

ŽILINSKÁ UNIVERZITA V ŽILINE
FAKULTA RIADENIA A INFORMATIKY

DIPLOMOVÁ PRÁCA

Študijný odbor: **Informačné systémy - Spracovanie dát**

Olga Chovancová

Nástroj pre fuzzifikáciu numerických hodnôt

Vedúci: **Ing. Miroslav Kvaššay, PhD.**

Reg.č. 6/2016

Máj 2017

ZADANIE TÉMY DIPLOMOVEJ PRÁCE.

Študijný program : Informačné systémy

Zameranie: Spracovanie dát

Meno a priezvisko

Ol'ga Chovancová

Osobné číslo

556217

Názov práce v slovenskom aj anglickom jazyku

Nástroj pre fuzzifikáciu numerických hodnôt

Tool for fuzzification of numerical values

Zadanie úlohy, ciele, pokyny pre vypracovanie

(Ak je málo miesta, použite opačnú stranu)

Cieľ diplomovej práce:

Cieľom diplomovej práce je experimentálne porovnať algoritmy, ktoré slúžia pre fuzzifikáciu numerických hodnôt.

Obsah:

1. Oboznámenie sa s problematikou fuzzifikácie (transformácie numerických hodnôt na lingvistické).
2. Rozbor existujúcich algoritmov pre fuzzifikáciu numerických hodnôt.
3. Implementácia vybraných algoritmov fuzzifikácie v jazyku C++.
4. Experimentálne porovnanie implementovaných algoritmov na rôznych výstupných dátach.

Meno a pracovisko vedúceho DP: Ing. Miroslav Kvaššay, PhD., KI, ŽU

Meno a pracovisko tútora DP:

vedúci katedry
(dátum a podpis)

Čestné prehlásenie

Prehlasujem, že som diplomovú prácu *Nástoj pre fuzzifikáciu numerických hodnôt* vypracovala samostatne pod vedením ... , a uviedla v nej všetky použité literárne a iné odborné zdroje v súlade s právnymi predpismi, vnútornými predpismi Žilinskej univerzity a vnútornými aktmi riadenia Žilinskej univerzity a Fakulty riadenia a informatiky.

V Žiline, dňa XX.05.2017

Oľga Chovancová

Poďakovanie

Na tomto mieste by som chcela poďakovať vedúcemu diplomovej práce.... za cenné pripomienky a odborné rady, ktorými prispel k vypracovaniu tejto diplomovej práce. Taktiež dakujem môjmu Zároveň ďakujem mojej rodine a priateľom za ich nekonečnú podporu a trpezlivosť.

V Žiline, dňa XX.05.2017

Oľga Chovancová

Abstrakt

CHOVANCOVÁ OĽGA: *Nástroj pre fuzzifikáciu numerických hodnôt* [Diplomová práca]

Žilinská Univerzita v Žiline, Fakulta riadenia a informatiky, Katedra informatiky.

Vedúci: Ing. Miroslav Kvaššay, PhD.

Stupeň odbornej kvalifikácie: Inžinier** Informatiky

Cieľom diplomovej práce je experimentálne porovnať algoritmy, ktoré slúžia pre fuzzifikáciu numerických hodnôt.

1. Oboznámenie sa s problematikou fuzzifikácie (transformácie numerických hodnôt na lingvistické). 2. Rozbor existujúcich algoritmov pre fuzzifikáciu numerických hodnôt. 3. Implementácia vybraných algoritmov fuzzifikácie v jazyku C++. 4. Experimentálne porovnanie implementovaných algoritmov na rôznych výstupných dátach.

Kľúčové slová: fuzzy, TODO

Abstract

CHOVANCOVÁ OĽGA: *Tool for fuzzification of numerical values* [Diploma thesis]

University of Žilina, Faculty of Management Science and Informatics, Department of Informatics

Tutor: Ing. Miroslav Kvaššay, PhD.

Qualification level: Masters of Informatics

The aim of the thesis is to compare experimental algorithms that are used for Fuzzification numerical values.

TODO

1. This introduction to the Fuzzification (numeric values to transform linguistic). 2. Analysis of the existing algorithms for fuzzification numerical values. 3. Implementation of selected algorithms Fuzzification in C ++. 4. Experimental comparison algorithms implemented on different output data. Key words: fuzzy, entropy, TODO

Obsah

Úvod	14
1 Analýza súčasného stavu	15
2 Cieľ práce	16
3 Teoretické východiská práce	17
3.1 Získavanie znalostí z databáz	17
3.2 Fuzzy prístupy	17
3.2.1 Teória fuzzy množín	18
3.3 Fuzzifikácia	18
3.4 Fuzzy logika	18
3.5 Meranie Entropie	18
3.5.1 Shannonova Entropia	19
3.6 TODO INE	19
3.7 Záver	19
4 Analýza existujúcich algoritmov pre fuzzifikáciu numerických hodnôt	20
4.1 Fuzzy klasifikátor s možnosťou výberu na základe fuzzy entropie	20
4.1.1 Opis upravenej fuzzy entropie	20
4.1.2 Algoritmus FEBFC	21
4.2 Minimum description length partition	21
4.2.1 Algoritmus MDLP	21

4.3	Záver	21
5	Implementácia algoritmov pre fuzzifikáciu numerických hodnôt	22
5.1	Implementácia Algoritmu FEBFC	22
5.2	Modifikácia Algoritmu FEBFC	22
5.3	TODO = DRUHY algoritmus	22
5.4	TODO INE	22
5.5	Záver	22
6	Experimentálny výskum	23
6.1	Parametre, určujúce kvalitu fuzzifikácie	23
6.2	Priebeh experimentnov - použité dátové množiny	23
6.3	Výsledky a vyhodnotenie experimentov pre vybrané dátové množiny . . .	23
6.4	Zhrnutie výsledkov	23
6.5	Záver	23
7	Draft - todo - spracovanie jednotlivých kapitol	24

Zoznam obrázkov

Zoznam tabuliek

Zoznam skratiek

FEBFC Fuzzy entropy-based fuzzy classifier

Úvod

Cieľom diplomovej práce je experimentálne porovnať algoritmy, ktoré slúžia pre fuzzifikáciu numerických hodnôt.

Náplňou prvej časti práce je oboznámenie sa s problematikou fuzzifikácie.

Ďalšia časť sa zaoberá rozborom existujúcich algoritmov pre fuzzifikáciu numerických hodnôt na lingvistické.

Ďalšia kapitola popisuje spôsob implementácie vybraných algoritmov fuzzifikácie v jazyku C++.

Posledná časť práce experimentálne porovnáva dané implementácie na rôznych výstupných dátach.

Kapitola 1

Analýza súčasného stavu

Kapitola 2

Ciel' práce

Cieľom diplomovej práce je experimentálne porovnať algoritmy, ktoré slúžia pre fuzzifikáciu numerických hodnôt.

Postup práce

1. Oboznámenie sa s problematikou fuzzifikácie (transformácie numerických hodnôt na lingvistické).
2. Rozbor existujúcich algoritmov pre fuzzifikáciu numerických hodnôt.
3. Implementácia vybraných algoritmov fuzzifikácie v jazyku C++.
4. Experimentálne porovnanie implementovaných algoritmov na rôznych výstupných dátach.

Kapitola 3

Teoretické východiská práce

3.1 Získavanie znalostí z databáz

3.2 Fuzzy prístupy

Ludské znalosti sú niekedy neurčité, nepresné a ako aj nekonzistentné. Fuzzy prístupy umožňujú určitým spôsobom formalizovať a ďalej spracúvať vágne poznatky. Vágnosť možno považovať za typ nepresnosti. Takýto typ znalostí je ťažké a často aj nemožné vhodne formalizovať konvenčnými metódami. Fuzzy prístupy predstavujú jednu z možných ciest ako k nim pristupovať, a formalizovať neurčitost.

Teória fuzzy množín je zovšeobecnením klasickej teórie množín - fuzzy množiny sú vágne v tom či prvok patrí alebo nepatrí do množiny. Na fuzzy množinách možno vykonávať určité operácie čiastočne analogické s tými, ktoré sú v klasickej teórii množín.

Fuzzy logika predstavuje prístup, ktorý zovšeobecuje konvenčnú logiku a produkčné pravidlá zavedením tzv. lingvistických premenných a lingvistických pravidiel. Fuzzy logika umožňuje formulovať vágne pravidlá. Fuzzy aritmetika rozširuje princípy klasickej aritmetiky na vágne - fuzzy - čísla.

3.2.1 Teória fuzzy množín

V klasickej teórii množín prvok môže do množiny buď patriť alebo nepatriť. Pre klasické množiny možno definovať tzv. charakteristickú funkciu.

Charakteristická funkcia klasickej množiny S je priradenie typu:

$$\mu_S : U \longrightarrow \{0, 1\} \quad (3.1)$$

Priradenie hodnoty 0 - nepatrí, alebo hodnoty 1 - patrí - ku každému prvku $x \in U$, pričom definičný obor charakteristickej funkcie U sa nazýva univerzum. Univerzum je množina všetkých hodnôt, o ktorých rozhodujeme či do danej množiny patria, alebo nepatria. Platí $S \subseteq U$.

Charakteristickú funkciu klasickej množiny možno definovať nasledovne:

$$\mu_S(X) = \begin{cases} 1, & \text{ak } x \in S \\ 0, & \text{ak } x \notin S \end{cases} \quad (3.2)$$

V teórii fuzzy množín sa zavádza rozšírenie tohto konceptu - prvok môže do množiny patriť aj čiastočne: viac alebo menej. Vágnosť je teda v otázke príslušnosti prvku ku množine.

Stupeň príslušnosti a funkcia príslušnosti

3.3 Fuzzifikácia

3.4 Fuzzy logika

3.5 Meranie Entropie

Entropia je meraná množstvom neistoty výsledku náhodného experimentu, alebo ekvivalente, meraním informácií keď sa pozoruje výsledok. Tento koncept bol zadefinovaný rôznymi spôsobmi [25]–[30] a zovšeobecnený v rozličných aplikovaných oblastiach, ako napríklad teória komunikácie, matematiky, štatistickej termodynamike a ekonómii

[31]–[33]. Z pomedzi týchto rozličných definícií, Shannon prispel k najširšej a najfundamentálnejšej definícii entropie v informačnej teórii. V nasledujúcom texte najprv uvedieme Shannonovu entropiu a potom popíšeme štyry Luca-Termini axiómy [25], ktoré dobre-definovaná entropia musí spĺňať. Nakoniec navrhujeme meranie fuzzy entropie, ktoré je rozšírenie Shannonovej definície.

3.5.1 Shannonova Entropia

Za entropiu možno považovať meranie neistoty náhodnej premennej X . Nech X je náhodná spočítateľná premenná s konečnou N -znakovou abecedou danou \mathcal{X} . Ak výsledok x_j sa vyskytuje s pravdepodobnosťou $p(x_j)$, tak potom množstvo informácie spojené so známim výskytom výstupu x_j je definované ako:

1. TODO To znamená, že pre diskkrétne zdroje, informácie získané výberom symbolu sú bitové. V priemere, symbol bude vybratý n -krát z celkového počtu N výberov, takže priemerné množstvo informácie získanej z n zdrojových výsledkov je:

2. TODO

$$D_j = \frac{\sum_{r \in S_{C_j}(r_n)} \mu_{\tilde{A}}(r)}{\sum_{r \in X} \mu_{\tilde{A}}(r)} \quad (3.3)$$

Podelením (2.) číslom n získame priemerné množstvo informácie na symbol výstupu zdroja. To je známe ako priemerná informácia, neistota, alebo entropia definovaná nasledovne. Definícia 1: Entropia $H(X)$ náhodnej diskkrétnej premennej X je definovaná ako

3. TODO Alebo 4. TODO Kde Všimnite si, že entropia je funkcia distribúcie X . Nezáleží na skutočných hodnotách náhodnej premennej X , ale iba na pravdepodobnostiach. Preto entropiu možno zasísať ako $H(p)$.

3.6 TODO INE

3.7 Záver

Kapitola 4

Analýza existujúcich algoritmov pre fuzzifikáciu numerických hodnôt

4.1 Fuzzy klasifikátor s možnosťou výberu na základe fuzzy entropie

Táto kapitola popisuje efektívny fuzzy klasifikátor s možnosťou výberu založenom na meraní fuzzy entropie (FEBFC). Fuzzy entropia je použitá na vyhodnotenie informácie o distribúcii vzorov v priestore vzorov. S touto informáciou vedia rozdeliť priestor vzorov na disjunktné rozhodovacie regióny pre rozoznávanie vzorov. Vďaka tomu, že rozhodovacie regióny sú disjunktné, aj komplexnosť, aj výpočtová náročnosť je zredukovaná, a tým pádom aj čas trénovania a klasifikácie je extrémne krátka. Hoci rozhodovacie regióny sú rozdelené do disjunktných pod priestorov, môžu dosiahnuť kvalitnú klasifikáciu vďaka tomu, že pod priestory boli správne stanovené navrhovaným meraním fuzzy entropie. Okrem toho môžeme skúmať ďalšie využitie fuzzy entropie na vybraté prvky. Procedúra výberu prvkov nielenže znižuje dimenziu problému, ale aj redukuje šum, zbytočné a nedôležité prvky.

4.1.1 Opis upravenej fuzzy entropie

todo - word

4.1.2 Algoritmus FEBFC

Algoritmus FEBFC sa skladá z nasledujúcich krokov:

Krok 1. Zistenie počtu intervalov pre každú dimenziu.

Krok 2. Zistenie centra a šírku pre každý interval.

Krok 3. Priradenie funkcie príslušnosti pre každý interval.

Krok 4. Označenie tried pre každý rozhodovací región.

4.2 Minimum description length partition

1. MDLP method developed in the Fayyad, U. M. and Irani, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning, Artificial intelligence, 13, 1022-1027. I had an interactive version of the program, which lets you choose from several stopping criteria: 1) using the criteria proposed in the original work; 2) criteria for the number of partitions intervalov; 3) criteria for the threshold Gini index (which assesses the effectiveness of the partition at the point in terms of the decrease in entropy).

4.2.1 Algoritmus MDLP

4.3 Záver

Kapitola 5

Implementácia algoritmov pre fuzzifikáciu numerických hodnôt

5.1 Implementácia Algoritmu FEBFC

5.2 Modifikácia Algoritmu FEBFC

5.3 TODO = DRUHY algoritmus

5.4 TODO INE

5.5 Záver

Kapitola 6

Experimentálny výskum

- 6.1 Parametre, určujúce kvalitu fuzzifikácie
- 6.2 Priebeh experimentov - použité dátové množiny
- 6.3 Výsledky a vyhodnotenie experimentov pre vybraté dátové množiny
- 6.4 Zhrnutie výsledkov
- 6.5 Záver

Kapitola 7

Draft - todo - spracovanie jednotlivých kapitol

FEBFC algoritmus

FEBFC algoritmus pozostáva z týchto krokov:

Krok 1. Určenie počtu intervalov.

Krok 2. Určenie polohy intervalov.

Krok 3. Priradenie funkcie príslušnosti pre každý interval.

Krok 4. Vypočítanie fuzzy entropie pre každú položku cez sumarizáciu fuzzy entropie pre všetky intervaly pre dané dimenzie položky.

Krok 1. Určenie počtu intervalov

Počet intervalov pre každú dimenziu má účinok na učiacu efektívnosť a klasifikačnú presnosť. Ak je počet intervalov príliš veľký, tak to zaberie veľa času na dokončenie tréningu a klasifikačného procesu a môže vzniknúť preučenie. Na druhú stranu, ak je počet intervalov príliš malý, veľkosť pre každú rozhodovaciu oblasť môže byť príliš veľká pre danú distribúciu vstupných vzorov, a klasifikačný výkon môže byť pomalší.

Kroky na určenie počtu intervalov pre každú dimenziu sú nasledovné:

Krok 1.

Krok 2.

Krok 3.

Krok 4.

Záver

Literatúra

[1] <https://github.com/chovancova/diploma-thesis-fuzzification>

[2] Priezvisko M., *Názov knihy, článku*, Vydavateľstvo, 2013, ISBN 978-1-449-36536-3