|} \\ &= \frac{|Q(a) \cap Q(b)|}{|Q(a) \cap Q(b) + Q(a)\backslash Q(b)|} \\ &= \frac{|Q(a) \cap Q(b)|}{|Q(a) \cap Q(b) + Q(a)\backslash Q(b)|} \\ &= 1/\left(1 + \frac{|Q(a)\backslash Q(b)|}{|Q(a) \cap Q(b)|}\right). \end{aligned} \tag{10}$$

co jest tożsame z alternatywnym modelem proporcji (9) i monotonicznie związane z modelem proporcji (2) i modelem kontrastu (1), jeżeli przyjąć, że funkcja f jest addytywna oraz $\alpha = 0$ i $\beta = 1$.

Należy zauważyć, że otrzymaliśmy tutaj model zgodny z modelem Tversky'ego, pomijając teoretyczne problemy, jakie się z nim wiążą. W modelu wyprowadzonym z modelu Bayesowskiego nie pojawiają się kłopotliwe do interpretacji wolne parametry, a postać modelu wynika z ogólnej, wywiedzionej z uzasadnionych założeń teorii, a nie jest tylko formalnym odzwierciedleniem obserwowanych wyników.

Tak ogólny model można zastosować w bardzo wielu przypadkach. Jest oczywiste, że konstruując przestrzeń hipotez, nie musimy ograniczać się do przedziałów liczb całkowitych, ale nawet ciągłe przedziały wartości nie wyznaczają granic możliwości stosowania modelu Bayesowskiego. Pozostając na gruncie liczb, możemy sobie wyobrazić dowolne hipotezy wyrażające kategorie, w jakie liczby są ujmowane przez ludzi, podczas operowania w świecie, na przykład liczby parzyste, liczby „okrągłe", małe liczby itd. Można jednak pójść dalej i opisywać kategoryzowane obiekty na dowolnych skalach, na przykład koloru, posiadania czterech nóg lub „puchatości".

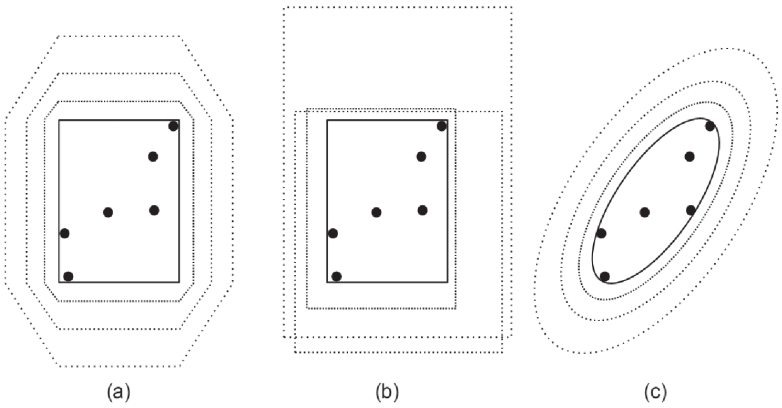
Do tej pory omawiana była elementarna wersja modelu Bayesowskiego, to znaczy funkcja podobieństwa była tworzona w jednowymiarowej dziedzinie dyskretnych wartości, na podstawie jednego elementu wzorcowego. Nietrudno jednak wyobrazić sobie znacznie ogólniejszą postać tego modelu. Po pierwsze więc, zastąpienie dyskretnej dziedziny funkcji dziedziną ciągłą wymaga jedynie dopuszczenia zbioru hipotez, których zarówno początki, jak i długości (jeżeli mówimy o hipotezach będących przedziałami) zawierają się w kontinuum. Zastosowanie modelu Bayesowskiego do takiego przypadku nie nastręcza żadnych problemów koncepcyjnych, należy tylko do obliczenia prawdopodobieństwa brzegowego danych użyć wzoru dla ciągłych wartości hipotez (wzór 6 zamiast wzoru 5). Dalej, stworzenie funkcji podobieństwa na podstawie więcej niż jednej obserwacji również jest możliwe. Autorzy proponują w tym celu zdefiniować wiarygodność wektora danych A jako iloczynu wiarygodności każdego z punktów danych a_i (wzór 11),

$$p(A|h) = \prod_{i=1}^{|A|}(p(a_i|h) \tag{11}$$

Nieco bardziej kłopotliwe jest rozszerzenie teorii Bayesowskiej na większą liczbę wymiarów. Autorzy teorii proponują dwa rozwiązania tego problemu. (a) Określenie funkcji podobieństwa niezależnie na każdym z wymiarów, a następnie scalenie predykcji (Tenenbaum i Griffiths, 2001, zob. też ryc. 3a). (b) Rozważanie jako hipotez wycinków przestrzeni obejmującej wszystkie rozważane wymiary (Tenenbaum, 1999, zob. ryc. 3b). Obydwa powyższe rozwiązania doskonale sprawdzają się dla przypadku problemu uczenia się prostokątów (*rectangle learning task*), to znaczy – problemu nabywania kategorii zdefiniowanej przez niezależne, ciągłe przedziały obserwowalnych wartości na dwóch wymiarach. Używając wcześniej stosowanego przykładu: o ka-

tegorii zdefiniowanej przez prostokąt możemy mówić w odniesieniu do sytuacji, kiedy owoce jadalne to te, których waga leży w przedziale od 200 do 450 gramów oraz czas ich pojawiania się zawiera się między kwietniem a czerwcem.

Nie można jednak wykluczyć, że pewne kategorie będą określone przez takie wartości na obserwowalnych wymiarach, które są od siebie zależne. Na przykład cechę kryterialną mogą posiadać obiekty, które są opisane za pomocą wysokiej lub niskiej wartości na obu rozważanych wymiarach, ale nie niskiej – na jednym wymiarze i wysokiej – na drugim. Zależność między wymiarami nie jest możliwa do odzwierciedlenia w modelu operującym rozłącznie na każdym z nich. Zastosowanie takiego modelu do kategorii określonej przez skorelowane wartości doprowadzi do stworzenia nieadekwatnej funkcji podobieństwa (por. rycina 3a i 3c). Opisanej sytuacji nie da się także wyrazić w modelu operującym na hipotezach „prostokątnych", jednak możliwe jest takie zastosowanie modelu, używającego hipotez wielowymiarowych, aby operował na innych niż prostokątne kształtach hipotez, w tym takich, które obejmowałyby relacje zależności pomiędzy wymiarami. Niestety tak elastycznie zdefiniowanych hipotez jest znacznie więcej niż hipotez prostokątnych, a że liczba hipotez rośnie wykładniczo wraz ze wzrostem liczby wymiarów, opisana metoda wykrywania zależności pomiędzy wartościami na różnych wymiarach nie jest możliwa do zastosowania w praktyce.



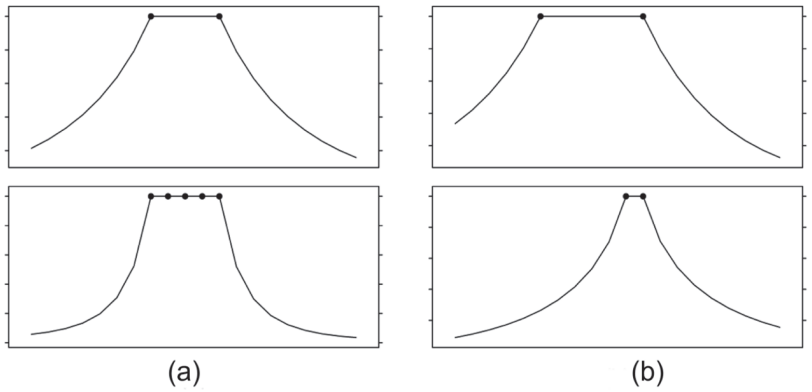
Ryc. 3. Funkcja podobieństwa dla dwóch wymiarów, określona niezależnie na każdym z wymiarów (a) i określona w przestrzeni hipotez dwuwymiarowych (b), oraz pożądany kształt funkcji podobieństwa uwzględniający zależność wymiarów (c)

Model Bayesowski przewiduje wiele jakościowych własności procesu oceny podobieństwa. Kilka z nich zasługuje na uwagę. Wpływ liczby obserwacji na kształt funkcji podobieństwa jest jedną z nich. Zarówno na intuicji, dostępnej prawdopodobnie każdemu, kto miał okazję eksplorować środowisko, jak i na analogii z wnioskowaniem statystycznym, tak Bayesowskim, jak częstościowym, można oprzeć sąd, że im więcej obserwacji danego zjawiska zebraliśmy, tym mniejsza jest nasza niepewność co do oszacowania właściwości populacji, z której pochodziła próba. Bayesowski model kategoryzacji odtwarza ten efekt. Tak więc im więcej zaobserwowanych obiektów, należących do kategorii jest podstawą wnioskowania, tym węższy przedział wartości jest objęty przez określoną część rozkładu wyznaczonego przez funkcję podobieństwa.

Aby wyjaśnić w jaki sposób powyższy efekt jest uzyskany przez model, należy przypomnieć, że wiarygodność zbioru danych A jest równa iloczynowi prawdopodobieństw dla każdego z obiektów w zbiorze (wzór 11). Zatem im więcej obserwacji, tym więcej czynników wchodzi w skład wyżej przedstawionego wyrażenia. Ponieważ prawdopodobieństwa prawdziwości hipotez są zawsze nie większe niż jeden, prawdopodobieństwo każdej z nich maleje (lub pozostaje stałe) wraz ze wzrostem liczby obserwacji. Chociaż prawdopodobieństwo węższych hipotez maleje szybciej niż szerszych, to jednak zawsze pozostaje większe niż dla hipotez szerszych (zob. wzór 8). Dalej, z faktu, że suma prawdopodobieństw marginalnych maleje w miarę wzrostu liczby obserwacji szybciej niż prawdopodobieństwo którejkolwiek z hipotez, wynika, że wraz ze wzrostem liczby obserwacji, stosunek prawdopodobieństwa generalizacji dla wartości objętych węższymi hipotezami, do prawdopodobieństwa generalizacji dla wartości objętych szerszymi hipotezami, rośnie. Efekt ten jest zobrazowany na rycinie 4a. Funkcja podobieństwa określona na podstawie pięciu obserwacji maleje szybciej w miarę oddalania się od zbioru obserwacji (dolny panel), niż funkcja określona na podstawie dwóch obserwacji (górny panel). Oznacza to, że jeżeli model Bayesowski jest trafny, to im więcej danych o obiektach należących do pewnej kategorii zostało zebranych z danego przedziału, tym mniej skłonni jesteśmy dopuszczać, że kategoria rozciąga się poza ten przedział.

Innym interesującym efektem, który model Bayesowski jest w stanie odtworzyć, jest efekt wpływu wariancji wartości obserwacji na pewność oszacowania. Wyniki badań pokazują, że im mniejsza zmienność danych, tym większa pewność osób badanych dotycząca wniosków opartych na tych danych (Fried i Holyoak, 1984; Osherson, Smith, Wilkie, Lopez i Shafir, 1990; Rips, 1989). Na gruncie podanego we

wstępie przykładu można ten efekt opisać następująco: jeżeli wszystkie zjedzone przez organizm owoce okazały się jadalne, to w przypadku kiedy wszystkie te owoce były koloru jasnoczerwonego, organizm powinien być mniej skłonny za jadalne uznać owoce o innych kolorach, niż wtedy gdy jadalnymi okazały się być owoce o szerszym spektrum kolorów (na przykład od żółtego po ciemnoczerwony).



Ryc. 4. Efekt wpływu liczby obserwacji (a) oraz wariancji obserwacji (b) na funkcję podobieństwa w modelu Bayesowskim

Wyjaśnienie tego efektu na gruncie teorii Bayesowskiej jest stosunkowo proste. Im mniej zróżnicowane dane, tym węższa jest hipoteza, która obejmuje wszystkie obserwacje. Ze wzoru (8) wynika, że im węższa hipoteza, tym większa wiarygodność danych, a więc, na mocy twierdzenia Bayesa, większe prawdopodobieństwo hipotezy i w konsekwencji – większa (relatywnie) wartość funkcji podobieństwa w danym punkcie.

Podsumowując tę krótką prezentację modelu Bayesowskiego, należy powiedzieć, że ten model kategoryzacji i oceny podobieństwa ma dwie ważne zalety. Pierwszą jest niesłychana ogólność, która pozwala wyrazić w tym modelu niemal każdy problem kategoryzacyjny oraz dowolny zbiór hipotez, a także uwzględnić aprioryczną wiedzę podmiotu wnioskującego dotyczącą zarówno struktury problemu, jak i oczekiwanych wyników wnioskowania. Skutkiem tej ogólności jest możliwość wywiedzenia z modelu Bayesowskiego dwóch ważnych modeli kategoryzacyjnych: modelu Tversky'ego i modelu Sheparda.

Drugą zaletą modelu Bayesowskiego jest to, że podobnie jak model Sheparda, wyraża on problem oceny podobieństwa wprost. Model Tenenbauma i Griffithsa jest optymalnym rozwiązaniem dobrze okre-

ślonego problemu. Działanie modelu nie jest efektem dopasowywania jego konstrukcji *ad hoc*, w celu odtworzenia obserwowanych wyników, ale racjonalnej analizy problemu, którego rozwiązanie jest celem modelowanego zachowania.

Literatura cytowana

Angus, I. (2001). An introduction to Erlang B and Erlang C. *Telemanagement*, *187*, 6–8.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, *53*, 370–418.

Blough, D.S. (1961). The shape of some wavelength generalization gradients. *Journal of the Experimental Analysis of Behavior*, *4*, 31–40.

Fried, L.S. i Holyoak, K.J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 234–257.

Gershman, S.J., Cohen, J.D. i Niv, Y. (2010). Learning to selectively attend. W: S. Ohlsson i R. Catrambone (red.), *Proceedings of the 32nd Annual Conference of Cognitive Science Society*, s. 1270–1275. Austin, TX: Cognitive Science Society.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *186* (1007), 453–461.

Jones, M. i nas, F.C. (2010). Integrating reinforcement learning with models of representation learning, w: S. Ohlsson i R. Catrambone (red.), *Proceedings of the 32nd Annual Conference of Cognitive Science Society*, s. 1258–1263. Austin, TX: Cognitive Science Society.

Maehigashi, A. i Miwa, K. (2010). Estimation of trade-off between costs of preprocessing and primary processing, w: S. Ohlsson i R. Catrambone (Red.), *Proceedings of the 32nd Annual Conference of Cognitive Science Society*, s. 943–948. Austin, TX: Cognitive Science Society.

McGuire, W.J. (1961). A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, *62* (4), 335–347.

Murphy, G.L. i Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92* (3), 289–316.

Newton, I. (1979). *Opticks*. New York: Dover Publications. (Oryginalna praca opublikowana w 1704).

Osherson, D.N., Smith, E.E., Wilkie, O., Lopez, A. i Shafir, E. (1990). Caegory based induction. *Psychological Review*, *97* (2), 185–200.

Paulewicz, B. (2010). *Interakcja i adaptacja – propozycja metateoretyczna w badaniach nad zachowaniem*. Niepublikowana praca doktorska, Kraków: Uniwersytet Jagielloński.

Rips, L.J. (1989). Similarity, typicality, and catrgorization. W: S. Vosniadou i A. Orton (Red.), *Similarity and analogical reasoning* (s. 21–59). New York: Cambridge University Press.

Rosch, E.H. (1978). Principles of categorization, w: E.H. Rosch i B.B. Lloyd (Red.), *Cognition and categorization* (s. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *8*, 382–439.

Rothkopf, Z.E. (1985). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, *53*, 94–101.

Shepard, R.N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65* (4), 242–256.

Shepard, R.N., Cermak, G.W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, *4* (3), 351–377.

Sheppard, R.N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Tenenbaum, J.B. (1999). Bayesian modeling of human concept learning. *Advances in Neural Information Processing Systems*, *11*, 59–65.

Tenenbaum, J.B., Griffiths, T.L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, *24*, 629–640.

Tenenbaum, J.B., Griffiths, T.L., Kemp, C. (2006). Theory-based bayesian models of inductive learning and reasoning. *TRENDS in Cognitive Sciences*, *10* (7), 309–318.

Torgerson, W.S. (1965). Multidimensional scaling of similarity. *Psychometrika*, *30*, 273–286.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Wish, M. (1967). A model for the prediction of morse code-like signals. *Human Factors*, *9*, 529–540.