

Rozmyte zapytania do Baz Danych

Krzysztof Dembczyński
Instytut Informatyki
Krzysztof.Dembczynski@cs.put.poznan.pl

Plan wykładu

- Wstęp
- Logika rozmyta w Bazach Danych
- Rozmyty Relacyjny Model Danych (RRMD)
- Operacje w Rozmytym Relacyjnym Modelu Danych
- Fuzzy SQL (SQLf)
- Rozszerzenie funkcjonalności RRMD
- Implementacja
- Projekt

Wstęp

- Tradycyjne bazy danych przechowują informację precyzyjną
- W świecie rzeczywistym częściej operujemy informacją nieprecyzyjną i niepewną
- Pozyskując wiedzę z danych precyzyjnych może nam zależeć na rozmytych odpowiedziach

Wstęp – informacja rozmyta

- Nieprecyzyjność w danych rzeczywistych
Dane podawane przez czujnik mogą być reprezentowane jako pewien rozkład
- Subiektywny osąd
Ocena filmu – pewnym osobom film może się bardzo podobać, innym wręcz przeciwnie
- Informacja, którą jesteśmy zainteresowani może nie być precyzyjna
Poszukujemy produkt nie droższy od 100 PLN, który spełnia wymogi bezpieczeństwa na poziomie lepszym niż 3 – interesującą odpowiedzią byłoby uzyskanie informacji o produkcie trochę droższym, ale o wyższym poziomie bezpieczeństwa np.cena - 105 PLN, poziom bezpieczeństwa: 5.

Logika rozmyta w Bazach Danych

- Przechowywanie, uaktualnianie i przetwarzanie informacji nieprecyzyjnej z natury
 - Nieprecyzyjne wartości atrybutów
 - Podobieństwo
 - Możliwość
 - Częściowe zawieranie się krotek w relacji – relacja rozmyta
- Przetwarzanie nieprecyzyjnych (rozmytych) zapytań

Logika rozmyta w Bazach Danych – nieprecyzyjne wartości atrybutów

- Podobieństwo – dotyczy atrybutów, których dziedziny są dyskretne, określa podobieństwo między wszystkimi wartościami atrybutu (np. podobieństwo pomiędzy zawodami)
- Możliwość – jest bardziej ogólna, może dotyczyć zarówno atrybutów ciągłych jak i dyskretnych, reprezentuje rozkład możliwości dla wartości atrybutów

Logika rozmyta w Bazach Danych – relacja rozmyta

- W klasycznych bazach danych krotka albo występuje w danej relacji albo nie
- W relacjach rozmytych dana krotka może występować w jakimś stopniu – występowanie częściowe, tzw. krotki ważone (ang. Weighted Tuples)

Logika rozmyta w Bazach Danych – rozmyte zapytania

- Zapytanie do Bazy Danych jest nieprecyzyjne (rozmyte) jeśli spełnia poniższe warunki:
 - Nieprecyzyjne (rozmyte) warunki
 - Znajdź wszystkich podatników, których zarobki były mniejsze od 30 000 PLN.
 - Znajdź wszystkich podatników, których zarobki były *małe*.
 - Nieprecyzyjne (rozmyte) operatory
 - Znajdź wszystkie kraje, których trzy importowane dobra były takie same jak trzy eksportowane dobra przez Polskę
 - Znajdź wszystkie kraje, których trzy importowane dobra były *prawie* takie same jak trzy eksportowane dobra przez Polskę
 - Nieprecyzyjne (rozmyte) kwantyfikatory
 - Znajdź firmy których wszyscy klienci są urzędnikami państwowymi.
 - Znajdź firmy których *prawie* wszyscy klienci są urzędnikami państwowymi.

Logika rozmyta w Bazach Danych – redundancja i zależność funkcyjna

- Redundancja i zależność funkcyjna w klasycznych Bazach Danych
- Kiedy dwie krotki w relacji rozmytej są takie same
- Jak wyznaczyć zależność funkcyjną w przypadku relacji rozmytych

Rozmyty Relacyjny Model Danych

- Atrybuty i krotki
- Dziedzina atrybutów może być dyskretna lub ciągła
- Relacja R jest podzbiorem iloczynu kartezjańskiego wszystkich kombinacji wartości na atrybutach relacji R .

Imię	Nazwisko	Wiek	Typ pracy	Wysztalczenie
Bob	McLendon	23	Akademicki	Sztuczna Inteligencja
Rob	Mucker	27	Przemysłowy	Systemy Ekspertowe
Nancy	McCay	40	Zarząd	Statystyka
John	Hunt	55	Zarząd	Robotyka

Pracodawca	Typ pracy	Wysztalczenie	Lokalizacja
Texas A&M	Akademicki	Sztuczna Inteligencja	Texas
SUN Microsystems	Przemysłowy	Systemy ekspertowe	California
Harvard University	Akademicki	Systemy ekspertowe	Boston
NASA Ames	Zarząd	Robotyka	California

Rozmyty Relacyjny Model Danych – podobieństwo (1)

- Wykorzystuje się zmienne lingwistyczne do opisu wartości atrybutu (np. nazwy zawodów)
- Nieprecyzyjność zmiennych jest przedstawiona za pomocą macierzy podobieństwa

	Robotyka	Systemy ekspertowe	Sztuczna inteligencja	Statystyka
Robotyka	1.0	0.6	0.6	0.2
Systemy ekspertowe	0.6	1.0	0.9	0.2
Sztuczna Inteligencja	0.6	0.9	1.0	0.2
Statystyka	0.2	0.2	0.2	1.0

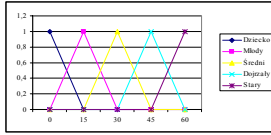
Logika rozmyta w Bazach Danych – podobieństwo (2)

- Relacja podobieństwa S na dziedzinie D jest relacją rozmytą, która przypisuje każdej parze wartości dziedziny D wartość z przedziału $[0, 1]$, tzn. $S: D \times D \rightarrow [0, 1]$. Relacja S spełnia trzy poniższe własności dla każdego x, y, z z należącego do D :
 - zwrotność: $S_D(x, x) = 1$
 - symetryczność: $S_D(x, y) = S_D(y, x)$
 - przechodność: $S_D(x, z) \geq \max_{y \in D} (\min(S_D(x, y), S_D(y, z)))$

Logika rozmyta w Bazach Danych - możliwość

- Wykorzystuje się rozkład możliwości do opisu wartości atrybutu
- Nieprecyzyjność może być modelowana jako różnego rodzaju rozkłady możliwości

Zmienna lingwistyczna	Wartość
Dziecko	0
Młody	15
Średni	30
Dojrzały	45
Stary	60



Operacje w Rozmytym Relacyjnym Modelu Danych (1)

- Rozmyta algebra relacyjna

- Suma

$$\mu_{R \cup S}(t) = \mu_R(t) \oplus \mu_S(t)$$

- Różnica

$$\mu_{R \setminus S}(t) = \mu_R(t) \oplus (1 - \mu_S(t))$$

- Iloczyn kartezjański

$$\mu_{R \cdot S}(t_1, t_2) = \mu_R(t_1) \otimes \mu_S(t_2)$$

Operacje w Rozmytym Relacyjnym Modelu Danych (2)

- Rozmyta algebra relacyjna

- Projekcja
 - Ograniczenie za pomocą progu przynależności do relacji
 - Wybór n krotek o największej przynależności
 - Wyświetlenie wszystkich krotek z niezerową przynależnością
- Selekcja
 - Ograniczenie za pomocą progu przynależności do relacji
 - Wybór n krotek o największej przynależności
 - Wyświetlenie wszystkich krotek z niezerową przynależnością
- Połączenie
 - Poprzez atrybuty dyskretne oraz ciągłe

$$\mu_{R \bowtie S}(t_1, t_2) = \min(\mu_R(t_1), \mu_S(t_2), \mu_0(R.a(t_1), S.b(t_2)))$$

gdzie R, S są relacjami rozmytymi, a, b są atrybutami połączeniowymi, a, t_1 jest krotką łączącą t_1 i t_2

Operacje w rozmytym relacyjnym modelu danych (3)

- Redundancja

Niech t_1 i t_2 będą dwoma krotkami w relacji o atrybutach a_1, a_2, \dots, a_n . Stopień równości pomiędzy dwoma rozmytymi krotkami t_1 i t_2 , oznaczany $Equal(t_1, t_2)$, jest rozmytą koniunkcją rozmytych miar równości pomiędzy odpowiednimi wartościami atrybutów, tzn:

$$Equal(t_1, t_2) = \bigotimes_{i=1}^k Eq(a_i(t_1), a_i(t_2))$$

Gdzie Eq jest odpowiednio miarą podobieństwa lub możliwości pomiędzy i -tymi wartościami atrybutu.

Fuzzy SQL (1)

```
SELECT 0.6 Imie, Nazwisko, Wykształcenie
FROM Pracownicy
WHERE Wykształcenie = 'Systemy Ekspertowe'
```

Imię	Nazwisko	Wykształcenie	Podobieństwo
Bob	McLedon	Sztuczna Inteligencja	0.9
Rob	Mucker	Systemy Ekspertowe	1.0
John	Hunt	Robotyka	0.6

Fuzzy SQL (2)

```
Select 0.4 Imie, Nazwisko, Wiek
From Pracownicy
Where Wiek = 'Średni'
```

Imię	Nazwisko	Wiek	Możliwość
Bob	McLedon	23	0.8
Rob	Mucker	27	0.53

Fuzzy SQL (3)

```
Select 2 Imię, Nazwisko, Wykształcenie
From Pracownicy
Where Wykształcenie = 'Systemy Ekspertowe'
```

Imię	Nazwisko	Wykształcenie	Podobieństwo
Rob	Mucker	Systemy Ekspertowe	1.0
Bob	McLendon	Sztuczna Inteligencja	0.9

Fuzzy SQL (4)

```
SELECT 0.7 Nazwisko, Pracownik.Wykształcenie,
Praca.Wykształcenie, Pracodawca, Lokalizacja, Praca.TypPracy
FROM Pracownicy
WHERE Pracownicy.TypPracy = Praca.TypPracy
and Pracownicy.Wykształcenie = Praca.Wykształcenie
```

Nazwisko	Pracownik. Wykształcenie	Praca. Wykształcenie	Pracodawca	Lokalizacja	Praca. TypPracy	Podobieństwo
McCay	Statystyka	Statystyka	NIH	DC	Zarząd	1.0
McLendon	Sztuczna Inteligencja	Sztuczna Inteligencja	Texas A&M	Texas	Akademicki	1.0
McLendon	Sztuczna Inteligencja	Systemy Ekspertowe	Harvard University	Boston	Akademicki	0.9
Hunt	Robotyka	Robotyka	NASA	California	Zarząd	1.0

Rozszerzenie funkcjonalności RRMD (1)

- Inne możliwości zapytań:
SELECT 0.7 Imię, Nazwisko, #Wykształcenie
FROM Pracownicy

Imię	Nazwisko	Wykształcenie	Podobieństwo
Rob	Mucker	Systemy Ekspertowe	1.0
Rob	Mucker	Sztuczna Inteligencja	0.9
Bob	McLendon	Sztuczna Inteligencja	1.0
Bob	McLendon	Systemy Ekspertowe	0.9
John	Hunt	Robotyka	1.0
Nancy	McCay	Statystyka	1.0

Rozszerzenie funkcjonalności RRMD (2)

- Rozmyte funkcje agregujące
- Rozmyte grupowanie
SELECT 0.6 count(*)
FROM Pracownicy
GROUP BY #Wykształcenie

Wykształcenie	Count(*)
Systemy Ekspertowe	3
Sztuczna Inteligencja	3
Robotyka	3
Statystyka	1

Rozszerzenie funkcjonalności RRMD (3)

- Różne reprezentacje rozkładu możliwości
- Możliwość definiowania relacji rozmytych typu: około, mniejsze lub równe
- Możliwość zastosowania algorytmów generujących reguły asocjacyjne
- Możliwość modelowania preferencji

Implementacja

- Wszystko od początku
 - Najtrudniejsze ale można najbardziej zoptymalizować
- Klasyczne zapytania do Bazy Danych i przetwarzanie odpowiedzi
 - Prędkość działania, przetwarzanie na serwerze czy u klienta
- Translacja zapytań z Fuzzy SQL do SQL
 - Czy jest w ogóle możliwa?

Projekt

- Rok temu powstały dwa konkurencyjne projekty, których idee należałoby powiązać ze sobą oraz dalej rozwinąć
- Lepiej zrobić jeden bardzo dobry projekt, niż 10 mało znaczących i mało rozwijających
- Kontakt
Krzysztof.Dembczynski@cs.put.poznan.pl

Podziękowania

- Dla mgr inż. Ani Pankowskiej za pomoc w przygotowaniu wykładu oraz wspólną realizację projektu
- Dr inż. Maciejowi Hapke za zaproszenie i umożliwienie mi wygłoszenia wykładu