Žilinská Univerzita v Žiline
Fakulta riadenia a informatiky

# Diplomová práca

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

**Oľga Chovancová**

**Nástroj pre fuzzifikáciu numerických hodnôt**

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

Reg.č. 6/2016 Máj 2017

**ŽILINSKÁ UNIVERZITA V ŽILINE, FAKULTA RIADENIA A INFORMATIKY.**

**ZADANIE TÉMY DIPLOMOVEJ PRÁCE.**

**Študijný program : Informačné systémy**

**Zameranie: Spracovanie dát**

---

**Meno a priezvisko**

Oľga 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

Zadanie zaregistrované dňa 31. 10. 2016 pod číslom 6/2016 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 ďakujem 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 OLGA: *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

# 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

# Cieľ 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 Fuzzy dáta a expertné odhady

Metódy a algoritmy využívané v súčasných systémoch pre podporu rozhodovania by mali brať do úvahy možnú nestochastickú neurčitosť vstupných dát, spôsobenú nedostatočnou presnosťou merania. Jeden z prístupov, ktorý berie do úvahy neurčitosť, je vyjadrenie vstupných hodnôt pomocou neurčitých fuzzy dát alebo pomocou viachodnotových dát. Matematickým aparátom spracovania neurčitých dát a viachodnotových dát je fuzzy logika a viac hodnotová logika. [1]

Neurčité dáta sa reprezentujú pomocou lingvistických premenných, ktorých hodnoty patria do konečnej množiny. [1] Lingvistické premenné sa javia ako jeden z často používaných spôsobov vyjadrenia hodnôt, opisovaných kvantitatívnymi alebo kvalitatívnymi veličinami. Kvalitatívne veličiny sa v mnohých prípadoch javia ako výsledok formalizácie expertných odhadov. [1] Každý objekt alebo proces sa opisuje skupinou ukazovateľov. V modeloch pre podporu rozhodovania sa využívajú ukazovatele, ktoré nadobúdajú hodnoty z množiny reálnych čísel. Dôvody, ktoré sťažujú použitie reálnych čísiel sú nasledovné: [1]

1. Zložitosť presného merania hodnôt ukazovateľa.

2. Nie sú algoritmy a metódy výpočtu presných hodnôt ukazovateľa.

3. Zložitosť dostupnosti nameraných dát, súvisiaca s radom objektívnych a subjektívnych faktorov.

4. Absencia nevyhnutnosti poznať presné hodnoty.

5. Relatívne vysoké náklady na zmeranie presných hodnôt ukazovateľov.

Reálne hodnoty jednotlivých ukazovateľov bez prihliadnutia na iné ukazovatele jednej strane nesú pre výber rozhodnutia zbytočne podrobnú informáciu o tomto ukazovateli, na druhej strane nemôžu byť základom pre výber rozhodnutia. [1] Z toho vyplýva, že sa ukazuje ako užitočné využiť približné, neurčité fuzzy hodnoty vstupných dát. Použitie podobných fuzzy hodnôt umožňuje zaviesť do výskumu kvalitatívne opisy. Vo výsledku sa berie do úvahy neurčitosť úlohy rozhodovania, zabezpečuje sa adekvátny opis všetkých faktorov, ktoré majú vzťah danej, ale nie je možné zabezpečiť ich presný kvantitatívny opis.

Spracovanie neurčitej fuzzy informácie v rozhodovacích úlohách sa zabezpečuje aplikovaním lingvistického prístupu.[1] Matematickým aparátom ich formalizácie je teória fuzzy množín.

Lingvistický prístup pri konštrukcii modelov rozhodovania umožňuje[1]:

- využiť na opis elementov úlohy rozhodovania subjektívne odhady expertov, vyjadrené pomocou neurčitých fuzzy pojmov, vzťahov a výrokov profesionálneho jazyka;

- formalizovať neurčité fuzzy opisy pomocou fuzzy množín a lingvistických premenných;

- spracovávať neurčité fuzzy opisy prostredníctvom matematického aparátu teórie fuzzy logiky.

Základ tohto prístupu tvoria pojmy neurčitej fuzzy premennej a lingvistickej premennej. Medzi základné pojmy patrí definícia fuzzy množiny, fuzzy premennej, lingvistickej premennej, funkcie príslušnosti.[1]

Nech $X = \{x\}$ je množina prvkov x.

**Definícia 1.** Fuzzy množina $A \subset X$ je predstavovaná množinou dvojíc $\{(x, \mu_A(x))\}$, kde $x \in X$ a $\mu_A : X \longrightarrow \langle 0, 1 \rangle$ je funkcia príslušnosti, ktorá predstavuje subjektívnu mieru príslušnosti elementu x k množine A. Veličina $\mu_A(x)$ nadobúda hodnoty od nuly, ktorá označuje absolútnu nepríslušnosť po hodnotu jedna, ktorá hovorí o absolútnej príslušnosti elementu x do fuzzy množiny A. [1, 2]

Ak je fuzzy množina A definovaná na konečnej univerzálnej množine
$X = \{x_1, x_2, ..., x_i, ..., x_n\}$, potom je vhodné označiť ju nasledovne

$$A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), ..., (x_i, \mu_A(x_i)), ..., (x_n, \mu_A(x_n))\},$$

kde $(x_i, \mu_A(x_i))$ - je dvojica tvorená elementov $x_i$ a jeho funkciou príslušnosti, nazývaná singleton. [1]

**Definícia 2.** Fuzzy premenná je definovaná trojicou $(\alpha, X, A)$, kde $\alpha$ - je meno fuzzy premennej, $X = \{x\}$ - je množina, tvoriaca definičný obor premennej x, A - je fuzzy podmnožina fuzzy množiny X, pre každý prvok ktorej je definovaná funkcia $\mu_A(x)$, udávajúca stupeň príslušnosti daného elementu x do množiny A. [1]

**Definícia 3.** Lingvistická premenná je definovaná päticou $(\beta, T, X, G, M)$, kde $\beta$ - je meno lingvistickej premennej; T - je množina jej hodnôt (*termov*), z ktorých každá je fuzzy premennou na množine X; G - je syntaktické pravidlo pre tvorbu nových mien hodnôt lingvistickej premennej $\beta$; M - je sémantická procedúra, umožňujúca transformovať novú hodnotu premennej $\beta$, určenú procedúrou G, na fuzzy premennú, t.j. vytvoriť zodpovedajúcu fuzzy množinu. [1]

**Definícia 4.** Funkcia príslušnosti $\mu_A(x)$ kvantitatívne určuje príslušnosť elementov základnej množiny uvažovaného priestoru $x \in X$ k fuzzy množine A. Hodnota A tejto funkcie značí, že element nepatrí do fuzzy množiny, hodnota 1 opisuje úplne patriaci element. Hodnoty medzi 0 a 1 charakterizujú neurčito zaradené elementy. [1, 3, 4, 5]

Lingvistická premenná sa od číselnej premennej líši tým, že jej hodnotami nie sú čísla, ale slová alebo výroky prirodzeného alebo formálneho jazyka. Je zrejmé, že takýto kvalitatívny popis s využitím slov je menej presný ako pomocou čísel. Napriek tomu použitie lingvistickej premennej umožňuje približne opísať zložité javy, ktoré nie je možné opísať pomocou obvyklých kvalitatívnych termínov. Dôležitý aspekt lingvistickej premennej spočíva v tom, že táto premenná má vyššiu úroveň ako fuzzy premenná v tom zmysle, že hodnotami lingvistickej premennej sú fuzzy premenné. [1]

## 3.2 Fuzzy prístupy

Fuzzy prístupy možno považovať za odpoveď na požiadavku spracovania neurčitosti, resp. nepresnosti. Táto požiadavka je veľmi rozšírená - v určitej forme sa vyskytuje prakticky v každom reálnom systéme aplikujúcom metódy umelej inteligencie - predovšetkým sa však objavuje tam, kde systémy nejakým spôsobom interagujú s človekom alebo využívajú ľudské znalosti. V súvislosti s neurčitosťou sa rozlišujú tri základné pojmy: [8, 6]

- **Neurčitosť** - vyplýva z nedostatočnej znalosti faktorov alebo udalosti. O neurčitosti hovoríme, že je aleatórna, ak pramení z vnútorných vlastností nejakého náhodného javu - t.j. ju principiálne nemožno odstrániť. Ak neurčitosť vyplýva z neznalosti, tak je epistemická.

- **Nepresnosť** - O nepresnosti hovoríme, ak je znalosť faktov a udalosti taká kompletná ako len môže byť, ale spôsob ich vyjadrenia nie je presný alebo jednoznačný.

- **Nekonzistentnosť** - O nekonzistentnosti hovoríme, ak si znalosti, resp. známe fakty navzájom odporujú.

Ľudské znalosti vo väčšine prípadov zahŕňajú neurčitosti, nepresnosť, niekedy môžu byť 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 použiť na formalizáciu neurčitosti. [6, 8]

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 teórie množín. Fuzzy logika predstavuje prístup, ktorý zovšeobecňuje 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. [6]

### 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 [6]

$$\mu_S : U \longrightarrow \{0,1\} \tag{3.1}$$

Priradenie hodnoty 0 - nepatrí, alebo hodnoty 1 - patrí - ku každému prvku x ∈ 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 $S \subseteq U$. [6]

Charakteristickú funkciu klasickej množiny možno definovať nasledovne [6, 9]

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

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. [6]

#### Stupeň príslušnosti a funkcia príslušnosti

Mieru do akej prvok patri do fuzzy množiny sa vyjadruje stupňom príslušnosti. Nech A je fuzzy množina. Stupeň príslušnosti prvku x ku množine A označujeme $\mu_A(x)$. Hovoríme tiež, že $\mu_A(x)$ je funkcia príslušnosti fuzzy množiny A. [6, 9] Funkcia príslušnosti je priradenie

$$\mu_A : U \longrightarrow \langle 0,1 \rangle \tag{3.3}$$

Obor hodnôt je teda v tomto prípade

$$\mu_A(X) : U \in \langle 0,1 \rangle \tag{3.4}$$

Pritom rozlišujeme nasledujúce prípady:

- ak $\mu_A(x) = 0$, hovoríme, že prvok do množiny A nepatrí,

- ak $\mu_A(x) = 1$, hovoríme, že prvok do množiny A patrí,
- ak $\mu_A(x) \in (0, 1)$, hovoríme, že prvok patrí do množiny A čiastočne, so stupňom príslušnosti ak $\mu_A(x)$.

Aj v tomto prípade sa dá použiť značenie $A \subseteq U$, čím sa rozumie, že množina A je definovaná na univerze U. [6]

### Spojité a diskrétne fuzzy množiny

Fuzzy množiny možno rozdeliť podľa spojitosti na spojité a diskrétne. V prípade spojitých fuzzy množín je univerzum spojité. T.j. aj funkcia príslušnosti je spojitá. Naopak v prípade diskrétnych fuzzy množín sú univerzum aj funkcia príslušnosti diskrétne. Obor funkcie príslušnosti je spojitý v oboch prípadoch. [6]

### Spôsoby zápisu fuzzy množín

Fuzzy množiny možno zapísať buď diskrétne alebo spojito.

V prípade, že ide o diskrétnu fuzzy množinu, možno použiť nasledujúci zápis [7, 6]

$$A = \{\mu_A(x_1)/x_1, \mu_A(x_2)/x_2, ..., \mu_A(x_n)/x_n\}, \tag{3.5}$$

kde n je počet prvkov, $x_i inU : \forall_i = 1, 2, ..., n$ sú prvky univerza a $\mu_A(x_i)$sú ich stupne príslušnosti.

Ďalšia konvencia zápisu diskrétnych fuzzy množín je [6]

$$A = \{\mu_A(x_1)/x_1 + \mu_A(x_2)/x_2, ... + \mu_A(x_n)/x_n\}, \tag{3.6}$$

$$A = \{(x_1; \mu_A(x_1)), (x_2; \mu_A(x_2)), ..., (x_n; \mu_A(x_n))\}, \tag{3.7}$$

$$A = \sum_{i=1} \mu_A(x_i)/x_i. \tag{3.8}$$

Spojité fuzzy množiny možno reprezentovať výrazom v tvare [7]

$$A = \int_U \mu_A(x)/x dx. \tag{3.9}$$

#### Singleton

Špeciálnym typom fuzzy množiny je tzv. singleton. Ide o taký typ fuzzy množiny, pre ktorý iba jeden bod univerza má stupeň príslušnosti väčší ako 0. Nech teda A je fuzzy množina definovaná na univerze U. Potom fuzzy množinu A považujeme za singleton, ak existuje bod $x_0 \in U$ taký, že platí [6]

$$\mu_S(X) = \begin{cases} b & x = x_0 \\ 0 & inak \end{cases} \tag{3.10}$$

$$b \in (0,1\rangle$$

Singleton možno zapísať obdobným spôsobom ako diskrétnu fuzzy množinu a to je [6]

$$A = \{b/x_0\}, \tag{3.11}$$

$$A = \{(x_0;b)\} \tag{3.12}$$

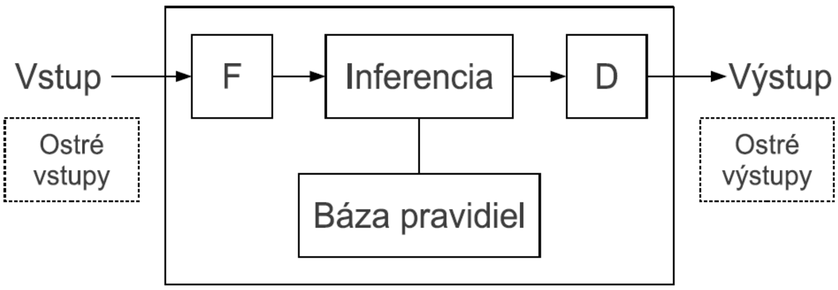
### 3.2.2 Fuzzy logika

#### Lingvistická premenná

Lingvistická, resp. jazyková premenná je zvláštnym typom premennej, ktorá sa od numerických premenných odlišuje tým, že jej hodnoty - tzv. lingvistické hodnoty - nie sú čísla, ale slovné výrazy. Pritom každej lingvistickej hodnote je priradený význam, t. j. určitá fuzzy množina definovaná na spoločnom univerze. [6]

#### Fuzzy inferenčný systém

Lingvistické premenné možno využiť na odvodzovanie a získať z nich výstupné ostré hodnoty (t.j. presné číselné hodnoty). Inferenčný systém je štruktúra umožňujúca odvodzovať na základe vopred daných pravidiel a známych faktov nové fakty. V prípade fuzzy inferenčného systému (FIS) možno pri formulácii pravidiel navyše využiť lingvistické premenné a hodnoty. Medzi najznámejšie metódy fuzzy inferencie patrí Mandaniho inferencia. [6] V prípade, že sa má fuzzy inferenčný systém použiť ako jadro regulátora, je potrebné, aby jeho vstupmi a výstupmi boli ostré hodnoty. FIS umožňujú aj prevod z ostrých (presne číselných) vstupov na lingvistické hodnoty a naopak. [6] Bloková schéma na Obr. 3.1 zobrazuje hlavné komponenty fuzzy inferenčného systému: F - fuzzifikácia, inferencia, báza pravidiel, D - defuzzifikácia.




Obr. 3.1: Fuzzy inferenčný systém - bloková schéma.[6]

**Báza pravidiel**

Fuzzy inferenčný systém pri odvodzovaní nových faktorov vychádza z určitých vopred daných pravidiel. Tie sú združené v báze pravidiel. Pravidlá majú vo všeobecnosti nasledujúci tvar [6, 10, 7]

**AK** (predpoklad) **POTOM** (dôsledok), resp.
**IF** (antecedent) **THEN** (consequent).

Predpoklad sa pritom skladá z termov nasledujúceho tvaru [6]

$$X \ is \ A_i, \tag{3.13}$$

kde X je lingvistická premenná a $A_i$ je i-tá lingvistická hodnota tejto premennej.

**Fuzzifikácia**

Z blokovej schémy na Obr. 3.1 je zrejmé, že do fuzzy inferenčného systému môžu vstupovať ostré hodnoty. Aby bolo možné vykonať porovnanie podľa (3.13), ostrú vstupnú hodnotu je potrebné previesť na fuzzy množinu. Tento proces sa nazýva fuzzifikácia.

Fuzzifikácia je prevod ostrej hodnoty x na fuzzy množinu $L_x$, pričom výstupom je často singleton [6]

$$L_x = \{1/x\}\,. \tag{3.14}$$

Majme term v tvare podľa (3.13). Ak ostrá hodnota premennej je x a jej fuzzifikovaná hodnota $L_x$, potom takýto term možno vyhodnotiť jednoducho pomocou zvolenej T-normy (analógia prieniku). Ako T-norma sa v tomto prípade spravidla používa operátor minima $T_m(a,b) = min(a,b)$. Ak sa aplikuje na dané operandy vznikne [6]

$$\alpha_x = min(L_x, A^i) \forall x \in U, \tag{3.15}$$

kde $\alpha_x$ možno považovať za mieru platnosti tvrdenia (3.13). Vzhľadom na vlastnosti singletonov možno tento postup zjednodušiť na

$$\alpha_x = \{\mu_{A^i}(x)/x\}\,, \tag{3.16}$$

Keďže $\alpha_x$ je tiež singleton, tak stačí, ak sa vyjadrí iba ako $\mu_{A^i}(x)$, čo je ostré číslo. Fuzzifikácia v praxi je vyhodnotenie miery platnosti predpokladu X is A (3.13) na základe znalosti ostrej hodnoty x premennej X. Miera platnosti takého predpokladu je pritom ekvivalentná stupňu príslušnosti ostrej hodnoty x ku fuzzy množine A, čiže $\mu_A(x)$. [6]

## 3.3 Transformácia číselných hodnôt na lingvistické premenné

Transformácia číselných premenných na lingvistické predpokladá vzťah medzi kombináciou ich hodnôt a stupnicou s konečným počtom m intervalov a nazýva sa diskretizáciou. Pritom sa každý interval asociuje s hodnotou lingvistickej premennej (termom). Proces diskretizácie predpokladá transformáciu kvantitatívnych dát na kvalitatívne. [1, **?**]

Vo všeobecnom prípade diskretizácia zmenšuje objem dát a nevedie k zníženiu klasifikačnej presnosti a spoľahlivosti hodnôt, využívaných v systémoch pre podporu rozhodovania. Naopak, takéto diskrténe hodnoty adekvátne ohodnocujú ukazovatele a sú stabilnejšie vzhľadom na zmeny a metodiky merania. [1, **?**]

Úloha diskretizácie sa formuluje nasledovne.

### 3.3.1 Definícia úlohy diskretizácie

Majme množinu pozostávajúcu z N príkladov. Každý príklad obsahuje množinu vstupných číselných premenných. Nech niektorá vstupná číselná premenná X v týchto príkladoch nadobúda hodnoty v rozsahu od $x_min$ po $x_max$. Potom diskretizáciou tejto spojitej premennej sa nazýva proces rozkladu jej hodnôt na m diskrétnych intervalov

$$D = \{\langle d_0, d_1), \langle d_1, d_2), ..., \langle d_{m-1}, d_m\rangle\}, \tag{3.17}$$

kde $d_0$ je minimálna hodnota $x_min$ tejto premennej, $d_m$ je maximálna hodnota $x_max$ premennej a platí $d_i < d_{i+1}$, pre i = 0, 1, ... m-1. Pritom, $P = \{d_1, d_2, ..., d_{m-1}\}$ je množina bodov rezu premennej X. [1, **?**]

### 3.3.2 Metódy diskretizácie

Získanie optimálnej diskretizácie prestavuje NP-zložitú úlohu. [1, **?**] Existencia obrovského množstva metód diskretizácie je vysvetliteľná rozmanitou povahovou vstupných dát a požiadaviek na ich spracovanie. Výber použitej metódy diskretizácie určuje úspešnosť ich budúceho spracovania. Na výber metódy má vplyv množstvo parametrov. [1]

1. **Počet intervalov rozkladu.** Príliš malý počet intervalov spôsobuje nepresnosti a chyby vo vstupných dátach a vedie k hrubému vyjadreniu výsledku. Na druhej strane rozklad vstupných dát na príliš veľký počet intervalov spôsobuje prílišnú detailizáciu a následne vedie k tomu, že spracovanie takýchto dát je pomalé a neefektívne [1, **?**]

2. **Nesúlad s výsledkom.** Nesúlad s výsledkom sa objasňuje vznikom neočakávaných chýb v procese diskretizácie. Tieto chyby súvisia s nepresnosťami diskretizácie vstupných dát a vo vzťahu medzi týmito diskretizovanými hodnotami a hodnotami výstupného atribútu. [1]

3. **Presnosť rozkladu.** Predpokladá, že úspešný algoritmus diskretizácie, skonštruovaný na základe výberu tréningových dát pracuje bez podstatného zníženia kvality aj na všetkých nasledujúcich dátach. [1]

4. **Časové ohraničenia.** V prípade statických procesov, kedy požiadavka diskretizácie výučbovej množiny vzniká iba raz, nie je čas výpočtu dôležitým parametrom. Avšak pri dynamických procesoch, keď sa etapa diskretizácie výučbovej množiny opakuje mnohokrát, je čas spracovania kritickým parametrom. [1]

Proces transformácie reálnych hodnôt na konečný počet intervalov a reprezentovanie každého intervalu s diskrétnou hodnotou predstavuje proces, ktorý je už dosť dobre preskúmaný. Vykonaná analýza odhalila rad základných kritérií klasifikácie existujúcich metód diskretizácie. Podrobnejšie je prehľad a analýza existujúcich metód diskretizácie uvedená v prácach [**?**, **?**, **?**, **?**, **?**, **?**, **?**] .

### Statické a dynamické metódy diskretizácie

Kritérium uvažuje klasifikáciu metód na základe opakovateľnosti procesu diskretizácie.

Statické metódy sa realizujú v tvare predbežnej samostatnej etapy spracovania vstupných dát a nezávisia od spôsobu následného využitia diskretizovaných hodnôt.

Dynamická diskretizácia je obvykle zabudovaná do mechanizmu inteligentného spracovania dát a využíva sa napríklad pri konštrukcii rôznych klasifikátorov. [1]

#### Globálne a lokálne metódy diskretizácie

Kritérium berie do úvahy úplnosť množiny a dostupnosť vstupných hodnôt, použitých v procese diskretizácie.

Globálne metódy diskretizácie spracovávajú vstupnú množinu hodnôt číselného atribútu počas etapy predbežného spracovania.

Lokálne metódy predpokladajú diskretizáciu dát súčasne s inými metódami ich spracovania. K tomuto typu patria dynamické metódy diskretizácie, ktoré zisťujú body rezu v rámci vnútorných operácií algoritmov spracovania dát a nemajú prístup k úplnej množine vstupných dát. [1]

#### Metódy jednorozmernej, mnohorozmernej diskretizácie

Kritérium slúži na klasifikáciu metód diskretizácie na základe počtu súčasne spracovaných atribútov.

V jednorozmernej diskretizácii sa každý číselný atribút transformuje na lingvisticky nezávisle od hodnôt iných atribútov.

Metódy mnohorozmernej diskretizácie predpokladajú súčasné spracovanie všetkých číselných atribútov. Vo výsledku rozklad hodnôt číselných atribútov na intervaly vykonáva s prihliadnutím na množný vzájomný vplyv vstupných atribútov jedného na druhý. [1]

#### Diskretizácia bez učitela, s učiteľom

Kritériium rozlišuje neriadenú diskretizáciu bez učiteľa alebo riadenú diskretizáciu s učiteľom.

Diskretizácia bez učiteľa určuje body rezu iba na základe analýzy vstupných atribútov. Hodnoty výstupného atribútu sa pritom neberú do úvahy.

Diskretizácia s učiteľom pri rozklade vstupného atribútu na niekoľko intervalov navyše zisťuje vzťah medzi hodnotami vstupných atribútov a im zodpovedajúcimi hodnotami výstupných atribútov. Výsledkom použitia diskretizácie s učiteľom ej možnosť takého rozdelenia vstupného atribútu, kedy rôznym intervalom zodpovedajú rôzne hodnoty výstupného atribútu. Využitie diskretizácie s učiteľom umožňuje automaticky určovať najlepší počet intervalov a body rozdelenia pre každý atribút s perspektívou realizácie budúcej klasifikácie alebo klusterizácie výstupného atribútu. [1]

**Priame a inkrementálne metódy diskretizácie**

Kritérium predpokladá klasifikáciu podľa spôsobu získania intervalov.

Metódy priamej diskretizácie naraz rozdelia vstupnú množinu na m intervalov. Využitie týchto metód na začiatku vyžaduje určenie veličiny m. V každom kroku spracovania vyberajú tieto metódy niekoľko bodov rezu.

Na druhej strane metódy postupnej diskretizácie predpokladajú hľadanie najlepšieho kandidáta na bodu rezu. [1]

### 3.3.3 Neparametrické a parametrické metódy diskretizácie

Kritérium predstavuje variant predchádzajúceho kritéria a určuje spôsob získania počtu intervalov pre každý atribút.

Neparametrické metódy určujú najlepší počet intervalov pre každý atribút.

Parametrické metódy predpokladajú, že počet intervalov je už apriórne zadaný. [1]

### 3.3.4 Metódy zlučovacej a rozdeľovacej diskretizácie

Kritérium klasifikuje metódy diskretizácie v závislosti od spôsobu spracovania vstupných hodnôt - zlučovanie alebo rozdeľovanie.

Zlučovacia (merging) diskretizácia predpokladá postupné pridávanie bodov rezu a jeho výsledkom je rozklad na menšie intervaly.

Rozdeľovacia (splitting) diskretizácia spočíva v odstraňovaní bodov rezu, čo vedie k postupnému zlučovaniu skupiny susedných intervalov do väčšieho intervalu.

Opakovanie procesov rozdeľovania alebo zlučovania sa riadi ukončovacím kritériom.

Ako primitívne kritérium ukončenia vystupuje napríklad apriórne zadaný počet intervalov.

Ako zložité kritérium sa používa napríklad minimálna chyba klasifikácie. Toto kritérium je prípustné iba pre metódy inkrementálnej diskretizácie.

Sú známe aj metódy diskretizácie, ktorých činnosť je založená na zlučovaní alebo rozdeľovaní niekoľkých intervalov. Hybridné metódy počas činnosti striedajú spájanie a rozdeľovanie. [1]

### 3.3.5 Metódy diskretizácie podľa použitého parametra

Kritérium klasifikuje metódy diskretizácie podľa použitého parametra porovnania rozličných variantov rozdelenia.

Metóda založená na informačných ukazovateľoch často využíva pojem entropie a iné ukazovatele, založené na pojmoch teórie informácie.

Štatistické metódy berú do úvahy mieru závislosti (korelácie) medzi atribútmi.

Metódy, berúce do úvahy frekvenčné charakteristiky, patria k najjednoduchším metódam diskretizácie. V týchto metódach a každý interval určuje vopred, určením apriórne zadaného počtu hodnôt.

Kolekcia metód využívajúcich chybu klasifikácie, presnosti klasifikácie.

### 3.3.6 Metódy diskretizácie vytvárajúce nepretínajúce a pretínajúce sa intervaly

Klasifikácia na základe vlastností získaných intervaloch.

K metódam diskretizácie vytvárajúce nepretínajúce sa intervaly patria doteraz všetky vymenované metódy diskretizácie. Tieto metódy vytvárajú v procese diskretizácie m diskrétnych intervalov $\langle d_0, d_1), \langle d_1, d_2), ..., \langle d_{m-1}, d_m\rangle$ pre ktoré platia nerovnosti $d_i < d_{i+1}$, pre i = 0, 1, ... m-1.

Pre metódy diskretizácie, ktoré pretínajú intervaly nemusí platiť nerovnosť. K takýmto metódam patrí fuzzy diskretizácia.

### 3.3.7 Príklady metód

Tabuľka metód 3.1....

## 3.4 Meranie Entropie

Entropia je meraná množstvom neistoty výsledku náhodného experimentu, alebo equivalene, 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 naj fundamentálnejšej definícii entropie v informačnej teórii. V nasledujúcom texte najprv uvedieme Shannonovu entropiu a potom popíšeme štyri Luca-Termini axiómy [25], ktoré dobre-definovaná entropia musí spĺňať. Nakoniec navrhneme meranie fuzzy entropie, ktoré je rozšírenie Shannonovej definície.

### 3.4.1 Shannonova Entropia

Za entropiu možno považovať meranie neistoty náhodnej premennej X. Nech X je náhodná spočítateľná premenná s konečnou N-znakovou abecedou danou . 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 diskrétne zdroje, informácie získané výberom symbolu sú bitové. V priemere, symbol bude vybratý -krát z celkového počtu N výberov, takže priemerné množstvo informácie získanej z nzdrojových výsledkou je:

2. TODO

$$D_j = \frac{\sum\limits_{r \in S_{C_j}(r_n)} \mu_{\tilde{A}}(r)}{\sum\limits_{r \in X} \mu_{\tilde{A}}(r)} \tag{3.18}$$

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

## 3.5 TODO INE

## 3.6 Záver

| **Metóda diskretizácie** | **Príklady** |
|---|---|
| Statická | ChiMerge, MDLP, Chi2, RFDisc, UnDisc, CDisc, MEMDL, ImformDisc, ImpMDLP |
| Dynamická | ID3, C45, MultiBayesian, MODLEM, ITFP |
| Globálna | EqualWidth, EqualFrequency, ChiMerge, Chi2, CAIM, RFDisc, UnDisc, ME-MDL, EBDA, ImforDisc |
| Lokálna | MDLP, ID3, C4.5, ImpMDLP |
| Jednorozmerná | Equal-Width, EqualFrequency, MDLP, UnDisc, EBDA, ME-MDL, ImformDisc |
| Mnohorozmerná | MVD, Multi-MDL, UCPD, MIDCA, WEDA, RFDisc, SMD, CBD |
| Neriadená (bez učiteľa) | EqualWidth, EqualFrequency, MVD, UnDisc, FFD |
| Riadená (s učiteľom) | ChiMerge, MDLP, ID3, C45, Chi2, WEDA, RFDisc, CDisc, EBDA, ME-MDL, ImforDis, ImpMDLP |
| Priama | UCPD, WEDA |
| Inkremetálna (postupná) | ChiMerge, ID3, CD45, Chi2, MVD, RFDisc, EBDA, ImpMDLP |
| Neparametrická | MDLP, USD, CAIM |
| Parametrická | ChiMerge, CADD |
| Zlučovacia | ChiMerge, EBDA |
| Rozdeľovacia | MDLP, RFDisc, CDisc, ME-MDL, ImpMDLP |
| Súčasná | HBD, IDD |
| Hybridná | CADD, WEDA, ImforDisc |
| Informačných ukazovateľov | ID3, C4.5, MDLP, DbEr, ME-MDL, ImforDisc, ImpMDLP, GINI |
| Štatistických ukazovateľov | ChiMerge, Zeta, Chi2, EBDA, MODL |
| Frekvenčných charakteristík | EqualWidth, EqualFrequency, UnDisc |
| Presnosti klasifikácie | Valley, MBIterative, RFDisc |

Tabuľka 3.1: Tabuľka príkladov algoritmov jednotlivých metód diskretizácie.

# 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 môžem 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

Algorimus FEBFC sa skladá z nasledujúcich krokoch:

**Krok 1.** Zistenie počtu intervalov pre každú dimenziu.

**Krok 2.** Zistenie centra a šírku pre každý interval.

**Krok 3.** Priradenie funkcie príslúšnosti 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 intevalov; 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 experimetnov - 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 trénovania 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] LEVASHENKO V. - ZAITSEVA E. - KOVALÍK Š. *Projektovanie systémov pre podporu rozhodovania na základe neurčitých dát.* Žilinská univerzita v Žiline/EDIS, 2013. ISBN 978-80-554-0680-0.

[2] Zadeh L., *Fuzzy sets.* Information and Control, vol.8, 1965, pp. 338-353.

[3] Kaufmann A., Gupta M., *Induction to fuzzy arithmetic: theory and applications.* New York : Van Nostrand Reinold Co., 1985, 361 p.

[4] Klir G., Yuan B., *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995, 591 p.

[5] Navara M., *Computation with fuzzy quantities.* Proc. of the 7th Conf. of the European Society for Fuzzy Logic and Technology (EUSFLAT), Aix-les-Bains, France, 2011, pp. 209-214.

[6] GREGOR M., *Umelá inteligencia 1* , CEIT, 2014, ISBN 978-80-971684-1-4.

[7] SPALEK.J - JANOTA. A - BLAŽOVIČOVÁ, M. - PŘIBYL, P. *Rozhodovanie a riadenie s podporou umelej inteligencie.* Žilinská univerzita v Žiline/EDIS, 2005. ISBN 80-8070-354-X.

[8] NICKLES. M - SOTTARA, D. *Approaches to Uncertain or Imprecise Rules - A Survey.* In Rule Interchange and Applications, vol. 5858 of *Lecture Notes in Computer Science*, pp. 323-336. Springer, 2009. ISBN 9783642049842.

[9] ROSS, T.J. *Fuzzy Logic with Engineering Applications..* John Wiley & Sons, 2004, second edition ed. ISBN 0-470-86075-8.

[10] PASISNO, K. M. - YURKOVICH, S. *Fuzzy control*, vol.42. Addison Wesley Longman, 1998. ISBN 0-201-18074-X.