



**Wydział Matematyki
i Nauk Informatycznych**

POLITECHNIKA WARSZAWSKA

Systemy ekspertowe

Biuro podróży i turystyka

Laura Korona
Karol Denst
Jakub Borek
Stanisław Durka

10 maja 2024

Spis treści

1	Opis projektu	2
2	Baza wiedzy	2
2.1	Parametry podlegające rozmyciu	2
2.2	Pozostałe parametry	2
2.3	Decyzje	3
3	Sposób działania systemu	3
3.1	Zbiory przybliżone	3
3.2	Zbiory rozmyte	4
3.3	Sposoby wnioskowania	4
3.4	Rozmywanie	5
3.4.1	Funkcja przynależności trapezoidalna	5
3.4.2	Funkcja przynależności trójkątna	6
3.5	Wyostrezanie	6
3.6	Generowanie reguł	7
3.7	Zbiory przybliżone	7
3.7.1	Dolne przybliżenie	7
3.7.2	Górne przybliżenie	7
3.8	Reguły minimalne	7
4	Bibliografia	10

1 Opis projektu

Tematem projektu jest doradczy system ekspertowy związany z turystyką, który może być wykorzystywany przez biura podróży. Jego działanie będzie polegać na pobieraniu od klienta jego preferencji dotyczących wakacji, takich jak czas trwania i zakres cenowy, i sugerowanie na ich podstawie najlepiej dopasowanych wycieczek znajdujących się w bazie danych. Finalny system będzie składał się z następujących modułów:

1. Baza danych i baza wiedzy - zbiór wszystkich ofert biura podróży zawierający informacje o dostępnych wycieczkach oraz zbiór reguł umożliwiający klasyfikację ofert do poszczególnych kategorii.
2. Moduł wnioskowania - moduł odpowiedzialny za przeprowadzenie wnioskowania i przedstawienie propozycji wycieczek na podstawie danych wprowadzonych do systemu przez użytkownika. Moduł ten będzie wykorzystywał zarówno klasyczną logikę boolowską jak i logikę rozmytą.
3. Interfejs użytkownika - tekstowy interfejs umożliwiający klientowi interakcję z systemem. Aby zapewnić jak największy komfort korzystania z programu, interfejs oraz zadawane pytania powinny być jak najprostsze i konkretne. W interfejsie pojawią się zarówno opcje jednokrotnego, jak i wielokrotnego wyboru.

2 Baza wiedzy

Projektowany system będzie proponował klientom wycieczki na podstawie dopasowania dostępnych ofert do ich preferencji. Implementacja systemu obejmie zarówno parametry boolowskie jak i rozmyte.

2.1 Parametry podlegające rozmyciu

- Czas trwania wycieczki - opisany za pomocą następujących wartości: *bardzo krótki, krótki, średni, długi*.
- Cena - opisana za pomocą następujących wartości: *bardzo niska, niska, umiarkowana, wysoka, bardzo wysoka*

2.2 Pozostałe parametry

- Kraj/region - w systemie dostępna będzie większość krajów europejskich.
- Sposób transportu - *samolot, pociąg, statek, autobus*
- Standard noclegu - *textit*zadowalający, dobry, wysoki, luksusowy
- Wyżywienie - *w cenie, we własnym zakresie, all inclusive*
- Typ - *góry, morze, miasto*

- Przyjazny dla dzieci - *tak, nie*
- Przyjazny dla zwierząt - *tak, nie*
- Odległość od Polski - Wartość numeryczna podana w kilometrach
- Stopień zatłoczenia turystycznego - *niski, umiarkowany, wysoki, bardzo wysoki*
- Liczba sklepów z pamiątkami na kilometr kwadratowy - wartość numeryczna.

2.3 Decyzje

Decyzjami systemu będą dostępne w ofercie biura wycieczki. Przykładowe oferty w bazie to:

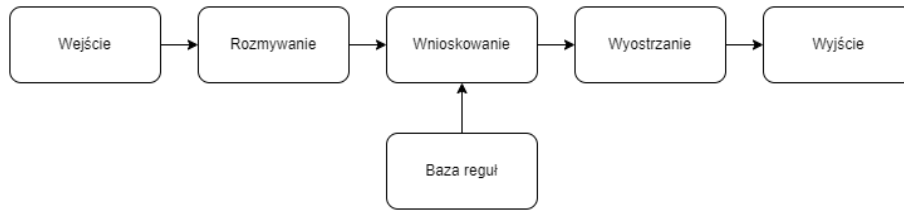
- Santorini, Grecja - *cena bardzo wysoka, długość wycieczki średnia, standard luksusowy, klimat gorący, samolot, morze, nieprzyjazny dla zwierząt*
- Sztokholm, Szwecja - *cena umiarkowana, długość wycieczki krótka, standard dobry, klimat chłodny, samolot, przyjazny dla dzieci, zatłoczenie niskie*
- Zakopane, Polska - *cena niska, długość wycieczki bardzo krótka, standard zadowalający, klimat umiarkowany, autobus, góry, przyjazny dla dzieci, przyjazny dla zwierząt, zatłoczenie wysokie*
- Split, Chorwacja - *cena umiarkowana, bardzo długa, standard zadowalający, klimat gorący, samolot, morze*

3 Sposób działania systemu

Działanie systemu będzie się opierać na kilku różnych metodach reprezentacji preferencji klientów i wnioskowania na ich podstawie - wykorzystywane będą między innymi zbiory przybliżone i rozmyte oraz dwa rodzaje pytań - zamknięte i otwarte.

3.1 Zbiory przybliżone

Wykorzystywane będą zbiory przybliżone w celu obsłużenia przypadków, gdy klient nie podaje preferencji na dany temat. Korzystają one z logiki trójwartościowej, dzięki czemu można reprezentować brak odpowiedzi klienta na dane pytanie. Zbiory przybliżone definiowane są za pomocą dwóch pojęć - górnego i dolnego przybliżenia. Przybliżenie dolne zbioru tworzą elementy, które na pewno należą do tego zbioru, zaś przybliżeniem górnym - elementy, których nie można wykluczyć, według obecnej wiedzy, jako elementy zbioru. Różnicę dolnego i górnego przybliżenia zbioru nazywa się obszarem brzegowym zbioru - są to elementy, dla których nie mamy pewności, czy należą do zbioru, czy nie.



Rysunek 1: Proces wnioskowania rozmytego

3.2 Zbiory rozmyte

Teoria zbiorów rozmytych jest rozszerzeniem klasycznej teorii zbiorów. Pojęcie zbioru rozmytego oznacza wprowadzenie funkcji przynależności do danego zbioru.

Niech X będzie uniwersum, czyli zbiorem wszystkich możliwych wartości dla danej cechy lub zmiennej w kontekście danego problemu. Zbiór rozmyty A w przestrzeni X to para uporządkowana (X, μ_A) , gdzie $\mu_A : X \rightarrow [0, 1]$ jest funkcją przynależności, która określa stopień przynależności każdego elementu x zbioru X do zbioru rozmytego A . Funkcję μ_A charakteryzują następujące warunki:

1. $\mu_A(x)$ reprezentuje stopień przynależności elementu x do zbioru A i może przyjmować wartości z przedziału $[0, 1]$.
2. Dla każdego $x \in X$, $\mu_A(x)$ odpowiada stopniowi przynależności elementu x do zbioru A .
3. $\mu_A(x) = 0$ oznacza, że element x nie należy do zbioru A , $\mu_A(x) = 1$ oznacza, że element x w pełni należy do zbioru A , a wartości pośrednie odpowiadają stopniom przynależności między tymi dwoma skrajnościami.

Zbiory rozmyte mogą być wykorzystywane w systemach ekspertowych do modelowania nieprecyzyjnych czy niejednoznacznych danych oraz do podejmowania decyzji na podstawie tych danych. W systemach ekspertowych funkcje przynależności mogą być definiowane na podstawie eksperckiej wiedzy lub danych empirycznych, co pozwala na elastyczne reprezentowanie złożonych zależności i podejmowanie wniosków w warunkach niepewności.

Proces wnioskowanie rozmytego został przedstawiony na 1.

3.3 Sposoby wnioskowania

Moduł wnioskowania w naszym systemie będzie wykorzystywał dwa rodzaje logiki – logikę boolowską oraz logikę rozmytą. W przypadku logiki boolowskiej system będzie odpowiadał zero-jedynkowo na pytanie, czy dana wycieczka pasuje do preferencji klienta. Na przykład, jeśli klient zaznaczył, że chciałby wyjechać do Hiszpanii, to system zwróci wartość 1 dla wszystkich wycieczek do tego państwa oraz 0 dla wszystkich innych. Logika rozmyta zostanie zaś użyta w przypadkach, gdy nie da się w ten sposób wyznaczyć odpowiedzi. Przykładowo, jeśli klient

zaznaczy, że chciałby spędzić wakacje w ciepłym miejscu, nie da się powiedzieć, czy dana wycieczka będzie się odbywać gdzieś, gdzie jest ciepło, czy nie. Można tylko powiedzieć, że jest bardziej ciepło lub mniej ciepło. Dlatego wykorzystana będzie logika rozmyta, która zwraca wartości od 0 do 1, gdzie 1 oznacza perfekcyjne dopasowanie do preferencji klienta, a 0 kompletny brak dopasowania. Innymi słowy, zadaniem bloku wnioskowania jest określenie stopnia przynależności odpowiedzi systemu ekspertowego do odpowiednich zbiorów reprezentujących wartości lingwistyczne. W tym celu stosowane są implikacje rozmyte, na przykład:

- Implikacja Łukasiewicza

$$I(x, y) = \min(1 - x + y, 1)$$

- Implikacja Reichenbacha

$$I(x, y) = 1 - x + xy$$

Wnioskowanie będzie odbywać się na podstawie bazy wiedzy, której każda wycieczka będzie miała opisane swoje cechy (zarazem rozmyte jak i nierozmyte).

3.4 Rozmywanie

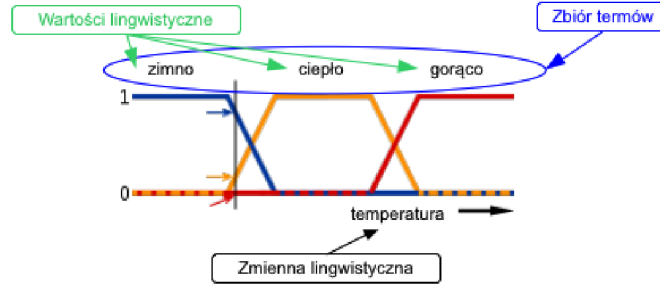
Proces rozmywania ma na celu zamianę wartości liczbowych na stopniowane wartości lingwistyczne (jakościowe). Przyporządkowanie jest wynikiem funkcji przynależności zdefiniowanej w rozdziale 3.2.

3.4.1 Funkcja przynależności trapezoidalna

Trapezoidalna funkcja przynależności dla wyznaczonego odcinka $[m, n]$ opisana jest wzorem

$$\mu(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{m-a} & \text{if } x \in [a, m] \\ 1 & \text{if } x \in [m, n] \\ \frac{b-x}{b-n} & \text{if } x \in [n, b] \\ 0 & \text{if } x > b \end{cases}$$

Ta funkcja przynależności może zostać użyta do przyporządkowania wartościom liczbowym odpowiednim grupom wartości lingwistycznych opisujących zakresy wartości, na przykład grupując temperaturę lokacji wycieczki na kategorie, jak widać na Rysunku 2.



Rysunek 2: Funkcje przynależności trapezoidalne dla stopniowania temperatury – "zimno", "ciepło", "gorąco"

3.4.2 Funkcja przynależności trójkątna

Trójkątną funkcję przynależności stosuje się do określenia jak blisko konkretnej oczekiwanej wartości znajduje się punkt w zbiorze:

$$\mu(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{c-a} & \text{if } x \in [a, c] \\ \frac{b-x}{b-c} & \text{if } x \in [c, b] \\ 0 & \text{if } x > b \end{cases}$$

3.5 Wyostrzanie

Wyostrzanie zostanie zastosowane w celu obliczenia wartości liczbowej na podstawie funkcji zdaniowej uzyskanej podczas wnioskowania. Używane będą dwie popularne metody:

- metoda środka ciężkości

$$y^{COG} = \frac{\sum_{i=1}^n y_i \mu(y_i)}{\sum_{i=1}^n \mu(y_i)},$$

gdzie $\mu(y_i)$ oznacza stopień przynależności elementu y_i do rozważanego zbioru

- metoda średniego maksimum

$$y^{MOM} = \frac{1}{|M|} \sum_{y_i \in M} y_i,$$

gdzie M jest zbiorem maksimumów.

3.6 Generowanie reguł

System decyzyjny to system informacyjny przydzielający obiekty do pewnych klas określonych za pomocą jednego z atrybutów, zwanego *atrybutem decyzyjnym*, zależnym od tzw. *atrybutów warunkowych*. Wnioskując ze zbioru krotek w bazie danych, celem jest stworzenie rozłącznych klas abstrakcji zbioru obiektów U , opartych na relacji równoważnościowej - tzw. relacji *nierozróżnialności*. Klasa abstrakcji dla obiektu $x \in U$ relacji $IND(B)$ wygląda następująco:

$$[x]_{IND(B)} = \{y \in U, \quad \forall a \in B, (a(x) = a(y))\}$$

3.7 Zbiory przybliżone

Szukając podzbioru atrybutów systemu informacyjnego determinującego podział obiektów na rozłączne klasy abstrakcji, można natknąć się na niespójności w danych. Do radzenia sobie z takimi przypadkami przydaje się teoria zbiorów przybliżonych, która wprowadza pojęcia górnego oraz dolnego przybliżenia zbioru.

3.7.1 Dolne przybliżenie

Dolne przybliżenie $\underline{B}X$ to zbiór elementów, które w całości zawierają się w zbiorze X . Jest to rodzaj minimalnego zbioru elementów, które *na pewno* należą do zbioru X . Dolne przybliżenie służy do reprezentowania najbardziej pewnych lub istotnych elementów, na podstawie dostępnych danych.

$$\underline{B}X = \{Y \in IND(B) : Y \subseteq X\}$$

3.7.2 Górne przybliżenie

Górne przybliżenie definiuje się jako zbiór elementów z $IND(B)$, które wystarczy, że mają część wspólną ze zbiorem X . Jest to maksymalny zbiór, o którego elementach mówimy, że *być może* należą do danej klasy decyzyjnej.

$$\widetilde{B}X = \{Y \in IND(B) : Y \cap X \neq \emptyset\}$$

3.8 Reguły minimalne

Do generowania reguł minimalnych na podstawie tablicy decyzyjnej zostanie wykorzystany algorytm wyglądający następująco:

1. Doprowadź tablicę decyzyjną do spójności
2. Usuń identyczne obiekty
3. Utwórz macierz nierozróżnialności
4. Dla każdej wartości atrybutu decyzyjnego:
 - (a) Utwórz uogólnioną macierz nierozróżnialności względem decyzji

- (b) Zapisz funkcję nierozróżnialności, zminimalizuj ją
- (c) Zapisz regułę decyzyjną na podstawie zminimalizowanej funkcji rozróżnialności.

Każdy z tych kroków jest opisany w dalszej części tej sekcji.

Doprowadź tablicę decyzyjną do spójności

Ten krok polega rozwiązaniu sprzeczności w danych, tak aby każdy zestaw atrybutów warunkowych prowadził do jednoznacznie określonego wyniku decyzyjnego. Wykorzystana będzie w tym celu *metoda uogólnionego atrybutu decyzyjnego*.

U	a	b	c	d
1	0	0	0	0
2	0	1	0	1
3	0	0	1	0
4	1	1	0	1
5	1	1	1	1
6	1	1	1	0
7	0	0	1	2
8	1	0	1	2

Tabela 1: Przykładowa tablica decyzyjna

U	a	b	c	δ_a
1	0	0	0	{0}
2	0	1	0	{1}
3	0	0	1	{0,2}
4	1	1	0	{1}
5	1	1	1	{0,1}
6	1	1	1	{0,1}
7	0	0	1	{0,2}
8	1	0	1	{2}

Tabela 2: Doprowadzenie do spójności poprzez wprowadzenie uogólnionego atrybutu decyzyjnego δ_a

Usuń identyczne obiekty

Celem tego etapu jest redukcja zbioru danych poprzez wyeliminowanie duplikatów (duplikaty tworzą zbiory obiektów, które dla każdego atrybutu mają takie same wartości, np. w powyższym przykładzie obiekty 3. i 7. tworzą parę duplikatów). Ułatwia to dalszą analizę i zmniejsza obciążenie obliczeniowe.

U	a	b	c	δ_a
1	0	0	0	{0}
2	0	1	0	{1}
3	0	0	1	{0,2}
4	1	1	0	{1}
6	1	1	1	{0,1}
8	1	0	1	{2}

Tabela 3: Usunięcie identycznych obiektów z tabelki 2

Utwórz macierz nierozróżnialności

Dla danego systemu informacyjnego $SI = (U, A)$, gdzie $U = \{x_1, x_2, \dots, x_n\}$ jest zbiorem obiektów, zaś A zbiorem atrybutów, oraz podzbioru atrybutów $B \subseteq$

A macierz nierozróżnialności $M(SI) = [c_{ij}]_{n \times n}$ definiowana jest następująco:

$$c_{ij} = \{a \in A : a(x_i) \neq a(x_j), i = 1, 2, \dots, n\}.$$

Innymi słowy, każdy element macierzy c_{ij} jest zbiorów atrybutów rozróżniających i -ty i j -ty obiekt z U .

	1	2	3	4	6	8
1	\emptyset	b	c	ab	abc	ac
2	b	\emptyset	bc	a	ac	abc
3	c	bc	\emptyset	abc	ab	a
4	ab	a	abc	\emptyset	c	bc
6	abc	ac	ab	c	\emptyset	b
8	ac	abc	a	bc	b	\emptyset

Tabela 4: Macierz nierozróżnialności dla tabelki 3

Utwórz uogólnioną macierz nierozróżnialności względem decyzji

Dla każdej wartości atrybutu decyzyjnego tworzona jest macierz, która bierze pod uwagę tylko te atrybuty, które wpływają na daną decyzję. To pozwala na bardziej szczegółową analizę w kontekście każdej z możliwych decyzji.

	1	2	3	4	6	8
1	\emptyset	b	c	ab	abc	ac

Tabela 5: Uogólniona macierz nierozróżnialności dla reguły 1

Zapisz funkcję nierozróżnialności, zminimalizuj ją

Funkcja nierozróżnialności tworzona jest na podstawie uogólnionej macierzy nierozróżnialności, a następnie jest minimalizowana w celu znalezienia najprostszych, ale jednocześnie kompletnych reguł, które opisują dane.

Funkcja nierozróżnialności i jej zminimalizowana forma dla reguły 1:

$$f(a, b, c) = bc(a + b)(a + b + c)(a + c) = bc \quad (1)$$

Zapisz regułę decyzyjną na podstawie zminimalizowanej funkcji rozróżnialności.

Na końcu, na podstawie zminimalizowanej funkcji nierozróżnialności, formułowane są reguły decyzyjne, które mogą być stosowane do klasyfikacji nowych danych.

Reguła decyzyjna na podstawie funkcji rozróżnialności dla reguły 1:

$$b = 0 \wedge c = 0 \implies \delta_a = 0. \quad (2)$$

4 Bibliografia

- [1] http://bcpw.bg.pw.edu.pl/Content/2031/zbiory_przyblizone_nowa_mat._metoda_analizy_danych.pdf