

DAUGELIO ELEMENTŲ PASKIRSTYMO PROBLEMA: PIRMUMO STRATEGIJOS ANALIZĖ

Gediminas Murauskas¹, Marijus Radavičius²

Vilniaus Universitetas. Adresas: Naugarduko g. 24, Vilnius, 03225, Lietuva.
El. paštas: ¹gediminas.murauskas@mif.vu.lt, ²marijus.radavicius@mii.vu.lt

Gauta: 2016 m. rugpjūtis Pataisyta: 2016 m. spalio Paskelbta: 2016 m. lapkritis

Santrauka. Straipsnyje nagrinėjamas atskiras daugelio elementų paskirstymo uždavinio atvejis – studentų dalykų pasirinkimo (paskirstymo) problema. Darbe analizuojama Vilniaus Universiteto bendrųjų universitetinių studijų įgyvendinimo praktika iki 2016 metų. Tiriamosios analizės rezultatai parodo, kad taikomas dalykų paskirstymas pirmumo principu nėra efektyvus. Norint palyginti dalykų paskirstymo algoritmus reikia turėti duomenų aibę su studentų preferencijomis. Darbe, naudojant cenzūruotos regresijos modelį, turimų realių duomenų pagrindu pasiūlytas studentų preferencijų generavimo algoritmas. Imituotų preferencijų duomenys panaudoti lyginant kelių plačiai taikomų praktikoje paskirstymo algoritmų charakteristikas.

Reikšminiai žodžiai: daugelio elementų paskirstymas, pirmumo principas, preferencijos, cenzūruotoji regresija.

1. Įvadas

Su daugelio elementų paskirstymo problema susidūrėme nagrinėdami Vilniaus universiteto (VU) studentų studijų dalykų pasirinkimo uždavinį. Ši problema yra svarbi ir iki šiol teoriškai neišspręsta, t. y. formaliai neegzistuoja efektyvi, teisinga ir strategijų atžvilgiu indiferentiška paskirstymo schema, jeigu resursai yra riboti ir patvirtinto pasirinkimo negalima pakeisti kitu. Straipsnyje nagrinėjama studentų dalykų pasirinkimo problema yra atskiras daugelio elementų paskirstymo uždavinio atvejis.

Studijų kokybės ir mokslo rezultatyvumo gerinimas yra vienos iš pagrindinių universitetų veiklos kryptių. Studijų kokybei įtakos turi studijų programų atnaujinimas, restruktūrizavimas, naujų studijų organizavimo formų paieška. Pagal pasirinkimo tipą universitete (-uose) studijų programos sudarytos iš dėstomųjų dalykų, kurie skirstomi į *privalomuosius* ir *pasirenkamuosius*. Natūralu, kad yra pasirenkamųjų dalykų kvotos ir ne visi studentai gali pasirinkti norimus dalykus. Pasirenkamųjų dalykų teisingas ir efektyvus paskirstymas yra svarbus uždavinys. Yra nemažai darbų, kuriuose aptariami dalykų paskirstymo algoritmai [7, 2, 6].

Straipsnyje nagrinėjama tik vienos pasirenkamųjų dalykų grupės – bendrųjų universitetinių studijų (BUS) – įgyvendinimo praktika Vilniaus universitete iki 2016 m. Naudojant sukauptus duomenis apie studentų BUS dalykų pasirinkimus, empiriškai įvertintos pasirinkimų registracijos pirmumo principu algoritmo charakteristikos, atlikta studentų registravimosi laiko tiriamoji analizė.

Panaudojus tiriamosios analizės rezultatus bei cenzūruotos regresijos modelį, sudarytas studentų preferencijų generavimo algoritmas. Jis buvo panaudotas imitaciniams duomenims su preferencijomis gauti, nes dalis dalykų paskirstymo algoritmų naudoja informaciją apie studentų preferencijas (tiesioginės informacijos apie studentų pageidavimus turimuose duomenyse nėra). Gauti duomenys buvo panaudoti lyginant pirmumo principu pagrįstą algoritmą su kitais žinomais dalykų paskirstymo algoritmais.

Kitame skyriuje trumpai aprašyti dalykų paskirstymo algoritmai ir jų savybės. Trečias skyrius skirtas turimų duomenų apie VU studentų BUS dalykų pasirinkimus tiriamajai analizei. Jos rezultatai taikomi 4 skyriuje konstruojant cenzūruotos regresijos modelį. Šiame skyriuje pirmumo principu pagrįstas algoritmas lyginamas su kitais žinomais dalykų paskirstymo algoritmais. Pabaigoje tyrimo rezultatai apibendrinami ir pateikiamos išvados.

2. Dalykų paskirstymo algoritmai

Nagrinėjama studentų dalykų pasirinkimo (paskirstymo) problema. Kaip jau minėta, šis uždavinys yra atskiras daugelio elementų paskirstymo (*multi-unit assignment*) uždavinio atvejis. Formaliai uždavinį galima užrašyti taip:

Studentų aibė $S = \{s_1, \dots, s_n\}$, kiekvienam studentui $s \in S$ reikia pasirinkti m dalykų¹. Dalykų aibė $C = \{c_1, \dots, c_k\}$ su atitinkamomis užpildymo kvotomis $Q = \{q_1, \dots, q_k\}$. Taip pat yra duota studentų pageidavimų (preferencijų) aibė $\{\succ_s, s \in S\}$: $c \succ_s c'$ reiškia, kad studentui s dalykas c yra labiau pageidautinas negu c' . Kiekvienam studentui, atsižvelgiant į jo preferencijas, reikia priskirti m dalykų taip, kad nebūtų viršytos dalykų kvotos. Papildomai gali būti apibrėžiama prioritetų tvarka: $\forall c \in C$ ir kiekvienai studentų porai (s', s'') egzistuoja ryšys $s' \succ_c s''$ arba $s'' \succ_c s'$.

Atliekant kiekybinę analizę vietoje preferencijų yra patogiau naudoti rangus $r_s(c), s \in S, c \in C$. Čia $r_s(c)$ yra studento s dalykui c priskiriamas rangas – dalyko numeris dalykų, išdėstytų preferencijų \succ_s mažėjimo tvarka, eilutėje. Bendru atveju rangai nebūtinai yra natūralieji skaičiai.

Resursų trūkumo sąlyga: egzistuoja bent vienas dalykas c_j , kurį daugiau kaip q_j studentų įtraukė į savo m dalykų pasirinkimo sąrašus.

Nagrinėjant dalykų paskirstymo algoritmus, vartojamos *teisingumo* ir *efektyvumo* sąvokos. Teisingumo reikalavimai formuluojami taip:

- indiferentiškumas strategijų atžvilgiu: studentas klaidingai nurodydamas savo preferencijas negali pagerinti pagal algoritmą jam priskiriamų dalykų rinkinio;
- „pavydo kriterijus“: kiekvienas studentas turi turėti galimybę pasirinkti dalykus tiek pat geriamus kaip ir kito studento pasirinkti dalykai.

Antrasis reikalavimas yra sunkiai įvertinamas, jei yra skirtingas pasirenkamų dalykų skaičius (taip yra VU atveju). Taisyklės a) ir b) nėra suderinamos daugiau kaip vieno dalyko pasirinkimo atveju.

Efektyvumas suprantamas taip: *nėra kito dalykų paskirstymo, kuriame visų studentų pasirinkimai yra ne „blogesni“ ir bent vieno studento yra „geresnis“*.

Vieno pasirinkimo nustatytu (paskirtu) laiku (*choose one at the time*) ir siūlomų taškų (*Bidding Points, BP*) algoritmai yra dažniausiai naudojami ir nagrinėjamos paskirstymo algoritmų grupės. Pirmosios grupės algoritmai dar vadinami rikiavimo diktatūros (*Serial Dictatorship, SD*) algoritmais. Daugelio elementų rikiavimo diktatūros algoritmas: pagal tam tikrą tvarką sudaroma studentų eilė. Pirmasis studentas gauna jo labiausiai pageidaujamą (-us) dalyką (-us), antrasis studentas gauna jo labiausiai pageidaujamus dalykus iš likusiųjų su neužpildytomis kvotomis ir t. t. Jei studentų eilė (tvarka) yra atsitiktinė, turime RSD (*Random SD*) algoritmą.

Prie rikiavimo diktatūros algoritmų priskirtinas ir FCFS (*First-Come First-Served*) algoritmas [4], kurio pavadinimas nusako jo esmę – anksčiau atėjęs, anksčiau ir aptarnaujamas.

Jei studentų tvarka yra atsitiktinė, bet 1, 3, 5, ... etapuose studentai tiesiogine tvarka renkasi po vieną dalyką, o 2, 4, 6, ... etapuose studentai renkasi po vieną dalyką atvirkštine tvarka, tada turime HBS algoritmą [2].

Antrosios grupės algoritmų tipinis pavyzdys yra atidėtojo priskyrimo (*Deferred Acceptance, DA*) algoritmas [8], kuris naudoja informaciją ir apie prioritetų tvarką $\succ_c, c \in C$:

1 žingsnis: Pirmiausia į dalykus registruojami studentai (jų pageidavimai) su aukščiausiais prioritetais. Jei kurio nors dalyko kvota viršijama, tai atmetami studentai, kurių pageidavimai turi žemiausią prioritetą, o neatmestų studentų pageidavimai priskiriami galutinai. Likę studentai registruojami tik preliminariai, jų pageidavimai gali būti atmesti kituose etapuose. Būtent ši galutinio priskyrimo atidėjimo taisyklė ir įvardyta algoritmo pavadinime.

Žingsnis t ($t \geq 2$): Kiekvienas neregistruotas studentas iš naujo dalyvauja registracijoje su naujais pageidavimais (preferencijomis). Lyginami visi studentai – ir naujai užsiregistravę, ir registruoti preliminariai. Jiems kartojama ta pati procedūra. Jei kurio nors dalyko kvota viršijama, tai atmetami studentai, kurių pageidavimai turi žemiausią prioritetą, neatmestų studentų pageidavimai priskiriami galutinai, o likę studentai registruojami preliminariai.

¹Nagrinėjamuose algoritmuose m laikomas pastoviu, nors praktikoje jis dažnai yra kintamas. Pavyzdžiui, Vilniaus universitete BUS dalykų, kuriuos studentas turi pasirinkti, skaičius priklauso nuo specialybės ir semestro.

Priskyrimo procesas baigiamas, kai nelieka atmestų studentų. Visi preliminariai registruoti studentai tampa registruotais.

Atidėtojo priskyrimo algoritmas taip pat žinomas kaip optimalus studentų paskirstymo stabilusis algoritmas [4].

Praktiniuose dalykų pasirinkimo uždaviniuose yra daug apribojimų, kurie ne visada tiksliai aprašomi modeliuose. Pavyzdžiui, tvarkaraščiai turi įtakos ne tik pasirinkimo kombinacijų skaičiui ir prioritetams, bet ir pasirinkimų keitimų būtinybei. Todėl teoriniai modeliai ir algoritmai yra tik „bazė“ praktiniams sprendimams. Tačiau formalūs kriterijai ir matai yra svarbūs, nes naudojant empirinius duomenis juos galima „objektyviai“ įvertinti tuo pagrindžiant vieno ar kito algoritmo (mechanizmo) privalumus ir trūkumus.

3. Studijų dalykų pasirinkimo problema Vilniaus universitete

Pagal pasirinkimo tipą niversitete (-uose) studijų programos sudarytos iš dėstomųjų dalykų, kurie skirstomi į *privalomuosius* ir *pasirenkamuosius*. Privalomieji dalykai sudaro pasirinktos studijų krypties išsilavinimo pagrindus, o pasirenkamieji užtikrina bendrąjį universitetinį lavinimą ir iš esmės papildo pasirinktos pagrindinės ir (ar) gretutinės studijų krypties (šakos) išsilavinimą. Pasirenkamuosius dalykus galima rinktis iš tam tikro sąrašo tuose semestruose, kuriuose jie yra numatyti pagal studijų planą. VU registracija į BUS dalyką (-us) vykdoma *pirmumo principu* (beje, šis pasirinkimo principas taikomas ir kituose universitetuose, pavyzdžiui LSMU). BUS tvarka VU buvo patvirtinta 2012 m. balandžio 19 d.

Pirmumo principą atitinka *FCFS algoritmas* daugelio elementų paskirstymo uždaviniuose [4]. Šioje dalyje aprašomos registracijų į BUS dalyką (-us) charakteristikos, pagrindinės paskirstymo savybės.

Duomenys. Analizėje naudojami BUS penkių semestrų dalykų (2013 m. pavasaris – 2015 m. pavasaris) pasirenkamumo duomenys, t. y. 15 164 nuasmeninti studentų registracijos įrašai (14 141 – patvirtinta registracija ir laikė dalyko egzaminą).

Tyrimo kintamieji: SEMESTR – studento mokymosi semestras; STUDIJUPROGR – studento studijų programa; SESVID – studento ankstesnės sesijos pažymių vidurkis; MAXSTUD – kvota, t. y. kiek studentų gali registruotis; IVERTNM – studento pasirinkto dalyko pažymys; RN – kiek studentų klausytis dalyko registravosi patys; VISOSK – kiek studentų užsiregistravo (buvo užregistruoti) į dalyką; KARTAS – kiek kartų studentas laikė dalyko egzaminą; HOURS (MINUTES) – studento registravimosi laiko tarpas nuo registracijos pradžios (į atitinkamą dalyką); KOMBI – studijų programos pasirinkimų (kombinacijų) skaičius atitinkamą semestrą.

Per 5 semestrus studentai galėjo pasirinkti iš 283 (pasirinko 281; skaičiuojama, kur vietų skaičius >1). 72 (26,9 %) dalykų kvotos buvo užpildytos. Taigi, dalis „populiarių“ (tų, kurių vietos užpildomos) dalykų tarp siūlomų dalykų yra nedidelė. 1 lentelėje yra pateiktas kiekvieno semestro bendras BUS dalykų skaičius ir BUS dalykų, kurių kvotos buvo užpildytos, skaičius bei procentinė dalis.

1 lentelė. Kvotų užpildymo lentelė

Semestras	Dalykų skaičius	Užpildyta (%)
2013 pav.	30	5 (16,7 %)
2013 rud.	47	6 (12,8 %)
2014 pav.	71	20 (28,2 %)
2014 rud.	56	17 (30,4 %)
2015 pav.	79	24 (30,4 %)

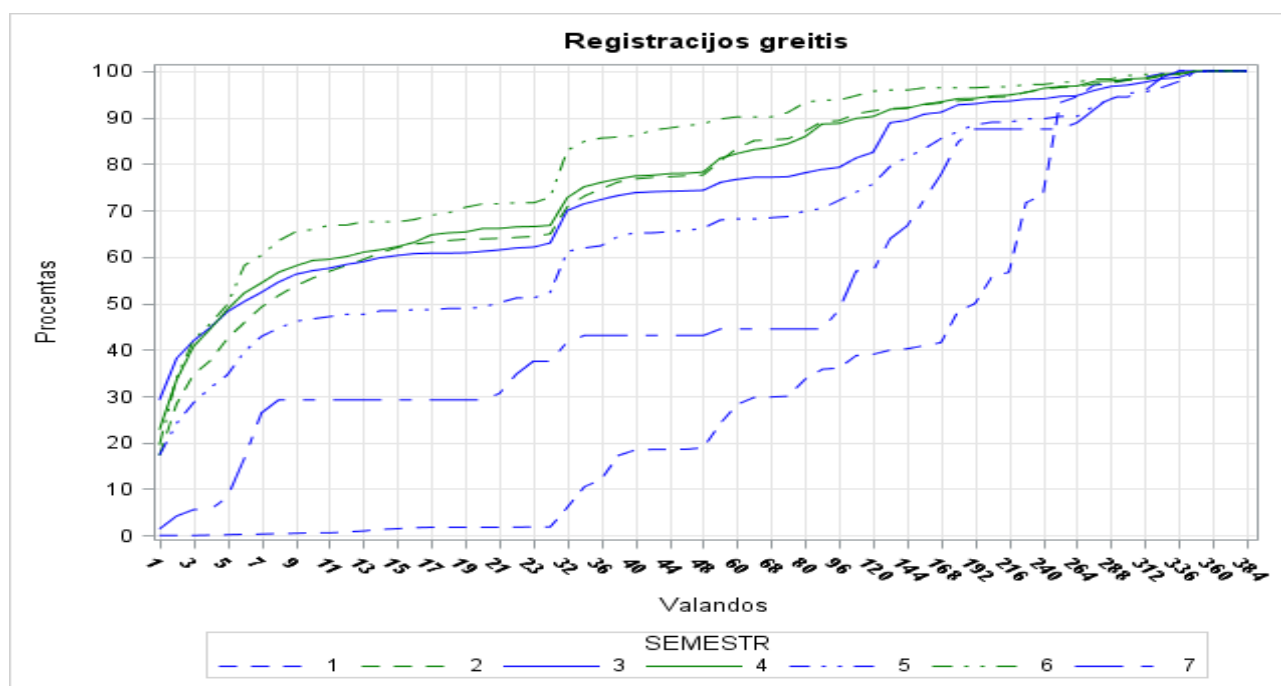
Nors, atrodytų, kad visiškai užpildomų dalykų kvotų nėra daug, tačiau pasirinkimo kombinacijų, užpildžius vieno ar kelių dalykų kvotą, gerokai sumažėja. Taigi, *resursų trūkumo* apribojimas BUS pasirenkamumo uždavinyje egzistuoja.

Per dvi savaites (tiek laiko skiriama studentams) nuo registracijos pradžios buvo 11 400 registracijų, tai sudaro 77,86 % nuo visų 14 641 patvirtintų registracijų. Taigi, yra nemažai studentų, kurie „nesivargina“ registruotis – už juos tai padaro administratoriai.

Kaip jau minėjome, *rikiavimo diktatūros* algoritmuose pakeitimas (perleidimas) nėra leidžiamas. Deja, tai praktiškai nėra realizuojama. Yra daug situacijų (netinka tvarkaraštis, studijų stabdymas

dėl ligos), kai studentas atsisako savo pasirinkimo. Tuo atveju jo pasirinkti dalykai vėl yra laisvi. Per nagrinėjamą laikotarpį VU apie 7 % BUS registracijų buvo pakartotinės, t. y. neišvengiamas *teisingumo principo* pažeidimas.

Per pirmąsias dvi registracijos valandas būna 20 % registracijų (būtų ir daugiau, bet riboja techninės galimybės). Per dvi paras buvo 7 436 registracijos, t. y. 50,79 % visų registracijų. Tam tikrų dalykų laisvos vietos užpildomos daug greičiau, vėliau bandantys registruotis jau turi mažesnes pasirinkimo galimybes. Pavyzdžiui, septynių (8,86 %) 2015 pavasario dalykų kvotos buvo užpildytos per keturias valandas. Natūralu, kad studentai stengiasi kuo greičiau prisijungti prie sistemos, tačiau tai tik trikdo sistemos funkcionavimą, bet neišsprendžia „teisingo“ dalykų paskirstymo problemos. Toliau grafike (žr. 1 pav.) pateikiamas dalykų kvotų užpildymo greitis skirtinguose semestruose besimokančių studentų. Mažas pirmo kurso studentų aktyvumas registracijos pradžioje, tikėtina, yra nulemtas to, kad registracija prasideda šventinę rugsėjo 1-ąją dieną.



1 pav. Registracijos greitis

Pasirinkimo galimybės priklauso nuo specialybės – specialybės programa nustato, kiek BUS dalykų galima pasirinkti vieną ar kitą semestrą. Dalykų pasiūla taip pat skiriasi įvairių specialybių studentams. Todėl daug charakteristikų yra skaičiuojama atsižvelgiant į studentų specialybę. Apskaičiuota koreliacija tarp kiekvienos studijų programos studentų vidutinio registravimo laiko (atskiros eilutės semestrams) ir pasirinkamų jose dalykų skaičiaus (tiksliau, kiek skirtingų dalykų jose buvo pasirinkta) aiškiai rodo priklausomybę tarp skubėjimo ir pasirinkimų skaičiaus (žr. 2 lent.). Mažas pasirinkimas – lėtesnė registracija.

2 lentelė. Koreliacijos koeficientai

Pirsono (Spearman) koreliacijos koeficientai, N=268, p-reiškė<0,0001		
	Me(min)	Vid(min)
kombinac	-0,32627 (-0,36041)	-0,30520 (-0,29993)

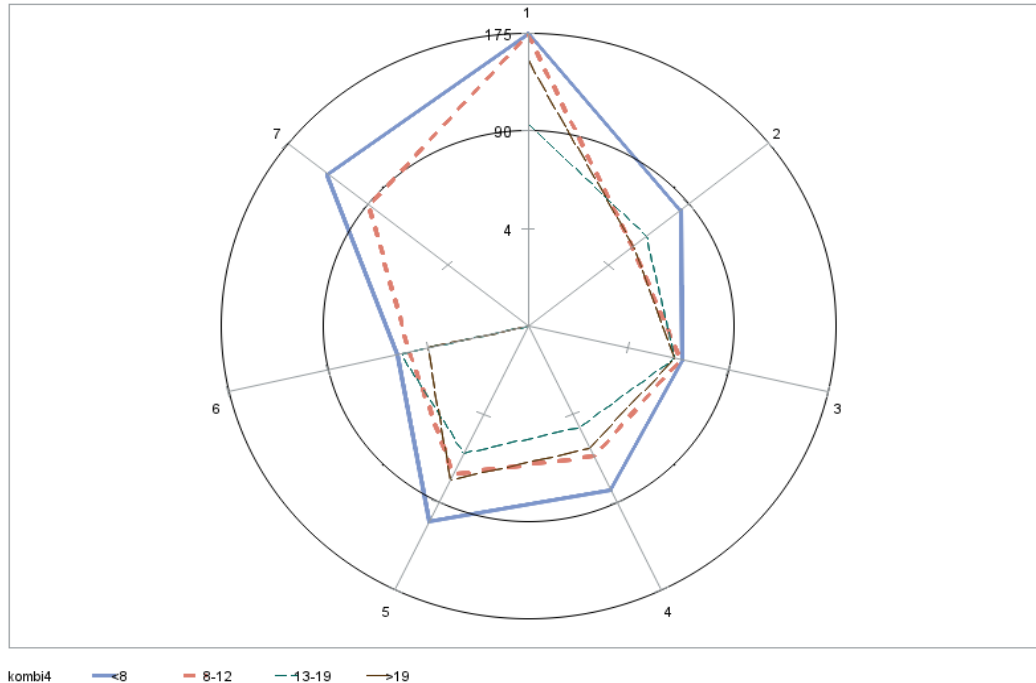
Žvaigždinė diagrama (žr. 2 pav.) rodo, kad greičiau registruojasi studentai, kurie turi daugiau pasirinkimų. Taip pat joje matyti priklausomybė nuo studentų mokymosi semestro.

Yra silpna priklausomybė tarp registracijos greičio ir pažangumo. Tendencija: pažangesni studentai greičiau registruojasi (žr. 3 lent.). Koreliacija tampa praktiškai nereikšminga, jei nagrinėjama priklausomybė tarp registracijos greičio ir studentų, užsiregistravusių per 24 valandas, pažangumo.

Ar galima įtarti, kad viena iš skubėjimo priežasčių yra noras patekti į lengvą dalyką? Tam reiktų apibrėžti lengvo dalyko sąvoką. Šio tyrimo autoriai lengvą dalyką apibrėžia taip: jei visi dalyko egzamino pažymiai yra teigiami (pirmo laikymo rezultatai, neįskaitant dėl vienokių ar kitokių priežasčių

nelaikiusių egzamino), tai laikoma, kad dalykas *lengvas*, kitais atvejais jis nėra *lengvas* (kintamasis LENGVAS). Pagal šį kriterijų tarp analizuojamų dalykų 149 (59 % nuo visų dalykų) yra *lengvi*, o likę 132 – ne.

Reiktų pastebėti, kad BUS dalykų sąrašas (t. y. skirtingų pavadinimų dalykai) kiekvienais metais keičiasi labai nedaug, taigi studentai žino (yra „girdėję“) praeitų metų atsiskaitymų rezultatus.



2 pav. Registravimosi vidutinis laikas atsižvelgiant į semestrus ir kombinacijų skaičių

3 lentelė. Koreliacijos koeficientai

Pirsono (Spearman) koreliacijos koeficientai, N=10697, p-reiškė<0,0001		
	SESVID	SESSVERTVID
minutes	-0,16581 (-0,15790)	-0,16263 (-0,15779)

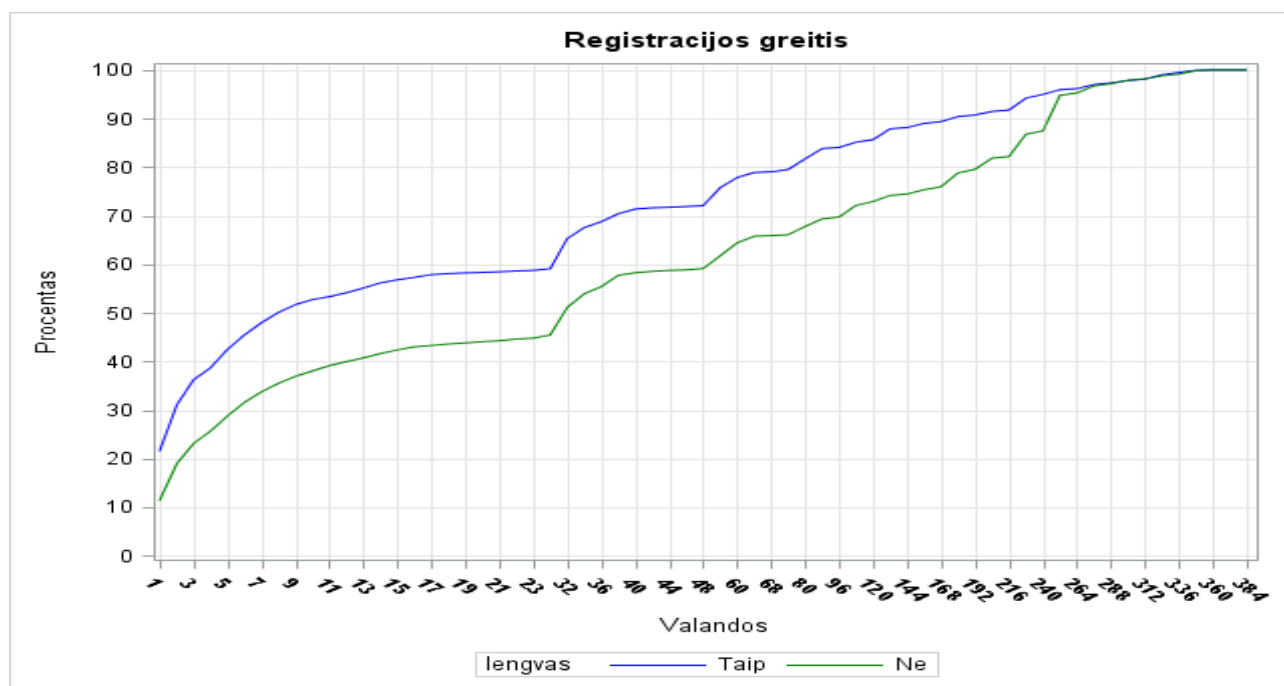
Ar studentai „intuityviai“ skuba rinktis „lengvesnį“ dalyką įvertinti sunku, tačiau į dalykus, kurių rezultatuose nebuvo neigiamų pažymių, nemažiau kaip 50 % užsiregistravo per 8 val., o į kitus – per 31 val. (žr. 3 pav.). Iš sąrašė pateiktų „lengvų“ dalykų, per tiriamąjį laikotarpį, užpildytos kvotos buvo į 31 dalyką (20,8 %), kitų – 13 (9,8 %) (pačių studentų registracijos).

„Lengvumas“ nėra labai tinkama sąvoka. Dalykas įdomus ir geras dėstytojas – visi studentai gerai išlaiko egzaminą; dalykas sudėtingas ir nuobodus, bet dėstytojas „neskriaudžia“ – visi studentai gerai išlaiko egzaminą. Yra 14 dėstytojų, kurie per nagrinėjamą laikotarpį ne mažiau kaip per trijų dalykų egzaminus neparašė neigiamų pažymių. Registracijos greitis į jų dėstomus dalykus, palyginti su kitų dėstytojų dėstomais dalykais, yra didesnis.

Studentų pasirinkimų „akademiniio pažangumo naudą“ galima matuoti įvairiai. Pavyzdžiui, skirtumu tarp studento pasirinkto dalyko pažymio ir jo sesijos pažymių vidurkio (kintamasis DIFFER). Taip pat galima ieškoti ir kitų motyvų studentams skubėti pasirenkant studijų dalyką. Buvo sudaryta keliolika išvestinių kintamųjų. Visus kintamuosius galima suskirstyti į šias grupes:

Akademiniio pažangumo nauda, kurią atspindi dalyko pažymių vidurkis, jau minėti kintamieji LENGVAS, DIFFER ir kiti.

Dalyko populiarumas, deficitu lygis, kuriuos matuoja, pavyzdžiui, kintamieji VISOFPRC – kokią dalį tarp pasirinkusiųjų dalyką sudaro tą dalyką kuruojančio fakulteto studentai, KSTUD – kiek studentų vidutiniškai pasirenka konkretaus dėstytojo konkretų dėstomą dalyką, MAXSTUD – dalyko kvota, LIKO (LIKOPRC) – konkrečios pasirinktų dalykų kombinacijos likusių neužpildytų vietų skaičius (jų



3 pav. Registracijos greitis į skirtingo lengvumo dalykus

procentinė dalis).

Pasirinkimo galimybių įvairovė („pasirinkimo laisvė“), kurios poveikį aprašo kintamasis KOMBI ir įvairūs iš jo sudaryti ranginiai kintamieji.

Norint įvertinti jau aprašytų grupių kintamųjų įtaką registracijos greičiui buvo sudaryti regresijos modeliai parenkant regresorius pagal Mallows C_p kriterijų. Buvo pastebėta, kad studentų dalykų pasirinkimo sąlygos labai priklauso ne tik nuo fakulteto, bet ir nuo semestro. Todėl regresiniai modeliai buvo sudaromi kiekvienam fakultetui ir semestrai atskirai. Atsako kintamasis (priklausomas kintamasis) šiuose modeliuose yra $y = \log_2(t) - \text{studento registravimosi laiko minutėmis logaritmas pagrindu 2}$ (kai t yra lygus vienai valandai, $y \approx 6$).

Rezultatai:

- Regresorių kiekis sudarytuose modeliuose – nuo 2 iki 9.
- Didelė dalis sudarytų modelių – visuose nemažuose fakultetuose bent viename semestre – turėjo R^2 didesnę už 0,4, kai kada R^2 siekė 0,6 (atvejais su mažai stebinių atmeti). O sudaryto bendro visiems didesniems fakultetams ir visiems semestrams, išskyrus I-ąjį kursą, regresinio modelio determinacijos koeficientas $R^2 < 0,14$.
- Pats studento „uolumas“, matuojamas to studento sesijos vidurkiu ir išvestais iš jo kintamaisiais, buvo reikšmingas faktorius, skatinantis skubėti registruotis, bet tai pasireiškė ne visais atvejais ir ne visada teigiamai.

Atliktos tiriamosios statistinės analizės pagrindu rasti regresoriai, turintys reikšmingos įtakos registracijos laiko skirstiniui, yra naudojami statistiškai vertinant sąlyginį studentų preferencijų tikimybinių skirstinį (žr. kitą skyrių).

4. Paskirstymo algoritmų palyginimas

Norint palyginti paskirstymo algoritmus reikia įvertinti jų teisingumą, efektyvumą, naudingumą, ekonomiškumą. Viena iš problemų yra praktinis išvardytų charakteristikų apibrėžimas ir skaičiavimas. Be to, reikia turėti lyginamų algoritmų panaudojimo tai pačiai studentų aibe duomenis. Taigi, lyginant algoritmus naudojama simuliacija ir lyginamos tik kelios algoritmų charakteristikos [5].

4.1. Studentų preferencijų generavimo algoritmas

Rinkdamasis studijų dalyką studentas tuo pačiu renkasi ir tam tikrą to dalyko užsiėmimo laiką, vietą ir dėstytoją. Su konkrečiais užsiėmimais susietą studijų dalyką vadinsime *moduliu*.

Daugelio elementų paskirstymo algoritmų apžvalga rodo, kad algoritmų palyginimo kriterijai turėtų remtis studentų preferencijomis ir matuoti, koku mastu tie algoritmai leidžia studentams pasirinkti pageidaujamus modulius. Tiesioginės informacijos apie studentų preferencijas turimuose duomenyse nėra. Tačiau, remiantis pirmumo principu pagrįstais pasirinkimais, galima susieti studento registracijos laiką su jo preferencijomis toliau aprašytu būdu.

Studentui s laike t registruojant savo pasirinkimą c yra tam tikra tikimybė $p(t; c) > 0$, kad modulio c kvota jau bus išnaudota. Jeigu modulis c turi aukštą rangą studento s preferencijų sistemoje, tai jis bus linkęs jį užregistruoti tada, kai $p(t; c)$ yra pakankamai maža, sakykime $p(t; c) \leq p_s(c)$. Tarkime, kad su studento s preferencijomis \succ_s yra susieti modulio rangai $r_s(c)$ taip, kad $c \succ_s c'$ (c geriau už c') $\Leftrightarrow r_s(c) < r_s(c')$ ir skirtingi moduliai negali turėti vienodų rangų. Tada $r_s(c) < r_s(c') \Leftrightarrow p_s(c) < p_s(c')$. Apibrėžkime optimalius pasirinkimo laikus $\tau_s(c)$ lygtimi

$$p(\tau_s(c); c) = p_s(c). \quad (1)$$

Darant prielaidą, kad $p(\cdot; c) \equiv p(\cdot)$ (t. y., studentas neturi apriorinės informacijos apie modulių populiarumą ir todėl visų modulių užpildymo tikimybės yra vienodos), gauname, kad $r_s(c) < r_s(c') \Leftrightarrow \tau_s(c) < \tau_s(c')$. Taigi, $\{\tau_s(c), c \in C\}$ apibrėžia tas pačias studentų preferencijas \succ_s kaip ir $\{r_s(c), c \in C\}$. Todėl bet kuri griežtai monotonišė pasirinkimo laikų $\{\tau_s(c), c \in C\}$ funkcija apibrėžia naujus rangus $\{\rho_s(c), c \in C\}$, suderintus su \succ_s . Toliau pateiktas studentų preferencijų generavimo metodas remiasi rangų

$$\rho_s(c) := \log_2(\tau_s(c)). \quad (2)$$

tikimybinio modeliu.

4.1 Pastaba. Verta pastebėti, kad rangai $\{\rho_s(c), c \in C\}$, skirtingai negu pradiniai rangai $\{r_s(c), c \in C\}$, yra susieti su bendra visiems s ir c laiko skale, priklauso ir nuo dalykų c kvotų užpildymo tikimybių $p(t; c)$. Todėl jie turi tam tikrą prasmę ir lyginant tarpusavyje skirtingų studentų ir modulių pasirinkimus. Jeigu $p_{s_1}(c_1) = p_{s_2}(c_2)$, bet $p(t; c_1) < p(t; c_2)$ su visais $t > 0$, tai $\rho_{s_1}(c_1) < \rho_{s_2}(c_2)$.

Darydami prielaidą, kad studentas yra apsisprendęs dėl $p_s(c)$, $p(\cdot; c) \equiv p(\cdot)$ ir yra žinoma, galime tarti, kad jis registravosi optimaliu laiku $\tau_s(c)$. Taigi, $t_s(c) = \tau_s(c)$. Čia $t_s(c)$ yra studento s realus registracijos laikas, kuris jau yra užfiksuotas turimuose duomenyse. Tačiau realiai tai pagrįsta tik tuo atveju, kai nė vieno (realiai galimo) modulio kvota dar nėra išnaudota. Jeigu laike t nors vienas modulis jau yra užpildytas, studentų pasirinkimo galimybės mažėja, todėl visos tikimybės $p(t, c)$ natūraliai išauga. Taigi, pakoreguotos tikimybės $\bar{p}(t, c) \geq p(t, c)$, ir remiantis 4.1 Pastaba išvedame, kad atitinkamai pakoreguoti optimalūs registracijos laikai $t_s(c)$, $\bar{p}(t_s(c), c) = p_s(c)$, tenkina nelygybę $t_s(c) \leq \tau_s(c)$. Vadinasi, mus dominanti reikšmė $\tau_s(c)$ duomenyse yra *cenžūruota iš dešinės*, vietoje jos duomenyse yra fiksuojama už ją mažesnė reikšmė $t_s(c)$.

Remiantis šiuo pastebėjimu, sudarant ir turimų duomenų pagrindų statistiškai vertinant rangų (2) (sąlyginį) tikimybinį skirstinį naudojamas *cenžūruotosios Gauso regresijos modelis* (žr., pvz., [3, 9]).

Cenzūruotųjų iš dešinės stebinių atveju daroma prielaida, kad *pilnieji* duomenys $\{(y_j^*, u_j^*, x_j), j = 1, \dots, N\}$ nėra stebimi. Atsako y_j^* reikšmės stebimos tik tuo atveju, kai jos yra mažesnės už cenzūravimo iš dešinės kintamojo u_j^* reikšmes. Be to, yra žinoma, ar reikšmė y_j^* buvo cenzūruota, ir jeigu – taip, tai koku dydžiu u_j^* . Taigi, stebimi duomenys yra

$$\{(y_j, z_j, x_j), j = 1, \dots, N\}, \quad y_j := \min(y_j^*, u_j^*), \quad z_j := I(y_j^* > u_j^*).$$

Cenzūruotosios regresijos modeliuose įprasta reikalauti, kad atsitiktinės poros $\{(y_j^*, u_j^*), j = 1, \dots, N\}$ ir jų komponentės yra tarpusavyje sąlyginai nepriklausomos, kai yra žinomos regresorių x reikšmės $\{x_j, j = 1, \dots, N\}$. Be to, sąlyginiai $\{y_j^*, j = 1, \dots, N\}$ ir $\{u_j^*, j = 1, \dots, N\}$ skirstiniai, kai reikšmės $\{x_j, j = 1, \dots, N\}$ yra žinomos, neturi bendrų nežinomų parametrų. Cenzūruotosios Gauso regresijos atveju $\{(y_j^*, x_j), j = 1, \dots, N\}$ tenkina klasikinį Gauso regresijos modelį.

Dėl įvairių priežasčių (tvarkaraščio apribojimai, reikalavimai kreditams bei kvalifikacijos reikalavimai) studentas s realiai gali rinktis tik dalį iš visų modulių. Juos vadinsime *galimais moduliais* ir jų aibę, turinčią k_s elementų, žymėsime C_s . Studentų galimų modulių preferencijas nusakančių rangų $\rho_s(c)$ naujos reikšmės yra generuojamos pagal tokį algoritmą.

Algoritmas:

1. Pasirinktomis studijų programoms ir pasirinktiems semestrams sudaromas jiems galimų (realizuojamų) modulių sąrašas C ir randamas pirmojo užpildyto modulio užpildymo laikas τ_{min} . Visi modulių registracijos laikai $t_s(c)$ po laiko momento τ_{min} laikomi cenzūruotomis $\tau_s(c)$ reikšmėmis.

2. Naudojant turimus pradinius duomenis ir iš jų sudarytus papildomus regresorius, aprašytus 3 skyriuje, ir taikant cenzūruotosios Gauso regresijos modelį atsako kintamiesiems $y_s(c) := \log_2(t_s(c))$ įvertinamas rangų $\rho_s(c)$, apibrėžtų lygybe (2), sąlyginis tikimybinis skirstinys.

3. Ankstesniame etape parinkto cenzūruotosios Gauso regresijos modelio pagrindu kiekvienam studentui s ir moduliui $c \in C_s$ atsitiktinai generuojamos naujos rangų $\rho_s(c)$ reikšmės $\rho_s^*(c)$, $c \in C_s$, o tuo pačiu ir naujos studento s preferencijos. Faktiškai yra generuojamas porų rinkinys

$$\succ_s^* := (c_{s,i}, r_{s,i}), \quad i = 1, \dots, k_s, \quad r_{s,1} \leq \dots \leq r_{s,k_s}. \quad (3)$$

Jis gaunamas iš porų rinkinio

$$\{(c, \rho_s^*(c)), c \in C_s\}, \quad (4)$$

jas surikiavus rangų $\{\rho_s^*(c), s \in C_s\}$ didėjimo tvarka. Tas pačias preferencijas kaip (3) nusako ir poros

$$(c_{s,i}, i), \quad i = 1, \dots, k_s. \quad (5)$$

4. Pasirenkamas modulių paskirstymo algoritmas ir remiantis sugeneruotomis studentų preferencijomis (3) arba (5) imituojami studentų pasirinkimai. Apskaičiuojamos imituotų pasirinkimų rangų ar jų monotoninių transformacijų empirinės centrinės tendencijos ir sklaidos charakteristikos, detaliau žr. 4.2 poskyryje.

4.2 Pastaba. Kai kurių studijų programų kai kuriuose semestruose studentų galimų modulių aibė C_s yra labai maža, kartais turi tik vieną elementą. Todėl imituojant modulių pasirinkimus studentas kartais neturi iš ko rinktis. Tada studentui formaliai yra priskiriamas fiktyvus modulis su jo rangui priskirta fiktyvia rango reikšme, didesne už maksimalią realizuojamą rango reikšmę. Mažos galimų pasirinkimų aibės problema aktuali net ir tyrimuose, kurie remiasi studentų deklaruotomis preferencijomis, nes studentai yra linkę deklaruoti tik juos dominančių dalykų preferencijas [4].

4.2. Palyginimo rezultatai

Straipsnyje lyginami penki modulių paskirstymo algoritmai: RSD, HBS, dvi FCFS algoritmo modifikacijos, FCFSg ir FCFSr, bei rikiavimo diktatūros algoritmo modifikacija ASD (*Academic Serial Dictatorship*), kai studentų pasirinkimų registracijos eilė sudaroma pagal jų semestro pažymių vidurkių mažėjimo tvarką. Algoritme FCFSg studento registracijos laiką apsprendžia pirmojo pagal sugeneruotas jo preferencijas modulio apibendrintasis rangas: kuo jis mažesnis, tuo registracijos laikas ankstesnis. Algoritme FCFSr studento registracijos laikas sutampa su pradiniuose duomenyse fiksuotu studento registracijos laiku. Imituojant studentų pasirinkimus pagal sugeneruotas studentų preferencijas ir taikant tiriamus paskirstymo algoritmus daroma prielaida, kad visi studentai turi pasirinkti du modulius.

Lyginant paskirstymo algoritmus praktiniu požiūriu, paprastai pateikiamas studentų rangų, realizuotų vykdant tuos algoritmus, empirinis skirstinys (žr., pvz.[4]). Taip pateikti rezultatai užima daug vietos, be to, dėl gaunamų skirstinių įvairovės ir ryškios asimetrijos jie dažnai yra sunkiai suvokiami ir nevienareikšmiškai interpretuojami. Šiame tyrime, lygindami modulių paskirstymo algoritmus, naudojame logaritmuotų suvestinių rangų bendras centrinės tendencijos ir sklaidos empirines charakteristikas. Logaritmo transformacija žymiai sumažina skirstinių asimetriją. Kiekvieno studento dviejų modulių pasirinkimui suvestinis rangas apskaičiuojamas trimis būdais: rangų (5) aritmetinio vidurkio logaritmas (žymima, A – aritmetinis), jų geometrinio vidurkio logaritmas (žymima, G –

geometinis) ir apibendrintojo rango (2) aritmetinis vidurkis (žymima, B – apibendrintas). Suvestinių rangų vidurkiai, standartiniai nuokrypiai, medianos ir tarpkvartiliniai skirtumai pateikti 4 lentelėje.

4 lentelė. Suvestinių rangų empirinės centro ir sklaidos charakteristikos

Būdai	Vidurkis			Mediana			Stand. nuokrypis			Tarpkvart. skirtumas		
	A	G	B	A	G	B	A	G	B	A	G	B
Algoritm.												
RSD	1,98	1,88	7,54	1,32	1,29	6,89	1,69	1,68	3,79	2,32	2,20	4,12
HBS	1,90	1,48	7,28	2,00	1,50	7,24	1,01	0,75	2,28	1,70	1,25	3,02
FCFSg	1,77	1,70	7,25	1,32	1,29	6,71	1,15	1,18	4,05	2,12	2,16	4,69
FCFSr	2,01	1,89	7,44	1,32	1,29	6,79	1,64	1,62	3,79	2,32	2,20	4,24
ASD	2,03	1,91	7,52	1,32	1,29	6,90	1,67	1,66	3,86	2,32	2,29	4,49

Kaip ir galima buvo tikėtis, mažiausi vidutiniai suvestiniai rangai gauti taikant algoritmus HBS ir FCFSg tuo atveju, kai suvestinis rangas yra geometrinis vidurkis. Apibendrintųjų rangų vidurkių skirtumai tarp algoritmų maži. Visų algoritmų suvestinių rangų medianos labai artimos išskyrus HBS algoritmą, kuriam medianos daug didesnės visiems suvestinių rangų skaičiavimo būdams.

Panašiai elgiasi ir suvestinių rangų standartiniai nuokrypiai ir tarpkvartiliniai skirtumai. Jų mažumu vėl išsiskiria HBS ir FCFSg algoritmai. Bet FCFSg turi didžiausią apibendrintojo suvestinio rango standartinį nuokrypį bei tarpkvartilinį skirtumą.

ASD algoritmo suvestinių rangų charakteristikos praktiškai nesiskiria nuo RSD algoritmo, kuris dėl savo gerų teorinių savybių dažnai naudojamas kaip etalonas, o taip pat nuo FCFSg, kuriame pasirinkimo registracijos laikas pagal pirmumo principą nebuvo tiesiogiai susietas su studento sugeneruotomis preferencijomis. Skirtingus modulius pasirinkusių studentų palyginimas, kurio rezultatai čia nepateikiami, rodo, kad paskirstymo pagal ASD algoritmą atveju yra ryški modulių diferenciacija pagal pažangumą.

5 lentelė. Suvestinių rangų empirinės centro ir sklaidos charakteristikos

Būdai	2013 m. rudenį						2014 m. pavasarį					
	Vidurkis			Stand. nuokrypis			Vidurkis			Stand. nuokrypis		
	A	G	B	A	G	B	A	G	B	A	G	B
Algoritm.												
RSD	0,84	0,73	6,24	0,52	0,46	2,49	2,63	2,53	7,93	1,90	1,91	4,48
HBS	0,76	0,62	6,15	0,27	0,20	2,40	2,54	1,98	7,68	0,94	0,70	2,31
FCFSg	0,83	0,73	6,24	0,42	0,42	2,65	2,29	2,23	7,60	1,23	1,26	4,74
FCFSr	0,86	0,75	6,29	0,58	0,52	2,64	2,64	2,52	7,79	1,84	1,85	4,44
ASD	0,86	0,74	6,27	0,56	0,49	2,57	2,67	2,55	7,86	1,85	1,85	4,49

Paskirstymo algoritmų rezultatams daro įtaką studentų galimybės rinktis. 5 lentelėje pateiktos 2013 m. rudens ir 2014 m. pavasario semestrų suvestinių rangų charakteristikos. Abiejuose semestruose visų modulių kvotos buvo tokios pačios, bet studentų, kuriems buvo sugeneruotos preferencijos, kiekis pirmuoju atveju buvo daugiau kaip 4 kartus mažesnis. Matyti, kad to paties paskirstymo algoritmo rezultatų skirtumai tarp semestrų yra daug didesni negu tų pačių semestrų skirtumai tarp algoritmų.

5. Išvados

Siūlomų BUS dalykų sąrašas yra gana pastovus todėl pagal sukauptus duomenis apie studentų pasirinkimus galima įvertinti jų pasirinkimo motyvus. Atlikta tiriamoji analizė parodė:

- Galimų pasirinkimų skaičius turi įtakos studentų aktyvumui – jie greičiau registruojasi, kai turi didesnę pasirinkimą.
- Yra tendencija greičiau registruotis į „lengvesnius“ dalykus.

- Pažangesni studentai registruojasi greičiau, jie uolesni, bet registracijos pradžioje pažangumas nėra reikšmingas faktorius.
- Viena iš svarbesnių problemų – tvarkaraštis. To paties pavadinimo dalykas gali tapti „populiarus“ arba „nepopuliarus“ vien tik dėl pasiūlyto tvarkaraščio.

Kuriant ir pasirenkant studijų dalykų paskirstymo pagal studentų preferencijas procedūras, svarbu atsižvelgti ir į motyvus, kuriais tos preferencijos remiasi. Tai yra susiję su paskirstymo algoritmų praktinio palyginimo problema. Įprasta studentų preferencijas laikyti pagrindine vertybe ir algoritmus lyginti pagal tai, kaip jiems pavyksta tas preferencijas realizuoti. Bet ne visi motyvai, lemiantys studento preferencijas, universitetui yra vienodai vertingi.

Tais motyvais remiamasi ir pasirenkant dalykus pagal pirmumo principą. Šiuo principu grindžiama dalykų paskirstymo procedūra nėra efektyvi ir neatitinka teisingumo kriterijų (studentų lūkesčių). Be to, ji nėra pilnai formalizuota. Pavyzdžiui, realiai studentai turi galimybę atsisakyti savo pasirinkimo, bet atsisakymo tvarka neaprašyta.

Kadangi studento pasirinkimo registravimo laiku dalies dalykų kvotos gali būti jau užpildytos, registravimo laiko duomenims buvo taikytas cenzūruotosios regresijos modelis. Įvertinto modelio pagrindu imituotoms studentų preferencijoms buvo pritaikyti keli žinomi dalykų paskirstymo algoritmai: RSD, HBS, FCSFg, FCSFr ir ASD (žr. 4 skyrių).

Taikant algoritmus HBS ir FCSFg studentų pasirinktų modulių rangų vidurkis ir standartinis nuokrypis, kaip ir tikėtasi, buvo pastebimai mažesni negu taikant kitus algoritmus. Pastariesiems minėtos charakteristikos praktiškai nesiskyrė. Taigi, algoritmams FCSFg ir FCSFr, realizuojantiems pirmumo principą, gauti skirtingi rezultatai. FCSFg atveju studento registracijos laikas yra suderintas su imituotom jo preferencijom: kuo mažesnis jo labiausiai pageidaujamo dalyko apibendrintasis rangas, tuo anksčiau jis registruojasi. FCSFr algoritme naudojamas realus (paimtas iš turimų duomenų) studento registracijos laikas.

Buvo paskaičiuoti studentų pasirinktų dalykų logaritmuotų rangų vidurkiai ir standartiniai nuokrypiai atskiruose semestruose. Esant toms pačioms dalykų kvotoms studentų skaičius skirtingais semestrais kartais skyrėsi daugiau kaip 4 kartus. Apskaičiuotų vidurkių ir standartinių nuokrypių skirtumai tarp tokių semestrų buvo žymiai didesni negu tarp taikytų skirtingų paskirstymo algoritmų. Vadinasi, su studentų poreikiais suderinta dalykų pasiūla ir jų kvotos gali turėti jų paskirstymui didesnę teigiamą efektą negu taikymas sudėtingo paskirstymo algoritmo.

Šiame darbe DA algoritmo empiriškai netyrėme dėl papildomų sudėtingų problemų, kylančių jį realizuojant. Greta gerų teorinių savybių DA tipo algoritmai turi ir praktinių privalumų. Pirma, studentai be apribojimų deklaruoja savo pradines preferencijas, kurių duomenis galima panaudoti gerinant dalykų pasirinkimo ir studijų procesą. Antra, prioritetų tvarką $\succ_c, c \in C$, kuri kartu su studentų preferencijom yra naudojama DA algoritme, galima sudaryti pagal studentų reitingus, kurie jau dabar yra skaičiuojami dėl rotacijos poreikio.

Literatūra

- [1] Abdulkadiroglu, A., and Sonmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica* 66(3), p.689-701.
- [2] Budish, Eric, and Estelle Cantillon. (2012) The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard *American Economic Review*, 102(5), p.2237-71.
- [3] Cox, D. R., and Oakes, D. (1984) *Analysis of Survival Data*. London: Chapman & Hall.
- [4] Franz Diebold, Haris Aziz, Martin Bichler, Florian Matthes*, Alexander Schneider. Course Allocation via Stable Matching [interaktyvus].2014. Žiūrėta [2016-08-01]. p.1-12. < http://dss.in.tum.de/files/bichler-research/2014_diebold_matching_course_allocation.pdf >
- [5] Richard Hoshino and Caleb Raible-Clark (2014) The Quest Draft: an Automated Course Allocation Algorithm AAAI Publications, Twenty-Sixth IAAI Conference Proceedings of the Twenty-Sixth Annual Conference on Innovative Applications of Artificial Intelligence <https://www.aaai.org/ocs/index.php/IAAI/IAAI14/paper/download/8341/8657>

- [6] Kojima, Fuhito (2013) Efficient resource allocation under multi-unit demand *Games and economic behavior*, Volume: 82, p.1-14.
- [7] Roth, A.E. (2008) Deferred Acceptance Algorithms: History, Theory, Practise, and Open Questions, *International Journal of Game Theory*, Special Issue in Honor of David Gale on his 85th birthday, 36, p.537 - 569.
- [8] Roth, A.E., Sotomayor, M.O. (1990) Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis Econometric Society Monographs Cambridge Univ. Press, Cambridge (1990)
- [9] SAS Institute Inc. (2013) *SAS/STAT 13.1 User's Guide*. Cary, NC: SAS Institute Inc.

MULTI-UNIT ASSIGNMENT PROBLEM: FCFS COURSE ALLOCATION SYSTEM DATA ANALYSIS

Gediminas Murauskas (VU), Marijus Radavičius (VU)

Abstract. We study the allocation of courses to students with multi-unit demand. Exploratory analysis of course allocation practices at Vilnius University (VU) is performed. The allocation of students to courses is done by a simple first-come first-served (FCFS) procedure. We show that FCFS procedure does not deliver desirable outcomes. Data with student preferences is needed in order to compare assignment algorithms. We use censored regression model to construct student preferences generation mechanism based on empirical data collected at VU. Using simulated data several multi-unit assignment algorithms are compared.

Key words: multi-unit assignment, FCFS, preferences, censored regression.