LMD mokslo darbai, 249–255
1998, Vilnius

Matematikos ir informatikos studijų VPU rezultatų analizė

G. Dzemyda, G. Leonavičius (VPU), A. Našlėnas (MII)

1. Įvadas

Vykdant mokyklos reformą bei jos kompiuterizacijos projektą vis reikšmingesnis tampa informatikos mokytojas. Todėl iškyla būtinybė paanalizuoti informatikos mokytojų rengimo ypatumus. Vilniaus pedagoginiame universitete (VPU) informatikos mokytojai buvo pradėti rengti matematikos fakultete nuo 1993 metų. Iki 1996 metų informatika buvo studijuojama matematikos specialybės studentų kaip pasirenkamoji (gretutinė), siekiant be matematikos bakalauro matematikos mokytojo kvalifikacijos įgyti ir informatikos mokytojo kvalifikaciją. Informatikos mokymo programa yra gana glaudžiai susipynusi su matematikos programa, nes tarp matematikos ir informatikos egzistuoja tamprūs tarpdalykiniai ryšiai. Todėl nuo 1994 metų buvo sudaryta vieninga matematikos bakalauro, matematikos ir informatikos mokytojo ruošimo programa.

Duomenis analizei sudarė 1992-1997 metų bėgyje studentų bakalaurinių studijų metu išlaikytų 25 egzaminų rezultatai. Studentai, nepasirinkę informatikos mokytojo specializacijos, laikė 6 egzaminais mažiau. Domėn buvo priimti tik studentų, sėkmingai baigusių studijas, egzaminų rezultatai. Šių rezultatų pagrindu apskaičiuotos disciplinų koreliacinės matricos.

Lentelė 1. Disciplinų sąrašas

Nr	Disciplina	Nr	Disciplina
1	Tikimybių teorija	13	Geometrija
2	Matematka ir jos dėstymo metodika (valstybinis egzaminas)	14	Psichologija
3	Geometrija	15	Algebra ir skaičių teor.
4	Pedagogika ir psichologija (valstybinis egz.)	16	Matematinė analizė
5	Geometrija	17	Informatika
6	Mat.analizė	18	Algebra ir skaičių teorija
7	Pedagogika	19	Matematinė analizė
8	Geometrija	20	Informatika (valstybinis egz.)
9	Psichologija	21	Mokomųjų pro-gramų kūrimas
10	Algebra	22	Informatikos dėstymo metodika
11	Matematinė analizė	23	Matematinių uždavinių sprendimo paketai
12	Užsienio kalba	24	Algoritmai
		25	Programavimo metodai

Disciplinų koreliacinių matricų analizei taikytas grupavimas (klasterizavimas) ir faktorinė analizė. Gauti rezultatai leidžia grupuoti disciplinas ir atskleisti tarpdisciplininius ryšius, įvertinant studentų gabumus ir mokymosi rezultatus.

Koreliacinėje analizėje dydžiai, kurių koreliacinės matricos yra analizuojamos, dažniausiai vadinami parametrais arba kintamaisiais.

Apskaičiuotos dvi koreliacinės matricos:

- 19 disciplinų, kurių egzaminus laikė visi tiriami studentai, koreliacinė matrica;
- 25 disciplinų, kurių egzaminus laikė studentai, pasirinkę informatikos mokytojo specializaciją, koreliacinė matrica.

Disciplinų sąrašas ir jas atitinkantys numeriai pateikti Lentelėje 1. 1-19 numeriais pažymėtų disciplinų egzaminus laikė visi studentai, o 20-25 numeriais pažymėtų disciplinų egzaminus laikė tik studentai, pasirinkę informatikos mokytojo specializaciją. Didesni numeriai tarp 1-19 ir 20-25 atitinka anksčiau laikytus egzaminus.

2. Parametrų grupavimo algoritmų taikymas

2.1. Parametrų grupavimo algoritmai. Darbuose [1]–[5] yra pasiūlytos bei ištyrinėtos ekstremalinio parametrų (kintamųjų) grupavimo (klasterizavimo) strategijos. Tikslas yra išskirti žinias, glūdinčias parametrų $x_1,\dots,x_n$ koreliacinėje matricoje $R=\{r_{x_i x_j}, i,j=\overline{1,n}\}$, apie jų grupavimąsi. Tyrinėti metodai yra skirti koreliacinių matricų analizei, t.y., parametrų grupavimui į fiksuotą skaičių p nesikertančių grupių $A_1,\dots,A_p$ tų parametrų koreliacijų pagrindu. Sukurtos parametrų grupavimo deterministinės strategijos [1], [2], [4] yra grindžiamos nuoseklia parametrų peržiūra ir grupės, į kurią reikia perkelti konkretų parametrą, nustatymu naudojantis tam tikrais kriterijais ir siekiant maksimizuoti funkcionalą $I_1(A_1,\dots,A_p)=\sum_{L=1}^{p}\sum_{x_i\in A_L} r^2_{x_i F_L}$, čia F_L yra faktorius, kurio dispersija lygi vienam ir kuris turi maksimizuoti $\sum_{x_i\in A_L} r^2_{x_i F_L}, L=\overline{1,p}$; $r_{x_i F_L}$ yra x_i ir F_L koreliacijos koeficientas.

Deterministinių algoritmų tyrimas [4] parodė, kad geresnės grupavimo efektyvumą aprašančio funkcionalo I_1 reikšmės yra gaunamos naudojant algoritmus, kurie skaičiuoja koreliacinių matricų $R_L=\{r_{x_i x_j}, x_i, x_j\in A_L\}$ maksimalias nuosavas reikšmes (algoritmas A2 [2], [4]). Darbe [4] pasiūlyti du parametrų pradinio suskaidymo algoritmai H1 ir H2. Deterministiniai algoritmai dažnai randa tik lokalinį I_1 maksimumą. Todėl darbe [3] yra pasiūlyta parametrų grupavimo uždavinį formuluoti kaip kombinatorinio optimizavimo uždavinį ir ieškoti globalinio I_1 maksimumo. Tuo būdu sukurta keletas modeliuojamo atkaitinimo (simulated annealing) algoritmų (SA1, SA2, SA3, SB1, SB2 ir SB3). Modeliuojamo atkaitinimo algoritmuose pastebėtas ryškus parametrų suskaidymo į grupes gerėjimas pradinėse modeliuojamo atkaitinimo algoritmų iteracijose. Todėl tikslinga naudoti keletą modeliuojamo atkaitimo algoritmo iteracijų pradiniam parametrų suskaidymui, o po to naudoti deterministinius algoritmus. Šiame darbe bus naudojamos dvi algoritmo SA3 iteracijos.

Darbe [4] atskleistas ryšys tarp parametrų grupavimo jų koreliacinės matricos pagrindu ir vienetinės sferos S^n vektorių klasterizavimo, leidžiantis naudoti paprastesnius parametrų grupavimo efektyvumą aprašančius funkcionalus. Tam, kad parametrų grupavimui būtų galima naudoti vektorių (objektų) klasterizavimo funkcionalus, įrodyta galimybė įvesti vektorių sistemą $Y_1,\ldots,Y_n \in S^n$, t.y. $\|Y_i\| = 1, i = \overline{1,n}$, kuri taip siejasi su parametrų sistema $x_1,\ldots,x_n$: $\cos(Y_i, Y_j) = r^2_{x_i x_j}$. Vektoriai $Y_1,\ldots,Y_n \in S^n$ gali būti klasterizuojami naudojant klasikinių klasterizavimo algoritmų (pvz. K-means) modifikacijas. Pažymėkime šio tipo algoritmą KM. Tai K-means algoritmo [6] realizacija. Žinoma, kad jos grupavimo gerumo kriterijus skirsis nuo I_1 (žr. [4]).

Koreliacinės matricos analizuotos naudojantis dviem strategijomis:

STR1) pradinis suskaidymas naudojantis algoritmu A1, o po to SA3; funkcionalo I_1 maksimizavimas algoritmu A2;

STR2) funkcionalo I_1 maksimizavimas strategija STR1, o po to panaudojimas algoritmo KM.

2.2. Disciplinų grupavimo rezultatai. Analizės rezultatai pateikiami lentelėse 2 ir 3.

Disciplinų, kurių egzaminus laikė visi studentai, analizė. Disciplinų skaidymas į dvi grupes ryškiai atskiria matematines disciplinas nuo nematematinių. Deja, bet informatika pakliūna į nematematinių disciplinų grupę (kartu su pedagogika, psichologija ir užsienio kalba). Informatika lieka vienoje grupėje su nematematinėmis disciplinomis, kuomet disciplinos skaidomos ir į 3, 4, 5 ar 6 grupes. Tik skaidant į 7 ir daugiau grupių informatika sudaro savarankišką grupę. Suskaidymuose į 6 ir 7 grupes matomos dvi stambesnės matematinių disciplinų grupės – matematika vyresniuosiuose kursuose (disciplinų grupė 1, 2, 3, 5, 6, 11) ir matematika jaunesniuosiuose kursuose (10, 15, 16, 18, 19). Pradedant suskaidymu į 5 grupes, savarankišką grupę sudaro pedagogikos ir psichologijos valstybinis egzaminas. Tai atspindi VPU pedagoginę orientaciją.

Disciplinų, kurių egzaminus laikė studentai, pasirinkę informatikos mokytojo specializaciją, analizė. Disciplinų skaidymas į dvi grupes, naudojant antrąją strategiją irgi išskiria tradicines matematines disciplinas - prie jų prisišlieja tik algoritmų disciplina, tačiau naudojantis pirmąja strategija, atskiros grupės pagrindą sudaro valstybiniai egzaminai. Ši grupė išlieka ir skaidant į didesnį grupių skaičių abiem strategijomis – tik esant didesniam grupių skaičiui, informatikos disciplinos atskyla, sukurdamos atskirą grupę. Visais atvejais pastebima tendencija grupuotis nematematinėms disciplinoms su informatikos disciplinomis. Matematinių disciplinų grupavimasis į jaunesniųjų ir vyresniųjų kursų grupes išlieka kaip ir anksčiau aprašytu visų studentų laikytų egzaminų atveju.

3. Faktorinės analizės taikymas

3.1. Centroidų metodas. Tarkime, turime atsitiktinių dydžių (parametrų) rinkinį $x_1, x_2, ..., x_n$. Faktorinė analizė iškelia hipotezę, kad yra tokie tarpusavyje nepriklausomi atsitiktiniai dydžiai $e_1, e_2, ..., e_n$, kad rinkinys $x_1 - e_1, x_2 - e_2, ..., x_n - e_n$ nuo jų nepriklauso. Be to, yra tokie atsitiktiniai dydžiai (faktoriai) $f_1, f_2, ..., f_k$, kad $x_i - e_i = \sum_{j=1}^{k} l_{ij} f_j$ ir $k<n$.

l_{ij} vadinami j-tojo faktoriaus svoriu i-tajam parametrui.

Vienas iš dažniausiai praktikoje naudojamų faktorinės analizės metodų - centroidų metodas (kartais vadinamas paprasto sumavimo metodu) [1], [7]. Jo esmė tokia: tariame, kad parametrus $x_1, x_2, ..., x_n$ atitinka n-matės erdvės vektoriai, kurių ilgiai - atitinkamų parametrų standartiniai kvadratiniai nuokrypiai, o kampo tarp vektorių kosinusai lygūs atitinkamų parametrų koreliacijoms.

Jeigu reikia, laikinai pakeičiame vektorių kryptis, kad kuo daugiau koreliacijų taptų teigiamomis. Tada vektoriai turės tendenciją grupuotis viena kryptimi į pluoštą. Tuomet pirmasis faktorius bus apibrėžiamas kaip vektorių pluošto suma ir eis (tam tikra prasme) per pluošto vidurį (centrą). Svoriai bus vektorių projekcijų faktoriuje ilgiai. Atmetus rastojo faktoriaus poveikį, analogiškai ieškomas kitas faktorius.

Lentelė 2. 19 disciplinų, kurių egzaminus laikė visi tiriami studentai, analizė

Grupių skaičius	Strategija	Grupės							
2	STR1	4, 7, 9, 12, 14, 17		1, 2, 3, 5, 6, 8, 10, 11, 13, 15, 16, 18, 19					
2	STR2	4, 7, 9, 12, 14, 17		1, 2, 3, 5, 6, 8, 10, 11, 13, 15, 16, 18, 19					
3	STR1	4, 7, 9, 12, 17		2, 3, 5, 6, 10, 11, 15, 16, 18, 19				1, 8, 13, 14	
3	STR2	4, 7, 9, 12, 17		2, 5, 6, 10, 11, 15, 16, 18, 19				1, 3, 8, 13, 14	
4	STR1	4, 9, 12		7, 14, 17		5, 10, 11, 15, 16, 18, 19		1, 2, 3, 6, 8, 13	
4	STR2	4, 5, 9, 12		7, 14, 17		10, 11, 15, 16, 18, 19		1, 2, 3, 6, 8, 13	
5	STR1	4	9, 12	7, 14, 17		5, 10, 11, 15, 16, 18, 19		1, 2, 3, 6, 8, 13	
5	STR2	4	5, 9, 12	7, 14, 17		10, 11, 15, 16, 18, 19		1, 2, 3, 6, 8, 13	
6	STR1	4	9, 12	7, 17		8, 13, 14	10, 15, 16, 18, 19	1, 2, 3, 5, 6, 11	
6	STR2	4	9, 12	7, 17		8, 13, 14	10, 15, 16, 18, 19	1, 2, 3, 5, 6, 11	
7	STR1	4	9, 12	7	17	8, 13, 14	10, 15, 16, 18, 19	1, 2, 3, 5, 6, 11	
7	STR2	4	9, 12	7	17	8, 13, 14	10, 15, 16, 18, 19	1, 2, 3, 5, 6, 11	
8	STR1	4	9, 12	7	17	8, 13, 14	10, 15, 16, 18, 19	1	2, 3, 5, 6, 11
8	STR2	4	9, 12	7	17	8, 13, 14	15, 16, 18, 19	1, 3	2, 5, 6, 10, 11

Lentelė 3. 25 disciplinų, kurių egzaminus laikė studentai, pasirinkę informatikos mokytojo specializaciją, analizė

Grupių skaičius	Strategija	Grupės							
2	STR1	4, 20, 21	1- 3, 5-19, 22-24, 25						
2	STR2	1-3, 5, 6, 8, 10, 11, 13, 15, 16, 18, 19, 24	4, 7, 9, 12, 14, 17, 20-23, 25						
3	STR1	1-3, 5, 6, 8, 10, 11, 13, 15-19, 24, 25	4, 20, 21	7, 9, 12, 14, 22, 23					
3	STR2	1-3, 5, 6, 8, 10, 11, 15, 16, 18, 19, 24, 25	4, 20, 21	7, 9, 12-14, 17, 22, 23					
4	STR1	1-3, 5, 6, 9-11, 15, 24	4, 20, 21	7, 14, 22, 23	8, 12, 13,16-19, 25				
4	STR2	1-3, 5, 6, 10, 11, 15, 24	4, 20, 21	7, 9, 12, 17, 22, 23, 25	8, 13, 14,16, 18, 19				
5	STR1	1-3, 5, 6, 10, 11, 15, 24	4, 20, 21	7, 14, 22, 23	9, 12	8, 13, 16-19, 25			
5	STR2	1-3, 5, 6, 10, 11, 15, 24	4, 20, 21	7, 14, 22, 23	9,12,17	8, 13, 16, 18, 19, 25			
6	STR1	1-3, 5, 6, 10, 11, 15, 24	4	20, 21	7, 14, 22, 23	9, 12	8, 13, 16-19, 25		
6	STR2	1-3, 5, 6, 10, 11, 15, 24	4, 20, 21	7, 14, 22, 23	9, 12	17, 25	8,13,16,18,19		
7	STR1	1-3, 6, 11, 13	5, 10, 15, 16, 24	4	20, 21	7, 14, 22, 23	9, 12	8, 17-19, 25	
7	STR2	1-3, 6, 11	5, 10, 15, 24	4, 20, 21	7, 14, 22, 23	9, 12	17, 25	8,13,16,18,19	
8	STR1	1-3, 6, 11	5, 10, 15, 24	4	20, 21	7, 14, 22, 23	9, 12	17, 25	8,13,16,18,19
8	STR2	1-3, 6, 11	5, 10, 15, 24	4	20, 21	7, 14, 22, 23	9, 12	17, 25	8,13,16,18,19

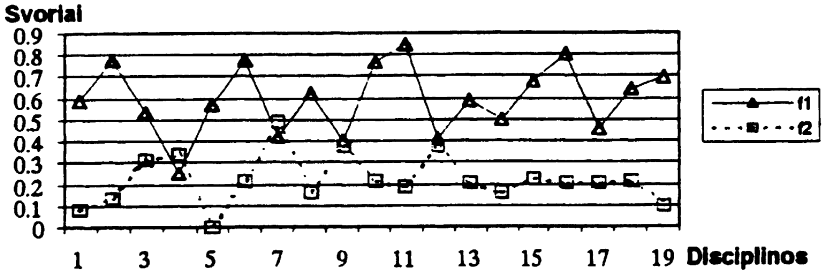
3.2. Disciplinų faktorinės analizės rezultatai. Pritaikę centroidų metodą disciplinų analizei, išskyrėme du faktorius (f_1 ir f_2), kurių įtaka pateikta lentelėje 4. Abiejų faktorių svorių pasiskirstymą kiekvieni disciplinai iliustruoja paveikslai 1 ir 2. Pagal gautus duomenis, atsižvelgdami į faktorių įtaką kiekvienai disciplinai, jas galima sugrupuoti. Šį sugrupavimą iliustruoja lentelė 5.

Lentelė 4. Faktorių f_1 ir f_2 įtaka visų faktorių sistemos atžvilgiu

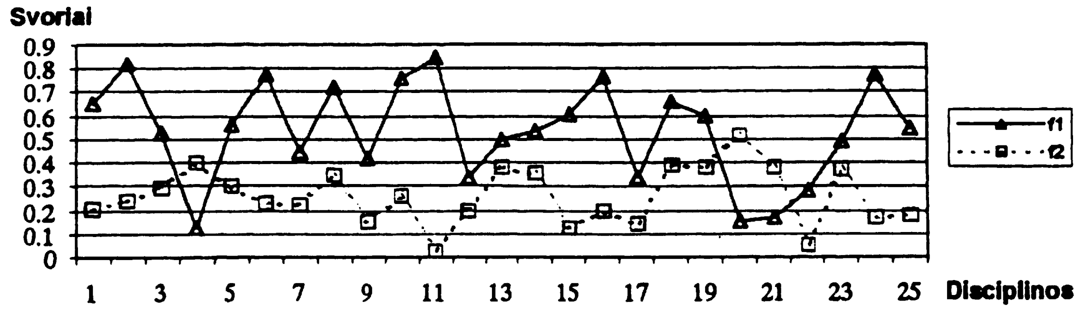
Faktorius	Visiems studentams	Studentams, pasirinkusiems informatiką
f_1	37.9 %	33.1 %
f_2	6.1 %	8.3 %

Lentelė 5. Disciplinų grupavimas pagal faktorių įtaką (maksimali, minimali, vidutinė)

	Visiems studentams		Studentams, pasirinkusiems informatiką	
Faktorius	Įtaka	Disciplinos	Įtaka	Disciplinos
f_1	max	2,6,10,11,15,16,19	max	1,2,6,8,10,11,15,16,18,19,24
	min	4,7,9,12,14,17	min	4,7,9,12,17,20,21,22
	vid.	1,3,5,8,13,18	vid.	3,5,13,14,23,25
f_2	max	3,4,7,9,12	max	4,8,13,14,18,19,20,21,23
	min	1,2,5,19	min	11,22
	vid.	6,8,10,11,13,15,16,17,18	vid.	1,2,3,5,6,7,9,10,12,15,16,17,24,25



Pav. 1. Faktorių svorių pasiskirstymas, analizuojant visų studentų egzaminų rezultatus.



Pav. 2. Faktorių svorių pasiskirstymas, analizuojant studentų, pasirinkusių informatikos mokytojo specializaciją, egzaminų rezultatus.

Esminę įtaką turi pirmasis faktorius (f_1). Didžiausius svorius šiame faktoriuje turi matematinės disciplinos. Nagrinėjant disciplinas, kurių egzaminus laikė informatiką pasirinkę studentai, pastebime, kad prie didžiausius svorius turinčių disciplinų prisijungia tik viena informatikos grupės disciplina – algoritmai. Mažiausią įtaką turi nematematinės disciplinos (pedagogika, psichologija, užsienio kalba). Į šią grupę patenka ir dalis informatikos disciplinų. Antrasis faktorius (f_2) savo esmingumu žymiai skiriasi nuo pirmojo faktoriaus, o jo prasmę yra sunku interpretuoti, kadangi negalima įžvelgti ryškesnių grupavimosi tendencijų.

4. Išvados

Pagrindinis šio tyrimo rezultatas yra tai, kad pasiūlytą metodiką (studentų egzaminų įverčių koreliacijų analizę specialiais statistinės duomenų analizės metodais) galima sėkmingai taikyti studijų programų analizei, jų kokybei tirti ir vertinti.

Tyrimai parodė, kad informatikos kursas VPU studentams, pradėjusiems studijas 1992 ir 1993 metais, nebuvo toks vienalytis, kaip matematikos, kuris turi gilias ilgametes tradicijas. Tam tikra prasme tai galima paaiškinti tuo, kad 1992-1993 mokslo metai – tai informatikos mokytojų ruošimo VPU pradžia, o informatika – viena iš labiausiai besivystančių mokslo šakų. Tyrimo metu pastebima tendencija grupuotis nematematinėms disciplinoms su informatikos disciplinomis rodo, kad šių disciplinų dėstymą reikėtų labiau matematizuoti, nes matematika ir informatika yra glaudžiai susijusios mokslo šakos. Taip pat tai įrodo, kad informatikos disciplinos pasižymėjo daugiau taikomaisiais bruožais, nei teoriniais. Todėl tikslinga mokymo programas papildyti naujomis teorinės informatikos disciplinomis.

LITERATŪRA

[1] Braverman, E.M., and I.B. Muchnik (1983). *The Structural Methods for Empirical Data Processing.* Nauka, Moscow (in Russian).

[2] Dzemyda, G. (1996). Clustering of parameters on the basis of correlations via simulated annealing. *Control and Cybernetics*, **25** (1), R.V.V.Vidal and Z.Nahorski (Eds.), *Special Issue on: Simulated Annealing Applied to Combinatorial Optimization*, 55-74.

[3] Dzemyda, G. (1997). Clustering of parameters on the basis of correlations: A comparative review of deterministic approaches. *Informatica*, 8(1), 83–118.

[4] Späth. H. (1980). *Cluster Analysis Algorithms for Data Reduction and Classification of Objects.* Ellis Horwood, Chichester.

[5] Lawley, D. N., and A. E. Maxwell (1963). *Factor Analysis as a Statistical Method.* Butterworths, London.

Analysis of the mathematics and informatics studies in Vilnius Pedagogical University

G. Dzemyda, G. Leonavičius, A. Našlėnas

The article deals with application of statistical methods (parameter clustering and factor analysis) to discover the knowledge on interaction of disciplines via analysis of the results of examinations. The experiments indicate that interactions of mathematical disciplines with disciplines of informatics are weak. This leads to necessary revision of the programs of studies.