POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

ROZMYTOŚĆ I BIPOLARNOŚĆ W INTELIGENTNYM WYSZUKIWANIU INFORMACJI

Sławomir Zadrożny

Warszawa 2013

POLSKA AKADEMIA NAUK
INSTYTUT BADAŃ SYSTEMOWYCH

Seria: BADANIA SYSTEMOWE
Tom 73

Redaktor naukowy:
Prof. dr hab. inż. Jakub Gutenbaum

Warszawa 2013

iBS PAN

POLSKA AKADEMIA NAUK
INSTYTUT BADAŃ SYSTEMOWYCH

Sławomir Zadrożny

ROZMYTOŚĆ I BIPOLARNOŚĆ W INTELIGENTNYM WYSZUKIWANIU INFORMACJI

Warszawa 2013

Autorzy:

Dr hab. Sławomir Zadrożny
Instytut Badań Systemowych Polskiej Akademii Nauk
ul. Newelska 6, 01-447 Warszawa
Slawomir.Zadrozny@ibspan.waw.pl

Recenzenci:

dr hab. inż. Maciej Krawczak
dr Marek Reformat

Skład: Aneta M. Pielak

Wydawca:

Instytut Badań Systemowych
Polskiej Akademii Nauk
Newelska 6, 01-447 Warszawa
www.ibspan.waw.pl

ISSN 0208-8029
ISBN 83-894-7551-0

Rozdział 5

Bipolarność informacji

Termin *bipolarność* będziemy rozumieć tu pragmatycznie jako występowanie informacji *pozytywnej* i *negatywnej*, która w różny sposób wpływa na nasze preferencje i przekonania. Ważną rolę w próbach formalizacji tego pojęcia odgrywa logika rozmyta. Intensywne badania w tym zakresie zainicjowali Dubois i Prade wraz ze współpracownikami. Dotyczyły one przede wszystkim szeroko rozumianej reprezentacji wiedzy i modelowania preferencji [10, 11, 12, 81, 95, 13, 97, 96, 98, 99, 100]. Inne prace Dubois i Prade'a ze Smetsem [101], Hájkiem [84] czy Ughetto [212], choć dotyczą bardziej ogólnych zagadnień, też stanowią bardzo ważne podstawy teoretyczne do analizy pojęcia bipolarności.

W obszarze wyszukiwania informacji bardzo ważną rolę odegrała praca Lacroix i Lavency [147]. Zaproponowano w niej dwie kategorie warunków w zapytaniu do bazy danych i choć podejście to nie odwołuje się wprost do pojęcia bipolarności, to można je łatwo zreinterpretować, przy czym pozytywna i negatywna informacja jest tu rozumiana w specjalny sposób. Idea ta została podjęta i przeniesiona na grunt logiki rozmytej we wczesnych pracach Bosca i Piverta [37, 39] i rozwinięta w ich i ich współpracowników późniejszych pracach [45, 151, 152, 153], a swoje pierwsze kompletne przedstawienie znalazła w pracy Dubois i Prade'a [94]. Zadrożny i Kacprzyk [245, 250, 251, 247, 253, 248, 135, 134, 252, 254, 104] w swoich pracach starają się najwierniej przenieść ideę podejścia Lacroix i Lavency na grunt logiki rozmytej. De Tré ze współpracownikami rozwija podejście do reprezentacji ogólniej rozumianej bipolarności w zapytaniach [162, 78, 161].

Z punktu widzenia baz danych bipolarność może być analizowana w dwóch aspektach. Po pierwsze, reprezentacja informacji bipolarnej w bazie danych wymaga opracowania sposobów jej modelowania i przetwarzania. Podejścia proponowane w literaturze w tym względzie dotyczą raczej ogólniejszego zagadnienia reprezentacji wiedzy. Jednak wiele propozycji z tego zakresu może znaleźć bezpośrednie zastosowanie w systemach zarządzania bazą danych. Jednocześnie wiele zagadnień szczegółowych dotyczy tu również drugiego aspektu, modelowania bipolarnych preferencji użytkownika w zapytaniach do bazy danych. Ten aspekt rozważany jest przez znacznie większą liczbę publikacji, które dotyczą już stricte zastosowań związanych z wyszukiwaniem informacji w bazach danych. Pierwszy punkt niniejszego rozdziału poświęcamy reprezentacji informacji bipolarnej w bazach danych, zaś pozostałe punkty dotyczą zagadnień związanych z pojęciem zapytania bipolarnego.

5.1 Bipolarność w reprezentacji danych

Bipolarność w rozważanym tu kontekście odnosi się do istnienia informacji *pozytywnej* i *negatywnej* dotyczącej danych reprezentowanych w bazie danych [10, 81, 96, 98, 99]. Interesuje nas szczególnie przypadek kiedy wartość atrybutu dla danej krotki nie jest precyzyjnie znana, ale pewne informacje co do niej są dostępne w formie zarówno *pozytywnych* jak i *negatywnych* stwierdzeń [98]. Pozytywna informacja może określać jakie wartości atrybutu dla danej krotki są *możliwe*, *dopuszczalne*, *satysfakcjonujące*, *pożądane* lub *akceptowalne*. Z drugiej strony informacja negatywna może określać jakie wartości są *niemożliwe*, *odrzucane* lub *zabronione*. Dubois i Prade [97, 96, 98, 99] wyróżniają różne typy tak rozumianej bipolarności informacji. Typ I to bipolarność *symetryczna na jednolitej skali bipolarnej* (ang. *symmetric univariate bipolarity*). Informacja pozytywna i negatywna traktowane są jako swoje dokładne dopełnienie. Egzemplifikację stanowi tu teoria prawdopodobieństwa. Na przykład jeśli prawdopodobieństwo, że dany obraz został namalowany w XVI wieku określono na 0.6 (informacja pozytywna), to prawdopodobieństwo, że nie został on namalowany w XVI wieku jest jednoznacznie określone i wynosi 0.4 (informacja negatywna). Taka prosta forma bipolarności informacji jest poprawnie reprezentowana i przetwarzana w tradycyjnych systemach informacyjnych. Jest ona wyrażana na *jednolitej skali bipolarnej* (ang. *bipolar univariate scale*). Zazwyczaj używany jest w tym celu przedział [0,1] lub [-1,1].

Typ II to bipolarność *symetryczna na podwójnej skali unipolarnej* (ang. *symmetric bivariate bipolarity*). Informacja pozytywna i negatywna traktowane są jako pojęcia dualne, wyrażane na dwóch oddzielnych skalach, ale oparte na tych samych danych (ang. *evidence*). Obydwie wartości są ze sobą powiązane pewną relacją. Egzemplifikację stanowi tu teoria intuicjonistycznych zbiorów rozmytych w sensie Atanassova (def. 2.16), gdzie stopnie przynależności i nieprzynależności powiązane są relacją (2.23). Na przykład dane mogą wskazywać w stopniu 0.6, że Rubens jest autorem danego obrazu (informacja pozytywna) i jednocześnie w stopniu 0.2, że nie jest on autorem tego obrazu (informacja negatywna). Ta forma bipolarnosci jest wyrażana na *podwójnej skali unipolarnej* (ang. *unipolar bivariate scale*), na którą składają się dwie skale unipolarne. Zazwyczaj używane są w tym celu dwa przedziały [0,1].

Typ III to bipolarność *asymetryczna/heterogeniczna* (ang. *asymmetric/heterogeneous bipolarity*). Informacja pozytywna i negatywna oparte są na odrębnych danych (ang. *separate bodies of evidence*), które są niezależne i mają różny charakter. Przyjmuje się, że może być narzucony pewien warunek gwarantujący niesprzeczność informacji obu typów, ale poza tym stwierdzenia pozytywne i negatywne są od siebie niezależne. Ta forma bipolarności jest wyrażana na takiej samej skali jak bipolarność typu II.

Bipolarność typu III jest szczególnie interesująca z punktu widzenia reprezentacji danych w bazie danych. Dubois i Prade [98] kładą szczególny nacisk na różną naturę obydwu rodzajów informacji. Przyjmuje się, że źródłem informacji negatywnej jest *wiedza* dotycząca reprezentowanych danych, która narzuca pewne ogólne ograniczenia na dopuszczalne wartości atrybutu. Z drugiej strony informacja pozytywna odnosi się do posiadanych *danych* dotyczących pokrewnych przypadków. Inaczej mówiąc informacja pozytywna wywodzi się z zaobserwowanych przypadków, które uzasadniają przyjęcie danego elementu jako wartości atrybutu. Ilustruje to przykład 5.1

Przykład 5.1 ([98]). *Rozważmy bazę danych o muzeach, które opisane są między innymi atrybutami* `godziny_wstępu` *i* `cena_biletu`. *Załóżmy, że w przypadku jednego z muzeów nie znamy wartości tych atrybutów. Możemy jednak posiadać informację, że w godzinach 14:00–16:00 jest ono na pewno otwarte (informacja pozytywna) i że z pewnością jest ono zamknięte w nocy (od 21:00 do 9:00) (informacja negatywna, wynikająca z ogólnych regulacji dotyczących godzin otwarcia muzeów). Warto zauważyć, że posiadana informacja nie wyklucza, że muzeum jest otwarte również w godzinach rannych, choć brak potwierdzenia tego faktu (brak informacji pozytywnej w postaci obserwacji otwarcia muzeum w tych godzinach).*

Podobnie możemy wiedzieć, że cena biletu jest nie mniejsza niż 5 PLN i jest nie większa niż 30 PLN, znów zgodnie z ogólnymi regulacjami.

Jednocześnie cena pomiędzy 10 i 15 PLN jest najbardziej wiarygodna, gdyż takie ceny obowiązują w znanych nam podobnych muzeach.

W nawiązaniu do modelu posybilistycznego (por. p. 3.2.2) Dubois i Prade (por. np. [98]) proponują reprezentować bipolarność typu III, dotyczącą wartości atrybutu A dla krotki t, z użyciem dwóch oddzielnych rozkładów możliwości $\delta_{A(t)}$ i $\pi_{A(t)}$, zdefiniowanych na dziedzinie dom_A atrybutu A. Rozkład możliwości $\pi_{A(t)}$, jak w oryginalnym modelu posybilistycznym, określa dla każdego elementu $x \in dom_A$ dziedziny atrybutu A stopień możliwości, że jest on wartością atrybutu A dla krotki t przy uwzględnieniu posiadanej wiedzy na ten temat. Stopień możliwości wyrażony jest na negatywnej skali unipolarnej opartej na przedziale [0,1]. Skrajne wartości rozkładu π_A mają następujące znaczenie:

- wartość $\pi_{A(t)}(x) = 1$ oznacza, że jest *potencjalnie* całkowicie możliwe, że x jest wartością atrybutu A dla krotki t; wartość 1 jest elementem *neutralnym* na tej skali unipolarnej;

- wartość $\pi_{A(t)}(x) = 0$ oznacza, że jest całkowicie *niemożliwe*, że x jest wartością atrybutu A dla krotki t; wartość 0 jest ekstremalną wartością *negatywną* na tej skali unipolarnej.

Rozkład możliwości $\delta_{A(t)}$ określa stopień w jakim posiadane dane (obserwacje) potwierdzają, że element $x \in dom_A$ jest dobrym kandydatem na wartość atrybutu A dla krotki t. W tym przypadku:

- $\delta_{A(t)}(x) = 0$ oznacza brak takiego potwierdzenia, ale jest traktowane jako element neutralny na pozytywnej skali unipolarnej;

- $\delta_{A(t)}(x) = 1$ oznacza pełne potwierdzenie; wartość 1 jest ekstremalną wartością *pozytywną* na tej skali unipolarnej.

Przykład 5.2 jeszcze raz uzasadnia praktyczne znaczenie bipolarności typu III.

Przykład 5.2. *Rozważmy ponownie bazę danych reprezentującą kolekcję obrazów (por. p. 3.2.2). Może się zdarzyć, że dokładne datowanie obrazu nie jest znane, ale wiadomo, że jego twórcą był Rubens i stąd obraz musiał powstać za jego życia w latach 1577-1640*[1] *(informacja negatywna, wykluczająca lata przed urodzeniem malarza i po jego śmierci). Dodatkowo*

[1] Z pewnością można ten przedział czasowy ograniczyć wyłącznie do lat twórczości artysty, ale dla uproszczenia przyjmiemy tu cały okres jego życia.

może być dostępna informacja pozytywna potwierdzająca jako datę powstania obrazu krótszy (rozmyty) przedział czasowy, np. "lata mniej więcej 1628-1629", w którym powstały obrazy Rubensa malowane podobną techniką i zbliżone tematycznie. Taką informację można więc reprezentować w postaci dwóch rozkładów możliwości, na przykład następującej postaci:

$$\pi_{\boldsymbol{data}(t)}(x) = \begin{cases} 0 & \text{dla } x \leq 1577 \\ 1 & \text{dla } 1577 < x \leq 1640 \\ 0 & \text{dla } x > 1640 \end{cases}$$

$$\delta_{\boldsymbol{data}}(x) = \begin{cases} 0 & \text{dla } x \leq 1625 \\ \dfrac{x - 1625}{3} & \text{dla } 1625 < x < 1628 \\ 1 & \text{dla } 1628 \leq x \leq 1629 \\ \dfrac{1632 - x}{3} & \text{dla } 1629 < x \leq 1632 \\ 0 & \text{dla } x > 1632 \end{cases}$$

Rozkłady możliwości $\delta_{A(t)}$ i $\pi_{A(t)}$ reprezentują, odpowiednio, zbiór *potwierdzonych/faktycznie możliwych* (ang. *guaranteed/actually possible*) wartości atrybutu A dla krotki t i zbiór *potencjalnie możliwych* (ang. *potentially possible*) wartości atrybutu A dla krotki t.

Nakłada się następujący warunek niesprzeczności obydwu rozkładów:

$$\delta_{A(t)}(x) \leq \pi_{A(t)}(x), \qquad \forall\, x \in dom_A \tag{5.1}$$

który wyraża fakt, że aby pewna wartość mogła być zaobserwowana ($\delta_{A(t)}$), to musi być najpierw dopuszczalna ($\pi_{A(t)}$). W (5.1) zakłada się, że wartości obydwu rozkładów są porównywalne (ang. *commensurable*). Można osłabić to założenie i zastąpić (5.1) następującym warunkiem (por. warunek (2.22) charakteryzujący podwójne zbiory rozmyte):

$$\forall x \in dom_A \quad \delta_{A(t)}(x) > 0 \Rightarrow \pi_{A(t)}(x) = 1$$

Stopień spełnienia zapytania "X jest F" względem posybilistycznej bazy danych oblicza się w formie pary miar $(\Pi_{A(t)}(F), \mathrm{N}_{A(t)}(F))$; por. (3.20)-(3.21). W przypadku wyżej omówionej reprezentacji informacji bipolarnej typu III używa się dodatkowo dwóch miar, które dla nierozmytego zbioru F wyrażają się następującymi wzorami:

$$\Delta_{A(t)}(F) = \inf_{x \in F} \delta_{A(t)}(x) \tag{5.2}$$

$$\nabla_{A(t)}(F) = 1 - \Delta_{A(t)}(\overline{F}) \tag{5.3}$$

Miary te noszą nazwy, odpowiednio, *gwarantowanej możliwości* (ang. *guaranteed possibility*) i *potencjalnej konieczności* (ang. *potential necessity*). Wzór (5.2) można zaadoptować na przypadek rozmytych argumentów F w następujący sposób:

$$\Delta_{A(t)}(F) = \inf_{x \in dom_A} \max(\delta_{A(t)}(x), 1 - \mu_F(x)) \tag{5.4}$$

Wzór (5.3) stosuje się do rozmytych argumentów odpowiednio.

5.2 Zapytania bipolarne

Formułowanie preferencji w szeroko rozumianych zagadnieniach podejmowania decyzji jest procesem złożonym. Odnosi się to również do formułowania zapytań do bazy danych. Zapytanie może być rozumiane jako opis cech poszukiwanego obiektu, w szerokim tego słowa znaczeniu[2]. Naturalnym jest wtedy rozpatrywanie zarówno *pozytywnych* cech takiego obiektu jak i, niezależnie, cech *negatywnych*. Oczekujemy, że system wyszuka dla nas obiekty, które posiadają te pierwsze cechy i nie posiadają tych drugich. Najprostszym sposobem wyrażenia tak zarysowanych preferencji wydawałoby się sformułowanie warunku zapytania w postaci koniunkcji elementarnych warunków specyfikujących cechy pozytywne i zanegowanych warunków specyfikujących cechy negatywne. Jednak w praktyce zależności pomiędzy tymi warunkami mogą być znacznie subtelniejsze. W szczególności, nieposiadanie (wszystkich) cech pozytywnych nie musi dyskwalifikować obiektu i jednocześnie nieposiadanie żadnej z cech negatywnych nie oznacza jeszcze, że obiekt spełnia oczekiwania użytkownika. Tak więc warunek zapytania powinien składać się z dwóch części: spełnienie przez krotkę pierwszej z nich, "pozytywnej", skłania system do zaliczenia tej krotki do odpowiedzi na zapytanie, zaś spełnienie drugiej z nich, "negatywnej", skłania system do pominięcia tej krotki przy formowaniu odpowiedzi na zapytanie. Takie zapytanie będziemy nazywać *zapytaniem bipolarnym*.

Podsumowując, bipolarność jest naturalną cechą preferencji użytkownika i powinna być uwzględniona przy wyszukiwaniu informacji, w szczególności w bazach danych. W niniejszej książce utożsamiamy zapytanie z określeniem warunków jakie powinny spełniać interesujące użytkownika dane. Przyjmujemy, że użytkownik postrzega zarówno *pozytywne* cechy

[2]Rozważamy tu oczywiście nietrywialne zapytania do bazy danych, które odnoszą się do wielu atrybutów relacji i nie są elementem rutynowego procesu przetwarzania danych.

poszukiwanych danych (np. niska cena domu), jak i ich *negatywne* cechy (np. położenie w nielubianej dzielnicy miasta). Cechy te są wyspecyfikowane w postaci odrębnie określonych warunków zapytania. Zależności pomiędzy obydwiema częściami zapytania bipolarnego, ich interpretacja i wpływ na ostateczną decyzję oraz ich konstrukcja mogą być analizowane z różnych punktów widzenia.

Pojęcie bipolarności preferencji z jednej strony oddaje ich emocjonalny charakter. Kwestia roli emocji w podejmowaniu decyzji i ich modelowania wchodzi w zakres szybko się ostatnio rozwijającego obszaru badań noszącego nazwę *obliczeń afektywnych* [179]. Z drugiej strony adekwatne uwzględnienie bipolarności preferencji w kontekście wyszukiwania informacji wymaga odpowiednich podstaw teoretycznych. W dalszej części niniejszego rozdziału omawiamy i systematyzujemy różne rozwiązania zaproponowane w tym względzie w literaturze.

5.2.1 Model ogólny

Ogólną postać zapytania bipolarnego można zdefiniować następująco[3].

Definicja 5.1. Zapytanie bipolarne to zapytanie, którego warunek ma dwie składowe:

$$(R, P) \tag{5.5}$$

przy czym warunek R określa *negatywne* cechy wyszukiwanych krotek (reprezentowanych przez nie obiektów świata rzeczywistego), zaś warunek P opisuje ich cechy *pozytywne*. Spełnienie przez krotkę warunku R stanowi przesłankę do jej odrzucenia, zaś spełnienie przez nią warunku P stanowi przesłankę do jej akceptacji.

W najogólniejszej interpretacji zapytań bipolarnych nie zakłada się nic ponad to co zostało określone w def. 5.1. Warto zwrócić uwagę, że def. 5.1 wprowadza trzy kategorie krotek: pozytywne (spełniające warunek P), negatywne (spełniające warunek R) i neutralne (niespełniające żadnego z tych warunków). Faktyczne zależności pomiędzy tymi kategoriami krotek są bardziej złożone – omawiamy je w dalszej części rozdziału. W każdym razie zakładamy, że użytkownik może krotkę akceptować (“lubić”), odrzucać (“nie lubić”) lub być neutralnym wobec niej.

[3] Definicję tę wyrażamy w terminach krotek. W dalszej części rozdziału analizujemy również bipolarność preferencji na poziomie dziedzin atrybutów, ale nie będzie naruszać to ogólności podanej tu definicji.

Wykonanie tak rozumianego zapytania bipolarnego sprowadza się w najprostszym przypadku do obliczenia dla każdej krotki stopnia spełnienia przez nią obydwu warunków: pozytywnego i negatywnego. Jedyne co pozostaje do ustalenia to porządek w jakim krotki będą przedstawione użytkownikowi w odpowiedzi na zapytanie. Określając porządek należy uwzględnić specyfikę tych dwóch stopni spełnienia: fakt, że odnoszą się one do pozytywnego i negatywnego warunku. Można przyjąć, że mamy tu do czynienia z sytuacją zbliżoną do rozważanej w zakresie *podejmowania decyzji w warunkach ryzyka*. Decydent *niechętny podejmowaniu ryzyka* (ang. *risk-averse*) unika akcji prowadzących z pewnym niezerowym prawdopodobieństwem do straty, natomiast decydent *skłonny do podejmowania ryzyka* (ang. *risk-prone*) może ignorować ryzyko nawet poważnych strat o ile tylko są duże szanse na zysk. Podobnie, w przypadku zapytań bipolarnych można oczekiwać, że niektórzy użytkownicy będą większą wagę przykładali do negatywnego warunku i nie będą zainteresowani krotkami spełniającymi ten warunek nawet w niewielkim stopniu. Inni użytkownicy mogą większą wagę przywiązywać do spełnienia warunku pozytywnego w wysokim stopniu i mogą być gotowi do zaakceptowania faktu, że krotka spełnia również w pewnym stopniu warunek negatywny. Tak więc zapytania bipolarne, rozumiane w tak ogólnym sensie, powinny być interpretowane przez system w sposób silnie uzależniony od nastawienia użytkownika. W przypadku ekstremalnych nastawień typu "risk-averse" i "risk-prone" uszeregowanie krotek w odpowiedzi na zapytanie będzie miało porządek leksykograficzny. Na przykład, dla nastawienia typu "risk-averse" będzie to uporządkowanie przede wszystkim niemalejąco według stopnia spełnienia warunku negatywnego, a w drugiej kolejności nierosnąco według stopnia spełnienia warunku pozytywnego. Dla pośrednich typów nastawienia użytkownika zastosowanie mogą znaleźć różne operatory agregacji pozwalające określić skalarną ocenę łącznego stopnia spełnienia przez krotkę warunków zapytania bipolarnego.

Najprostszą formę bipolarności, określoną mianem "symetrycznej na jednolitej skali bipolarnej" w p. 5.1 na s. 94, można odnaleźć już w klasycznych zapytaniach do bazy danych. Rozważmy zapytanie względem bazy danych nieruchomosci z jednym prostym warunkiem o następującej postaci:

$$\texttt{cena} \;\leq\; \texttt{500 000 PLN} \tag{5.6}$$

lub, ogólniej, postaci (por. (2.119) w p. 2.3.1):

$$X \text{ jest } F \tag{5.7}$$

Wtedy te wartości z dziedziny atrybutu `cena`, które są większe od 500 000 są odrzucane (są negatywne), zaś wartości mniejsze lub równe 500 000 są akceptowane (są pozytywne). Jest to jednak bardzo specyficzna, w pewnym sensie zdegenerowana forma bipolarności, która będzie nas mniej interesowała przynajmniej z dwóch powodów:

1. odrzucenie i akceptacja krotek mają charakter binarny,
2. nie ma tu kategorii wartości "neutralnych" i, co jest z tym ściśle związane, krotka która nie jest negatywna jest bezwarunkowo pozytywna i na odwrót.

Użycie zapytań nieprecyzyjnych (por. p. 4.3) w oczywisty sposób przezwycięża pierwsze ograniczenie, ale drugie nadal występuje: jeśli element x dziedziny dom_X jest zaklasyfikowany jako pozytywny w stopniu $\mu_F(x)$, to jednocześnie jest on traktowany jako negatywny w stopniu $1 - \mu_F(x)$.

Można analizować różne aspekty bipolarności rozważanej w kontekście wyszukiwania informacji, w tym:

- przestrzeń rozważań, w której wyrażane są preferencje,
- rodzaj skali stosowany do ich wyrażania,
- interpretację warunków pozytywnych i negatywnych.

W p. 5.2.2 omawiamy poziomy danych na jakich manifestować się może bipolarność preferencji.

W p. 5.1 omówiliśmy różne typy bipolarności w kontekście reprezentacji informacji bipolarnej w bazie danych i związane z nimi *skale*, na których informacja bipolarna jest wyrażana. Prowadzone tam rozważania znajdują bezpośrednie zastosowanie również w przypadku modelowania preferencji w ramach wyszukiwania informacji. W p. 5.2.3 ilustrujemy na przykładach zastosowanie *jednolitej skali bipolarnej* i *podwójnej skali unipolarnej* do modelowania preferencji bipolarnych w zapytaniach.

W p. 5.3 omawiamy pewną szczególną semantykę warunków pozytywnych i negatywnych, która przyjmowana jest w większości prac dotyczących zapytań bipolarnych.

Różne aspekty bipolarności zapytań zilustrujemy na przykładzie hipotetycznej bazy danych agencji nieruchomości (por. tabl. 3.1). Domy znajdujące sie w jej ofercie scharakteryzowane są z użyciem takich atrybutów jak: `cena`, `lokalizacja` i `powierzchnia` (mieszkalna, w m^2). Rozważmy klienta agencji, który poszukuje *taniego* domu. Stosując zapytanie nieprecyzyjne (por. p. 4.3) klient ten może wyrazić swoje preferencje w postaci warunku:

$$\texttt{cena} = \mathit{niska} \tag{5.8}$$

Zapytanie takie dogodnie będzie dla dalszych rozważań przedstawić w ogólniejszej formie jako *wyrażenie lingwistyczne* (por. def. 2.28):

$$X \text{ jest } F \tag{5.9}$$

gdzie X oznacza zmienną lingwistyczną reprezentującą atrybut (np. `cena`), zaś F oznacza zbiór rozmyty reprezentujący termin lingwistyczny występujący w zapytaniu (np. "niska").

5.2.2 Poziomy wyrażania preferencji bipolarnych

Bipolarne preferencje użytkownika mogą być wyrażone:

- *na poziomie dziedziny* danego atrybutu lub
- *na poziomie całościowej oceny krotki* (ang. *comprehensive evaluation*) [116].

Pierwszy przypadek ma miejsce kiedy użytkownik postrzega (klasyfikuje) poszczególne elementy dziedziny danego atrybutu A jako, *do pewnego stopnia pozytywne*, *negatywne* lub *neutralne*. Całościowa ocena danej krotki t, z uwzględnieniem preferencji użytkownika wyrażonych względem wartości innych jeszcze atrybutów, powinna być uzależniona od tego, do której z tych trzech klas i w jakim stopniu zalicza się $t(A)$. W drugim przypadku użytkownik określa względem krotek warunki pozytywne i negatywne, odnoszące się w ogólności do wielu atrybutów. Tak więc, w tym przypadku pewne kombinacje wartości wybranych atrybutów są postrzegane jako pozytywne, negatywne bądź neutralne. Pierwszy przypadek może być traktowany jako szczególny przypadek drugiego, ale ich rozróżnienie ma walor praktyczny i jednocześnie ułatwia analizę podejść do modelowania zapytań bipolarnych zaproponowanych w literaturze. Zilustrujemy to na przykładach.

Przykład 5.3. *Rozważmy klienta agencji nieruchomości o następujących preferencjach dotyczących dziedziny atrybutu* `cena`*:*

(a) *cena powyżej 700 000 PLN jest zdecydowanie negatywna,*

(b) *cena poniżej 500 000 PLN jest zdecydowanie pozytywna,*

(c) *pozostałe ceny są neutralne, ani pozytywne ani negatywne.*

W przykładzie 5.3 bipolarność wyrażona jest *binarnie* na poziomie dzidziny atrybutu `cena`. W kolejnym przykładzie preferencje są bardziej złożone.

Przykład 5.4. *Rozważmy innego klienta. Jego preferencje można podsumować następująco:*

(a) nieruchomości droższe niż 700 000 PLN i jednocześnie o powierzchni mniejszej niż 120 m^2 są zdecydowanie negatywne,

(b) nieruchomości zlokalizowane w południowej części Śródmieścia są zdecydowanie pozytywne,

(c) pozostałe nieruchomości są neutralne.

W tym przypadku bipolarność preferencji odnosi się do kombinacji wartości atrybutów czy, równoważnie, do całych krotek reprezentujących nieruchomości dostępne w ofercie agencji. W przykładzie 5.4 –jak również w przykładzie 5.3, co jest jednak mniej oczywiste– warunki (a) i (b) wymagają dodatkowej dyskusji dotyczącej ich wzajemnej niesprzeczności. Taka dyskusja zostanie przeprowadzona w dalszej części tego rozdziału, przy opisie alternatywnych sposobów formalnej reprezentacji bipolarności w ramach zapytania (por. p. 5.3).

W obydwu przykładach 5.3-5.4 bipolarne preferencje mają charakter binarny, ale w realiach wyszukiwania domów w ofercie agencji nieruchomości będą one miały w sposób naturalny charakter rozmyty. Występujące w nich precyzyjne ograniczenia dotyczące wartości atrybutów `cena` i `powierzchnia` będą zazwyczaj zastąpione, subiektywnie rozumianymi przez klienta, terminami lingwistycznymi takimi jak: "bardzo drogi", "mała" itp. Warto zwrócić uwagę, że preferencje klienta przedstawione w tych przykładach, zarówno w wersji binarnej jak i rozmytej, nie mogą być adekwatnie wyrażone z użyciem klasycznych zapytań do baz danych czy zapytań nieprecyzyjnych.

5.2.3 Skala oceny

Bipolarna ocena dopasowania krotki do preferencji użytkownika może być wyrażona za pomocą różnych skal omawianych w p. 5.1. Zilustrujemy to teraz przede wszystkim na przykładach dotyczących wyrażania preferencji na poziomie dziedziny atrybutu.

Jednolita skala bipolarna (ang. *univariate bipolar scale*)

W przypadku użycia tej skali zakłada się, że użytkownik po określeniu swojej oceny pozytywnej i negatywniej (elementy dziedziny atrybutu lub całej krotki) jest w stanie je zagregować i wyrazić swoją łączną oceną w pewnym przedziale liczbowym. Najczęściej przyjmuje się przedziały [-1, 1] i [0, 1]. W pierwszym przypadku liczba 0 określa poziom oceny neutralnej, zaś w przypadku drugim rolę tę pełni liczba 0.5. Przeanalizujemy teraz zastosowanie tej skali w przypadku wyrażania oceny na poziomie dziedziny atrybutu.

Załóżmy, że użytkownik chce wyrazić swoją bipolarną ocenę poszczególnych elementów dziedziny dom_X pewnego atrybutu X. Dogodnie będzie przyjąć, że ocenę tę można wyrazić następująco: (por. np. [116]):

$$\xi_X : dom_X \to [-1, 1] \tag{5.10}$$

przy czym $\xi_X(x) > 0$ oznacza, że element $x \in dom_X$ jest oceniany pozytywnie w kontekście danego zapytania i $\xi_X(x)$ wyraża stopień tej pozytywnej oceny; $\xi_X(x) < 0$ oznacza, że element $x \in dom_X$ jest oceniany negatywnie w kontekście danego zapytania i $\xi_X(x)$ wyraża stopień tej negatywnej oceny; natomiast $\xi_X(x) = 0$ oznacza, że element $x \in dom_X$ jest oceniany neutralnie w kontekście danego zapytania.

W takim przypadku preferencje użytkownika można formalnie modelować z użyciem *podwójnego zbioru rozmytego* (por. def. 2.15) zdefiniowanego w dziedzinie dom_X atrybutu X. Zapytanie bipolarne można wtedy wyrazić stosując (5.9), przy czym przyjmuje się, że F oznacza teraz podwójny zbiór rozmyty o następującej interpretacji:

- funkcja przynależności π_F służy reprezentacji negatywnej oceny elementów dziedziny dom_X; jej wartości są obliczane na podstawie funkcji ξ_X (5.10) następująco:

$$\pi_F(x) = \min(1 + \xi_X(x), 1) \tag{5.11}$$

- funkcja przynależności η_F służy reprezentacji pozytywnej oceny elementów dziedziny dom_X; jej wartości są obliczane na podstawie funkcji ξ_X (5.10) następująco:

$$\eta_F(x) = \max(\xi_X(x), 0) \tag{5.12}$$

Z własności podwójnych zbiorów rozmytych (por. def. (2.15)) wynika, że istnieje wzajemnie jednoznaczne przekształcenie stopni oceny ξ_X zdefiniowanych w (5.10) w wartości funkcji przynależności odpowiadającego im podwójnego zbioru rozmytego. Przekształcenie $\xi_X(x)$ na parę wartości funkcji przynależności $(\pi_F(x), \eta_F(x))$ dane jest wzorami (5.11)-(5.12). Odwrotne przekształcenie ma następującą postać:

$$(\pi_F(x), \eta_F(x)) \rightarrow \xi_X(x) \tag{5.13}$$

$$\xi_X(x) = \begin{cases} \eta_F(x) & \text{dla } \pi_F(x) = 1 \\ \pi_F(x) - 1 & \text{wpp} \end{cases} \tag{5.14}$$

Powyższe założenia i przyjęty sposób modelowania preferencji prowadzą do następującego scenariusza. Użytkownik wyraża swoje bipolarne preferencje względem wartości atrybutu X stosując jednolitą skalę bipolarną. W tym celu użytkownik wybiera ze słownika (por. p. 4.3) dwa terminy lingwistyczne, które reprezentowane są przez zbiory rozmyte tworzące wspólnie podwójny zbiór rozmyty, czyli ich funkcje przynależności spełniają warunek (2.22).

Przykład 5.5. *Rozważmy klienta, który nie lubi małych domów i najbardziej byłby zadowolony z domu o powierzchni około 350 m^2. Może on wyrazić swoje preferencje definiując lub wybierając ze słownika systemu wyszukiwania dwa terminy lingwistyczne "mała" oraz "około 350" i tworząc z nich podwójny zbiór rozmyty F o następujących funkcjach przynależności $(\pi_F(x), \eta_F(x))$:*

$$\begin{aligned} \pi_F(x) &= 1 - \mu_{\text{“mała”}}(x) \\ \eta_F(x) &= \mu_{\text{“około 350”}}(x) \end{aligned}$$

przy czym zakłada się, że nośnik (por. def. 2.4) zbioru rozmytego reprezentującego termin lingwistyczny "około 350" jest podzbiorem rdzenia (por. def. 2.5) zbioru rozmytego będącego dopełnieniem zbioru rozmytego reprezentującego termin lingwistyczny "mała".

Sprawdzenie czy funkcje przynależności $\pi_F(x)$ and $\eta_F(x)$ spełniają warunek charakteryzujący podwójne zbiory rozmyte (2.22) jest trywialne, zakładając że ich funkcje przynależności mają na przykład postać trapezoidalną. Można więc przyjąć, że użytkownik odpowiednio wsparty w ramach interfejsu systemu wyszukiwania będzie w stanie dobrać lub zdefiniować odpowiednie terminy lingwistyczne.

Podwójna skala unipolarna (ang. *unipolar bivariate scale*)

W tym przypadku zakłada się, że użytkownik formułuje odrębnie ocenę pozytywną oraz negatywną i nie jest dostępna jego zagregowana ocena na jednolitej skali bipolarnej. Oceny mogą być ponownie wyrażone na poziomie dziedziny wybranych atrybutów lub na poziomie zbioru krotek. Przeanalizujemy teraz dokładniej ten pierwszy przypadek.

Zakłada się, że użytkownik wyraża swoją bipolarną ocenę względem elementów dziedziny dom_X wybranego atrybutu X. Każdy element dziedziny może posiadać pewne cechy pozytywne i jednocześnie pewne cechy negatywne. Można to też wypowiedzieć w ten sposób, że użytkownik może jednocześnie "lubić" i "nie lubić" danej wartości, z różnych powodów[4]. Dogodnie będzie przyjąć, że taka bipolarna ocena wyrażona jest z użyciem dwóch funkcji:

$$\xi_X^+ : dom_X \rightarrow [0, 1] \tag{5.15}$$

$$\xi_X^- : dom_X \rightarrow [0, 1] \tag{5.16}$$

gdzie $\xi_X^+(x)$ i $\xi_X^-(x)$ określają w jakim stopniu, odpowiednio, "pozytywny/lubiany" i "negatywny/nielubiany" jest dany element dziedziny $x \in dom_X$. Zilustrujemy to na przykładzie.

Przykład 5.6. *Rozważmy klienta agencji nieruchomości, który jest szczególnie zainteresowany lokalizacją domu, podaną w bazie danych w formie nazwy dzielnicy miasta, w której dany dom się znajduje. Dziedziną atrybutu* `lokalizacja` *jest więc zbiór nazw dzielnic. Dla każdej z dzielnic użytkownik jest w stanie określić listę czynników przemawiających* <u>*za*</u> *wyborem domu tam zlokalizowanego, jak i listę czynników przemawiających* <u>*przeciw*</u> *takiemu wyborowi. Każdy taki czynnik może mieć przypisany stopień ważności. Na przykład dzielnica "Śródmieście" jest dobrze skomunikowana z resztą miasta, ale notowany jest tam wysoki poziom przestępczości. Na podstawie sporządzonych list "za" i "przeciw" użytkownik jest w stanie określić oddzielnie swoją pozytywną i negatywną ocenę danego elementu dziedziny atrybutu.*

Taka ocena poszczególnych elementów dziedziny danego atrybutu na dwóch skalach unipolarnych ma uzasadnienie głównie wtedy, kiedy istnieje pewien zestaw kryteriów, które mają wpływ na tę ocenę i jednocześnie nie znajdują one swojej bezpośredniej reprezentacji w bazie danych. Jeśli lokalizacja domów w bazie danych wspomnianej w przykładzie 5.6

[4]Za Grabischem, Greco i Pirlot [116] można przytoczyć następujący przykład ilustrujący taką sytuację. Niech atrybut opisuje napój, który ma stanowić element zamawianego posiłku. Do dziedziny tego atrybutu niech należy między innymi "gorąca czekolada". Klient określający swoje preferencje może wyrazić pozytywną ocenę tego napoju dla jego walorów smakowych. Jednocześnie ten sam klient może oceniać ten napój negatywnie ze względu na jego wysoką kaloryczność.

byłaby reprezentowana na bardziej szczegółowym poziomie, na przykład z uwzględnieniem stopnia skomunikowania z resztą miasta, poziomu przestępczości itp., to preferencje użytkownika można by zapewne lepiej wyrazić używając jednolitej skali bipolarnej lub nawet pojedynczej skali unipolarnej względem dziedzin atrybutów stanowiących taką bardziej szczegółową reprezentację informacji o lokalizacji domu.

Interesującą analizę agregacji list argumentów “za” i “przeciw” Czytelnik może znaleźć w pracy [82].

5.3 Model “wymagane/postulowane”

W literaturze poświęconej zapytaniom bipolarnym najczęściej przyjmuje się specjalną interpretację warunków pozytywnego i negatywnego. Polega ona na traktowaniu warunku negatywnego jako *wymaganego* (ang. *required*), zaś warunku pozytywnego jako *postulowanego/pożądanego* (ang. *desired*). Dokładniej, warunek wymagany stanowi dopełnienie warunku negatywnego R i musi być bezwzględnie spełniony. Warunek postulowany odpowiada wprost warunkowi pozytywnemu i musi być spełniony tylko wtedy *jeśli jest to możliwe*. Spełnienie warunku postulowanego jest pożądane, ale jeśli występuje konflikt ze spełnieniem warunku wymaganego, to priorytet ma ten drugi. Dopełnienie warunku R oznaczamy jako C, $C = \overline{R}$, i całe zapytanie bipolarne w tym modelu oznaczamy jako parę (C, P). W p. 5.4.2 wprowadzimy jeszcze inne oznaczenie na specjalny przypadek z jawnie podanym operatorem agregacji obydwu warunków.

Warunek wymagany traktowany jest jako *ograniczenie* (ang. *constraint*), które określa zbiór dopuszczalnych rozwiązań – w naszym przypadku krotek. Warunek postulowany określa zaś pożądane cechy poszukiwanych krotek, ale ich nieposiadanie przez krotkę nie dyskwalifikuje jej bezwarunkowo.

Zilustrujmy tę specjalną semantykę zapytania bipolarnego następującym przykładem.

Przykład 5.7. Wyszukaj nieruchomości tańsze niż 500 000 PLN *i jeśli to możliwe* zlokalizowane nie dalej niż 300 m od stacji kolejowej.

W przykładzie tym warunek bezwzględnie *wymagany* (ang. *required*) C odnosi się do ceny, a warunek *postulowany* (*pożądany*) (ang. *desired*) P odnosi się do lokalizacji nieruchomości względem stacji kolejowej. Nieprecyzyjna wersja takiego zapytania może przybrać następującą postać.

Przykład 5.8. *Wyszukaj* tanie *nieruchomości* i jeśli to możliwe *położone* niedaleko *stacji kolejowej.*

Semantyka zapytania bipolarnego takiego jak w przykładzie 5.7 jest następująca:

- preferowane są krotki spełniające obydwa warunki C i P,
- w drugiej kolejności do odpowiedzi na zapytanie należą krotki spełniające jedynie warunek C.

Możliwa jest też nieco inna interpretacja, która jest bliższa koncepcji Lacroix i Lavency [147], którą przyjmujemy w niniejszej książce jako podstawę omawianej semantyki zapytań bipolarnych. Interpretację tę można wyrazić następująco:

- jesli istnieją krotki, które spełniają jednocześnie obydwa warunki C i P, to *wyłącznie* one należą do odpowiedzi na zapytanie,
- jeśli takie krotki nie istnieją, to do odpowiedzi należą krotki spełniające warunek C.

Warto zwrócić uwagę na istotne znaczenie nieprecyzyjności (rozmytości) drugiego z przykładów 5.8. Dla przykładu 5.7 uzasadnione może być następujące dwuetapowe postępowanie przy realizacji tego zapytania:

- najpierw wybierz nieruchomości tańsze niż 500 000 PLN,
- następnie, spośród nich, wybierz te położone nie dalej niż 300 metrów od stacji (o ile takie istnieją).

Dla drugiego przykładu pierwszy krok tego postępowania nie jest dobrze określony: nieprecyzyjna (rozmyta) natura pojęcia *tani* nie pozwala na proste rozróżnienie tanich nieruchomości od pozostałych. Powrócimy do tego zagadnienia w dalszych rozważaniach dotyczących zapytań bipolarnych.

Rozważmy jeszcze jeden przykład.

Przykład 5.9. *Wyszukaj obrazy namalowane w XVII wieku i, jeśli to możliwe autorstwa jednego z mistrzów flamandzkich.*

W przykładzie 5.9 warunek C to “`data(t)`$\in$ [1601,1700]” (oryginalny warunek negatywny to “`data(t)` $\notin$ [1601,1700]”), zaś warunek postulowany/pozytywny P to “`autor(t)`$\in$ {zbiór mistrzów flamandzkich}”.

Niezależnie od przyjętej w tym modelu specyficznej interpretacji warunków pozytywnego i negatywnego nadal realizacja zapytania bipolarnego polega na obliczeniu stopni spełnienia warunków C i P. Podobnie jak w przypadku ogólnym (por. p. 5.2.1) sposoby uporządkowania krotek w odpowiedzi na zapytanie są zasadniczo dwa: określenie porządku na parach stopni spełnienia obydwu warunków lub użycie pewnej ich skalaryzacji. W p. 5.4.2 omówimy szczegółowo ten drugi sposób, argumentując że lepiej reprezentuje on semantykę "wymagane/postulowane" rozumianą tak jak w przykładach 5.7-5.9, czyli z jawnym użyciem operatora "i jeżeli to możliwe". W dalszej części niniejszego punktu omówimy pokrótce użycie porządku leksykograficznego oraz przedstawimy krótki przegląd literatury dotyczącej tego modelu zapytań bipolarnych.

Pojęcie zapytania bipolarnego i jego model "wymagane/postulowane" wprowadzili Dubois i Prade w 2002 roku [94]. Dubois i Prade definiują zapytanie bipolarne jako zbiór par warunków (predykatów):

$$\{(C_i, P_i)\} \tag{5.17}$$

przy czym C_i i P_i określają odpowiednio wymagane i postulowane warunki, które powinna spełnić wartość atrybutu A_i. Warunki względem poszczególnych atrybutów agreguje się zgodnie z następującymi założeniami. Spełnienie zbioru par warunków (5.17) równoznaczne jest ze spełnieniem *wszystkich* (koniunkcji) warunków wymaganych i *przynajmniej jednego* (alternatywy) warunku postulowanego. Formalnie agregację warunków w podejściu Dubois i Prade'a zapisuje się następująco

$$(C, P) = (\times_i C_i, +_i P_i) \tag{5.18}$$

gdzie $\times_i C_i = C_1 \times C_2 \times \ldots \times C_k$, $+_i P_i = \overline{(\overline{P_1} \times \overline{P_2} \times \ldots \times \overline{P_k})}$ i $\overline{X}$ oznacza dopełnienie zbioru X. Charakter przyjętej agregacji, opartej na koniunkcji (alternatywie), staje się bardziej ewidentny, kiedy zapiszemy (5.18) równoważnie jako:

$$C(t) = \min_i C_i(t)$$
$$P(t) = \max_i P_i(t)$$

W dalszych rozważaniach traktuje się warunek C całościowo, jako warunek atomowy, chociaż w rzeczywistości może on mieć złożoną strukturę.

Przyjmijmy, że użytkownik wyraża swoje bipolarne preferencje względem dziedziny dom_X danego atrybutu X w ten sposób, że określa zbiór rozmyty R elementów *odrzucanych* i zbiór rozmyty elementów P *postulowanych/pożądanych*. Preferencje takiego użytkownika można wtedy formalnie opisać z użyciem dwóch funkcji przynależności:

$$\xi_X^R : dom_X \to [0, 1] \tag{5.19}$$

$$\xi_X^P : dom_X \to [0, 1] \tag{5.20}$$

takich, że dla $x \in dom_X$ wartość $\xi_X^R(x)$ określa stopień odrzucenia elementu x, zaś $\xi_X^P(x)$ oznacza stopień w jakim spełnia on oczekiwania użytkownika. Należy zauważyć, że jeśli zbiory R i P są swoimi dopełnieniami, to wystarczy określić tylko jeden z nich i mamy wtedy do czynienia z prostym zapytaniem nieprecyzyjnym (5.9). Funkcja ξ_X^P może wtedy odpowiadać na przykład funkcji przynależności terminu "niska" (5.8). Tak więc z interesującym przypadkiem bipolarności mamy do czynienia tylko wtedy, gdy $C = \overline{R} \neq P$ i takie zapytanie może być reprezentowane równoważnie przez pary zbiorów (predykatów) rozmytych (R, P) lub (C, P), jak to już wcześniej wskazywaliśmy.

Przyjęta semantyka "wymagane/postulowane" może skłaniać do przyjęcia, że zbiory rozmyte C i P powinny spełniać warunek spójności $P \subseteq C$ wyrażający następującą zależność: jeśli element dziedziny jest postulowany/pożądany, to nie może być jednocześnie odrzucany. Biorąc pod uwagę taki warunek, dobrym modelem preferencji użytkownika będzie więc zbiór intuicjonistyczny w sensie Atanassova (AIFS) (por. def. 2.16) określony na dziedzinie dom_X. Zapytanie bipolarne o takiej semantyce wyrazimy więc z użyciem szablonu "X jest F" (2.119), w którym zbiór rozmyty F jest zastąpiony zbiorem AIFS, który będzie interpretowany następująco:

- funkcja przynależności μ_A określa na ile dany element dziedziny jest postulowany/pożądany jako wartość atrybutu X:

$$\mu_A(x) = \xi_X^P(x)$$

- funkcja przynależności ν_A określa na ile dany element dziedziny jest odrzucany jako wartość atrybutu X:

$$\nu_A(x) = \xi_X^R(x)$$

Warunek spójności $P \subseteq C$, który można wyrazić równoważnie jako $P \subseteq \overline{R}$ lub $\xi_X^P(x) \leq 1 - \xi_X^R(x)$ pokrywa się wtedy z warunkiem (2.23), który charakteryzuje zbiory AIFS.

Użytkownik wyraża więc swoje preferencje na podwójnej skali unipolarnej. W tym celu wybiera dwa terminy lingwistyczne reprezentowane przez zbiory rozmyte R i P spełniające warunek (2.23). Ilustrację stanowi przykład 5.10, który jest zmodyfikowaną wersją przykładu 5.5.

Przykład 5.10. *Rozważmy klienta, który nie lubi małych domów i najbardziej byłby zadowolony z domu o powierzchni około 350 m^2. Może on wyrazić swoje preferencje definiując lub wybierając ze słownika systemu wyszukiwania dwa terminy lingwistyczne "mała" oraz "około 350" i tworząc z nich zbiór AIFS F o następujących funkcjach przynależności i nieprzynależności:*

$$\begin{aligned}\nu_F(x) &= \mu_{\text{``small''}}(x)\\ \mu_F(x) &= \mu_{\text{``around 350 sq. m.''}}(x)\end{aligned}$$

przy czym zakłada się, że dla wszystkich elementów dziedziny atrybutu `powierzchnia` *suma wartości ich funkcji przynależności i nieprzynależności nie przekracza 1.*

Dla tak doprecyzowanej semantyki "wymagane/postulowane" Dubois i Prade [94] przyjmują jako wynik realizacji zapytania bipolarnego *leksykograficzne uporządkowanie* $\preceq$ krotek, określone następująco:

$$t_1 \preceq t_2 \Leftrightarrow (C(t_1) < C(t_2)) \vee ((C(t_1) = C(t_2)) \wedge (P(t_1) \leq P(t_2))) \quad (5.21)$$

W ogólnym przypadku, niemożliwe jest określenie funkcji, która przypisuje poszczególnym krotkom stopnie *spełnienia* zapytania (C, P) w sposób odpowiadający uporządkowaniu leksykograficznemu w przestrzeni $\mathcal{R} \times \mathcal{R}$ (por. np. [106]). Przy pewnych dodatkowych założeniach[5], racjonalnych w przypadku rozważanych tu zapytań do bazy danych, można jednak określić taką funkcję. Szerszą dyskusję tego zagadnienia można znaleźć w pracy Fishburna [106], a także odnośnie do zapytań do bazy danych w pracy Bosca i Piverta [37]. Za tą ostatnią pracą można przytoczyć przykład takiej funkcji, którą oznaczymy jako ψ:

$$\psi(C, P, t) = (2 * C(t) + P(t))/3 \quad (5.22)$$

gdzie $\psi(C, P, t)$ oznacza stopień spełnienia przez krotkę t zapytania bipolarnego o warunkach C i P.

Wtedy $\psi(C, P, t) \in [0, 1]$ i dla *nierozmytych* warunków C i P otrzymuje się

$$t_1 \preceq t_2 \Leftrightarrow \psi(C, P, t_1) \leq \psi(C, P, t_2) \quad (5.23)$$

gdzie $\preceq$ oznacza uporządkowanie leksykograficzne określone we wzorze (5.21).

[5]Sprowadzają się one w praktyce do przyjęcia pewnego skończonego podzbioru przedziału [0,1] dla reprezentacji stopni spełnienia zapytania.

Dubois i Prade [94] rozważają również inne funkcje określające skalarny stopień spełnienia zapytania przez wiersze.

Spójność warunku wymaganego (C lub R) i postulowanego (P), wyrażona na przykład z użyciem zbiorów AIFS, nie jest jednoznacznie związana z semantyką “wymagany/postulowany”. W p. 5.4 omawiamy szerzej podejście, w którym charakter obu warunków jest zachowany, ale nacisk kładzie się na ich jawną agregację z użyciem operatora “and possibly”.

Podsumowując, źródeł modelu “wymagane/postulowane” zapytań bipolarnych należy upatrywać w pracy Lacroix i Lavency [147] z 1987 roku. Mimo, że nie jest tam używany termin “bipolarny” ani nie ma żadnych odniesień do pojęcia bipolarności, to związek z jawnie później wprowadzonym prze Dubois i Prade’a [94] pojęciem zapytań bipolarnych jest oczywisty. Lacroix i Lavency jako pierwsi zaproponowali zastosowanie zapytania z warunkiem o dwóch różnych składowych: C, której spełnienie jest wymagane, i P, której spełnienie jest postulowane/pożądane. Jest to przykład zapytania bipolarnego, w którym stosuje się podwójną skalę unipolarną. Bipolarność może być wyrażona na poziomie relacji ze specyficzną, omawianą wcześniej, interpretacją warunków pozytywnego i negatywnego. Może też być wyrażona na poziomie dziedziny atrybutu, jak w następującym przykładzie:

Przykład 5.11. *Wyszukaj nieruchomości o cenie poniżej 700 000 PLN i, o ile to możliwe, o cenie poniżej 500 000 PLN.*

Praktyczna przydatność zapytań w tym przypadku wydaje się mniejsza, ale warto przeanalizować i ten przykład. Bipolarność w przykładzie 5.11 wydaje się wprowadzać jedynie pewne osłabienie warunku narzuconego na cenę: żądanie, żeby była ona mniejsza niż 500 000 PLN nie jest tak kategoryczne i również nieruchomości droższe, o cenie z przedziału [500 000, 700 000] mogą znaleźć się w odpowiedzi na to zapytanie. Jednak, zgodnie z oryginalną semantyką modelu “wymagane/postulowane” pochodzącą od Lacroix i Lavency, stanie sie tak wyłącznie wtedy, gdy nie ma nieruchomości o cenie poniżej 500 000 PLN. To cecha, na którą warto zwrócić uwagę raz jeszcze. Jeśli istnieją krotki, które spełniają obydwa warunki C i P, to zgodnie z tą semantyką wyłącznie one znajdą się w odpowiedzi na zapytanie. W innych podejściach (por. np. [100]) do modelowania pary warunków wymaganego i postulowanego przyjmuje się, że krotki spełniające wyłącznie warunek C też znajdą się w odpowiedzi na zapytanie, tyle że z mniejszym stopniem spełnienia niż te, które spełniają obydwa warunki lub na odleglejszym miejscu na liście krotek zwracanych jako odpowiedź na zapytanie. Takie podejście ma swoje zalety, ale przede wszystkim wtedy, gdy bezpośrednim odbiorcą wyników zapytania bipolarnego jest użytkownik. Jeśli natomiast para warunków (C, P) stanowi jedynie element większego zapytania, to oryginalna semantyka Lacroix i Lavency modelu "wymagane/postulowane" może być właściwsza.

5.4 Agregacja ocen bipolarnych

Zastosowanie podwójnej skali unipolarnej wydaje się być najbardziej elastyczną formą reprezentacji preferencji użytkownika, pozwalającą na ich wierniejsze wyrażenie. W efekcie realizacji zapytania z tak wyrażonymi preferencjami uzyskujemy więc dla każdej krotki dwa stopnie spełniania przez nią, osobno, warunku pozytywnego i negatywnego. Należy określić sposób uporządkowania krotek w odpowiedzi na zapytanie. W ramach badań nad zapytaniami bipolarnymi poświęcono temu zagadnieniu wiele miejsca. W dalszej części niniejszego rozdziału przedstawimy podstawowe podejścia zaproponowane w tym względzie, ponownie z rozróżnieniem ogólnej koncepcji zapytań bipolarnych (por. def. 5.1) i ich modelu "wymagane/postulowane" (por. p. 5.3), któremu poświęcono szczególnie wiele prac.

Zasadniczo w literaturze rozważa się dwa rozwiązania. J e d n o z nich polega na określeniu uporządkowania wprost na parach liczb. Stosuje się wtedy przede wszystkim porządek leksykograficzny, uwzględniając najpierw stopień spełnienia warunku negatywnego lub najpierw stopień spełnienia warunku pozytywnego. Wybór jednej z tych dwóch opcji może zależeć od dyskutowanego już wcześniej (por. p. 5.2.1; s. 100) nastawienia użytkownika. Jeśli bardziej zależy mu na unikaniu spełnienia warunku negatywnego to wybrana powinna być opcja pierwsza, natomiast jeśli ważniejsze jest jak najlepsze spełnienie warunku pozytywnego, to wybrana powinna być opcja druga. Przyjęcie porządku leksykograficznego oznacza faktyczne odejście od równorzędnego traktowania warunku pozytywnego i negatywnego. Znajduje ono więc większe uzasadnienie w przypadku semantyki "wymagane/postulowane", gdzie takie nierównorzędne traktowanie występuje niejako z definicji (por. p. 5.3). D r u g i e rozwiązanie problemu określenia uporządkowania krotek w odpowiedzi na zapytanie polega na zagregowaniu stopni spełnienia warunku pozytywnego i negatywnego. Głównie temu zagadnieniu poświęcimy dalszą część niniejszego punktu.

5.4.1 Przypadek ogólny

Warto zauważyć, że zarówno ocena pozytywna, jak i negatywna elementu dziedziny atrybutu lub krotki może być wyrażona względem pewnej liczby kryteriów (w postaci pewnej liczby niezależnie określonych warunków) i wtedy zagadnienie ich agregacji wymaga dodatkowej analizy. Dla modelu "wymagane/postulowane" taką analizę przeprowadzili Dubois i Prade (por. 5.18 i [94]). Teraz krótko omówimy podejście zaproponowane w pracy [116], które znajduje zastosowanie w przypadku rozważanego tu ogólnego modelu zapytania bipolarnego.

W podejściu Grabischa, Greco i Pirlot [116] bipolarna ocena każdej krotki[6] t wyrażona jest w postaci następującego wektora:

$$t \mapsto [(\xi_1^+, \xi_1^-), \ldots, (\xi_n^+, \xi_n^-)]$$

gdzie n oznacza liczbę kryteriów[7] rozważanych przy ocenie krotek, zaś ξ_i^+ i ξ_i^- oznaczają, odpowiednio, pozytywną i negatywną ocenę krotki t względem i-tego kryterium.

Dogodnie jest przedstawić ten wektor w innej postaci:

$$t \mapsto [\xi_1^+, \ldots, \xi_n^+, \xi_1^-, \ldots, \xi_n^-]$$

W pracy [116] definiuje się różne typy agregacji dla tak wyrażonej oceny bipolarnej: *zagregowaną ocenę pozytywną* (ang. *comprehensive positive evaluation*), $CPE(t)$, *zagregowaną ocenę negatywną* (ang. *comprehensive negative evaluation*), $CNE(t)$, oraz *ocenę globalną* (ang. *comprehensive evaluation*), $CE(t)$:

$$CPE(t) = F^+(\xi_1^+, \ldots, \xi_n^+, \xi_1^-, \ldots, \xi_n^-) \tag{5.24}$$

$$CNE(t) = F^-(\xi_1^+, \ldots, \xi_n^+, \xi_1^-, \ldots, \xi_n^-) \tag{5.25}$$

$$CE(t) = G(F^+(\xi_1^+, \ldots, \xi_n^+, \xi_1^-, \ldots, \xi_n^-), F^-(\xi_1^+, \ldots, \xi_n^+, \xi_1^-, \ldots, \xi_n^-)) \tag{5.26}$$

[6]Dla uproszczenia notacji odnosimy się tu tylko do wyrażania preferencji bipolarnych na poziomie zbioru krotek, ale podobne rozważania można przeprowadzić dla przypadku, gdy są one wyrażane na poziomie dziedziny wybranego atrybutu. W [116] rozważa się zadanie wielokryterialnego podejmowania decyzji i stosuje się tam do opisu obiektu poddawanego ocenie określenia "opcja" lub "alternatywa".

[7]W najprostszym przypadku każde kryterium może być skojarzone z osobnym atrybutem.

gdzie $F^+, F^- : [0,1]^{2n} \longrightarrow [0,1]$, oraz F^+ i F^- są funkcjami niemalejącymi względem pierwszych n argumentów i nierosnącymi względem ostatnich n argumentów, zaś $G : [0,1]^2 \longrightarrow [0,1]$ jest funkcją niemalejącą względem pierwszego argumentu i nierosnącą względem drugiego argumentu. W tym bardzo ogólnym modelu agregacji przyjmuje się, że zarówno zagregowana ocena pozytywna, jak i zagregowana ocena negatywna zależą łącznie od wszystkich ocen pozytywnych i negatywnych względem poszczególnych kryteriów. Jako szczególny przypadek traktuje się wariant, w którym funkcje F^+ i F^- przyjmują następujacą postać:

$$\begin{aligned} CPE(t) &= F^+(\xi_1^+, \ldots, \xi_n^+) \\ CNE(t) &= F^-(\xi_1^-, \ldots, \xi_n^-) \\ CE(t) &= G(F^+(\xi_1^+, \ldots, \xi_n^+), F^-(\xi_1^-, \ldots, \xi_n^-)) \end{aligned}$$

Inne podejście do agregacji pozytywnej i negatywnej oceny w przypadku ogólnej postaci zapytania bipolarnego zaproponowano w pracach [162, 79, 161]. Para stopni spełnienia warunku pozytywnego i negatywnego (s, d) określana jest tu mianem *bipolarnego stopnia spełnienia* (ang. *bipolar satisfaction degree*) (BSD), przy czym składowa $s \in [0,1]$ nazywana jest *stopniem spełnienia* (ang. *satisfaction degree*), zaś składowa $d \in [0,1]$ *stopniem niespełnienia* (ang. *dissatisfaction degree*). Bipolarne stopnie spełnienia są semantycznie blisko związane z parą (π, ν) stopni przynależności i nieprzynależności w teorii intuicjonistycznych zbiorów rozmytych w sensie Atanassova (AIFS) (por. def. 2.16). Zasadnicza różnica polega na braku w przypadku BSD warunku (2.23) charakteryzującego AIFS. Zakłada się, zgodnie z ogólnym modelem zapytań bipolarnych, że stopnie spełnienia są od siebie całkowicie niezależnie określone, więc dopuszczalna jest sytuacja, w której $s + d > 1$.

Uporządkowanie krotek względem skojarzonych z nimi bipolarnych stopni spełnienia może być określone na kilka sposobów. Jedna z propozycji polega na skalaryzacji pary (s, d) z użyciem ich różnicy $s - d$. Nawiązując do omawianych wcześniej typów nastawienia użytkownika przy formułowaniu zapytań (por. p. 5.2.1; s. 100) można uznać, że w ten sposób modeluje się odpowiednik postawy użytkownika neutralnego wobec ryzyka (ang. *risk-neutral*). Taki użytkownik równorzędnie traktuje pozytywną i negatywną ocenę krotki.

Oryginalna koncepcja bipolarnych stopni spełnienia (BSD) przewiduje ich zastosowanie przede wszystkim na poziomie dziedziny atrybutu. Zakłada się, że są one określone dla zestawu wybranych przez użytkownika atrybutów, a następnie są agregowane dla uzyskania globalnego stopnia spełnienia zapytania zgodnie z zastosowanymi jawnie spójnikami logicznymi. W pracach [79, 161] proponuje się jak tę agregację przeprowadzać dla podstawowych spójnikow logicznych negacji, koniunkcji i alternatywy.

5.4.2 Agregacja w modelu "wymagane/postulowane"

W niniejszym punkcie przedstawimy podejście do jawnej agregacji stopni spełnienia warunku wymaganego $C(t)$ i postulowanego $P(t)$ zapytania bipolarnego w ramach modelu "wymagane/postulowane". Przyjmujemy, że preferencje użytkownika wyrażone są na podwójnej skali unipolarnej na poziomie relacji. Dodatkowo zakładamy, że obydwa warunki zapytania połączone są jawnie specjalnym spójnikiem "and possibly" (*"i jeżeli to możliwe"*). Takie zapytanie pokazuje przykład 5.7.

Dla wygody czytelnika przypomnimy teraz pokrótce podstawowe oznaczenia przyjęte w książce.

$T = \{t_j\}$ oznacza zbiór krotek (instancję relacji), względem którego formułowane są zapytania użytkownika; krotki będziemy też oznaczać symbolami x, $y, \ldots$ tam gdzie symbol t może być niejednoznacznie rozumiany,

C oznacza *warunek wymagany*, stanowiący składnik warunku zapytania bipolarnego; C jednocześnie oznaczać będzie zbiór rozmyty krotek spełniających ten warunek, jak również predykat rozmyty definiujący ten warunek; $C(t)$ będzie traktowane jako formuła logiki rozmytej, jak również jako wartość funkcji przynależności zbioru rozmytego C dla krotki $t \in T$ czy też, równoważnie, jako stopień spełnienia warunku C przez krotkę $t \in T$,

P oznacza *warunek postulowany*, stanowiący składnik warunku zapytania bipolarnego; P i $P(t)$ będą również, zależnie od kontekstu interpretowane analogicznie do C i $C(t)$,

(C, P) oznacza zapytanie bipolarne o warunkach wymaganym C i postulowanym P.

Zapytania bipolarne w modelu "wymagane/postulowane" będziemy również nazywać *zapytaniami ze spójnikiem "and possibly"* lub zapytaniami "C and possibly P" i zapisywać jako:

$$C \text{ and possibly } P \tag{5.27}$$

W oryginalnym nierozmytym podejściu Lacroix and Lavency [147] agregacja warunków C i P przebiega następująco. Krotka t należy do odpowiedzi na zapytanie (5.27), jeśli spełnia ona warunek wyrażony z użyciem następującej formuły logicznej [147]:

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \exists s\ (C(s) \wedge P(s)) \Rightarrow P(t) \tag{5.28}$$

Charakterystyczną cechą takiej interpretacji zapytania bipolarnego jest to, że uzależnia ona wynik działania operatora "and possibly" od możliwości jednoczesnego spełnienia warunku C i warunku P przez krotkę $t \in T$. Powiemy, że pomiędzy warunkami C i P występuje *konflikt* jeśli takie jednoczesne spełnienie warunków nie jest możliwe, to znaczy jeśli w całym zbiorze krotek T nie istnieje krotka jednocześnie spełniająca obydwa warunki. Jeśli pomiędzy warunkami nie występuje konflikt, to warunek zapytania bipolarnego (5.27) zamienia się w zwykłą koniunkcję tych warunków, $C \wedge P$. Wynika to z własności operatora implikacji: jeśli jego poprzednik $\exists s\ (C(s) \wedge P(s))$ jest prawdziwy, to wartość logiczna całej implikacji $(C(s) \wedge P(s)) \Rightarrow P(t)$ jest równa wartości logicznej następnika $P(t)$. Z kolei, jeśli występuje konflikt pomiędzy warunkami, to warunek zapytania bipolarnego zamienia się w warunek wymagany C.

Bardzo ważna jest tu obserwacja, że spełnianie bądź niespełnianie przez daną krotkę zapytania bipolarnego zależy od postaci całego zbioru krotek T. Pokazuje to brak ekstensjonalności (ang. *truth-functionality*) operatora "and possibly", co czyni go istotnie różnym od popularnych operatorów agregacji. Warunek postulowany odgrywa w zapytaniu w pewnym sensie drugorzędną rolę. Należy jednak podkreślić, że ten drugorzędny charakter nie da się wyrazić z użyciem klasycznie rozumianych *stopni ważności*. Spróbujmy zastosować najogólniejszy schemat agregacji ważonej, czyli operator przedstawiony na s. 52 wzorem (2.144). Zapiszmy wzór (2.144) dla przypadku dwóch wyrażeń C i P[8]:

$$\min((w_1 \to C(t)), (w_2 \to P(t))) \tag{5.29}$$

Yager zauważył [230, 231], że schemat (5.28) można zinterpretować jako specjalny przypadek (5.29), w którym waga $w_1 = 1$, zaś w_2 jest zależne od *kontekstu* i jest równe wartości logicznej formuły $\exists_{s \in T}(C(s) \wedge P(s))$:

$$\min((1 \to C(t)), (w_2 \to P(t))) \tag{5.30}$$

[8]Przy czym dla operatora implikacji przyjmujemy notację infiksową i używamy dla oznaczenia go symbolu $\to$ zamiast `i` oraz używamy operatora minimum w roli `t`-normy.

Tak więc można zinterpretować (5.28) jako ważoną koniunkcję wyrażeń, przy czym stopień ważności warunku wymaganego C jest równy 1, zaś stopień ważności warunku postulowanego P zależy od zawartości bazy danych. Zapytanie bipolarne "C and possibly P" można więc traktować jako pewien szablon zapytań, wyrażony w formie (5.30), który przybiera postać konkretnego zapytania –z ustaloną wartością w_2– zależną od zawartości bazy danych. Wzór (5.30) jawnie wskazuje, że wartość w_2 określa stopień w jakim warunek postulowany P "nie jest drugorzędny": dla $w_2 = 0$ jest on całkowicie ignorowany zaś dla $w_2 = 1$ staje się on równie wymagany jak warunek C. Najbardziej niejednoznaczne i interesujące są oczywiście przypadki pośrednie kiedy $w \in (0, 1)$.

Semantyka modelu "wymagane/postulowane" wywodzi się z pracy Lacroix i Lavency [147] i określona jest przez operator "and possibly" (5.27)-(5.28). Operator agregacji o takich właściwościach został mniej więcej w tym samym czasie niezależnie zaproponowany przez Yagera [234] oraz Dubois i Prade'a [87] w kontekście wnioskowania przez domniemanie (ang. *default reasoning*), jak również przez Yagera [228, 230, 229, 231] w obszarze wielokryterialnego podejmowania decyzji. Yager [231] wprowadza pojęcie *posybilistycznie warunkowanego kryterium* (ang. *possibilistically quantified criterion*), które intuicyjnie charakteryzuje jako kryterium, które powinno być spełnione, chyba że jego spełnienie jest w konflikcie ze spełnieniem innych kryteriów. Stanowi to istotę semantyki operatora "and possibly" (5.27)-(5.28), tak jak ją przyjmujemy w tej książce.

Operator "and possibly" o takiej semantyce stosowany był również przez Bordognę i Pasi [25] w obszarze wyszukiwania informacji tekstowej.

Lacroix i Lavency [147] rozważają jedynie nierozmyte postacie warunków C i P. Wtedy zapytanie bipolarne z operatorem "and possibly" (C, P) może być realizowane z użyciem strategii:

(i) najpierw wybierz krotki spełniające warunek C,

(ii) następnie uporządkuj je względem (binarnego) spełniania przez nie warunku P.

Tak więc relacja wynikowa może być określona z użyciem złożenia dwóch operacji wyboru algebry relacji: najpierw z warunkiem C, a następnie z warunkiem P. Konieczna jest jednak pewna modyfikacja tego złożenia: jeśli wynikiem drugiej operacji wyboru jest pusty zbiór krotek, to przyjmuje się, że tym wynikiem jest wynik pierwszej operacji wyboru. Ta strategia stanowi punkt wyjścia większości podejść, w których dopuszcza się rozmytą postać warunków C i P. Dotyczy to pierwszych prób Bosca i Piverta [37, 39], jak również podejścia Dubois i Prade'a [96]. W podejściach tych główna uwaga skupiona jest na agregacji cząstkowych warunków wymaganych C_i i postulowanych P_i (por. (5.18)), a następnie na parach zagregowanych globalnych warunków C i P stosowany jest po prostu porządek leksykograficzny. Dotyczy to również nowszych podejść Bosca i in. [45], Lietarda i Rocachera [151] czy Lietarda i in. [152, 153]; por. również przegląd takich podejść autorstwa Dubois i Prade'a [100]. Zadrożny i Kacprzyk [245, 250, 252, 253, 248, 135, 254] zaproponowali i przeanalizowali rozmytą wersję podejścia Lacroix i Lavency; por. p. 5.4.3.

W literaturze znane są inne podejścia do modelowania operatora takiego jak "and possibly". Dujmović [103] wprowadził pojęcie operatora *częściowej absorpcji* (ang. *partial absorption function*), który może być używany do agregacji wartości dwóch wyrażeń w taki sposób, że wartość pierwszej ze zmiennych kontroluje wpływ wartości drugiego wyrażenia na wynik agregacji. Dzięki temu można wyrazić taki tryb agregacji, w którym wysoka wartość całego wyrażenia wymaga wysokiej wartości pierwszego wyrażenia. Wysoka wartość drugiego wyrażenia jest pożądana ale niewymagana. Jeśli zastosuje się ten operator do stopni spełnienia warunków $C(t)$ i $P(t)$, to uzyskuje się efekt zbliżony do osiąganego z użyciem formuły Lacroix i Lavency (5.28). Jednak brak tu zależności od zawartości całej bazy danych. Podejście Dujmovića stanowi implementację idei *dynamicznych* stopni ważności agregowanych wyrażeń, gdzie stopnie te same zależą od agregowanych wartości. Podobne podejście zaproponowali później Dubois i Prade; por. np. [92].

Tudorie [208] zaproponowała operator "among", którego zastosowanie w zapytaniu daje efekt podobny do operatora "and possibly". Tudorie rozważa zapytania typu:

> Wyszukaj krotki spełniające warunek P wśród tych, które spełniają warunek C,

które są zasadniczo równoważne zapytaniom bipolarnym z operatorem "and possibly" (5.27)–(5.28). Na przykład zapytanie bipolarne pokazane w przykładzie (5.8) można wyrazić w następującej postaci z użyciem operatora "among".

Przykład 5.12. *Wyszukaj nieruchomości położone niedaleko (P) stacji kolejowej wśród (among) tanich nieruchomości (C).*

Realizacja takich zapytań przebiega jednak inaczej. Polega ona na "przeskalowaniu" terminów lingwistycznych występujących w warunku

P i ich użyciu do obliczenia stopnia spełnienia zapytania w następujący sposób:

(1) Najpierw wybierz krotki, które spełniają warunek C w stopniu większym od zera.

(2) Następnie przeskaluj funkcje przynależności zbiorów rozmytych reprezentujących terminy lingwistyczne występujące w warunku P, takie jak "niedaleko" z uwzględnieniem zmiany zakresu wartości odpowiednich atrybutów (`od_stacji`; por. tab. 3.1) w zbiorze krotek wybranych w kroku (1) w porównaniu z zakresem ich wartości w całym zbiorze krotek. Na przykład, jeśli oryginalnie odległość dwóch kilometrów należy do zbioru rozmytego reprezentującego termin "niedaleko" w stopniu 0.5 i okazuje się, że jest to najkrótsza odległość od stacji kolejowej dla nieruchomości wybranych w kroku (1) (czyli, w przykładzie 5.12, dla nieruchomości *tanich* w niezerowym stopniu), to stopień przynależności dla liczby 2 może być podniesiony na przykład z 0.5 do 1.0.

(3) Oblicz stopień spełnienia zapytania traktowanego jako koniunkcja warunku C i warunku P zmodyfikowanego tak, że występujące w nim terminy lingwistyczne reprezentowane są przez zbiory rozmyte z funkcjami przynależności przeskalowanymi w opisany wyżej sposób.

Warto zauważyć, że jeśli nie ma konfliktu pomiędzy spełnieniem obydwu warunków –w sensie przyjętym przy omawianiu zapytań bipolarnych z operatorem "and possibly"– to funkcje przynależności terminów lingwistycznych występujących w P nie zostaną przeskalowane, gdyż nie wystąpi zmiana zakresu wartości odpowiadających im atrybutów. W związku z tym całe zapytanie będzie realizowane tak jakby użyta została w nim zwykła koniunkcja warunków $C \wedge P$ – dokładnie tak jak ma to miejsce w przypadku zapytań bipolarnych z operatorem "and possibly".

Zapytanie bipolarne (C, P) z operatorem "and possibly" można również w następujący sposób wyrazić z użyciem operacji algebry relacji rozmytych *winnow* (por. p. 4.2, s. 86) [63, 250]. Niech G oznacza rozmytą relację preferencji zdefiniowaną następująco:

$$G(s,t) \Leftrightarrow P(s) \wedge \neg P(t) \tag{5.31}$$

przy czym $G(\cdot,\cdot)$ i $P(\cdot)$ oznaczają zarówno predykaty rozmyte, jak i funkcje przynależności skojarzonych z nimi zbiorów rozmytych, zależnie od kontekstu użycia. Wtedy zapytanie bipolarne z operatorem "and possibly" można wyrazić jako złożenie $\omega_R(\sigma_C(T))$ rozmytej operacji wyboru $\sigma_C(\cdot)$ (por. (4.8)) i rozmytej operacji *winnow* $\omega_R(\cdot)$ (4.18):

$$\mu_{\omega_R(\sigma_C(T))}(t) = \mathrm{truth}(C(t) \wedge \forall_s \ (C(s) \rightarrow (\neg P(s) \vee P(t)))) \qquad (5.32)$$

przy czym truth($\mathcal{Z}$) oznacza wartość logiczną formuły $\mathcal{Z}$ oraz zakładamy, że zbiór krotek T jest zbiorem nierozmytym[9].

5.4.3 Własności operatora "and possibly"

Podstawową charakterystykę operatora "and possibly" stanowi formuła (5.28). Nazwiemy ją *kanoniczną interpretacją* tego operatora. Występujące tam spójniki logiczne można zinterpretować w kontekście logiki rozmytej na wiele różnych sposobów (por. p. 2.2.2). Tak więc formuła (5.28) stanowi raczej pewien szablon, który przybiera różne postacie w zależności od dobranej interpretacji spójników logicznych. Co więcej, operator "and possibly" można scharakteryzować za pomocą formuł innej postaci, które w przypadku logiki klasycznej są równoważne (5.28), ale w przypadku logiki rozmytej ta równoważność nie jest już zachowana. W niniejszym punkcie analizujemy własności operatora "and possibly" biorąc pod uwagę różne, powyżej nakreślone, możliwości jego charakteryzacji.

Interpretacja kanoniczna i różne trójki De Morgana

Zastosowanie kanonicznej formuły (5.28) pozwala wyrazić dla każdej krotki ocenę spełniania przez nią zapytania bipolarnego z operatorem "and possibly" w postaci liczby z przedziału [0,1]. Dzięki temu uzyskuje się naturalne uporządkowanie krotek w odpowiedzi na zapytanie. Jednak interpretując formułę (5.28) z użyciem różnych trójek De Morgana i skojarzonych z nimi operatorów implikacji uzyskuje się nie tylko różne wartości liczbowe stopni spełniania, ale również różne wynikowe uporządkowania krotek.

Tabela 5.1 przedstawia różne interpretacje formuły (5.28) uzyskane dla trzech trójek De Morgana M, Π i W (por. (2.83)), skojarzonych z nimi operatorów implikacji R-implikacji bądź S-implikacji (por. (2.91)-(2.90)) oraz kwantyfikatorów (2.80). Wyrażenie określające wartość logiczną formuły (5.28) dla danej kombinacji trójki De Morgana i operatora implikacji oznaczone jest symbolem γ z odpowiednim indeksem

[9]Wzór (5.32) łatwo uogólnia się na przypadek rozmytego zbioru krotek T, ale w dalszej części książki interesujący będzie dla nas przede wszystkim przypadek nierozmytego zbioru T.

$\gamma_{min,S}$	$\min(C(x), \max(1 - \max_{y \in X} \min(C(y), P(y)), P(x)))$
$\gamma_{min,R}$	$\begin{cases} C(x) & \text{jeśli } \max_y \min(C(y), P(y)) \leq P(x) \\ \min(C(x), P(x)) & \text{wpp} \end{cases}$
$\gamma_{\Pi,S}$	$C(x) \cdot (1 - \exists_\Pi(C(y_i) \cdot P(y_i)) \cdot (1 - P(x)) =$ $C(x) \cdot (\prod_i (1 - C(y_i) \cdot P(y_i)) \cdot (1 - P(x)) + P(x))$
$\gamma_{\Pi,R}$	$\begin{cases} C(x) & \text{jeśli } \exists_\Pi(C(y_i) \cdot P(y_i)) = 0 \\ C(x) \cdot \min(\frac{P(x)}{\exists_\Pi(C(y_i) \cdot P(y_i))}, 1) & \text{wpp} \end{cases}$
γ_W	$t_W(C(x), i_W(\exists_W\ t_W(C(y), P(y)), P(x)))$

Tablica 5.1: Wyrażenia określające wartość logiczną formuły (5.28) dla różnych kombinacji trójek De Morgana i operatora implikacji

dolnym[10]. Na przykład $\gamma_{M,S}$ oznacza wyrażenie uzyskane dla trójki De Morgana M oraz operatora S-implikacji (2.88) skojarzonego z tą trójką.

W tab. 5.2 zestawiono wartości wyrażeń pokazanych w tab. 5.1 dla czteroelementowego zbioru krotek. Przy obliczaniu wartości tych wyrażeń dogodnie jest najpierw obliczyć wartość logiczną formuły $\exists y\ C(y) \wedge P(y)$, określającej stopień konfliktu pomiędzy spełnieniem warunku C i warunku P. Dla poszczególnych kombinacji interpretacji spójników logicznych otrzymuje się następujące wyniki:

$$\exists_M y\ t_{min}(C(y), P(y)) = 0.8 \tag{5.33}$$

$$\exists_\Pi y\ t_\Pi(C(y), P(y)) = 0.96 \tag{5.34}$$

$$\exists_W y\ t_W(C(y), P(y)) = 1.0 \tag{5.35}$$

Formuła $\exists y\ C(y) \wedge P(y)$ bardzo często pojawia się w dalszych rozważaniach, a jej wartość logiczna dla danego zbioru krotek T i wybranej trójki De Morgana jest ustalona. Będziemy ja więc dla skrócenia notacji oznaczać jako $\exists CP$.

Na podstawie tab. 5.2 łatwo zauważyć potwierdzenie faktu, że różne interpretacje spójników logicznych w (5.28) prowadzą do różnych stopni spełnienia poszczególnych krotek (wartości wyrażeń oznaczonych symbolem γ), jak również do różnego wynikowego uporządkowania krotek.

[10] W tym punkcie oznaczamy krotki symbolami $x, y, \ldots$ zamiast $t, s \ldots$ dla uniknięcia konfliktu z odwołaniami do operatorów t-normy i t-konormy.

	C	P	I		II		III		IV		V	
			$<$	$\gamma_{M,S}$	$<$	$\gamma_{M,R}$	$<$	$\gamma_{\Pi,S}$	$<$	$\gamma_{\Pi,R}$	$<$	γ_W
1	0.9	0.7	2	0.7	2	0.7	3	0.64	3	0.63	2	0.6
2	0.8	0.8	1	0.8	1	0.8	2	0.65	2	0.67	2	0.6
3	0.7	1.0	2	0.7	2	0.7	1	0.7	1	0.7	1	0.7
4	1.0	0.0	4	0.2	4	0.0	4	0.04	4	0.0	4	0.0

Tablica 5.2: Porównanie wyników spełnienia zapytania według wzoru (5.28) dla relacji zawierającej cztery krotki, dla różnych interpretacji spójników logicznych; kolumny z nagłówkami C i P zawierają stopnie spełnienia tych warunków przez poszczególne krotki; kolumna z symbolem "$<$" w nagłówku pokazuje miejsce krotki w uporządkowaniu wynikowym.

Analiza charakterystyk różnych interpretacji spójników logicznych pozwala sformułować kilka interesujących własności formuły (5.28).

Własność 1 Jeśli w zbiorze T istnieje krotka x taka, że obydwa warunki C i P są całkowicie spełnione, $C(x) = P(x) = 1$, to dla dowolnego wyboru $t-$normy, $t-$conormy oraz $S-$implikacji lub $R-$implikacji takiej, że $i(1, x) = x$, formuła (5.28) jest równoważna koniunkcji $C(x) \wedge P(x)$, przy czym spójnik $\wedge$ jest interpretowany z użyciem danej t-normy.

Własność 1 wynika wprost z definicji operatorów t-normy, t-konormy, oraz S- i R-implikacji (por. p. 2.2.2). Warto zauważyć, że S-implikacje i R-implikacje skojarzone z trójkami De Morgana M, Π i W spełniają warunek wymieniony we własności 1. Niezależnie od wyboru konkretnych operatorów do modelowania spójników logicznych w ramach rozważanych trójek De Morgana, charakterystyczna cecha modelu "wymagane/postulowane" zostanie zachowana: jeśli istnieje krotka spełniająca obydwa warunki, to warunek zapytania staje się zwykłą koniunkcją.

Własność 2 Jeśli dla danej krotki $x \in T$ zachodzi $P(x) = 1$, to dla dowolnego wyboru t-normy, t-conormy i S-implikacji lub R-implikacji stopień spełnienia zapytania przez krotkę x równy jest $C(x)$.

Własność 2 jest prostą konsekwencją własności operatorów implikacji i t-normy.

W pracy [251] Czytelnik może znaleźć bardziej szczegółowe omówienie własności kanonicznej formuły (5.28).

Trzy formuły charakterystyczne

Na podstawie dotychczasowych rozważań formułę charakteryzującą operator "and possibly" można wyprowadzić na trzy następujące sposoby:

- jako bezpośrednią interpretację oryginalnej formuły Lacroix i Lavency (5.28) w terminach logiki rozmytej;

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge (\exists s \ (C(s) \wedge P(s)) \Rightarrow P(t)) \tag{5.36}$$

- jako bezpośrednią interpretację nierozmytej operacji *winnow* (4.11) w terminach logiki rozmytej i zastosowanie jej z użyciem relacji preferencji (5.31):

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \neg\exists s \ ((C(s) \wedge P(s) \wedge \neg P(t))) \tag{5.37}$$

- jako rozmytą wersję operacji *winnow* przedstawioną w def. 4.4 i zastosowanie jej z użyciem relacji preferencji (5.31):

$$C(t) \textit{ and possibly } P(t) \equiv C(t) \wedge \forall s \ (C(s) \Rightarrow (\neg P(s) \vee P(t))) \tag{5.38}$$

Łatwo stwierdzić, że te trzy sformułowania są faktycznie równoważne na gruncie logiki klasycznej. Poddamy je teraz analizie w ramach logiki wielowartościowej (rozmytej) uwzględniając interpretację spójników logicznych z użyciem różnych operatorów negacji, t-normy, t-konormy oraz implikacji. W szczególności będziemy rozważać ich własności dla różnych trójek De Morgana, podobnie jak w przypadku analizy kanonicznej interpretacji (5.28).

Zaczniemy od obserwacji, że przy pewnym doborze operatorów reprezentujących spójniki logiczne, równoważność formuł (5.36)–(5.38) jest zachowana również w ramach logiki rozmytej. Wyraża to następująca własność.

Własność 3 [254] W przypadku użycia *dystrybutywnej* trójki De Morgana, to jest takiej dla której zachodzi $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$, i skojarzonej z nią S-implikacji, wszystkie trzy formuły (5.36)–(5.38) są równoważne.

Wśród trzech rozważanych trójek De Morgana tylko trójka M jest dystrybutywna. Dla trójki tej i skojarzonej z nią S-implikacji (2.88) wszystkie trzy formuły (5.36)–(5.38) są więc równoważne. Własność 3 określa jedynie warunek dostateczny dla takiej równoważności, ale dla pozostałych dwóch trójek można pokazać przykłady wskazujące, że w ich przypadku równoważność ta nie zachodzi.

Własność 3 pozwala stwierdzić, że w przypadku zastosowania trójki De Morgana M i skojarzonej z nią S-implikacji własność 1 obowiązuje również dla formuł (5.37) i (5.38). Można pokazać, że własność 1 obowiązuje również dla formuły (5.38) w przypadku zastosowania trójki De Morgana M wraz ze skojarzoną R-implikacją [254].

Jednocześnie własność 2 stosuje się również do formuł (5.37) i (5.38) w przypadku zastosowania dowolnej interpretacji spójników logicznych.

Własność 4 Załóżmy, że wartość logiczna $\exists y\ (C(y) \wedge P(y))$ jest dla danego zbioru krotek niezerowa oraz, że krotka x całkowicie spełnia warunek wymagany C, $C(x) = 1$, i zupełnie nie spełnia warunku postulowanego P, $P(x) = 0$. Wtedy wartość logiczna formuły (5.36) obliczona dla krotki x z użyciem trójki De Morgana z operatorem t-normy bez dzielnika zerowego i skojarzoną R-implikacją wynosi 0.

Własność 4 wynika z własności R-implikacji. Konsekwencje tej własności są istotne. Zauważmy najpierw, że w trójkach De Morgana M i Π występują t-normy bez dzielnika zerowego. Stopień spełnienia formuły (5.36) przez krotkę x o podanych cechach będzie mniejszy od stopnia jej spełnienia przez krotkę y, która spełnia obydwa warunki w stopniu ϵ, $C(y) = P(y) = \epsilon$, niezależnie od tego jak małą wartością jest ϵ. Takie zachowanie jest niepożądane, tak więc formuła (5.36) użyta z trójkami De Morgana M lub Π i skojarzoną R-implikacją nie stanowi atrakcyjnego modelu zapytania bipolarnego z operatorem "and possibly".

Podobna własność obowiązuje również dla formuły (5.38).

Property 5 Załóżmy, że w zbiorze krotek występuje przynajmniej jedna krotka y taka, że $C(y) > 0$ i $P(y) = 1$. Wtedy wartość logiczna formuły (5.38) dla krotki x takiej, że $C(t) = 1$ i $P(t) = 0$ obliczona z użyciem trójki De Morgana z operatorem t-normy bez dzielnika zerowego i skojarzoną R-implikacją wynosi 0.

Tak więc również formuła (5.38) użyta z trójkami De Morgana M lub Π i skojarzoną R-implikacją nie stanowi atrakcyjnego modelu zapytania bipolarnego z operatorem "and possibly".

Własności 4 i 5 wskazują, że jeśli formuła (5.36) lub formuła (5.38) mają stanowić model zapytania bipolarnego i spójnik implikacji ma być interpretowany z użyciem R-implikacji, to należy zastosować trójkę De Morgana W.

Trójka De Morgana M z jej skojarzoną S-implikacją posiada również pewne negatywne cechy jako interpretacja zapytania bipolarnego z operatorem "and possibly". Mianowicie, dla formuły (5.36) można wykazać następującą własność, która na mocy własności 3 obowiązuje również dla dwóch pozostałych formuł (5.37)–(5.38).

Własność 6 Formuła (5.36) interpretowana z użyciem trójki De Morgana M i jej skojarzonej S-implikacji może być spełniona w tym samym stopniu przez dwie różne krotki x i y takie, że x ściśle dominuje y w sensie Pareto, $C(x) > C(y)$ i $P(x) > P(y)$.

Wymienione własności formuł (5.36)–(5.38) nie stanowią jeszcze pełnej charakteryzacji semantyki modelu "wymagane/postulowane". Pozwalają one jednak na sformułowanie pewnych ogólnych zaleceń co do doboru jednej spośród tych formuł i jej interpretacji z użyciem odpowiedniej trójki De Morgana wraz ze skojarzonym operatorem implikacji.

Zestawmy ogólne własności obowiązujące w przypadku wszystkich formuł (5.36)–(5.38). Będziemy je traktowali jako własności operatora "and possibly" oznaczanego dalej jako $\wedge_{possibly}$ (warto przypomnieć, że operator ten nie jest ekstensjonalny (ang. *truth-functional*)):

- monotoniczność (ale nie ścisła) względem obydwu argumentów,

- warunki graniczne: $1 \wedge_{possibly} 1 = 1$ i $x \wedge_{possibly} 1 = x$ (własność 2).

Podsumujemy teraz własności poszczególnych formuł (5.36)–(5.38) względem różnych trójek De Morgana.

Formuła (5.36) posiada własność 1 przy dowolnej interpretacji logicznej. Stanowi to z pewnością jej zaletę, gdyż własność ta stanowi fundamentalną cechę rozpatrywanej semantyki "wymagane/postulowane".

Własność 4 sugeruje, że nie należy używać $R-$implikacji w przypadku formuły (5.36). Nie dotyczy to przypadku zastosowania $R-$implikacji wraz z trójką De Morgana W, ale wtedy operator $R-$implikacji jest identyczny z operatorem $S-$implikacji, więc ogólne zalecenie unikania $R-$implikacji można uznać za uzasadnione.

Własność 6 wskazuje, że trójka De Morgana M nie jest zasadniczo dobrym wyborem w przypadku formuły (5.36).

Formuła (5.37) posiada pożądaną własność 1 tylko dla trójki De Morgana M, ale własność 6 w pewnym stopniu dyskwalifikuje tę trójkę.

Formuła (5.38) również posiada własność 1 tylko dla trójki De Morgana M, która nie jest i w tym przypadku najlepszym wyborem ze względu na własności 5 i 6.

Podsumowując, można sformułować następujące rekomendacje. Jeśli własność 1 ma być zachowana, co wydaje się być uzasadnionym postulatem, to najlepszy wybór reprezentacji semantyki zapytań bipolarnych z operatorem "and possibly" oraz jej logicznej reprezentacji stanowią formuła (5.36) i trójki De Morgana Π lub W wraz ze skojarzonym operatorem S-implikacji. Takie rozwiązanie wolne jest od negatywnych własności omawianych wcześniej. Należy przy tym zwrócić uwagę na pewną niedogodność związaną z zastosowaniem trójki De Morgana W. Występujący w niej operator t-normy daje mało zróżnicowane wyniki: dla każdej pary liczb $x, y \in [0, 1]$ takich, że $x + y < 1$ uzyskuje się $x \wedge_W y = 0$. Biorąc to pod uwagę należy uznać trójkę De Morgana Π za najlepszy wybór.

ISSN 0208-8029
ISBN 83-894-7551-0

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK
tel.: (+48) 22 3810246 / 22 3810277 / 22 3810241 / 22 3810273
e-mail: biblioteka@ibspan.waw.pl

ROZMYTOŚĆ I BIPOLARNOŚĆ W INTELIGENTNYM WYSZUKIWANIU INFORMACJI - S. Zadrożny