Obsah

PŘEDMLUVA 9

I. ZÁKLADNÍ POJMY SHLUKOVÉ ANALÝZY 11

1. Co je shluková analýza 11
1.1. Úvodem malý příběh 11
1.2. Vývoj metod shlukové analýzy 13
1.3. Klasifikace 14
1.4. Některé pokusy definovat shlukovou analýzu 15
1.5. Cíle a možnosti shlukové analýzy 16
1.6. Typy metod shlukové analýzy 17
1.7. Pojem shluku 17
2. Matice dat 18
2.1. Znaky a objekty 18
2.2. Typy znaků a jejich kódování 19
2.3. Standardizace dat 20
2.4. Normalizace objektů 21
2.5. Vzájemná závislost a nezávislost znaků 22
3. Geometrický model matice dat 23
3.1. Modely objektů a modely znaků 23
3.2. Hlavní komponenty 26
3.3. Procedura COMP 29
3.4. Dva příklady aplikace procedury COMP 35
4. Podobnostní vztahy mezi objekty 41
4.1. Přístupy k hodnocení podobnostních vztahů mezi objekty 41
4.2. Koeficienty asociace 42
4.3 Metriky 47
4.4. Koeficient korelace 51
4.5. Některé jiné způsoby vyjádření podobnostních vztahů objektů 53
4.6. Koeficient nepodobnosti objektů 55
4.7. Některé pokusy o definici shluku na základě koeficientu nepodobnosti objektů 60

II. METODY HIERARCHICKÉHO SHLUKOVÁNÍ 62

5. Hierarchická aglomerativní shlukovací procedura 62
5.1. Obecné vlastnosti hierarchických aglomerativních shlukovacích metod 62
5.2. Rekurzivní definice metody [HASP, *D*] 64
5.3. Koeficient nepodobnosti shluků 65
5.4. Nejznámější metody zavedení koeficientu nepodobnosti shluků 67
5.4.1. Metoda nejbližšího souseda 67
5.4.2. Metoda nejvzdálenějšího souseda 68
5.4.3. Centroidní metoda 68
5.4.4. Metoda průměrné nepodobnosti objektů 70
5.4.5. Medianová metoda 70

5.4.6. Lanceova a Williamsova „pružná strategie“ . . . 71
5.4.7. Wardova–Wishartova metoda . . . 72
5.5. Dvě uni-míry nepodobnosti shluků . . . 73
5.5.1. Hallova míra nehomogenity . . . 73
5.5.2. Objemová míra nepodobnosti . . . 74
5.5.3. Vlastnosti uni-měr nepodobnosti shluků . . . 75
5.6. Shlukování metodou [HASP, D] . . . 76
5.6.1. Procedura HASP . . . 76
5.6.2. Příklad použití procedury HASP . . . 83
6. Definitní hierarchická aglomerativní shlukovací procedura . . . 88
6.1. Definitní hierarchická aglomerativní shlukovací procedura a pojem μ-souvislých shluků . 88
6.2. Rekurzivní definice metody [DHASP, D] . . . 88
6.3. Koeficienty nepodobnosti shluků vytvořených metodou [DHASP, D] . . . 90
6.3.1. Koeficient $[Dnn/d]$ . . . 90
6.3.2. Koeficient $[Dfn/d]$ . . . 91
6.3.3. Koeficient $[Dc/d_E^2]$ . . . 91
6.3.4. Koeficient $[Dga/d]$ . . . 91
6.3.5. Koeficient $[Dm/d_E^2]$ . . . 92
6.3.6. Koeficient $[Dlw/d]$ . . . 92
6.3.7. Koeficient $[Dw/d_E^2]$ . . . 93
6.3.8. Uni-míry nepodobnosti shluků . . . 94
6.4. Shlukování metodou [DHASP, D] . . . 94
6.4.1. Procedura AGLO . . . 94
6.4.2. Shlukování dat zatížených chybou procedurou AGLO . . . 103
6.4.3. Příklady použití procedury AGLO . . . 105
7. Divizivní metody hierarchického shlukování . . . 117
7.1. Obecné vlastnosti hierarchických divizivních metod . . . 117
7.2. MacNaughton–Smithova divizivní metoda . . . 118
7.2.1. MacNaughton–Smithův algoritmus . . . 118
7.2.2. Algoritmus DIVI . . . 118
7.2.3. Rekurzivní definice metody [DIVI, d] . . . 123
7.2.4. Shlukování procedurou DIVI . . . 124
7.2.5. Příklady použití procedury DIVI . . . 126

III. METODY NEHIERARCHICKÉHO SHLUKOVÁNÍ . . . 133

8. Metody hledající optimální rozklad množiny objektů . . . 133
8.1. Funkcionály kvality rozkladu . . . 133
8.2. Problém počátečního rozkladu . . . 136
8.3. Metody zachovávající stanovený počet shluků . . . 141
8.3.1. Forgyova a Janceyova shlukovací metoda . . . 141
8.3.2. Definice Forgyovy a Janceyovy shlukovací metody . . . 141
8.3.3. Procedury FORG a JANC . . . 143
8.3.4. MacQueenova k-průměrová metoda a její konvergentní varianta . . . 148
8.3.5. Definice MacQueenovy a Wishartovy metody . . . 148
8.3.6. Procedury WISH a MACQ . . . 150
8.3.7. Optimalizace kritéria součtu čtverců chyb a konvergence uvedených metod . . . 152
8.3.8. Příklad použití metod, zachovávajících daný počet shluků . . . 153
8.4. Metody měnící počet shluků . . . 160
8.4.1. MacQueenova metoda se dvěma parametry . . . 160
8.4.2. Wishartova metoda RELOC . . . 162
8.4.3. Metoda ISODATA . . . 164

8.4.4. Metoda CLASS 167
8.4.5. Shlukování procedurou CLAS 175
9. Analýza modů 177
9.1. Frekvenční funkce náhodné veličiny a její modus 177
9.2. Frekvenční funkce mnohorozměrného náhodného vektoru a její modus 178
9.3. Hledání hustých oblastí prostoru modelů 180
9.4. Kittlerova metoda hledání modů 180
9.4.1. Metoda MODESP 180
9.4.2. Algoritmus a procedura ANMO 182
9.4.3. Příklad použití procedury ANMO 182

IV. DOPORUČENÍ UŽIVATELŮM SHLUKOVÉ ANALÝZY 186

10. Co a jak shlukovat 186
10.1. Tři problémy řešitelné metodami shlukové analýzy 186
10.2. Diskuse k výsledkům shlukové analýzy dat BRAN, SANQ a NUMT 187
11. Programový systém SHLUK 188
11.1. Základní verze systému 188
11.2. Užiti systému 191
11.3. Realizace programu 197
11.3.1. Popis vstupních dat 198
11.3.2. Seznam identifikátorů 198
11.3.3. Popis výstupů 199
11.3.4. Doba trvání 201
11.4. Systém SHLUK pro rozsáhlá data 202

LITERATURA 203

REJSTŘÍK 208