ZNIŽOVANIE DIMENZIE PRÍZNAKOVÉHO PRIESTORU

(DIMENSIONALITY REDUCTION OF FEATURE SPACE)

Eva OCELÍKOVÁ
Katedra kybernetiky a umelej inteligencie, Fakulta elektrotechniky a informatiky Technickej univerzity v Košiciach,
Letná 9, 042 00 Košice, tel. 055/602 2570, E-mail: ocelike@tuke.sk

SUMMARY

This paper deals with the problem of dimensionality reduction of feature space. Two major categories of dimensionality reduction methods, namely feature selection and feature extraction are described and their differences from the viewpoint of usability are discussed.

Feature selection methods can be defined as choose the best subset of size d from the given set of D features. The best subset optimized a criterion function $J()$ over all possible subsets of d features. Unlike feature selection methods, feature extraction methods are based on a transformation of the original features. The term feature transformation also applies to these methods.

Although from a purely mathematical point of view, feature selection is a special case of feature extraction (the transformation matrix has "1"s on the main diagonal and "0"s otherwise), the have significant practical differences. These differences will determine which approach better fits a task's requirements and goals. Feature preserves data interpretability; feature extraction causes a loss of interpretability. Feature selection has lower discriminative power then feature extraction does. These properties reveal that these methods are somewhat contradictory. Feature selection methods will be more suitable where the user wants to preserve the interpretability of the original data and prefers decision making on the basis of meaningful features. Furthermore, they will be suitable when the user wants to reduce the tediousness and costs of data acquisition by finding the data components that can be completely excluded from the remaining acquisition process. Feature extraction methods will be suitable when the data acquisition process does not represent any problem and there is no need to keep the original interpretability of data.

This article focuses on the feature selection methods, namely "exhaustive search" method and "d best features" method, which are discussed in more detail.

Keywords: *decision, dimensionality reduction, feature selection, feature extraction, criterion function*

1. ÚVOD

Je zrejmé, že kvalita rozhodovania je spojená s kvalitou a množstvom informácie, ktorá je k dispozícii. Pravdepodobnosť chyby klasifikátora je nepriamo úmerná množstvu informácie obsiahnutej vo vstupných dátach, čo vedie ku snahe zvyšovať počet príznakov, ktorými sú klasifikované objekty charakterizované. Pri veľkom počte príznakov preto hľadáme kompromis medzi veľkosťou chyby na jednej strane a zložitosťou rozhodovania pri veľkom počte príznakov, na strane druhej. Pre zníženie počtu príznakov existujú dve základné stratégie a to podľa toho, či sa jedná o výber z pôvodných alebo upravených (transformovaných) príznakov. Prvá stratégia – **výber príznakov** (feature selection) je orientovaná na identifikáciu významných, tzv. **informatívnych** príznakov, ktoré sa výrazne podieľajú na správnosti klasifikácie a to z množiny pôvodných príznakov. Druhá stratégia – **transformácia príznakov** (feature extraction) je orientovaná na transformáciu pôvodných príznakov pomocou určitého zobrazenia na nové príznaky, ktoré budú pre klasifikáciu efektívnejšie. Prv, ako sa budeme venovať podrobnejšie prvému z uvedených prístupov, stručne popíšeme základné princípy oboch postupov.

2. VÝBER PRÍZNAKOV

Ako bolo už uvedené, tento postup zníženia počtu príznakov je zameraný na identifikáciu informa-tívnych príznakov medzi pôvodnými príznakmi. Pri výbere príznakov je snaha eliminovať tie príznaky, ktoré sú redundantné alebo obsahujú malé množstvo relevantnej informácie. Problém výberu príznakov je definovaný ako snaha nájsť najlepšiu podmnožinu X z množiny Y všetkých D príznakov, ktorými popisujeme skúmané objekty.

$$X = \{x_i \mid i = 1, 2, ..., d, \quad x_i \in Y\} \tag{1}$$

$$Y = \{y_i \mid i = 1, 2, ..., D\} \tag{2}$$

Najlepšou podmnožinou rozumieme tú podmnožinu X, ktorá je najlepšia v zmysle zvolenej kriteriálnej funkcie $J()$ vzhľadom ku všetkým ostatným d-členným podmnožinám Z_d množiny Y.

Teda platí

$$J(X) = \max_{Z_d \subset Y} J(Z_d) \tag{3}$$

3. TRANSFORMÁCIA PRÍZNAKOV

Tento spôsob je orientovaný na transformáciu pôvodných príznakov pomocou určitého zobrazenia na príznaky, ktoré budú pre klasifikáciu dôležitejšie ako pôvodné. Pri tejto redukcii je teda snaha pretransformovať väčší počet menej efektívnych príznakov na menší počet efektívnejších príznakov. Problém redukcie príznakov môže byť teda sformulovaný ako úloha nájsť najlepšiu transformáciu $\mathbf{f}$

$$\mathbf{f}(\mathbf{y}) = (f_1(\mathbf{y}),...,f_d(\mathbf{y})), \quad (4)$$

ktorá transformuje D- rozmerný príznakový priestor vektorov $\mathbf{y} = (y_1, y_2,..., y_D)$ na d-rozmerný príznakový priestor vektorov $\mathbf{x} = (x_1, x_2,..., x_d)$ t.j.

$$x_i = f_i(y_1,..., y_D). \quad (5)$$

Najlepšou transformáciou rozumieme takú transformáciu $\mathbf{f}$, ktorá je najlepšia v zmysle určitej kriteriálnej funkcie $J(\,)$. vzhľadom ku všetkým ostatným možným transformáciám $\mathbf{g}$, t.j.

$$\mathbf{f}(\mathbf{y}) = \arg\max_{\forall g} J(\mathbf{g}(\mathbf{y})). \quad (6)$$

Vzhľadom na ľahké praktické použitie, obmedzujeme sa pri hľadaní transformácie $\mathbf{f}$ spravidla na lineárne zobrazenie. Toto je potom zadané určitou maticou transformácie $\mathbf{W}$ rádu *(D,d)* a vlastnú transformáciu potom vykonávame podľa vzťahu

$$\mathbf{x} = \mathbf{W}^T\mathbf{y}. \quad (7)$$

Z rozdielnych stratégií, ktoré obidva prístupy pri znižovaní počtu príznakov používajú, vyplývajú medzi výberom a transformáciou príznakov nasledujúce rozdiely:

- Pri transformácii príznakov sú nové príznaky kombináciou pôvodných príznakov. To znamená, že nové príznaky už nie sú intrepretovateľné, teda nepopisujú žiadny konkrétny atribút. To môže byť vážny nedostatok, pokiaľ dôvodom pre zníženie počtu príznakov bola snaha o odhalenie štruktúry problému, pretože nové príznaky nemajú pre expertov žiadny význam. Naproti tomu pri výbere príznakov nedochádza ku strate interpretovateľnosti príznakov.

- Z rovnakého dôvodu je transformácia príznakov nevhodná aj v prípade, keď požadujeme časové alebo ekonomické úspory z titulu menšieho počtu sledovaných príznakov. Transformácia príznakov vyžaduje totiž získanie všetkých pôvodných D príznakov. Naproti tomu, pri výbere príznakov sa môžeme obmedziť na získavanie menšej množiny príznakov, čo vedie k časovým i ekonomickým úsporám.

- Vzhľadom k tomu, že výber, projekcia je špeciálny prípad transformácie, transformácia príznakov je v porovnaní s výberom všeobecnejší postup. Môžeme preto očakávať, že bude dávať aj lepšie výsledky, čo sa týka kvality získanych výsledkov.

4. STRATÉGIA VÝBERU PRÍZNAKOV

Základné techniky pri výbere príznakov sa snažia eliminovať tie príznaky, ktoré sú redundantné alebo obsahujú málo relevantnej informácie. Hľadá sa teda d- členná podmnožina príznakov X, ktorá je najlepšia v zmysle zvolenej kriteriálnej funkcie $J(\,)$. O parametri d predpokladáme, že je vopred zadaný. Vieme, aký počet príznakov chceme vybrať. Táto hodnota vyjadruje kompromis medzi snahou o veľkú selekciu príznakov a jednoduchou klasifikáciou na jednej strane a medzi vzrastajúcou pravdepodobnosťou chybnej klasifikácie z dôvodu zníženia množstva informácie na strane druhej. Keďže kriteriálna funkcia $J(\,)$ má všeobecný tvar, nie je iná možnosť ako určiť optimálnu množinu X. ako tá, ktorou je postupné prehľadávanie všetkých d-členných podmnožín Z_d množiny Y. To znamená, že nie je možné pre tento problém sformulovať algoritmus, ktorý by vždy po rozumnom počte krokov dával optimálne riešenie. Algoritmy dávajúce optimálne riešenie sú cenné skôr z teoretického hľadiska. V praxi musíme často použiť algoritmy, ktoré vedú skôr k približnému riešeniu.

4.1 Úplné prehľadávanie (exhaustive search)

Úplné prehľadávanie je založené na postupnom preskúmavaní celého stavového priestoru úlohy, t.j postupné preskúmanie všetkých d- členných podmnožín množiny Y. Metódy, ktoré túto stratégiu využívajú sa nazývajú *optimálne*. Jeden z postupov používajúcich takúto stratégiu označujeme ako *úplné prehľadávanie*. Ide o postupné generovanie d-členných podmnožín, vyhodnotenie ich kriteriálnej funkcie a zapamätanie najlepšej z nich. Množiny môžeme generovať ľubovoľným spôsobom. Stavový priestor je možné reprezentovať stromom. Každý nekoncový vrchol okrem koreňa stromu predstavuje jeden odstránený príznak. Jednotlivé zostávajúce d-členné skupiny príznakov sú potom koncové uzly grafu (listy). Listov je toľko, koľko je d- členných

podmnožín z množiny Y o počte prvkov D. Vo všeobecnom prípade je počet listov daný kombinačným číslom

$$\binom{\mathrm{D}}{\mathrm{d}} = \frac{\mathrm{D!}}{\mathrm{d!(D-d)!}} \qquad (8)$$

Na generovanie *d*- členných podmnožín môžeme použíť aj iný spôsob, napr. generovanie podmnožín v lexikografickom poradí.

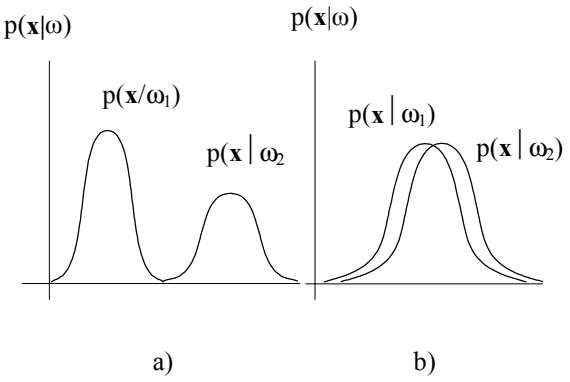
Zložitosť algoritmu úplného prehľadávania je exponenciálna, t.zn., že už pri relatívne malých hodnotách parametorov D a d (stavový priestor je rozsiahly) je nutné pre nájdenie optimálnej množiny X vykonať neúmerne veľa krokov, čo je veľká *nevýhoda úplného prehľadávania*. Z tohto dôvodu je metóda prakticky použiteľná iba pre určité hodnoty D a d a to v prípade, keď je aspoň jeden parameter malý alebo je malá hodnota rozdielu $D-d$. *Výhodou úplného prehľadávania* je, že vždy dospeje k optimálnemu riešeniu, pričom nezávisí na tvare kriteriálnej funkcie $J()$.

4.2 Metóda "d najlepších príznakov"

Časová zložitosť optimálnych algoritmov je vo všeobecnosti exponenciálna. Snahou je nájsť metódy, ktorých výpočtová zložitosť je podstatne nižšia a ktoré môžu byť prakticky využiteľné. Cenou za zjednodušenie výpočtu bude oproti optimálnym metódam nižšia kvalita výsledku. Metódy takéhoto typu nazývame *suboptimálne*. Jedným z takýchto algoritmov je metóda "*d najlepších príznakov*". Každý príznak je pomocou kriteriálnej funkcie ohodnotený individuálne, t.j bez vplyvu ostatných príznakov a cieľová množina X je zložená z d kriteriálnou funkciou *najlepšie ohodnotených* príznakov. Predpokladáme, že najlepšie príznaky môžu ako celok tvoriť najlepšiu podmnožinu. Nemusí to však tak byť, pretože sa pritom neberú do úvahy vzájomné vzťahy medzi príznakmi. Ak sú napríklad dva príznaky štatisticky závislé, obsahujú určitú časť informácie obidva príznaky. Ak je táto informácia podstatná, zaradí metóda "*d najlepších príznakov*" obidva príznaky do výslednej množiny. Týmto sa do cieľovej množiny vnáša redundancia, pretože z danej dvojice postačovalo vybrať iba jeden príznak. Výhodou tohto algoritmu je však jeho jednoduchá formulácia a implementácia, pričom ide o veľmi rýchly algoritmus. Nevýhodou je, že získané riešenie je takmer vždy približné a iba zriedka získame optimálne riešenie.

5. KRITERIÁLNE FUNKCIE

Jednou z často používaných kriteriálnych funkcií je miera založená na vzdialenosti pravdepodobností. Táto miera je založená na znalosti pravdepodobností, ktoré sa týkajú jednotlivých tried.. Predpokladáme teda, že poznáme apriórnu pravdepodobnosť $P(\omega_i)$ jednotlivých tried ω_i a rozdelenie podmienej pravdepodobnosti na triedach, $p(x/\omega_i),\ i=1,2,...,R,$ kde R je počet tried. Predpokladajme najprv situáciu, že realizujeme klasifikáciu do dvoch tried s rovnakými apriórnymi pravdepodobnosťami. Na vstupe máme konkrétny objekt x, o ktorom je potrebné rozhodnúť, do ktorej triedy patrí. Keďže sú známe podmienené pravdepodobnosti na obidvoch triedach, môžeme tieto hodnoty pre objekt x vypočítať. V ideálnom prípade by sme chceli, aby iba jedna z týchto hodnôt bola nenulová. Tá by potom určovala, do ktorej z dvoch tried objekt x jednoznačne patrí. Ak sú obe hodnoty nenulové, znamená to, že sa triedy prekrývajú a objekt x leží v ich prieniku. Čím viac sú teda tieto podmienené pravdepodobnosti $p(\mathrm{x}/\omega_1)$ a $p(\mathrm{x}/\omega_2)$ rozdielne, tým viac sú rozdielne samotné triedy, t.j chyba rozhodnutia sa zmenšuje. Uvedená skutočnosť je demonštrovaná na obr. 1.



Obr. **1a)** Dobre oddeliteľné (separovateľné) triedy
1b) Zle oddeliteľné triedy
Fig. 1a) Well separate classes
1b) Misseparate classes

Na obr. 1a) sú hodnoty podmienených pravdepodobností úplne rozdielne. Hustoty týchto pravdepodobností sa vzájomne neprekrývajú a triedy sú teda jednoznačne oddeliteľné, pretože majú prázdny prienik. Na obr. 1b) sa obe hustoty značne prekrývajú, čo zamená, že sa prekrývajú aj triedy samotné. Veľkosť prekryvu oboch hustôt pravdepodobností, pomocou ktorej budeme vyjadrovať "vzdialenosť" týchto hustôt, nám môže poslúžiť ako dobrá kriteriálna funkcia pre oddeliteľnosť tried.
Vo všeobecnosti može byť takouto kriterálnou funkciou každá funkcia J( ),

$$\mathrm{J}() = \int f(p(x|\omega_1), p(x|\omega_2), P(\omega_1), P(\omega_2))\,\mathrm{d}\,x, \qquad (9)$$

splňajúca nasledujúce podmienky:

1. $J(\,) \geq 0$.
2. $J(\,)$ dosahuje maximum, ak majú triedy ω_1 a ω_2 v priestore príslušnom objektom x prázdny prienik, t.j práve vtedy, ak pre každý objekt x platí, že $p(\mathrm{x}/\omega_1) = 0 \Leftrightarrow p(\mathrm{x}/\omega_2) > 0$.
3. $J(\,) = 0$, ak sa obe hustoty pravdepodobností, a teda aj triedy, dokonale prekrývajú, t.j ak pre každý objekt x platí, že $p(\mathrm{x}/\omega_1) = p(\mathrm{x}/\omega_2)$.

Každú funkciu $J(\,)$, ktorá spĺňa podmienky 1 až 3 budeme nazývať ***mierou vzdialenosti pravdepodobností.***

Uvedieme aspoň niektoré z týchto typov mier:

Chernoff

$$\mathrm{J}_C = -\ln \int p(\mathrm{x})^{\alpha}\, p(\mathrm{x}|\omega_1)\, p(\mathrm{x})^{1-\alpha}\, p(\mathrm{x}|\omega_2)\, dx, \qquad (10)$$

kde $\alpha \in \langle 0,1 \rangle$

Bhattacharyya

$$\mathrm{J}_B = -\ln \int \sqrt{p(\mathrm{x}|\omega_1)\, p(\mathrm{x}|\omega_2)}\, dx \qquad (11)$$

Divergencia

$$\mathrm{J}_D = \int \left(p(\mathrm{x}|\omega_1) - p(\mathrm{x}|\omega_2)\right) \ln \frac{p(\mathrm{x}|\omega_1)}{p(\mathrm{x}|\omega_2)}\, dx \qquad (12)$$

Uvedené vzťahy môžeme výrazne zjednodušiť, ak prijmeme určité predpoklady o podmienených hustotách pravdepodobnosti $p(\mathrm{x}/\omega_i)$. Často totiž prepokladáme, že pravdepodobnosti $p(\mathrm{x}/\omega_i)$ sa riadia známym parametrickým rozdelením, najčastejšie normálnym, t.j

$$\mathrm{p}(\mathbf{x}|\omega_i) = \frac{1}{\sqrt{(2\pi)^d\, |\Sigma_i|}} \exp^{-\frac{1}{2}(x-\mu_i)^{\mathrm{T}} \Sigma_i^{-1} (x-\mu_i)} \qquad (13)$$

kde μ_i je vektor stredných hodnôt triedy ω_i, Σ_i je kovariančná matica triedy ω_i a d je dimenzia vektorového priestoru objektu x. V takomto prípade je možné všeobecné tvary uvedených mier zjednodušiť na nasledujúce výrazy :

$$\mathrm{J}_C = \frac{1}{2}\alpha(1-\alpha)(\mu_2-\mu_1)^{\mathrm{T}} \left[(1-\alpha)\Sigma_2 + \alpha\Sigma_1\right]^{-1} (\mu_2-\mu_1) + \\ + \frac{1}{2} \ln \frac{\left|(1-\alpha)\Sigma_2 + \alpha\Sigma_1\right|}{|\Sigma_1|^{(1-\alpha)} + |\Sigma_2|^{\alpha}} \qquad (14)$$

$$\mathrm{J}_B = \frac{1}{8}(\mu_2-\mu_1)^{\mathrm{T}} \left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1} (\mu_2-\mu_1) + \\ + \frac{1}{2} \ln \frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1| + |\Sigma_2|}} \qquad (15)$$

$$\mathrm{J}_D = \frac{1}{2}(\mu_2-\mu_1)^{\mathrm{T}} \left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)(\mu_2-\mu_1) + \\ + \frac{1}{2} tr\left(\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 - 2I\right) \qquad (16)$$

Doposiaľ sme uvažovali o klasifikácii objektov do dvoch tried. V prípade, ak máme daných R tried $\omega_1, \omega_2, \ldots, \omega_R$, s apriórnymi pravdepodobnosťami $P(\omega_i)$ a hustotami $p((\mathrm{x}/\omega_i)$ môžeme vzťahy (14) až (16) zovšeobecniť pomocou váženého súčtu vzdialeností každej dvojice tried.
Ak označíme vzdialenosť medzi triedami ω_i a ω_j ako J_{ij}, môžeme súhrnnú vzdialenosť medzi všetkými triedami $\omega_1, \omega_2, \ldots, \omega_R$ vyjadriť ako

$$\mathrm{J} = \sum_{i=1}^{C} \sum_{j=1}^{C} P(\omega_i) P(\omega_j)\, \mathrm{J}_{ij}. \qquad (17)$$

6. ANALÝZA VÝSLEDKOV

Uvedené postupy výberu príznakov boli aplikované na trénovacie množiny s obrazmi z diaľkového prieskumu povrchu Zeme, ktoré mapovali oblasť Košíc. Jednotlivé vzorky boli charakterizované 7-rozmernými vektormi, ktorých zložky predstavovali jas v siedmich spektrálnych pásmach. Vstupný súbor dát obsahoval vzorky patriace do siedmich tried: staré mesto, nové mesto, poľnohospodárska pôda, nevyužitá pôda, banská oblasť, les a voda.
Ako kriteriálne funkcie pre hodnotenie vybraných príznakov bola použitá Bhattacharyyho miera a divergencia.

Príznak číslo	Hodnota krit. funkcie	Pôvodný príznak
1	55,7961	3
2	43,7569	2
3	39,8534	5
4	38,2734	7
5	28,6181	1
6	7,5426	6
7	6,7894	4

Tab. **1** Hodnoty **divergencie** metódy *"d najlepších príznakov"*
Tab. 1 Divergence values of *"d best features"* method

Pri výbere jediného informatívneho príznaku obe metódy, aj pri rozdielnych kriteriálnych funkciách, poskytovali rovnaké výsledky. So zvyšujúcim sa počtom *d*- vybraných príznakov sa výsledky mierne odlišovali. Tab. 1 uvádza výsledky metódy "*d najlepších príznakov*" s uvedením hodnôt divergencie ako kriteriálnej funkcie a s uvedením poradia príznakov od najinformatívnejšieho počnúc. Tab. 2 uvádza výsledky získané *metódou úplného prehľadávania* hodnotené rovnakou kriteriálnou funkciou.

Počet príznakov	Hodnota krit. funkcie	Vybrané príznaky
1	55,796	3
2	75,256	3, 6
3	96,649	2, 3, 6
4	111,471	3, 4, 6, 7
5	127,806	2, 3, 4, 6, 7
6	139,294	1, 2, 3, 4, 6, 7
7	147,872	1, 2, 3, 4, 5, 6, 7

Tab. **2** Výsledky *metódy úplného prehľadávania*
Tab. 2 Results of exhaustive search method

Pre overenie vhodného výberu podmnožín príznakov bola použitá zhlukovacia metóda CLASS, pomocou ktorej boli vzorky zatrieďované do tried a to na základe vybraných podmnožín príznakov.
Pri porovnaní s pôvodnými triedami vzoriek sa podľa očakávania potvrdila ako spoľahlivejšia metóda úplného prehľadávania. Je potrebné však uviesť, že výsledky metódy "*d najlepších príznakov"* boli uspokojivé, pretože chyba zhlukovania bola zrovnateľná s chybou, ktorú vykazovalo zhlukovanie na základe podmnožín priznakov vybraných metódou *úplného prehľadávania.* V tabuľke 3 sú uvedené chyby zhlukovania pri rôznom počte príznakov, vybraných jednotlivými metódami, kde E je chyba zhlukovania na základe všetkých príznakov metódou CLASS .

Počet príznakov	1	2	3	4	5	6	7
E= 1,28							
d najlepších príznakov							
Divergencia	0,06	0,09	0,48	0,69	0,79	0,83	1,28
Bhattacharyya	0,06	0,29	0,33	0,42	0,79	0,83	1,28
úplné prehľadávanie							
Divergencia	0,06	0,10	0,13	0,77	0,80	0,88	1,28
Bhattacharyya	0,06	0,10	0,15	0,72	0,85	1,24	1,28

Tab. **3** Priemerná chyba klasifikácie zhlukovacou metódou CLASS
Tab. 3 Average error of classification with cluster method CLASS

LITERATÚRA

[1] Hual, L.-Hiroshi, M.: Feature Extraction, Construction and Selection. Kluwer Academic Publishers, Boston, 1998, 440 pp.

[2] Ocelíková, E. - Zolotová, I.: Problem of High Data Dimension at their Classification. In: Proc. of 31th International Conference MOSIS 97, Hradec nad Moravicí, 1997, pp.92-97. ISBN 80-85988-16-X.

[3] Ocelíková, E. - Kočiš, M.: Porovnanie metód selekcie a metód extrakcie príznakov pri znižovaní dimenzie mnohorozmerných obrazových priestorov. In: Proc. of the 2nd International Conference „Informatics and Algorithms ´98“, Prešov, Sept. 3-4, 1998, pp.301-304. ISBN 80-96-593-8-8.

[4] Pudil, P. – Novovičová, J. - Kitler, J.: Floating Search Methods in Feature Selection. *Pattern Recognition Letters,* Vol.15,No. 11. 1994, pp. 1119-1125.

BIOGRAPHY

Eva Ocelíková defended her Ph.D. thesis, which dealt with multicriterial classification of situations in the complex system, in 1985 at the Slovak Technical University of Bratislava. She is working at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice as associate professor. Her research work includes problems of decision processes, especially the problems of multicriterial classifications designing and high dimensionality reduction of feature space of multidimensional data in decision. She is group leader of the VEGA project No. 1/6061/99: "Pattern Recognition on the Basis of the Intelligent and Information Technologies".