Liet. matem. rink., T. 43, spec. nr., 297–303

Duomenų gavybos technologijų panaudojimas jaunųjų kompiuterininkų mokyklos veiklos analizei

Sigita TURSKIENĖ, Genadijus KULVIETIS, Renata BURBAITĖ (ŠU)
el. paštas: sigita@fm.su.lt

1. Įvadas

Šiandien švietimo įstaigoms keliami uždaviniai: didinti darbo efektyvumą, racionaliai panaudoti resursus, teikti kokybiškas paslaugas. Mokyklų vadovai, priimantys sprendimus, privalo turėti galimybę realiu laiku naudoti turimus duomenis, juos apdoroti ir išgauti duomenyse esančią informaciją sprendimui priimti.

Šiaulių universiteto (ŠU) jaunųjų kompiuterininkų mokykla (JKM) savo veiklą pradėjo 1997 m. Mokyklos administracija elektroninėse laikmenose saugo didelius duomenų kiekius ne vienoje duomenų bazėje. Iki šiol duomenims apdoroti dažniausiai buvo naudojama statistinė analizė. Joje daug dėmesio skiriama vidutinėms imties charakteristikų vertėms, analizė užtrunka ilgai ir patvirtina arba paneigia išankstines hipotezes.

Vienas problemos sprendimo būdų – duomenų gavybos technologijų diegimas. Jos naudojamos daugelyje žmogaus veiklos sričių: telekomunikacijų, draudimo, medicinos, finansų kompanijose, bankuose. Duomenų gavybos technologijų kūrėjai pažymi sėkmingą technologijų taikymą švietimo srityje neįvardydami konkrečių sprendimų. Sėkmingai gavybos technologijos taikomos profesionaliajame sporte [5]; Indianos universitete efektyviam bendravimui ir paslaugų teikimui užtikrinti [8]. Bandymas panaudoti duomenų gavybos technologijas Lietuvos švietimo sistemos analizei – mokytojų kvalifikacijos, amžiaus kaitos bei jų skaičiaus dinamikos tyrimas taikant dirbtinius neuroninius tinklus [1]. Šio darbo tikslas – atlikti kelių duomenų gavybos programinių produktų galimybių analizę tiriant ŠU JKM veiklą, suformuluoti JKM veiklos rekomendacijas.

2. Duomenų gavybos technologijų galimybių analizė

Iš pradžių atlikta JKM veiklos tradicinė statistinių duomenų analizė, kuri išryškino bendriausius mokyklos veiklos bruožus. Ši analizė tik patvirtino arba paneigė iš anksto suformuluotus teiginius. JKM veiklos dėsningumams išryškinti panaudotos duomenų gavybos technologijos. Tirta: moksleivių bendrasis pasirengimas ir moksleivių lūkesčiai. Tyrimui naudotos šios programos: INTELLIGENT MINER for Data V.8.1.; XPERTRULE MINER 1.47; WIZWHY 3.0; STATGRAPHICS Plus 5.0.

Programa INTELLIGENT MINER for Data atlikta į JKM stojančiųjų bendrojo ir specialiojo pasirengimo klasterinė duomenų analizė. Atliekant pirmosios laidos moksleivių duomenų demografinę klasterizaciją sukonstruoti modeliai su skirtingu klasterių

▼ Details for Fields

Show Modal Value

▼ Cluster	įstojimas	lytis	mokykla	rajonas	[1užduotis]	[2užduotis]	[3užduotis]	[4užduotis]
[4] 3	0	v	Vaižganto	Mažeikių	0	1	0	0
[3] 2	1	v	Lieporių	Šiaulių	1	1	1	0
[2] 1	0	m	Žemynos	Šiaulių	0	1	0	0
[1] 0	1	m	Graičiūno	Kelmės	1	1	1	0
Total population	1	v	Lieporių	Šiaulių	0	1	0	0

1 pav. Pirmosios laidos stojimo demografinės klasterinės analizės detalizavimas.

kiekiu. Tai padėjo pasirinkti optimalų klasterių skaičių tolesnei analizei. Renkantis modelį įvertinamas klasterių homogeniškumas ir panašumas. Palyginus modelius, tolesnei analizei pasirinktas modelis, kuriame objektai suskirstyti į 4 klasterius. 1 pav. pateikiamas šio modelio detalusis vaizdas.

Išanalizavus modelio klasterius, modalines laukų vertes išryškėjo šios tendencijos:

- Į mokyklą neįstojo vaikinai ir merginos neišsprendę 1, 3 ir 4 užduočių. Vaikinai buvo iš Mažeikių, merginos – iš Šiaulių rajono.
- Į mokyklą daugiausia vaikinų įstojo iš Šiaulių Lieporių vidurinės mokyklos ir merginų iš Kelmės J. Graičiūno mokyklos.
- Moksleiviams sunkiausiai sekėsi spręsti 4 užduotį – dominuoja įvertinimas 0.

Iš šių tendencijų galime suformuluoti išvadas apie moksleivių pasirengimą:

1) 2 užduotis buvo lengviausia, nes teigiamas jos įvertis yra visuose klasteriuose;

2) 4 užduotis daugeliui moksleivių buvo sunki. Tokiems uždaviniams reikia skirti daugiau dėmesio.

Mokyklos darbuotojai klasterinės analizės duomenis gali naudoti užsiėmimų tvarkaraščiams sudaryti.

Analogiškai sukonstruotas 2 laidos moksleivių demografinės klasterizacijos modelis. Analizuodami šio modelio rezultatus pastebime, kad:

- 2 laidoje į mokyklą įstojo vaikinai iš Šiaulių Ragainės vidurinės mokyklos ir merginos iš Radviliškio J. Tumo Vaižganto gimnazijos.
- Tarp neįstojusių į mokyklą dominuoja merginos iš Šiaulių 12-osios ir vaikinai iš Šiaulių Medelyno mokyklos.
- 10 užduoties sprendimo rezultatai neturėjo įtakos stojamojo testo įvertinimui. Tai patvirtina ir aprašomosios statistikos rezultatai. Iš aprašomosios statistikos ir demografinės klasterizacijos modelio rezultatų matyti, kad
- Moksleiviai geriausiai sprendė 1, 2 ir 9 užduotis.
- Į mokyklą įstojo 3, 5 ir 7 užduotis išsprendusieji moksleiviai.
- 6 užduotį moksleiviai sprendė vidutiniškai – surinko pusę skiriamų balų.

Iš šių teiginių galima daryti išvadą, kad į mokyklą įstojo vidutiniškai pasirengę moksleiviai.

Demografinė klasterinė analizė atlikta tiriant 4 laidos moksleivių bendrąjį ir specialųjį pasirengimą. Sukonstruotas demografinės klasterizacijos modelis (2 pav.) išryškino stojančiųjų pasirengimą, jų lūkesčius: dauguma ketvirtosios laidos moksleivių turi darbo

Visible clusters:

Name	Size	Characteristics
[2] 1	51,28%	***Daugiaterpė aplinka*** is predominantly ne, ***Ekonomika*** is predominantly ne, ***Algoritmavimas*** is predominantly taip, ***Rajonas*** is predominantly Šiaulių, ***Raštvedyba*** is predominantly ne, ***[Fizika]*** happens to be medium, ***Mokykla*** is predominantly Gruzdžių A. Griciaus v. M., ***[Ką mokate?]*** happens to be predominantly raštvedyba, ***[Kas sunku, bet patinka?]*** happens to be predominantly programuoti, ***Lytis*** is predominantly v, ***[Kur naudosite žinias]*** happens to be predominantly studijos, ***[Ko nereikia?]*** happens to be predominantly visos reikalingos, ***[Patirtis]*** happens to be medium, ***[Lietuvių k]*** happens to be medium and ***[Kas patinka?]*** happens to be predominantly Praktinis darbas.
[3] 2	23,08%	***Daugiaterpė aplinka*** is predominantly taip, ***Ekonomika*** is predominantly taip, ***Rajonas*** is predominantly Radviliškio, ***Lytis*** is predominantly v, ***Algoritmavimas*** is predominantly taip, ***Raštvedyba*** is predominantly ne, ***[Internetas]*** happens to be predominantly taip, ***[Ko nereikia?]*** happens to be predominantly visos reikalingos, ***Mokykla*** is predominantly Vaižganto g., ***[Patirtis]*** happens to be high, ***[Kas patinka?]*** happens to be predominantly programavimas, ***[Fizika]*** happens to be medium, ***[Informatika]*** happens to be medium, ***[Matematika]*** happens to be medium and ***[Ko tikitės JKM]*** happens to be predominantly gerai programuoti.
[1] 0	17,95%	***Algoritmavimas*** is predominantly ne, ***Lytis*** is predominantly m, ***Raštvedyba*** is predominantly taip, ***[Fizika]*** happens to be high, ***[Lietuvių k]*** happens to be high, ***[Kas sunku, bet patinka?]*** happens to be predominantly programuoti, ***Ekonomika*** is predominantly taip, ***[Užsienio k]*** happens to be high, ***Mokykla*** is predominantly Kudirkos v. M., ***[Matematika]*** happens to be high, ***[Ko daugiau?]*** happens to be predominantly programavimo, ***[Patirtis]*** happens to be medium, ***[Chemija]*** happens to be high, ***[Ko tikitės JKM]*** happens to be predominantly dirbti kompiuteriu and ***[Internetas]*** happens to be predominantly taip.
[4] 3	7,69%	***Raštvedyba*** is predominantly taip, ***Rajonas*** is predominantly Šiaulių, ***[Ko tikitės JKM]*** happens to be predominantly dirbti kompiuteriu, ***[Chemija]*** happens to be low, ***Mokykla*** is predominantly Šalkauskio v. M., ***Daugiaterpė aplinka*** is predominantly taip, ***[Užsienio k]*** happens to be medium, ***[Kas patinka?]*** happens to be predominantly Praktinis darbas, ***Ekonomika*** is predominantly taip, ***Lytis*** is predominantly v, ***[Ko daugiau?]*** happens to be predominantly programavimo, ***[Patirtis]*** happens to be medium, ***[Ko nereikia?]*** happens to be predominantly visos reikalingos, ***[Ką mokate?]*** happens to be predominantly visko po truputį and ***[Informatika]*** happens to be medium.

2 pav. ŠU JKM 4 laidos moksleivių demografinės klasterizacijos modelis.

Tree	Node ID	Score	Record Count (% of all)
True	1	8.94871807098...	39 (1.00%)
Lietuvių k < 5,5	1.1	7.59999990463...	5 (13%)
Matematika < 6,5	1.1.1	6.5	2 (5%)
Matematika ≥ 6,5	1.1.2	8.33333301544...	3 (8%)
Lietuvių k ≥ 5,5	1.2	9.14705848693...	34 (87%)
Fizika < 8,5	1.2.1	8.90909099578...	22 (56%)
Fizika ≥ 8,5	1.2.2	9.58333301544...	12 (31%)

3 pav. Bendrasis JKM 4 laidos moksleivių pasirengimas.

kompiuteriu patirtį, gali atlikti kompiuteriu įvairius taikomuosius darbus, naudojasi internetu. Programavimo mokymas vidurinėse mokyklose kol kas netenkina moksleivių poreikių. Šį trūkumą gali kompensuoti JKM.

Moksleiviai JKM įgytas žinias ir įgūdžius žada naudoti studijuodami, todėl reikia juos skatinti naudoti kompiuterį kaip darbo įrankį gilinant matematikos [7], fizikos [7] ir kitų mokomųjų dalykų žinias. Moksleiviai galėtų atlikti kūrybines užduotis, parengę darbus juos pristatyti įvairiuose konkursuose [3, 4].

Klasifikaciniai modeliai sukonstruoti naudojant Sprendimų medžio gavybos funkciją. 3 pav. pateikiamas sprendimų medis, iliustruojantis 4 laidos moksleivių pasirengimą.

Duomenų gavybos modelis, parengtas Sprendimų medžio gavybos funkcija, yra lengvai interpretuojamas. Modeliuose gerai realizuotas duomenų vizualizavimas supaprastina duomenų analizę. Lygindami šį gavybos modelį su klasterinės analizės rezultatais pastebime, kad modeliai patvirtina ir papildo vienas kitą. Tai rodo, kad gavybos rezultatai yra patikimi.

Demografinės klasterizacijos rezultatų analizė rodo, kad:

- Windows operacine sistema, programa Norton Commander, archyvatoriais geriausiai naudojasi algoritmavimo specializaciją pasirinkę vaikinai.
- Merginos užduotis atlieka vidutiniškai.
- Visų specializacijų moksleiviai programas pristato vidutiniškai.
- Silpniausi moksleiviai pasirinko specializaciją „Daugiaterpė aplinka“.

Demografinės klasterizacijos rezultatus papildo Sprendimų medžio gavybos modeliai. Iš rezultatų seka, kad:

- merginų naudojimosi Windows operacine sistema įgūdžiai vertinami apie 8,5 balo;
- silpnus darbo programa Norton Commander įgūdžius įgijo specializacijos „Daugiaterpė aplinka“ moksleiviai. Jų įvertinimai apie 8 balus. Remdamiesi šiomis tendencijomis, JKM darbuotojai gali:
  - Koreguoti ir pildyti mokymo programas algoritmavimo specializacijai.
  - Daugiau valandų skirti programų pristatymams mokyti.
  - Koreguoti specializacijos „Daugiaterpė aplinka“ programą.

Statistinėje programoje STATGRAPHICS Plus 5.0 modelių kūrimo galimybes mažino tai, kad buvo naudojama demonstracinė šio produkto versija. Pastebėta, kad programoje STATGRAPHICS silpnai realizuotos darbo su kategoriniais duomenimis galimybės. Šis sistemos trūkumas labai apsunkina modelių kūrimą.

Lygindami moksleivių pasiskirstymą pagal rajonus, matome, kad pirmosios laidos moksleiviai buvo iš 12 rajonų, o 2 ir 4 laidų – iš 4 rajonų. Šiam pasiskirstymui įtakos turi ekonominė šalies padėtis, aukštųjų mokyklų išsidėstymas, nes beveik kiekviena aukštoji mokykla siūlo moksleiviams mokytis jos įsteigtose JKM.

Sistema WIZWHY išskiria visas IF-THEN taisykles, skaičiuoja kiekvienos taisyklės klaidos tikimybę, prognozuoja kiekvieną požymį ir apibrėžia prognozės paklaidas, išskiria duomenyse paslėptus ryšius. Šia sistema apdorota JKM 4 laidos moksleivių anketa ir sukonstruoti taisyklių rinkiniai. Labai įdomi yra viena taisyklė:

If **Chemija** *is* **6,00 ... 8,00** *(average =* **7,00***)*
and **Raštvedyba** *is* **ne**
and **Kur naudosite žinias** *is* **studijos**
Then
Ko norėtumėte išmokti JKM *is* **gerai programuoti**
Rule's probability: **0,800**
The rule exists in **12** *records.*
Significance Level: Error probability < 0,001
Positive Examples (records' serial numbers):
15, 16, 17, 18, 20, 23, 24, 29, 30, 33
Negative Examples

Išanalizavę šią taisyklę matome, kad moksleiviai, kuriems vidutiniškai sekasi chemija, nori išmokti gerai programuoti. Kitoje taisyklėje buvo informacija apie moksleivių, kuriems gerai sekasi chemija, lūkesčius.

Pastebėjome, kad moksleiviai, kuriems gerai sekasi chemija, nenori išmokti gerai programuoti. Jie ateityje tikriausiai rinksis specialybes, susijusias su chemija. Moksleiviams galima siūlyti alternatyvą „Eksperimento rezultatų apdorojimo metodai“. Taisyklės konstruojamos rankiniu būdu, kiekvieną kartą parenkant ir pažymint priklausomą kintamąjį. Sugeneruotos taisyklės pateikiamos tekstiniu sąrašu. Daug laiko užima sugeneruotų taisyklių analizė. Tai sistemos WIZWHY trūkumas. Taisyklės patvirtina ir papildo

INTELLIGENT MINER for Data sukurtų modelių rezultatus. Tai rodo rezultatų patikimumą.

Sistema XPERTRULE MINER 1.47 apdoroti JKM 4 laidos moksleivių atsakymai į anketos klausimus. XPERTRULE MINER 1.47 modelio rezultatai patvirtina IBM INTELLIGENT MINER for Data gautus rezultatus ir išryškina naują tendenciją:

- Šiaulių miesto ir rajono moksleiviai nenori rinktis specializacijų „Kompiuteriai ekonomikoje ir raštvedyboje“ ir „Daugiaterpė aplinka“.
  Tendenciją galima interpretuoti dvejopai:
  - Mokyklose gerai dėstomos šios temos ir moksleiviams nėra informacijos trūkumo;
  - Moksleiviai mano, kad žinių gali įgyti savarankiškai, ir renkasi kitas specializacijas.

Norint patvirtinti arba paneigti šias hipotezes, reikėtų tyrimą papildyti naujais duomenimis.

Be paminėtų modelių, naudojant sistemą XPERTRULE MINER 1.47 sugeneruoti sprendimų medžiai. Analizuojant medį, iliustruojantį moksleivių lūkesčius, išryškėja tendencijos:

- Didžioji dalis merginų į JKM įstoja norėdamos išmokti dirbti kompiuteriu (60%).
- Mažiau merginų norėtų išmokti programuoti ($\sim$40% iš vienos mokyklų grupės).
- Dauguma vaikinų stoja į JKM norėdami išmokti programuoti ($\sim$90%).
- Vaikinai, kuriems silpniau sekasi matematika, be programavimo, renkasi interneto puslapių kūrimą, nori pasirengti studijoms universitete ir išmokti dirbti kompiuteriu.

Apibendrinsime naudotų programinių produktų taikymo galimybes. 1 lentelėje pateikiami programiniai produktai ir duomenų gavybos modeliuose vartoti metodai.

Pastebime, kad daugiausiai analizės priemonių turi INTELLIGENT MINER for Data, todėl ja galima parengti išsamiausius duomenų gavybos modelius.

XPERTRULE MINNER konstruojami sprendimų medžiai, klasteriai ir asociacijų taisyklės lengvai interpretuojami, gerai realizuota grafinė vartotojo sąsaja. INTELLIGENT MINER modeliuose labai detaliai pateikiamos duomenų laukų charakteristikos, to pasigendama analizuojant XPERTRULE MINNER klasterizacijos rezultatus.

WIZWHY sistema generuoja IF-THEN taisykles. Taisyklės konstruojamos rankiniu būdu, kiekvieną kartą pasirenkant ir pažymint priklausomąjį kintamąjį. Sugeneruotos taisyklės pateikiamos tekstiniu sąrašu. Daug laiko užima sugeneruotų taisyklių analizė. Tai didelis sistemos trūkumas.

STATGRAPHICS, XPERTRULE MINNER ir WIZWHY modelių rezultatai patvirtina INTELLIGENT MINER for Data rezultatus. Tai rodo, kad gaunami patikimi rezultatai. Aukščiau išvardintomis duomenų gavybos priemonėmis gauti rezultatai patvirtina tradiciniais statistiniais metodais gautus rezultatus, sutrumpina duomenų analizės laiką, rezultatus pateikia keliomis formomis. Tiek galimybių neturi statistinės programos.

1 lentelė. Duomenų gavybos modeliuose naudoti metodai

Tyrimo tikslas	Programinis produktas			
	STATGRAPHICS Plus 5.0	WIZWHY 3.0	XPERTRULE MINER 1.47	INTELLIGENT MINER for Data V.8.1
Moksleivių bendrasis pasirengimas	Aprašomoji statistika, daugialypė regresinė analizė, grupės vidurkio metodu atlikta klasterizacija	IF-THEN taisyklės	Sprendimų medis	Aprašomoji statistika, demografinė klasterizacija, neuroninė klasterizacija, sprendimų medis
Moksleivių lūkesčiai	–	IF–THEN taisyklės	Sprendimų medis, asociacijų taisyklės ir klasterinė analizė	Aprašomoji statistika, demografinė klasterizacija, neuroninė klasterizacija, sprendimų medis

3. Rekomendacijos ir išvados

JKM veiklai tobulinti siūlomos šios **rekomendacijos**:

- Moksleivius pirmaisiais mokslo metais tikslinga skirstyti į grupes pagal bendrąjį ir specialųjį pasirengimą.
- Siūlant moksleiviams specializaciją, apie ją reikėtų išsamiau informuoti. Specializaciją moksleiviai galėtų rinktis antraisiais mokslo metais.
- Individualizuoti pirmųjų mokymosi metų programas pagal pasirengimo lygį.
- Moksleiviai, kuriems gerai sekasi chemija, nenori išmokti gerai programuoti. Jiems siūlyti kitas alternatyvas.
- Kadangi moksleiviai įgytas žinias planuoja naudoti studijuodami, tikslinga skirti daugiau kūrybinių-projektinių darbų, kur dėstytojai būtų konsultantais.

Išvados:

- ŠU JKM veiklai tirti naudoti keli programiniai produktai. Šiais produktais gautų rezultatų analizė rodo, kad išsamiausi gavybos ir tyrimų rezultatai gaunami sistema INTELLIGENT MINER for Data.
- Iš ŠU JKM veiklos tyrimo rezultatų lyginamosios analizės suformuluotos mokyklos veiklos tendencijos ir rekomendacijos mokyklos darbuotojams.
- Atliktą tyrimą galima traktuoti kaip išankstinį, atliktą nedideliam respondentų skaičiui. Ateityje reikėtų šį tyrimą pakartoti didesniam tiriamųjų skaičiui.

Literatūra

[1] G. Dzemyda, O. Kurasova, Dirbtinių neuroninių tinklų taikymas bendrojo lavinimo mokykloms palyginti, *Informacijos mokslai*, **22** (2002).

[2] J.F. Elder IV, W. Dean, Abbott elder research, a comparison of leading data mining tools, in: *Fourth International Conference Knowledge Discovery & Data Mining* (1998).
[3] *Nacionalinio jaunųjų mokslininkų konkurso darbų tezės*, Vilnius (2002).
[4] *Nacionalinio jaunųjų mokslininkų konkurso darbų tezės*, Vilnius (2003).
[5] *NBA Caches Score Big with IBM Data Mining Application*, IBM (2001).
[6] D. Skillicorn, Strategies for parallel data mining.
http://miles.cnuce.cnr.it/ palmeri/datam/articles/p4026.pdf
[7] S. Turskienė, Computer technology and teaching mathematics in secondary schools, *Informatics in Education*, **1**, 149–156 (2002).
[8] http://www.megaputer.com, 2002-12-03.
[9] В.И. Городецкий, В.В. Самойлов, А.О. Малов, *Современное состояние технологии извлечения знаний из баз и хранилищ данных*.
http://space.iias.spb.su/ai/doc/Gorodetski-DM-Overview.pdf

Application of data mining technologies for analysis of Šiauliai university's young computer users' school

S. Turskienė, G. Kulvietis, R. Burbaitė

The paper analyses possibilities of using data mining technologies for investigation of activity of Šiauliai university's young computer users' school. The application's domains, principal stages of mining model's creation are analyzed. Data mining models young computer users' school was made by using WIZWHY, XPERTRULE MINER, STATGRAPHICS, INTELLIGENT MINER SYSTEMS. The basic tendencies of young computer users' school activity are presented and the recommendations for school's activity improvement are formulated.