

VILNIAUS UNIVERSITETAS

Rimantas Kybartas

DAUGELIO KLASIŲ ATPAŽINIMAS NAUDOJANT KLASIFIKATORIUS
POROMS

Daktaro disertacijos santrauka
Fiziniai mokslai, informatika (09 P)

Vilnius, 2010

Disertacija rengta 2005-2010 metais Vilniaus universitete.

Mokslinis vadovas:

prof. habil. dr. Šarūnas Raudys (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P)

Disertacija ginama Vilniaus universiteto Informatikos mokslo krypties taryboje:

Pirmininkas

prof. dr. Romas Baronas (Vilniaus universitetas, fiziniai mokslai, informatika – 09P)

Nariai:

prof. habil. dr. Rimvydas Simutis (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07T)

doc. dr. Minija Tamošiūnaitė (Vytauto Didžiojo universitetas, fiziniai mokslai, informatika – 09P)

doc. dr. Rimantas Vaicekaskas (Vilniaus universitetas, fiziniai mokslai, informatika – 09P)

prof. dr. Olegas Vasilecas (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – 07T)

Oponentai:

prof. dr. Dalius Navakauskas (Vilniaus Gedimino technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07T),

doc. dr. Algirdas Bastys (Vilniaus universitetas, fiziniai mokslai, informatika – 09P)

Disertacija bus ginama viešame Informatikos mokslo krypties tarybos posėdyje 2010 m. rugsėjo 20 d. 10 val. Vilniaus universiteto Matematikos ir informatikos fakulteto Nuotolinių studijų centre

Adresas: Šaltinių g. 1A, LT – 03225, Vilnius, Lietuva

Disertacijos santrauka išsiuntinėta 2010 m. rugpjūčio mėn. 20 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje

VILNIUS UNIVERSITY

Rimantas Kybartas

MULTI-CLASS RECOGNITION USING PAIR-WISE CLASSIFIERS

Summary of doctoral dissertation
Physical sciences, informatics (09 P)

Vilnius, 2010

Dissertation work was carried out at Vilnius University from 2005 to 2010.

Scientific supervisor:

prof. habil. dr. Šarūnas Raudys (Vilnius University, physical sciences, informatics – 09 P)

The dissertation is being defended at the Council of Scientific Field of Informatics at Vilnius University:

Chairman

prof. dr. Romas Baronas (Vilnius University, physical sciences, informatics – 09 P)

Members:

prof. habil. dr. Rimvydas Simutis (Kaunas University of Technology, technological sciences, informatics engineering – 07T)

doc. dr. Minija Tamošiūnaitė (Vytautas Magnus University, physical sciences, informatics – 09 P)

doc. dr. Rimantas Vaicekaskas (Vilnius University, physical sciences, informatics – 09 P)

prof. dr. Olegas Vasilecas (Vilnius Gediminas Technical University, technological sciences, informatics engineering – 07T)

Opponents:

prof. dr. Dalius Navakauskas (Vilnius Gediminas Technical University, technological sciences, informatics engineering – 07T),

doc. dr. Algirdas Bastys (Vilnius University, physical sciences, informatics – 09 P)

The dissertation will be defended at the public meeting of the Council of Scientific Field of Informatics in the Distance Learning Center of VU MIF on the 20th of September, at 10:00.

Address: Šaltinių 1A, LT – 03225, Vilnius, Lithuania

The summary of dissertation was distributed on the 20th of August, 2010.

The dissertation is available at the library of Vilnius University

Turinys

Terminai ir apibrėžimai.....	vii
1. Įvadas	1
1.1 Tyrimų sritis	1
1.2 Problemos aktualumas	1
1.3 Tyrimo objektas	2
1.4 Tyrimo tikslai ir uždaviniai.....	3
1.5 Tyrimo metodai.....	3
1.6 Mokslinis naujumas	3
1.7 Praktinė darbo reikšmė.....	4
1.8 Ginamos tezės	4
1.9 Disertacijos struktūra	4
2. Disertacijos rezultatai.....	5
2.1 Daugelio klasių klasifikavimo uždavinys	5
2.2 Klasifikatoriai.....	6
2.3 Porinis „Fuzzy Templates“ apjungimo metodas.....	8
2.3.1 Porinio „Fuzzy Templates“ apjungimo metodo privalumai ir trūkumai	9
2.4 Klasifikavimo uždavinio problemos	10
2.5 Klasifikatorių poroms (VSP ir AV) sudėtingumas	11
2.6 Klasifikatorių palyginimo rezultatai	14
2.6.1 Duomenys.....	14
2.6.2 Mokymo procesas.....	15
2.6.3 Rezultatai.....	15
2.7 Taikomasis uždavinys.....	17
2.7.1 Uždavinio sprendimas	17
2.8 Kiti daugelio klasių klasifikavimo aspektai.....	19

2.8.1	Netikslaus optimizavimo kriterijaus ir duomenų kiekio įtaka klasifikavimo klaidai	19
2.8.2	Eksperimentų kiekio svarba	20
2.8.3	Apjungimo metodų pataisymas	21
3.	Rezultatai ir išvados	22
3.1	Rekomendacijos daugelio klasių uždavinio sprendimui	22
3.2	Pagrindinės išvados	22
3.3	Kiti rezultatai	23
	Publikacijos disertacijos tema	24
	Trumpos žinios apie autorių	25
	Summary	25
	Literatūra	27

Terminai ir apibrėžimai

Apibendrinančioji klasifikavimo klaidos tikimybė – angl. *generalization error*, kai kur tekste (pagal kontekstą) trumpinama iki „klasifikavimo klaida“.

Duomenų klasė – duomenų aibė, turinti bendrų savybių, aktualių taikomojo uždavinio dalykinei sričiai.

Duomenų vektorius – vienas duomenų aibės egzempliorius, sudarytas iš d požymių (atributų).

Klasifikatorius poroms – angl. *pair-wise classifier*, apmokytas klasifikatorius, gebantis duomenų vektorių priskirti vienai iš dviejų duomenų klasių.

Mokymo duomenys – duomenų aibė, skirta tinkamiems klasifikatoriaus parametrams parinkti.

Testavimo duomenys – duomenų aibė, skirta klasifikatoriaus veikimo tikslumui patikrinti.

Validavimo duomenys – duomenų aibė, naudojama parenkant klasifikatoriaus parametrus.

Įvadas

1.1 Tyrimų sritis

Kasdieniniame gyvenime kiekvienas žmogus susiduria su daugelio klasių atpažinimo (klasifikavimo) uždaviniais: skaitydamas priskiria grafinius simbolius raidėms ir atitinkamiems garsams, atlikdamas kasdieninius darbus pagal tam tikrus požymius skirsto juos į įvairaus svarbumo kategorijas ir pan. Daugybė daugelio klasių klasifikavimo uždavinių yra ir įvairiausiuose pramonės ar kuriose kitose veiklos srityse: mineralų klasifikavimas pagal jų išvaizdą bei cheminę sudėtį; augalų klasifikavimas pagal jų vaizdą, formą, vaisių skonį; ligos priskyrimas pagal ligonio išvardintus simptomus bei tyrimų rezultatus. Daugelį šių verslo bei pramonės uždavinių atlieka žmonės. Todėl ne retai ir gana paprastų uždavinių sprendimas brangiai kainuoja, gali pasitaikyti klaidų susijusių su žmogiškuoju faktoriumi, riboto darbo laiko, o kartais net yra pavojingas. Stengiantis išvengti šių trūkumų ir automatizuoti industrijos ar verslo uždavinių sprendimą, buvo pradėti naudoti tam tikri statistiniai klasifikavimo metodai. Ši sritis yra nuolat plėtojama, todėl dabar yra daugybė tiek statistinių, tiek empirinių daugelio klasių klasifikavimo uždavinių sprendimo metodų.

1.2 Problemos aktualumas

Sprendžiant pramonės, verslo, socialinius ar bet kokios kitos srities daugelio klasių klasifikavimo uždavinius, gali būti naudojami tokie klasikiniai statistiniai metodai kaip Fišerio diskriminantinė funkcija ar kvadratinė diskriminantinė funkcija. Tačiau šių metodų naudojimas duoda pakankamai gerus rezultatus tik tuomet, kuomet duomenys turi tam tikras statistines charakteristikas, kurioms atitinkami statistiniai klasifikatoriai yra optimalūs. Visgi taikomuosiuose uždaviniuose duomenys dažniausiai yra gana sudėtingi – ryšiai tarp požymių yra ne tiesiniai, skirtingos klasės yra smarkiai persidengusios ir pan. Tokiais atvejais gali būti naudojami ir statistiniai metodai – tik kiek sudėtingesni, pavyzdžiui, naudojantys branduolio funkcijas. Šalia statistinių metodų yra ir nemažai kitų daugelio klasių atpažinimo uždaviniui skirtų ne statistiniu pagrindu paremtų metodų pavyzdžiui, artimiausio vidurkio ar artimiausio kaimyno metodai. Pastaruoju metu dažniausiai (ypatingai tam tikrų dalykinių sričių taikymo uždaviniuose) yra naudojami dirbtiniai neuroniniai tinklai (DNT). Standartinis K -klasių neuroninis vieno sluoksnio perceptronų tinklas yra sudarytas iš vieno sluoksnio turinčio K vieno sluoksnio perceptronų (VSP) [1] – [3]. Tačiau tokia kompozicija veikia ne visais daugelio klasių atpažinimo uždavinio atvejais. Naudojant DNT su dar sudėtingesne architektūra, išskyla sunkumų, susijusių su DNT veikimo trukme bei generalizavimu. Buvo ir yra atliekama daugybė tyrimų, skirtų daugelio klasių atpažinimo uždavinio sprendimui pagerinti. Viena iš sėkmingiausių tokio pagerinimo kryptų yra dviejų etapų metodai, paremti klasifikatoriais poroms. Pirmame etape yra naudojami $K(K-1)$ klasifikatoriai poroms, sugebantys atpažinti tik dvi klases. Antrajame etape yra naudojamas tam tikras klasifikatorių poroms rezultatų apjungimo metodas galutiniam

atpažįstamo duomenų vektoriaus priskyrimui konkrečiai klasei. Dažniausiai pirmajame etape kaip klasifikatoriai poroms yra naudojami atraminiai vektoriai arba sprendimų medžiai [4] – [6]. Antrajame etape galutiniam duomenų priskyrimui tam tikrai klasei taip pat yra sudaryti ir specialūs algoritmai [7] – [10]. Kullback-Leibler atstumas [5], įvairios sąlyginių tikimybių kvadratų sumos [6], [11] ir kitos apytikslės klasifikavimo klaidos išraiškos yra naudojamos klasifikavimo klaidai daugelio klasių klasifikavimo uždavinyje sumažinti.

Daugelyje tyrimų skirtumai tarp nagrinėtų klasifikatorių poroms rezultatų apjungimo metodų buvo nežymūs. Dėl metodų sudėtingumo ar jų sprendimo priėmimo funkcijų netiesiškumo dažniausiai yra pernelyg sudėtinga teoriškai įvertinti metodų tikslumą. Dėl to dauguma tyrimų, lyginančių skirtingus metodus, tiesiog naudoja eksperimentinį būdą metodų tikslumui ir vieno prieš kitą pranašumui nustatyti. Labai dažnai gana mažas kiekis eksperimentų yra atliekamas net su mažų imčių duomenimis kuriuos padalijus į mokymo ir testavimo duomenis gaunama ne reprezentatyvi testavimo duomenų aibė. Nagrinėdami metodų tikslumą tyrėjai paprastai labiau akcentuoja skirtingų duomenų kiekį, tačiau dažnai pamiršta apie eksperimentų kiekio patikimumą. Jei duomenų nėra daug, atsitiktinis jų skaidymas į daugelio klasių klasifikatoriaus mokymo ir testavimo duomenis gali duoti skirtingus rezultatus. Dėl šios priežasties tokio metodo klaidos įvertinimas tampa nepatikimas.

Kita populiarī K klasių vieno sluoksnio perceptrono modifikacija yra K atskirų vieno sluoksnio perceptronų, kurių kiekvienas moka atskirti kažkurią tai vieną klasę nuo visų kitų. T.y. taip vadinama „vienas prieš visus“ (angl. „one-against-all“) strategija. Tačiau buvo parodyta [4], kad dviejų etapų klasifikavimas naudojant klasifikatorius poroms yra potencialiai geresnė daugelio klasių klasifikavimo strategija.

Dėl egzistuojančios klasifikatorių poroms ir jų apjungimo metodų gausos nėra aišku, kada ir kuriuos klasifikatorius poroms ir kuriuos jų apjungimo metodus naudoti tam, kad būtų gautas kiek įmanoma optimalus konkretaus daugelio klasių klasifikavimo uždavinio sprendimo metodas, duodantis kuo mažiau klasifikavimo klaidų.

Egzistuojančios teorijos (neretai kaip prielaidos) tinkamas ir aiškus išaiškinimas yra kur kas vertingesnis nei dar vieno naujo metodo pristatymas be tinkamo teorinio paaiškinimo. Dviejų etapų klasifikavimo metodai, paremti klasifikatoriais poroms, yra viena iš tokių iki galo neišaiškintų sričių. Dviejų etapų algoritmai yra perspektyvūs, nes tyrėjai gali išnaudoti dviejų klasių tarpusavio savybes – parinkti skirtingus požymius, skirtingo sudėtingumo klasifikatorius. Be to, tampa kur kas paprasčiau išspręsti nesubalansuotų mokymo duomenų klasių problemą. Taigi, yra reikalingas teorinis išaiškinimas, kodėl daugelio klasių klasifikavime dviejų etapų algoritmai, paremti klasifikatoriais poroms, yra tokie perspektyvūs.

1.3 Tyrimo objektas

Tyrimo objektas yra daugelio klasių klasifikavimas naudojant dviejų etapų klasifikavimo metodus, kurių pirmajame etape yra naudojami specializuoti klasifikatoriai poroms, o antrajame etape yra apjungiami pirmojo etapo klasifikatorių rezultatai ir gražinamas galutinis sprendimas – duotojo duomenų vektoriaus priskyrimas iš anksto apibrėžtai (naudojant mokymo duomenis, suskirstytus į klases) klasei.

1.4 Tyrimo tikslai ir uždaviniai

Šių tezių tikslas yra pasiūlyti daugelio klasių klasifikavimo metodą, kuris būtų kiek galima universalus mažų imčių ir nesubalansuotų duomenų klasių atvejais su nežinomomis apriorinėmis klasių pasiskirstymo tikimybėmis bei pasiskirstymo funkcijomis. Taip pat turi būti pateikti ir išsamūs šio metodo privalumų ir trūkumų paaiškinimai tam, kad tai nebūtų eilinis dar vienas daugelio klasių klasifikavimo metodas. Šiam tikslui pasiekti yra iškeliami tokie uždaviniai:

1. Išaiškinti daugelio klasių klasifikavimo sudėtingumo ir duomenų kiekio aspektus metoduose, paremtuose dviejų etapų klasifikatoriais, pirmajame etape naudojančiais klasifikatorius poroms.
2. Išaiškinti netikslaus klasifikavimo klaidos įvertinimo kriterijaus, naudojamo antrajame etape apjungiant klasifikatorių poroms rezultatus, įtaką pačiai klasifikavimo klaidai.
3. Išsamiai palyginti skirtingus klasifikatorių poroms apjungimo metodus.
4. Remiantis tyrimų rezultatais pateikti rekomendacijas daugelio klasių klasifikavimo uždavinio sprendimo konstravimui.

1.5 Tyrimo metodai

Paprasti tiesiniai klasifikatoriai buvo parinkti kaip klasifikatoriai poroms tam, kad būtų galima teoriškai išaiškinti bet kokių klasifikatorių poroms apjungimo teikiamą pranašumą. Tyrimo metu buvo atlikta tiek teorinė analizė, tiek eksperimentiniai tyrimai ir su realiais, ir su generuotais duomenimis. Analizės metu buvo pabrėžiami duomenų kiekio, nevienodų klasių imčių (nesubalansuotų klasių) bei klasifikatorių poroms apjungimo kriterijų aspektai. Analitiškai gauti rezultatai buvo panaudoti sudėtingame realiame uždavinyje su mineralų duomenų klasifikavimu. Tam, kad būtų gauti patikimi eksperimentiniai rezultatai, įvertinimai buvo atliekami panaudojant duomenų permaišymus (N_e) 250, 500 ar net 1000 kartų. Metodų įvertinimui buvo naudojamas klaidingai suklasifikuotų testavimo duomenų vektorių procentas (klasifikavimo klaida).

1.6 Mokslinis naujumas

Šio darbo mokslinis naujumas yra tai, kad yra aiškiai parodyta, kodėl dviejų etapų klasifikatoriai, paremti paprastais klasifikatoriais poroms, yra gera alternatyva sudėtingiems daugelio klasių klasifikatoriams. Buvo pasiūlytas tiesinio vieno sluoksnio perceptrono naudojimas vietoje populiarių atraminių vektorių klasifikatorių (tiesinių variantų) parodant pirmųjų pranašumą tiek teoriškai, tiek eksperimentiškai. Buvo pasiūlytas naujas klasifikatorių poroms apjungimo metodas (paremtas analogišku sprendimu daugelio klasių klasifikatoriams), pateikiant šio metodo privalumų ir trūkumų analizę. Darbe taip pat buvo išaiškintas skirtumas tarp konkuruojančių Hastie ir Tibshirani bei Wu, Lin ir Weng pasiūlytų klasifikatorių poroms apjungimo būdų. Buvo

parodyta, kad klasifikatorių poroms apjungimo taisyklės pataisymas naudojant triukšmo pridėjimą („naujiems“ duomenims gauti) yra perspektyvesnis nei analitinis pataisymas. O galiausiai buvo atliktas platus įvairių tipų klasifikatoriais poroms paremtų dviejų etapų daugelio klasių klasifikavimo metodų palyginimas.

1.7 Praktinė darbo reikšmė

Šios disertacijos praktinė reikšmė yra ta, kad remiantis joje aprašytų tyrimų rezultatais yra pateikiamos rekomendacijos daugelio klasių klasifikavimo uždavinio sprendimui. Praktiniuose uždaviniuose dažnai kylančios problemos, susijusios su duomenų imtimi, nesubalansuotumu ar tinkamos daugelio klasių klasifikatoriaus architektūros parinkimo problemos yra sprendžiamos pasiūlant dviejų etapų klasifikavimo metodą, paremtą vieno sluoksnio perceptronais kaip klasifikatoriais poroms bei jų rezultatų apjungimu.

1.8 Ginamos tezės

1. Dviejų etapų klasifikavimo metodai, paremti klasifikatoriais poroms gali būti tinkama alternatyva sudėtingiems algoritmams daugelio klasių klasifikavimo uždavinyje.
2. Vieno sluoksnio perceptronas yra tinkamas pasirinkimas kaip klasifikatorius poroms. Jis daugeliu atveju duoda geresnius rezultatus nei tiesinis atraminių vektorių klasifikatorius.
3. Porinis *Fuzzy Templates* metodas gali klasifikuoti su mažesne klasifikavimo klaida negu kiti daugelio klasių klasifikatoriai tada, kai statistiniai klasių porų parametrai žymiai skiriasi. Tuo tarpu šis metodas nėra rekomenduotinas, kai klasės yra išsidėsčiusios panašiais atstumais, simetriškos. Klasifikatoriaus poroms gražinamos reikšmės dydį reguliuojančio parametro tinkamas parinkimas gali pagerinti porinio *Fuzzy Templates* metodo veikimą abiem minėtais atvejais.

1.9 Disertacijos struktūra

Disertacija susideda iš septynių skyrių. Pirmajame skyriuje yra pateikta tezių įžanga, kur yra nusakomi darbo tikslai, uždaviniai, aktualumas, mokslinių tyrimų naujumas ir pan.

Antrajame skyriuje yra apibrėžiamas daugelio klasių klasifikavimo uždavinys, akcentuojamos pagrindinės šio uždavinio problemos. Šiame skyriuje taip pat pateikiama trumpa esamų daugelio klasių klasifikavimo uždavinio sprendimo būdų – daugelio klasių klasifikatorių bei jų apjungimo – apžvalga.

Trečiajame skyriuje yra pristatomi dviejų etapų klasifikatoriai paremti klasifikatoriais poroms. Parodoma, kodėl šie klasifikatoriai gali būti geresni nei

standartiniai vieno etapo daugelio klasių klasifikatoriai. Trečiajame skyriuje atliekama pakankamai išsami antrojo etapo – klasifikatorių poroms apjungimo metodų – apžvalga. Šalia egzistuojančių šio tipo algoritmų, pateikiamas ir naujas – porinis *Fuzzy Templates* metodas.

Ketvirtajame skyriuje yra nagrinėjamos mažų imčių bei nesubalansuotų duomenų klasių problemos bei pademonstruojama, kaip jos yra sprendžiamos dviejų klasių atveju.

Penktajame skyriuje yra nagrinėjama netikslaus klasifikatorių poroms apjungimo optimizavimo kriterijaus įtaka klasifikavimo klaidai. Lyginami du metodai paremti tokiais kriterijais. Skyriuje taip pat pateikiama nepakankamo eksperimentų kiekio problema bei parodoma, kodėl ir kokios yra jos pasekmės. Penktajame skyriuje taip pat yra pateikiami visų nagrinėtų dviejų etapų klasifikatorių, naudojančių tiek vieno sluoksnio perceptroną, tiek atraminių vektorių klasifikatorių kaip klasifikatorių poroms. Šalia palyginimui pateikiami ir standartinių daugelio klasių klasifikatorių rezultatai. Atliekama pateiktų rezultatų analizė. Skyriaus pabaigoje pateikiami bandymai pagerinti klasifikatorių poroms apjungimo algoritmą, įvedant naujus validavimo duomenis.

Šeštasis skyrius yra skirtas sudėtingo praktinio uždavinio (geologinių duomenų klasifikavimo) sprendimui. Sprendžiant šį uždavinį yra panaudojami disertacijoje gauti rezultatai. Papildomai pasitelkus panašumo požymius yra gaunami rezultatai, kurie gali būti praktiškai pritaikomi industrijoje.

Septintajame skyriuje yra pateikiamas rekomendacijų rinkinys daugelio klasių uždavinio sprendimui. Šiame skyriuje taip pat pateikiami pagrindiniai bei papildomi disertacijos rezultatai

Disertacijos rezultatai

2.1 Daugelio klasių klasifikavimo uždavinys

Pirmiausia apibrėškime daugelio klasių klasifikavimo uždavinį. Tarkime, kad turime duomenų aibę sudarytą iš vektorių turinčių d atributų (t.y. dimensija yra d) $x = \{x_1, x_2, \dots, x_d\}$. Kiekvienas duomenų vektorius yra priskirtas vienai ir tik vienai iš anksto dalykinės srities eksperto apibrėžtų klasių $\Pi_1, \Pi_2, \dots, \Pi_K$, kur $K \geq 3$. Kadangi klasės yra iš anksto apibrėžtos, duomenų nebereikia klasterizuoti. Daugelio klasių klasifikavimo uždaviniu vadinsime naujo duomenų vektoriaus x priskyrimą vienai iš K klasių, pasinaudojant informacija, gauta iš jau turėtų duomenų bei atitinkamų vektorių priklausymo klasėms. Tokio uždavinio sprendimo korektiškumas bus matuojamas klaidos santykiu (klasifikavimo klaida), kuris bus gaunamas padalinus naujų duomenų vektorių, klaidingai priskirtų klasėms, kurioms jie faktiškai nepriklauso kiekį iš visų naujų duomenų vektorių kiekio.

Duomenys, kurie jau yra priskirti klasėms, bus vadinami mokymo, o duomenys pagal kuriuos nustatoma klasifikavimo klaida, – testavimo duomenimis.

2.2 Klasifikatoriai

Bene populiariausia priemonė daugelio klasių klasifikavimo uždaviniui spręsti yra dirbtiniai neuroniniai tinklai (DNT) [1], [2], [3], [12]. DNT sukonstravimo idėja yra paimta iš neuronų esančių biologinio organizmo smegenyse. Buvo parodyta [13], kad standartinis daugiasluoksnis neuroninis tinklas su tam tikromis sąlygomis aktyvavimo funkcijai gali bet koku tikslumu aproksimuoti bet kokią tolydžią funkciją. Tačiau praktiškai šis teiginys yra sunkiai pritaikomas, nes kiekvienam uždaviniui reikia parinkti konkrečią neuroninio tinklo architektūrą, kas taip pat yra sudėtingas uždavinys. Dalinai fiksuotą architektūrą turi radialinių bazinių funkcijų tinklai [1], [2], [3], [14].

Dirbtiniai neuroniniai tinklai, jei jie tarpusavyje nekombinuojami, atstovauja vieno etapo sprendimo priėmimo strategiją. Alternatyva šiai strategijai yra dviejų etapų klasifikavimo strategija, kur pirmajame etape yra naudojami tam tikri klasifikatoriai, o tada jų rezultatai yra tam tikru būdu apjungiami. Pirmame etape naudojami klasifikatoriai gali klasifikuoti tiek K klasių, tiek dvi klases. Nuo to dažniausiai priklauso ir tai, kokia rezultatų apjungimo taisyklė gali būti naudojama. Šioje disertacijoje pirmajame etape dėl lankstumo, geresnių prisitaikymo prie statistinių duomenų savybių bei galimybių toliau tobulinti yra siūloma naudoti klasifikatorius klasifikuojančius dvi klases, t.y. klasifikatorius poroms.

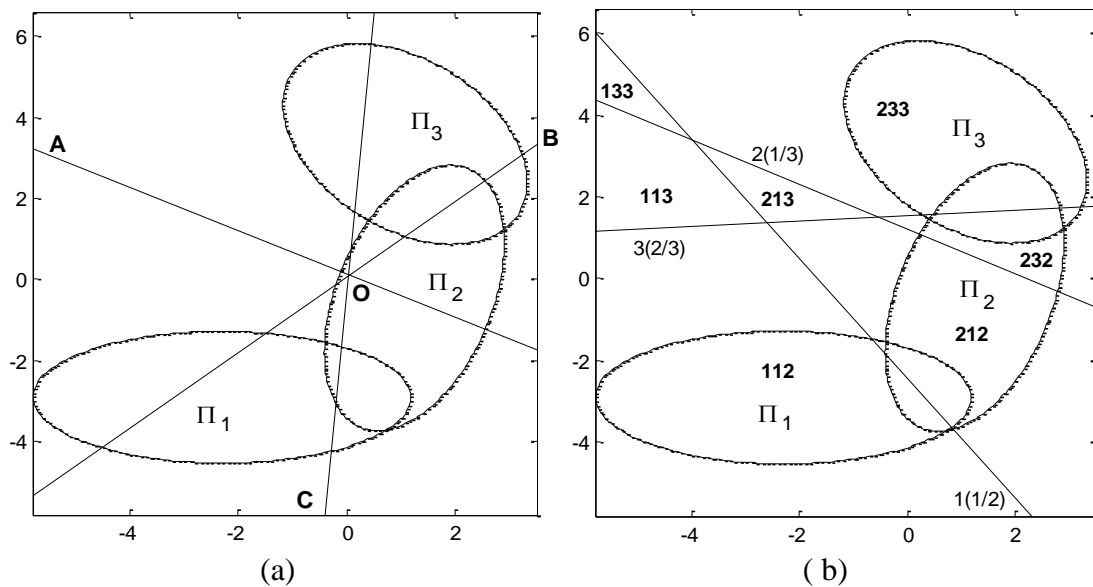
Panagrinėkime neuroninį tinklą sudarytą iš K perceptronų. Toks tinklas grąžina K rezultatų $o_i = f(v^T x + w_0) = f(w^T z)$, ($i = 1, 2, \dots, K$), kur v yra d -matis svorių vektorius, w_0 yra bazinis vektorius, $w = (w_0, v^T)^T$, $z = (1, x^T)^T$, o $f(s)$ yra netiesinė sigmoidinė aktyvavimo funkcija, t.y.

$$f(s) = 1/(1+\exp(-s)) \quad (1)$$

Svoriams gauti yra minimizuojama nuostolių funkcija

$$cost = \frac{1}{N_1 + N_2 + \dots + N_K} \sum_{h=1}^K \sum_{j=1}^K \sum_{s=1}^{N_j} [t_{jh} - f(\mathbf{z}_{js}^T \mathbf{w}_h)]^2 \quad (2)$$

kur t_{jh} yra reikalinga grąžinti reikšmė, o N_k , $k=1, \dots, K$ yra klasių duomenų kiekiai. Sigmoidinei aktyvavimo funkcijai reikalingos grąžinti reikšmės parenkamos taip: $t_{jh} = 1$ kai $h = j$, ir $t_{jh} = 0$ kai $h \neq j$. Daugelio klasių atveju formuojant svorius, skirtus konkrečios klasės atskyrimui, įtakos turi vis duomenys, kai tuo tarpu dviejų klasių atveju svorių optimizavime dalyvauja tik tų dviejų klasių duomenys, kurių klasifikavimui ir bus skirtas klasifikatorius. Dėl šios priežasties tinklas, sudarytas iš K perceptronų gali duoti žymiai prastesnius rezultatus nei tinkamai apjungti $K(K-1)/2$ perceptronų skirtų klasifikuoti klasių poras. Tokia situacija pavaizduota paveiksle nr. 1.



Paveikslas 1. Klasifikavimo regionai suformuoti a) neuroninio tinklo iš trijų vieno sluoksnio perceptrono; b) trijų skirtingų vieno sluoksnio perceptronų poroms.

Paveiksle 1(a) matosi, jog klasifikatoriai susikirsdami vienoje vietoje (dėl funkcijos (1)) atskiria klases neoptimaliai, kai tuo tarpu klasifikatoriai poroms (paveikslas 1(b)) optimizuoja klasių tarpusavio atskyrimą.

Kaip klasifikatoriai poroms šiame darbe yra nagrinėjami vieno sluoksnio perceptronai [1], [3] bei atraminių vektorių klasifikatoriai [15]. Pirmieji buvo pasirinkti dėl jų savybės mokymo metu evoliucionuoti per septynis skirtingus statistinius klasifikatorius bei didelių galimybių tobulinti įvedant papildomus parametrus. Ten, kur VSP reikšmė buvo reikalinga tik klasifikavimui buvo naudojama tik svorių suma priskiriant vienai iš klasių poros ji buvo lyginama su slenksčiu 0,5. Ten, kur reikėjo aktyvavimo funkcijos (pvz.: poriniame „*Fuzzy Templates*“ apjungimo algoritme) kaip aktyvavimo funkcija buvo naudojama sigmoidinė funkcija (1).

Atraminiai vektoriai buvo pasirinkti dėl didelio jų populiarumo pastaruoju metu. Kadangi vieno sluoksnio perceptronas pagal savo struktūrą klasifikavimo uždavinyje yra tiesinis tam, kad palyginimo sąlygos būtų vienodos, buvo pasirinktas ir tiesinis atraminių vektorių klasifikatorius. Visi kiti mokymo parametrai taip pat buvo maksimaliai suvienodinti. Eksperimentuose naudoti duomenys bei jų permaišymai buvo identiški.

Klasifikatorių poroms apjungimui yra sukurta nemažai strategijų. Šiame darbe klasifikatorių poroms rezultatams apjungti buvo pasirinkti du balsavimo, du tikimybiniai ir vienas gražinamų reikšmių panašumu pagrįstas metodas. Vienas iš balsavimo metodų – paprastas balsavimas, kai sprendimas dėl naujo vektoriaus priklausomybės priskyrimo konkrečiai klasei yra priimamas pagal balsų daugumą (dar žinomas kaip „*Max Wins*“ metodas). T.y. kiekvienas klasifikatorius poroms pasako, kuriai iš dviejų klasių, kuriai jis yra skirtas klasifikuoti, priklauso duotasis vektorius. Vektorius yra priskiriamas tai klasei, kuri surenka daugiausia balsų. Galimas atvejis, kuomet tą patį maksimalų balsų skaičių surenka kelios klasės. Tokiu atveju sprendimas yra priimamas atsitiktinai parenkant kažkurią iš šių klasių. Kitas pasirinktas balsavimo metodas buvo kiek kitoks balsavimo metodas, kai iš klasifikatorių poroms yra sukonstruojamas kryptinis ne ciklinis grafas (angl. *directed acyclic graph*) (DAG). Duotas naujasis vektorius yra

perduodamas šakniniam klasifikatoriui ir priklausomai nuo jo sprendimo perduodamas tolimesniam klasifikatoriui, kol galiausiai priklausomybę klasei identifikuoja klasifikatorius poroms, kuris yra kaip grafo lapas. Šis metodas buvo pateiktas darbui su atraminiais vektoriais [9]. Šioje disertacijoje vietoje atraminių vektorių taip pat buvo naudojami ir vieno sluoksnio perceptronai.

Kiti du klasifikatorių poroms rezultatų apjungimo metodai – Hastie-Tibshirani (H-T) [5] ir Wu, Lin, Weng (WLW) [6] – yra tikimybiniai. Pirmasis jų (pasirinktas dėl didelio populiarumo) yra paremtas Kullback-Leibler atstumo (3) tarp tikimybių įvertinimo \hat{r}_{ij} ir tikrų tikimybių μ_{ij} minimizavimu:

$$D_{KL}(p_1, p_2, \dots, p_K, \mathbf{x}_z) = \sum_{i \neq j} (N_i + N_j) \hat{r}_{ij} \log(\hat{r}_{ij} / \mu_{ij}), \text{ su sąlyga, kad } \sum_{i=1}^K p_i = 1, \quad (3)$$

$$p_i \geq 0.$$

Čia p_i yra aposteriorinė klasės Π_i tikimybė. Kitas tikimybinis metodas – WLW – buvo pasirinktas dėl jo autorių teigiamo pranašumo prieš H-T metodą. WLW metodas taip pat vertina tikimybes, naudodamas tam tikrą kriterijų – pasvertų sąlyginių porinių tikimybių skirtumų kvadratų sumą:

$$D_{WLW2}(p_1, p_2, \dots, p_K, \mathbf{x}_z) = \sum_{i \neq j} (p_i \hat{r}_{ji} - p_j \hat{r}_{ij})^2, \text{ su sąlyga, kad } \sum_{i=1}^K p_i = 1, p_i \geq 0. \quad (4)$$

Paskutinis pasirinktas klasifikatorių poroms rezultatų apjungimo metodas buvo metodas, paremtas klasifikatorių poroms gražinamų skaitinių reikšmių panašumu į šabloną, sudarytą iš tų pačių klasifikatorių gražintų reikšmių su mokymo duomenimis. Plačiau šis metodas yra aprašytas kitame skyrelyje.

2.3 Porinis „Fuzzy Templates“ apjungimo metodas

Kaip jau ir buvo minėta, disertacijos autoriui nepavyko aptikti klasifikatorių poroms apjungimo metodų, kurie naudotų klasifikatorių poroms gražinamas skaitines reikšmes ir jų panašumus su anksčiau gautomis skaitinėmis reikšmėmis. Tokie antro etapo apjungimo metodai egzistuoja ir yra sėkmingai taikomi, kai pirmajame etape yra naudojami daugelio klasių klasifikatoriai [16], [11]. Todėl tokio tipo klasifikatorių poroms apjungimą teko sukonstruoti tezių rengimo metu.

Kaip pagrindas buvo paimtas Liudmilos Kunčevos pristatytas metodas, skirtas daugelio klasių klasifikatorių apjungimui [11] ir pritaikytas fiksuoto kiekio $(K(K-1)/2)$ klasifikatorių apjungimui. Pirmiausia buvo pakeisti klasių Π_i „neryškūs“ (angl. *fuzzy*) šablonai iš matricų į vektorius $F_i = \{f_i(l)\}$ su $K-1$ atributų, kur

$$f_i(l) = \frac{\sum_{z=1}^{N_k} C_{m,n}(x_z)}{N_k} \quad (5)$$

visiems $l=1..K-1$ klasifikatoriams $C_{m,n}$ kur $m=i$ arba $n=i$, $\{\mathbf{x}_z, z=1..N\}$ yra aiškiai klasėms priskirti mokymo duomenys, N_k yra mokymo vektorių kiekis klasėje Π_k , $C_{m,n}(x_z)$ yra klasifikatoriaus klasių porai m ir n gražinama reikšmė. Tik klasifikatoriai su klasių numeracija

$$(i, j) = (1, 2), (1, 3) \dots (K-1, K) \quad (6)$$

yra aktualūs, nes $C_{j,i}(\mathbf{x}_z) = 1 - C_{i,j}(\mathbf{x}_z)$, o kiekvienai duomenų klasei yra $K-1$ unikalių klasifikatorių (nekartojant klasių poros). „Neryškaus“ šablono vektoriaus reikšmės yra išrikiuojamos pagal klasifikatorių eilės tvarka nurodyta (6) formulėje.

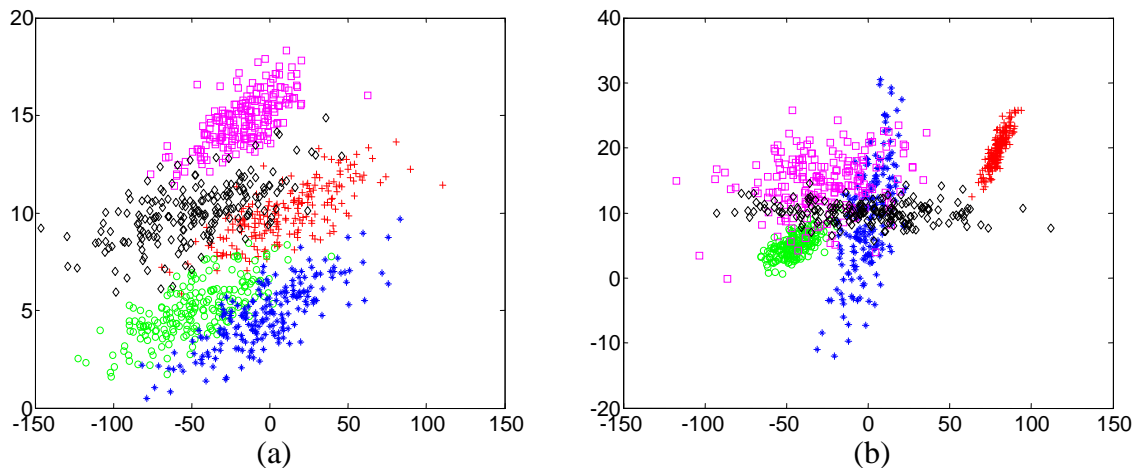
Kai klasifikavimui yra paduodamas naujas klasifikavimo vektorius \mathbf{x}_z jam yra suskaičiuojamas kiekvienos klasės sprendimo profilis $DP_i(\mathbf{x}_z) = \{C_{m,n}(\mathbf{x}_z)\}$, kur $m = i$ arba $n = i$, išrūšiuotas pagal formulę (6).

Galutinis sprendimo priėmimas yra atliekamas pagal funkciją

$$\max_i (S(F_i, DP_i(\mathbf{x}_z))), \text{ kur } S(F_i, DP_i(\mathbf{x}_z)) = 1 - \frac{1}{K-1} \sum_{l=1}^{K-1} (f_i(l) - DP_{i,l}(\mathbf{x}_z))^2 \quad (7)$$

2.3.1 Porinio „Fuzzy Templates“ apjungimo metodo privalumai ir trūkumai

Kadangi porinio *Fuzzy Templates* metodo rezultatas tiesiogiai priklauso nuo klasifikatorių poroms reikšmių, šis metodas yra tuo efektyvesnis kuo šios reikšmės labiau skiriasi viena nuo kitos.



Paveikslas 2. Penkios dvimatės Gauso pasiskirstymą turinčios klasės a) išdėstytos simetriškai su tomis pačiomis kovariacinėmis matricomis; b) išdėstytos atsitiktine tvarka su skirtingomis kovariacinėmis matricomis.

Išnagrinėkime du dvimačių duomenų atvejus. Pirmuoju atveju yra sugeneruotos penkios skirtingos simetriškai išsidėsčiusios duomenų klasės su vienodomis kovariacinėmis matricomis (žr. paveikslą 2(a)). Šioje situacijoje daugumos klasių porų optimalus klasifikatorius turėtų būti Fišerio diskriminantinė funkcija. Tad ir vieno sluoksnio perceptronas dėl jo sugebėjimo mokymo metu evoliucionuoti per septynis skirtingus statistinius klasifikatorius [17] turėtų sustoti ties ta vieta, kurioje jis elgiasi kaip Fišerio diskriminantinė funkcija. Dėl tos priežasties ir daugumos klasifikatorių poroms pagrįstų VSP grąžinamos reikšmės turėtų būti tarpusavyje apytiksliai panašios. Antroje duomenų klasių aibėje yra sugeneruotos penkios skirtingos duomenų klasės su skirtingomis kovariacinėmis matricomis ir skirtingai išsidėsčiusios viena kitos atžvilgiu

(žr. paveikslą 2(b)). Šiuo atveju kiekvienai klasių porai optimalus turėtų būti vis kitoks statistinis klasifikatorius, o tai reiškia, kad ir VSP mokymo metu turėtų būti sustota vis kitame etape. Eksperimentų su šiais duomenimis rezultatai yra pateikti lentelėje nr. 1. Kaip ir buvo galima atikėtis, pirmuoju atveju porinis „Fuzzy Templates“ algoritmas veikė pakankamai blogai palyginus su kitais klasifikatorių poroms apjungimo algoritmais. Tai yra dėl to, kad VSP mokymosi metu sustoja daugmaž tose pačiose vietose ir funkcija (1) visada gražina panašią reikšmę (reikšmių vidurkis yra 0,6519, t.y. arčiau 1/2 nei 1, kas reiškia, jog svoriai dar nespėjo išaugti), dėl ko yra sunku atskirti kuriai iš klasių priklauso naujas duomenų vektorius. Panaudojus atraminius vektorius (AV) kaip klasifikatorius poroms, ir be jų gražinamai svorių sumai taip pat pritaikius sigmoidinę funkciją (1), rezultatai buvo kiek geresni. Tai įvyko dėl to, kad prieš pritaikant sigmoidinę funkciją atraminių vektorių sumų reikšmės buvo pakankamai plačiai išsibarsčiusios.

Duomenys	Klasifikatorius	Porinis FT	H-T	Balsavimas	DAG
Aibė (a)	VSP	0.294	0.115	0.121	0.121
	AV	0.119	0.117	0.118	0.118
Aibė (b)	VSP	0.193	0.210	0.257	0.266
	AV	0.175	0.194	0.237	0.230

Lentelė 1. Dviejų generuotų duomenų aibių rezultatai panaudojant skirtingus klasifikatorius poroms ir skirtingus jų apjungimo metodus (geriausi rezultatai pažymėti pastorintu šriftu)

Iš lentelės nr. 1 matosi, kad prielaidos apie duomenų aibės (b) tinkamumą poriniam „Fuzzy Templates“ metodui pasiteisino. Šiuo atveju kaip klasifikatorius poroms naudojant tiek VSP, tiek atraminius vektorius pristatomas metodas davė geriausius rezultatus. Nors ir šiuo atveju atraminiai vektoriai pagal rezultatus pranoko vieno sluoksnio perceptroną kaip klasifikatorių poroms, tai neturėtų būti priimama kaip taisyklė – viskas priklauso nuo duomenų. Taip pat sigmoidinėje funkcijoje (1) įvedus naują mastelio keitimo parametą α , t.y. $f(s) = 1/(1+\exp(-\alpha s))$, ir tinkamai jį parinkus, porinis „Fuzzy Templates“ metodas su VSP kaip klasifikatoriumi poroms duoda ne prastesnius rezultatus nei su atraminiais vektoriais. Taip pat tinkamai parinktas parametras padeda pagerinti rezultatus ir tais atvejais, kai duomenų išsidėstymas yra nepalankus poriniam „Fuzzy Templates“ metodui. Tačiau tinkamo parametro α parinkimo metodai reikalauja papildomų tyrimų ir disertacijoje nebuvo nagrinėjami.

2.4 Klasifikavimo uždavinio problemos

Viena iš problemų, susijusių su daugelio klasių klasifikavimu ir sudėtingų mokymo algoritmų naudojimu, yra tai, kad sudėtingais algoritmais apmokytos sudėtingos struktūros (pvz. daugiasluoksnis neuroninis tinklas) yra linkusios prisitaikyti prie mokymo duomenų, tokiu būdu prarasdamos savo apibendrinamąsias savybes.

Kai žymiai skiriasi duomenų klasių kiekiai bei pasiskirstymas, tokia duomenų aibė yra vadinama nesubalansuota. Nesubalansuotų duomenų problema yra viena iš kelių kritiškiausių dirbtinio intelekto problemų. Didesnis klasių duomenų nesubalansuotumas padidina daugelio klasių klasifikavimo uždavinio sudėtingumą [18]. Ypač stipriai ši

problema veikia metodus, kuris buvo sukonstruoti darant prielaidą, kad duomenų klasių pasiskirstymas yra subalansuotas. Empiriniai tyrimai [18] parodė, kad daugelis metodų veikia dviejų klasių atveju, kai tuo tarpu didžioji dauguma jų nebeveikia daugelio klasių klasifikavimo uždavinyje. Ir nepaisant to, kad daug tyrėjų teigia, jog jų išvados susijusios su nesubalansuotomis klasėmis, gali būti taikomos ir daugelio klasių klasifikavimo uždaviniui [19], iš tiesų tėra vos keletas tyrimų (pvz. [18]), susijusių su nesubalansuotomis klasėmis daugelio klasių klasifikavimo uždavinyje.

Vadinamoji mažos duomenų imties problema atsiranda tada, kai klasių dydžiai (duomenų vektorių kiekiai) $N_k, k=1, \dots, K$ yra maži lyginant su duomenų dimensija d . Tuomet kovariacinių matricų, naudojamų standartiniuose statistiniuose metoduose (pvz. tiesinėje ar kvadratinėje diskriminantinėse analizėse), įvertinimas pasidaro labai netikslus. O kai duomenų kiekis yra mažesnis už dimensiją, t.y. $N_k < d$, net ne visi matricos įvertinimui naudojami parametrai gali būti apskaičiuojami. Šios problemos efektą yra aiškiai pristatęs J. H. Friedmanas [20].

2.5 Klasifikatorių poroms (VSP ir AV) sudėtingumas

Dviejų etapų klasifikavimo algoritmai suteikia galimybę pirmajame etape pasirinkti klasifikatorius, kurie yra optimalūs kiekvienai klasių porai. Dėl šios priežasties vieno sluoksnio perceptronas, kurį mokant jo sudėtingumas vis didėja, atrodo labai patrauklus klasifikatorių poroms. Šiame skyrelyje bus analizuojama hipotetinė situacija, kai daugiamatės duomenų klasės turi Gausinius pasiskirstymus su vienodomis kovariacinėmis matricomis. Tokiems duomenims asimptotškai (kai $N_1 \rightarrow \infty$ ir $N_2 \rightarrow \infty$) optimalus klasifikatorius yra Fišerio diskriminantinė funkcija, kurią vieno sluoksnio perceptronas imituoja pradinėse mokymosi iteracijose. Porinės Fišerio diskriminantinės funkcijos klasifikavimo klaida gali būti apskaičiuota pagal šią asimptotinę formulę [3], [21]

$$EP_N^F \approx q_i \Phi \left\{ -\frac{1}{2} \delta_{ij} T_{\mathbf{m}_i/ij} T_{\Sigma} \right\} + q_j \Phi \left\{ -\frac{1}{2} \delta_{ij} T_{\mathbf{m}_j/ij} T_{\Sigma} \right\} \quad (8)$$

kur $\Phi\{a\} = \int_{-\infty}^a (2\pi)^{-1/2} \exp(-t^2/(2\sigma^2)) dt$ yra standartinė Gauso suminė pasiskirstymo funkcija,

$$\begin{aligned} T_{\mathbf{m}_i/ij} &= \frac{1 - d/(N_i \delta_{ij}^2) + d/(N_j \delta_{ij}^2)}{\sqrt{1 + d/(N_i \delta_{ij}^2) + d/(N_j \delta_{ij}^2)}} \\ T_{\mathbf{m}_j/ij} &= \frac{1 + d/(N_i \delta_{ij}^2) - d/(N_j \delta_{ij}^2)}{\sqrt{1 + d/(N_i \delta_{ij}^2) + d/(N_j \delta_{ij}^2)}} \\ T_{\Sigma} &= \sqrt{\frac{N_i + N_j - d}{N_i + N_j}} \end{aligned} \quad (9)$$

δ_{ij} yra Mahalanobio atstumas tarp klasių Π_i ir Π_j

$$\delta_{ij} = ((\Delta_{ij})^T \Sigma^{-1} \Delta_{ij})^{1/2} \quad (10)$$

$\Delta_{ij} = m_i - m_j$, m_i , m_j yra klasių vidurkių vektoriai, o Σ yra jų „bendra“ kovariacinė matrica.

Formulėje (8) daugikliai $T_{mi/ij}$ ir $T_{mj/ij}$ atsiranda dėl netikslaus vidurkių vektorių įvertinimo iš mokymo duomenų imties. Daugiklis T_{Σ} atsiranda dėl netikslios kovariacinės matricos (daroma prielaida, kad ji yra bendra abiemis klasėms Π_i ir Π_j) įvertinimo iš mokymo duomenų imties. Euklido atstumo klasifikatoriuje kovariacinės matricos įvertinti nereikia. Todėl formulėje (8) nebereikia daugiklio T_{Σ} . Pastaruoju atveju standartinis Mahalanobio atstumas neapibrėžia asimptotinės klasifikavimo klaidos tikimybės. Vietoje to yra naudojamas „efektyvus atstumas“ ir „efektyvi dimensija“ ([3], 3 skyrius):

$$\begin{aligned} \delta^*_{ij} &= (\Delta_{ij})^T \Delta_{ij} ((\Delta_{ij})^T \Sigma \Delta_{ij})^{-1/2} \\ d^*_{ij} &= ((\Delta_{ij})^T \Delta_{ij})^2 \text{tr} \Sigma^2 ((\Delta_{ij})^T \Delta_{ij})^{-1/2} \end{aligned} \quad (11)$$

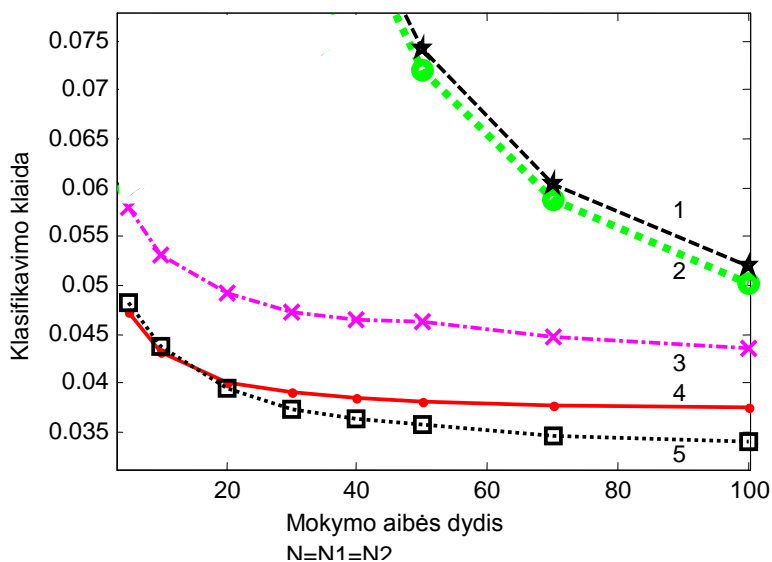
Mahalanobio atstumas Δ_{ij} ir „efektyvūs“ parametrai δ^*_{ij} , d^*_{ij} yra specifiniai kiekvienai klasių porai. Paprastai $\delta^*_{ij} < \delta_{ij}$, $1 < d^*_{ij} < \infty$ [3].

Atvejis, kai $N_2 = N_1$. Jeigu klasių duomenų kiekiai yra vienodi, t.y. $N_2 = N_1 = N$, o apriorinės tikimybės taip pat yra vienodos, t.y. $q_2 = q_1 = 0.5$, išraiškos (9) pasidaro paprastesnės. Tolimesniame pavyzdyje buvo naudoti 30-mačiai duomenys turintys Gauso pasiskirstymą. Jie buvo sugeneruoti pasitelkus nekilnojamo turto duomenų (žr. skyrelį 2.6.1) klasių vidurkius ir kovariacines matricas. Kadangi šie duomenys turi 17 požymių, likę 13 požymių buvo pridėti tokie, kad būtų neinformatyvūs, su dispersija $\sigma^2 = 0.1^2$. Sugeneruotų duomenų klasės sudaro ovalus ir yra išsidėsčiusios lanku. Dvi gretimos klasės Π_1 ir Π_2 bei Π_2 ir Π_3 šiek tiek persidengia (žr. paveikslą nr. 1). Tokiems duomenims Euklido atstumo bei Fišerio diskriminantinės funkcijos klasifikatorių asimptotinė klasifikavimo klaida (P_{∞}^E ir P_{∞}^F) yra beveik vienoda, todėl skirtumui tarp šių klasifikavimo klaidų padidinti buvo įvestos tam tikros kovariacinių matricų modifikacijos – $\Sigma_{1\text{new}} = \Sigma_1/\theta$ ir $\Sigma_{2\text{new}} = \Sigma_2 \times \theta$. Skaičiavimuose buvo naudojama nauja matrica: $\Sigma = 1/2(\Sigma_{1\text{new}} + \Sigma_{2\text{new}})$. Lentelėje nr. 2 parodoma kaip nuo to pasikeitė parametrai δ_{ij} , δ^*_{ij} , d^*_{ij} ir P_{∞}^F , P_{∞}^E .

Ekspirimentas	δ_{ij}	δ^*_{ij}	P_{∞}^F	P_{∞}^E	d^*_{ij}
Paveikslas nr. 3, $\theta = 0.7$	3.646	3.612	0.0342	0.03557	1.90
Paveikslas nr. 4, $\theta = 0.4$	3.234	3.030	0.0530	0.06492	1.84

Lentelė 2. Trijų 30-mačių Gausinių klasių duomenų parametrai

Naudotame pavyzdyje efektyvi dimensija buvo $d_{12}^* = 1.90$. Tai yra kur kas mažiau nei originali dimensija $d=30$. Tai konstatuoja faktą, jog Euklidinio atstumo klasifikatoriaus jautrumas mokymo duomenų kiekiui šiuo konkrečiu atveju yra mažas. Atlikus minėtų duomenų skaitinę analizę pasimatė (žr. paveikslą nr. 3), kad mažų mokymo duomenų imčių atveju Fišerio klasifikatoriaus apibendrinančioji klaidos tikimybė (juoda brūkšniuota kreivė - 1) yra didesnė nei asimptotinė klaidos tikimybė (žalia taškuota kreivė - 2).



Paveikslas 3. Apibendrinančiosios klasifikavimo klaidos kaip klasių dydžio funkcijos: 1 – Fišerio DF (eksperimentinė), 2 – Fišerio DF (teorinė, formulės (8), (9)), 3 – atraminių vektorių klasifikatorius, 4 - Euklido atstumo klasifikatorius, 5 – pseudo-validavimo būdu sustabdytas vieno sluoksnio perceptronas

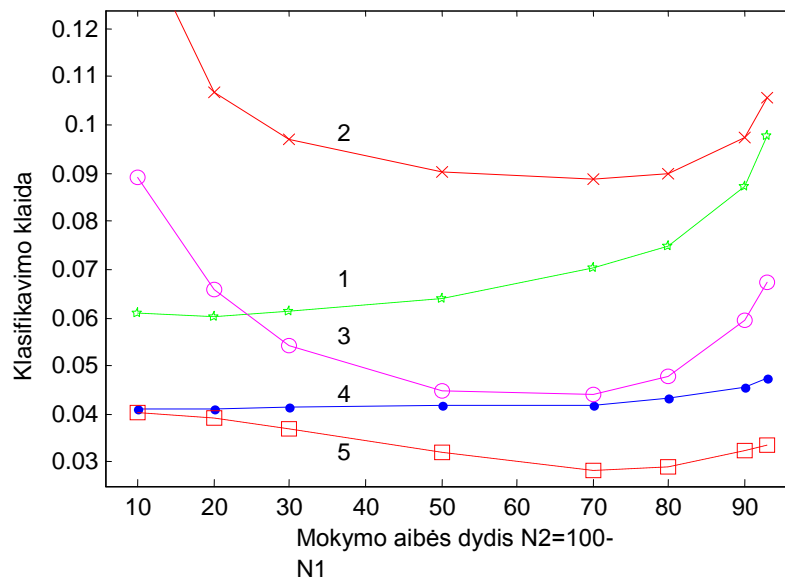
Paveiksle nr. 3 taip pat parodytos ir iš 1000 eksperimentų įvertintos kitų algoritimų apibendrinančios klasifikavimo klaidos: Euklidinio atstumo klasifikatoriaus (raudona ištiesinė kreivė – 4), naudojant validavimo duomenis (prie mokymo duomenų pridėjus spalvotą triukšmą) sustabdytas vieno sluoksnio perceptronas (juodi kvadratėliai – 5) ir atraminių vektorių klasifikatorius (rausvi kryžiukai - 3) kuriems parametras C buvo parinktas naudojant taip pat validavimo duomenis.

Šis duomenų modelis (naudojant validavimo duomenis reguliarizavimo parametro C nustatymui) yra nepalankus atraminių vektorių klasifikatoriui. Naudojant standartinį parametru C ($C = 1$) atraminių vektorių klasifikatorius parodė pastebimai geresnius rezultatus – labai artimus Euklidiniam atstumo klasifikatoriui. Tuo tarpu panaudojus validavimo duomenis vieno sluoksnio perceptrono sustabdymui buvo gauti geriausi rezultatai iš visų naudotų algoritimų.

Atvejis, kai $N_2 \neq N_1$. Kai duomenų apriorinės tikimybės yra skirtingos, o mokymo duomenų kiekiai nesubalansuoti, (9) lygybėse esantys daugikliai $d/(N_i \delta_{ij}^2)$, $d/(N_j \delta_{ij}^2)$ ir netiesinė suminė pasiskirstymo funkcija lygybėje (8) lemia tai, kad nei Euklidinis atstumo klasifikatorius, nei Fišerio klasifikatorius nėra optimalūs. Šis neoptimalumas taip pat yra paveldimas ir atraminiais vektoriais bei vieno sluoksnio perceptronais paremtuose klasifikatoriuose.

Paėmus ir išanalizavus daugelio nesubalansuotų klasių atvejį matytųsi, kad didinant vienos duomenų klasės duomenų kiekį, Fišerio diskriminantinės funkcijos klaida mažėja. Tarkime, turime dviejų klasių atvejį, kur duomenų klasių kiekiai N_1 , N_2 neatspindi tikrųjų apriorinių tikimybių q_1 ir q_2 . Tarkime, kad $q_2 = 0.75$, tačiau N_2 yra kintantis dydis. Nuolatinis bendras abiejų klasių duomenų kiekis lieka nepakitęs: $N = N_1 + N_2 = 100$.

Paveiksle nr. 4 pateikiamos klasių Π_1 , Π_2 poros (iš anksčiau minėto duomenų pavyzdžio, klasių kovariacinės matricos $\Sigma_2 \neq \Sigma_1$), apibendrinančios klasifikavimo klaidos kaip funkcijos nuo N_2 Euklidinio atstumo, Fišerio, atraminių vektorių bei vieno sluoksnio perceptronų. Tačiau šį kartą vertinant VSP nuostolių funkciją (2) kiekvienos klasės suma yra padauginama iš tos klasės apriorinės tikimybės ir duomenų kiekio santykio. Paveiksle nr. 4 yra pavaizduoti 1000 eksperimentų rezultatai. Duomenys buvo naudoti tie patys kaip ir prieš tai buvusiame pavyzdyje, tik šį kartą kovariacinių matricių skirtumo parametras buvo $\theta = 0.4$ (žr. lentelę nr. 2). Optimaliai sustabdytas VSP buvo beveik nejautrus nesubalansuotiems duomenų klasių kiekiams (N_1 , N_2). Visais atvejais tai buvo geriausias klasifikatorius. Jeigu vietoj apriorinių tikimybių būtų naudojami klasių kiekių N_i santykiai su visos duomenų aibės kiekiu N , klasifikavimo rezultatai būtų buvę blogesni.



Paveikslas 4. Dviejų klasių apibendrinančios klasifikavimo klaidos funkcijos nuo antros klasės duomenų kiekio: 1 – Fišerio DF (eksperimentinė), 2 – Fišerio klasifikatorius (teorinė, formulės (8), (9)), 3 – atraminių vektorių klasifikatorius, 4 – Euklidinio atstumo klasifikatorius, 5 – naudojant validavimo duomenis sustabdytas VSP su modifikuota nuostolių funkcija.

2.6 Klasifikatorių palyginimo rezultatai

2.6.1 Duomenys

Eksperimentams su realiais duomenimis buvo naudotos šios duomenų aibės:

- Chromosomų duomenys – 30 požymių, 24 klasės, kurių kiekviena turi po 500 duomenų vektorių.
- Iris – duomenys nusakantys gėlės žiedo dydį [22] – 3 klasės po 50 keturmačių vektorių.
- Nekilnojamo turto duomenys – trys duomenų klasės su atitinkamais 17-mačiais duomenų kiekiais (118, 160 ir 94).
- Satelito duomenys – šešios skirtingų dydžių klasės sudarytos iš 36-iamačių duomenų vektorių, aprašančių skirtingų pasėlių nuotraukas.
- Kviečių duomenys – 12-mačiai duomenys, aprašantys po 80 penkių rūšių kviečių geometrines savybes.
- Mielių duomenys – 1500-mačiai duomenys, aprašantys 10 skirtingų mielių infekcijų. Duomenų klasių kiekiai skiriasi nežymiai.
- Vyno duomenys – 13-mačiai duomenys, nusakantys tris skirtingas vyno rūšis. Kiekvienos klasės duomenų kiekiai lyginant su dimensija negausūs bei skirtingi
- Bakterijų *Ecoli* duomenys – aštuonmačiai duomenys, nusakantys penkias skirtingas proteinų lokalizacijos vietas gram-teigiamose bakterijose.

2.6.2 Mokymo procesas

Tam, kad gauti patikimus rezultatus eksperimentai buvo atliekami 250×2 kartų: kaskart klasių viduje permaišant duomenis ir pusę jų naudojant mokymui bei validavimo duomenų sudarymui, o kitą pusę – testavimui. Kitoje iteracijoje mokymui ir tertavimui naudoti duomenys buvo sukeisti. Tokia procedūra buvo kartojama $N_e = 250$ kartų (dėl pernelyg didelio duomenų kiekio tik 25×2 eksperimentų buvo atlikta su chromosomų duomenimis). Prieš mokant klasifikatorius, duomenys buvo normalizuojami naudojant standartinius kiekvieno požymio nuokrypius. Tada buvo pritaikoma transformacija paremta principinėmis komponentėmis ir tikrinėmis reikšmėmis. Taip pat prieš mokant klasifikatorius poroms, kiekvienos klasių poros vidurkiai kaskart būdavo pastumiami į nulį. VSP mokymo sustabdymui nustatyti ar atraminių vektorių reguliarizavimo parametrui C parinkti buvo naudojami validavimo duomenys – prie mokymo duomenų pridėtas spalvotas triukšmas. Remiantis patirtimi, buvo pridėtas triukšmas su tokiais standartiniais parametrais: artimiausių kaimynų kiekis $k = 2$; standartinis triukšmo vektorių nuokrypis $\sigma_{\text{noise}} = 1.0$; triukšmo vektorių kiekis kiekvienam duomenų vektoriui $n_{i_{\text{nn}}} = 2$.

2.6.3 Rezultatai

Eksperimentų rezultatai pateikiami lentelėje nr. 3. Pirmame stulpelyje yra išvardinti klasifikavimo metodai. Pirmose trijose eilutėse pateikti klasifikavimo klaidų rezultatai standartiniams vieno etapo daugelio klasių klasifikatoriams. Likusiose eilutėse pateikiami dviejų etapų klasifikatorių rezultatai, kai kaip klasifikatoriai poroms buvo naudojami atraminiai vektoriai (viršutinė eilutės dalis) bei vieno sluoksnio perceptronai

(apatinė eilutės dalis). Paskutinėje eilutėje (pažymėtoje σ^N) pateikiamas geriausio metodo klasifikavimo klaidos standartinis nuokrypis σ , padalintas iš $\sqrt{N_e}$ ($N_e = 250$; $N_e = 25$ chromosomų duomenims).

Metodai	Chromo- somų	Iris	Nekiln. turt	Satelito	Kviečių	Mielių	Vyno	Ecoli
<i>KSLP</i>	0.262	0.097	0.078	0.177	0.088	0.136	0.025	0.125
<i>KDA</i>	0.270	0.033	0.074	0.214	0.098	0.131	0.023	0.136
<i>RBF</i>	0.202	0.031	0.053	0.149	0.073	0.190	0.024	0.129
<i>+Balsav.</i>	0.195	0.042	0.056	0.147	0.069	0.149	0.032	0.138
	0.198	0.035	0.045	0.142	0.071	0.137	0.032	0.131
<i>+H-T</i>	0.193	0.038	0.056	0.144	0.067	0.146	0.032	0.147
	0.200	0.027	0.043	0.147	0.063	0.129	0.024	0.136
<i>+WLW</i>	0.193	0.038	0.056	0.145	0.067	0.147	0.032	0.148
	0.194	0.027	0.043	0.147	0.063	0.130	0.024	0.136
<i>+Fuzz.</i>	0.197	0.040	0.055	0.144	0.069	0.144	0.031	0.138
	0.200	0.038	0.098	0.141	0.197	0.150	0.095	0.348
<i>+DAG</i>	0.197	0.042	0.056	0.148	0.069	0.150	0.031	0.140
	0.199	0.036	0.045	0.143	0.072	0.140	0.034	0.133
σ^N	$8*10^{-4}$	10^{-3}	$7*10^{-4}$	$4*10^{-4}$	$9*10^{-4}$	10^{-3}	$9*10^{-4}$	10^{-3}

Lentelė 3. Įvairių daugelio klasių klasifikavimo algoritmų eksperimentų su realiais duomenimis klasifikavimo klaidų rezultatai

Kaip matoma iš klasifikavimo klaidų su realiais duomenimis rezultatų, dauguma atvejų geriausius rezultatus pateikė dviejų etapų algoritmai, paremti vieno sluoksnio perceptronais kaip klasifikatoriais poroms.

2.7 Taikomasis uždavinys

Gavus disertacijos rezultatus buvo iškeltas uždavinys pritaikyti juos realiame gyvenime, t.y. su duomenimis, kurie yra naudojami praktikoje realiame uždavinyje (duomenys, aprašyti 2.6.1 yra plačiai naudojami algoritmų testavimui). Tokius duomenis pateikė kolegos iš Turkijos, Selčuko universiteto. Duomenis sudaro 6-mačiai duomenų vektoriai, nusakantys dviem skirtingais apšvietimo tipais apšviestų mineralų pikselius. Patys duomenys yra gana sudėtingi – duomenų kiekiai klasėse skirtingi, skirtingų klasių duomenys persidengę (žr. paveikslą 5(a)).

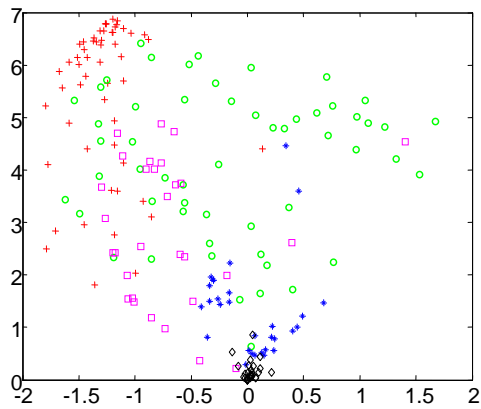
Geologiniams duomenims paprastai yra naudojami dirbtiniai neuroniniai tinklai [23], [24]. Tačiau dažniausiai jų naudojimas nėra tinkamas, nes kiekvienam duomenų modeliui reikia parinkti atskirą neuroninio tinko architektūrą. Tyrimuose paprastai yra apsiribojama vieno sluoksnio neuroniniu tinklu su fiksuotu neuronų kiekiu paslėptame sluoksnyje. Atliekant šį tyrimą pačioje pradžioje taip pat buvo bandyta naudoti dirbtinį neuroninį tinklą, tačiau dėl jo gražinamos didelės klasifikavimo klaidos jo buvo atsisakyta.

2.7.1 Uždavinio sprendimas

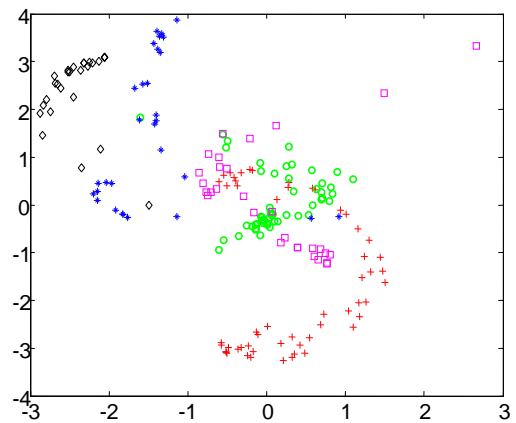
Pirmiausia duomenims supaprastinti buvo panaudoti panašumo požymiai. Šį kartą i -tąjį duomenų vektorių pažymėkime x_i . Tada jo panašumo požymiai yra išreiškiami formule:

$$s_i^j = \exp(-\alpha * \sqrt{\sum_{k=1}^d (x_i^k - x_j^k)^2}) \quad (12)$$

kur d yra duomenų dimensija, j yra panašumo požymio indeksas, duomenų vektoriaus viršutinis indeksas k žymi k -ąjį originalių duomenų atributą, o α yra normalizavimo koeficientas. Panašumo požymiai testiniams duomenims sudaromi analogiškai, tik naudojant ne testavimo, o mokymo duomenų atributus. Tokiu būdu yra gaunama nauja dimensija lygi mokymo duomenų kiekiui. Nors nagrinėjamu atveju dimensija išaugo, tačiau padidėjo duomenų klasių atsiskyrimas (žr. paveikslą 5(b)).



(a)



(b)

Paveikslas 5. Originalūs penkių mineralų klasių duomenys (a) bei duomenys, transformuoti su panašumo požymiais (b). Duomenys atvaizduoti dimensijos sumažinimui panaudojus principinių komponentių analizę.

Lentelėje nr. 4 pateikiami rezultatai tiek su originaliais, tiek su modifikuotais duomenimis. Palyginimui buvo naudojami trys standartiniai daugelio klasių klasifikatoriai ir trys dviejų etapų, naudojantys klasifikatorius poroms.

Duomenys	Vieno etapo klasifikatoriai			Dviejų etapų klasifikatoriai		
	RBF	KDA	K-SLP	Balsavimas	DAG	H-T
Originalūs	0.189	0.226	0.252	0.212/0.211	0.215/0.211	0.226/0.218
Panašumo pož.	0.177	0.212	0.174	0.227/0.174	0.238/0.174	0.173/0.183

Lentelė 4. Originalių bei modifikuotų duomenų klasifikavimo rezultatai su standartiniais daugelio klasių klasifikatoriais (RBF – radialinių bazinių funkcijų neuroniniu tinklu; KDA – branduolio funkcijos diskriminantine funkcija (Parzeno langas); K-klasių vieno sluoksnio perceptronu) bei dviejų etapų klasifikatoriais (balsavimu, kryptiniu acikliniu grafu (DAG) ir Hastie-Tibshirani apjungimo metodu), naudojančiais atraminiais vektorius (pirma reikšmė) bei vieno sluoksnio perceptronus (antra reikšmė) kaip klasifikatorius poroms.

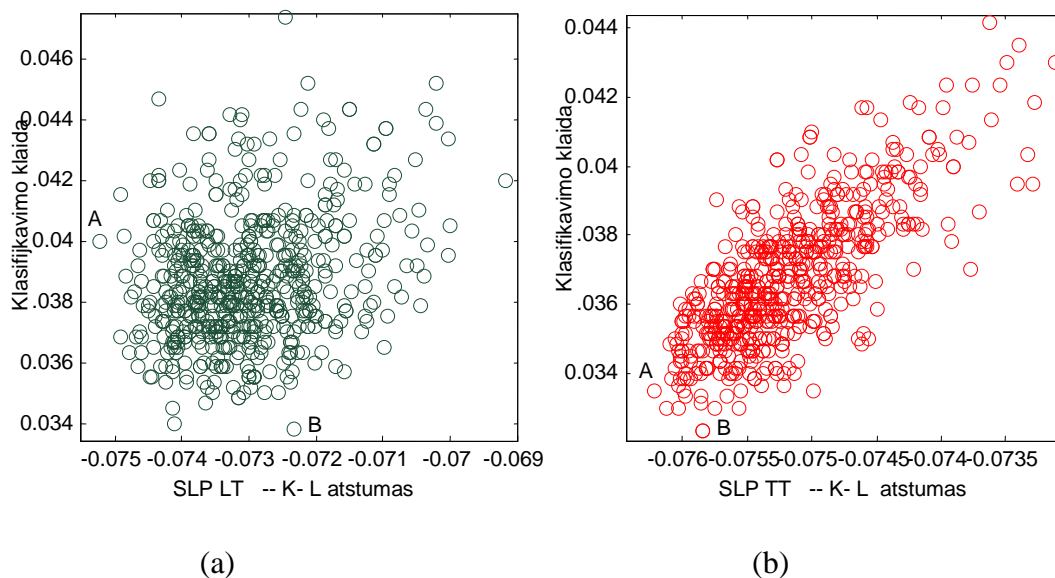
Rezultatai, pateikti lentelėje nr. 4, rodo, kad panašumo požymių pritaikymas pasiteisino. Rezultatai taip pat patvirtina ir disertacijoje pateikiamus teiginius – dviejų etapų klasifikatoriai, naudojantys klasifikatorius poroms sudėtingame uždavinyje nenusileido standartiniais daugelio klasių klasifikatoriams. Nors vidutinė visų metodų klasifikavimo klaida yra gana nemaža, ji yra priimtina ir taikytina realiame mineralų klasifikavimo uždavinyje.

2.8 Kiti daugelio klasių klasifikavimo aspektai

2.8.1 Netikslaus optimizavimo kriterijaus ir duomenų kiekio įtaka klasifikavimo klaidai

Vertinant klasifikatorių tikslumą yra galimi du klaidų šaltiniai: a) netikslius klasifikavimo tikslumo įvertinimo kriterijus; b) klasifikatorių konstravimas yra paremtas mokymo duomenimis, kai tuo tarpu klasifikatoriaus tikslumui nusakyti yra naudojami testiniai duomenys. Jų įtakai nustatyti buvo atlikti eksperimentai su dirbtiniais 30-mačiais duomenimis, suskirstytais į tris klases ir turinčiais tas pačias statistines charakteristikas kaip ir šios santraukos 2.5 skyriuje naudotieji.

Pirmiausia klasifikatorių mokymui buvo sugeneruota 600 klasių trejetų, turinčių po 50 30-mačių duomenų vektorių. Tada su tais duomenų rinkiniais buvo apmokyta 600 vieno sluoksnio perceptronų trejetų, skirtų klasifikavimui poroms. Testavimui ir Hastie-Tibshirani apjungimo taisyklės konstravimui buvo sugeneruotos dvi naujos duomenų aibės atitinkamai po atitinkamai 2000 ir 50 duomenų vektorių kiekvienai iš trijų duomenų klasių. Tada apmokyti 600 klasifikatorių trejetų buvo apjungiami H-T taisykle, sukonstruota naudojant vienintelius tos taisyklės sukonstravimui skirtus duomenis, bei testuojami naudojant tuos pačius testinius duomenis (trys klasės po 2000 vektorių). Tam, kad būtų ištirta optimizavimo kriterijaus įtaka, lygiagrečiai kiekvieno klasifikavimo klaidos vertinimui buvo skaičiuojama ir vidutinė Kullback-Leibler (K-L) atstumo (3), naudojamo H-T apjungimo taisyklės konstravime, vidutinė reikšmė. Rezultatai yra pateikti paveiksle nr. 6.



Paveikslas 6. Klasifikavimo klaidos ir Kullback-Leibler atstumo vidurkis naudojant Hastie-Tibshirani apjungimą apmokyta su (a) mokymo duomenimis ir (b) testiniais duomenimis.

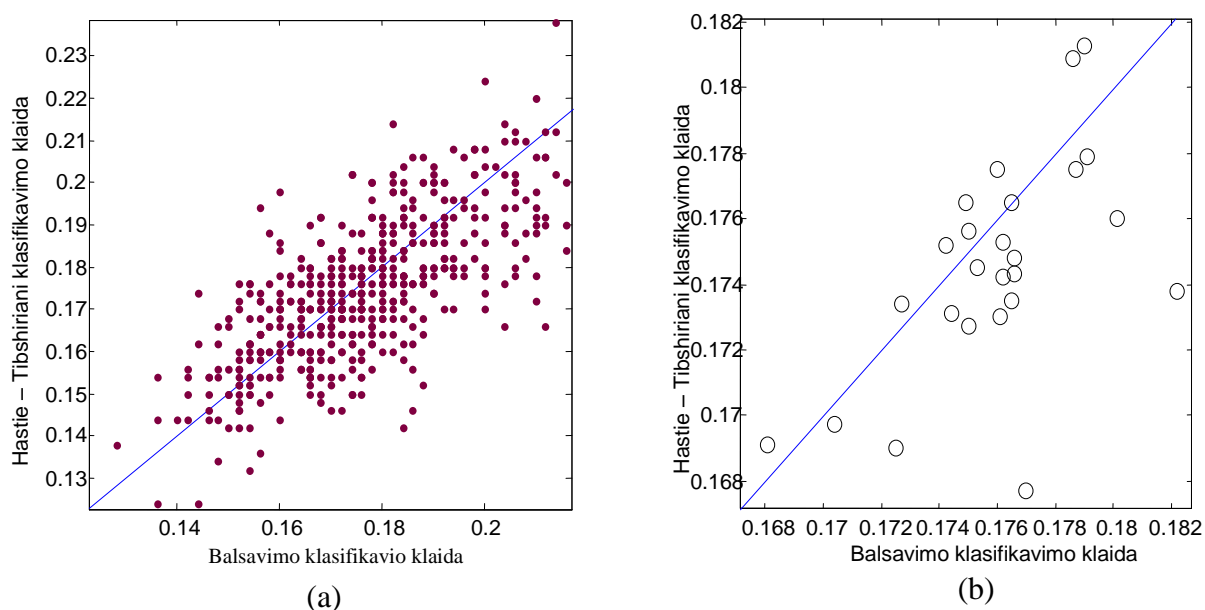
Paveiksle 6(a) matosi, kad klasifikavimo klaidos ir K-L atstumų vidurkiai labai silpnai koreliuoja (koreliacijos koeficientas $\rho = 0.263$). Maža koreliacija yra dėl priežasčių (a) ir (b), paminėtų šio skyrelio pradžioje. Kitame eksperimente H-T

parametrų nustatymui buvo naudojami testavimo duomenys. Paveiksle 6(b) matosi, kad šį kartą koreliacija yra kur kas didesnė (koreliacijos koeficientas $\rho = 0.757$). Iš šių dviejų eksperimentų galima daryti išvadą, kad mažas mokymo duomenų kiekis (šiuo atveju trys 30-mačių duomenų klasės po 50 vektorių) turi didesnę įtaką klasifikavimo klaidai nei skirtumai tarp klasifikavimo klaidos ir optimizavimo kriterijaus (šiuo atveju K-L atstumo).

2.8.2 Eksperimentų kiekio svarba

Tinkamam praktinio konkretaus klasifikatoriaus tinkamumo įvertinimui, reikalinga atlikti eksperimentus. Atliekant eksperimentus neretai yra kreipiamas didelis dėmesys į skirtingų duomenų rinkinių kiekį (pvz. [6]), pamirštant eksperimentų su tais duomenimis kiekio svarbą. Eksperimentų kiekio poreikio ne retai nepaisoma ir sprendžiant tam tikrus dalykinės srities uždavinius [24], [25]. Žinoma, kai duomenų yra pakankamai daug, kad juos padalinus į mokymo ir testavimo aibes, jos lieka pakankamai reprezentatyvios, pakanka atlikti ir vieną eksperimentą. Tačiau dažniausiai naudojamų duomenų aibės nėra labai didelės [22] o ir dėl nežinomo jų pasiskirstymo yra sunku įvertinti jų reprezentatyvumą.

Paveiksle nr. 7 parodomas nepakankamo eksperimentų kiekio efektas.



Paveikslas 7. Klasifikavimo klaidų išsibarstymas (a) atlikus 500 eksperimentų ir (b) paėmus vidurkį iš eilės einančių 20-ies eksperimentų vidurkį.

Paveiksle nr. 7 pavaizduotos dviejų etapų klasifikavimo metodų Hastie-Tibshirani ir balsavimo klasifikavimo klaidos klasifikuojant satelitų duomenis (žr. skyrių 2.6.1). Paveiksle 7(a) matosi, jog atlikus pakankamai didelį kiekį eksperimentų išaiškėja, kad du nagrinėjami metodai yra apytiksliai vienodi. Tačiau jei darytumėme išvadas iš kurių nors 20-ies iš eilės einančių eksperimentų, tai tokį atsakymą gautumėme tik vienu atveju (žr.

paveiksle 7(b) tašką, esantį ant įstrižainės). Visais kitais atvejais rezultatai gali būti atsitiktiniai tiek vieno, tiek kito metodo naudai.

2.8.3 Apjungimo metodų pataisymas

Tais atvejais, kai klasifikatorių poroms apjungimui yra naudojami tokie apmokomi metodai kaip Hastie-Tibshirani [5] ar Wu, Lin, Weng [6], jie gali būti linkę prisitaikyti prie mokymo duomenų kaip ir patys klasifikatoriai, jei tiek klasifikatorių poroms mokymui, tiek apjungimo metodų mokymui yra naudojami tie patys mokymo duomenys.

Klasifikatorių poroms prisirišimui prie mokymo duomenų sumažinti dažniausiai yra naudojamos analitinės priemonės darant prielaidą, kad klasifikatorius yra artimas Fišerio ar Euklido statistiniam klasifikatoriui. Daugiau apie šį metodą žr. disertacijos 5.4 skyriuje arba [3].

Disertacijoje yra siūlomas validavimo duomenų panaudojimas klasifikatorių poroms apjungimo taisyklių prisirišimo prie mokymo duomenų sumažinimui. Validavimo duomenys yra gaunami prie mokymo duomenų pridendant spalvotą triukšmą (žr. disertacijos skyrių 3.3.4). Lentelėje nr. 5 pateikiami rezultatai, kai kaip klasifikatoriai poroms yra naudojami tiek atraminiai vektoriai, tiek vieno sluoksnio perceptronai, o jų rezultatų apjungimui naudojamas Hastie-Tibshirani metodas. Rezultatai pateikiami generuotiems duomenims, naudotiems ir šios santraukos skyriuje 2.5.

Apjungimo būdas	VSP	AV
<i>Mokymo duomenys</i>	0.0446	0.0793
<i>Euklido klas. pataisymas</i>	0.0443	0.0787
<i>Fišerio klas. pataisymas</i>	0.0440	0.0783
<i>Naudoti validavimo duom.</i>	0.0439	0.0772
<i>Nauji validavimo duom.</i>	0.0439	0.0769
<i>Testavimo duomenys</i>	0.0424	0.0741

Lentelė 5. H-T apjungimo metodo klasifikavimo klaidos naudojant skirtingus pataisymus.

Lentelėje nr. 5 yra pateikti pataisymai su penkiais skirtingais H-T apjungimo metodo pataisymo būdais. Pirmojoje eilutėje H-T metodas yra mokomas su mokymo duomenimis, t.y. joks pataisymas neatliekamas. Antroje ir trečioje eilutėse H-T mokymo duomenų parametrai yra analitiškai pataisomi darant prielaidą, kad klasifikatorius yra

atitinkamai Euklido arba Fišerio statistinis klasifikatorius. Ketvirtoje ir penktoje eilutėse H-T apjungimas yra mokomas naudojant atitinkamai validavimo duomenis naudotus klasifikatorių poroms parametrų nustatymui, bei naujai sugeneruotus validavimo duomenis, skirtus tik apjungimo mokymui. Paskutinėje eilutėje pateikiami galimi idealūs rezultatai, kai H-T apjungimas buvo mokomas naudojant testinius duomenis.

Nors nominaliai pagerėjimas naudojant validavimo duomenis apjungimo taisyklei sudaryti ir yra mažas, tačiau santykinai žymus ir vertas dėmesio tolimesniems tyrimams.

Rezultatai ir išvados

3.1 Rekomendacijos daugelio klasių uždavinio sprendimui

Pirmiausia duomenys turi būti normalizuoti – visų vektoriaus atributų reikšmių režiai turi būti vienodi. Tada turi būti atlikta kita tinkama duomenų transformacija. Disertacijoje buvo naudojamos duomenų transformacijos pasitelkiant ypatingųjų reikšmių dekompoziciją (angl. *singular value decomposition*) ir panašumo požymius.

Jeigu yra žinomi klasių pasiskirstymai, tuomet turi būti naudojami specialūs – tokiems pasiskirstymams optimalūs – statistiniai klasifikatoriai. Priešingu atveju po duomenų transformacijos vieno sluoksnio perceptronas turėtų būti naudojamas kaip klasifikatorius poroms. Vieno sluoksnio perceptronas turi savybę mokymo metu sustoti toje vietoje, kurioje yra pasiekęs būseną, artimą optimaliam statistiniam klasifikatoriui.

Eksperimentai parodė, kad dauguma atvejų Hastie-Tibshirani siūloma klasifikatorių poroms apjungimo taisyklė yra geriausia. Tačiau nepaisant to, jei klasių statistinės charakteristikos tarpusavyje (tarp klasių porų) yra labai skirtingos, reikėtų naudoti klasifikatorių apjungimo metodą, naudojantį skaitines klasifikatoriaus reikšmes, pvz. kaip disertacijoje pristatomą *Pair-wise Fuzzy Templates*.

Patikimiems rezultatams mažų ir vidutinių imčių atveju gauti yra būtina atlikti kuo daugiau eksperimentų. Jei duomenų kiekis yra pakankamas reprezentatyvioms mokymo, testavimo ir valiavimo duomenų aibėms suformuoti, eksperimentų kiekis neturi būti labai didelis.

3.2 Pagrindinės išvados

Standartinė daugelio klasių vieno sluoksnio perceptrono nuostolių funkcija tiesiogiai nemažina klasifikavimo klaidos. Todėl toks algoritmas neleidžia gauti optimalių klasifikatorių net tada, kai klasių apriorinės tikimybės yra vienodos, o mokymo duomenys yra pakankamai subalansuoti. Vieninteliai klasifikatoriai, kurie tikslingai skirti gauti sprendimo priėmimo taisyklės, kurios minimizuoja klasifikavimo klaidą (su sąlyga, kad yra pakankamas mokymo duomenų kiekis), yra klasifikatoriai pagrįsti statistine sprendimo priėmimo teorija. Klasifikavimo klaidai minimizuoti disertacijoje buvo išanalizuotos dviejų etapų sprendimo priėmimo taisyklių optimizavimo savybės. Tokių taisyklių pirmajame etape yra gaunami optimalūs dviejų klasių klasifikatoriai, o antrajame etape tokių klasifikatorių poroms rezultatai yra

apjungiami naudojant tinkamą apjungimo taisyklę. Taigi, pagrindiniai disertacijos rezultatai yra šie:

1. Buvo teoriškai parodyta, kodėl dviejų etapų sprendimo priėmimo taisyklės gali duoti geresnius rezultatus nei vieno etapo sudėtingos taisyklės. Buvo išaiškinta, jog tai yra dėl šių pagrindinių priežasčių: (a) neuroninių tinklų atveju yra atsisakoma tradicinės K-klasių nuostolių funkcijos, (b) galima gauti optimalius ar beveik optimalius tiesinius klasifikatorius tam tikru būdu naudojant vieno sluoksnio perceptronus bei atraminius vektorius ir (c) tokio tipo klasifikatoriai gali išsaugoti svarbią klases skiriančią informaciją, kurią išsaugo tinkamai apjungiami pirmojo etapo klasifikatoriai. Iš tiesų ši sprendimo priėmimo schema perkelia dalį daugelio klasių klasifikatoriams kylančių sudėtingų problemų (pvz.: nesubalansuotos klasės, neoptimali nuostolių funkcija) į geriau išnagrinėtus ir geriau veikiančius dviejų klasių klasifikatorius.
2. Buvo parodyta, kad sėkmingai sustabdyti dvi klases klasifikuojantys vieno sluoksnio perceptronai yra tinkamas metodas pirmame sprendimo priėmimo etape. Tai yra dėl žinomos VSP savybės mokymo metu įgyti septynių skirtingų statistinių klasifikatorių elgseną ir sustoti (nutraukti mokymąsi) būtent tuo momentu, kada yra pasiekta optimalaus statistinio klasifikatoriaus elgsena. Jei tinkamai naudojama tinkama apjungimo taisyklė, tuomet uždaviniuose su vidutinio dimensiškumo duomenimis VSP yra palyginami ir dažnai net veikia geriau nei bene populiariausi atraminiai vektoriai.
3. Disertacijoje pristatytas klasifikatorių poroms apjungimo metodas *Fuzzy Templates* yra tinkamas apjungimo metodas, kai klasių porų tarpusavio statistinės charakteristikos yra pakankamai skirtingos (skiriasi klasių kovariacinės matricos, tarp klasių yra skirtingi atstumai ir pan.). Eksperimentai parodė, kad naudojant šį apjungimo metodą yra geriau naudoti atraminius vektorius nei vieno sluoksnio perceptronus, kai yra naudojami standartiniai parametrai. Tačiau į perceptrono aktyvavimo funkciją įvedus tam tikrą mastelio keitimo parametras α , VSP naudojimas tampa lygiavertis atraminių vektorių naudojimui.

3.3 Kiti rezultatai

1. Buvo pademonstruota, jog dvi klasifikatorių poroms apjungimo strategijos – Kullback-Leibler atstumu pagrįstas Hastie-Tibshirani pristatytas metodas ir Wu, Lin, Wen pristatytas bei jų pačių pristatomu kriterijumi pagrįstas metodas veikia tokiu pat tikslumu, kai mokymo duomenų imtys yra nedidelės. Tai paaiškinama teoriniu faktu, kad perdėtas netikslaus kriterijaus minizavimas gali būti žalingas (dirbtinio neuroninio tinklo permokymas ir perteklinis požymiu parinkimas [26]). Buvo atlikta skaitinė lygiagrečių netikslumo šaltinių – klasifikatoriaus veikimo supaprastinto matavimo ir duomenų kiekio tinkamumo – analizė. Buvo gautas rezultatas, kad duomenų dydžio įtaka buvo pagrindinis klaidų šaltinis konstruojant dviejų klasių klasifikatorių rezultatų apjungimo taisyklę.
2. Pakartotinis duomenų panaudojimas apjungimo taisyklei konstruoti padaro algoritmą labai prisirišusį prie konkrečių duomenų ir tokiu būdu pablogina jo gražinamą rezultatų tikslumą. Pseudo-validavimo duomenimis pagrįstas empirinis

prisirišimo algoritmas pasirodė esąs efektyvesnė kovos su duomenų prisirišimu priemonė nei teoriniai daugiamačiai metodai, pagrįsti Gauso pasiskirstymo modeliu. Šiame kontekste dar kartą buvo parodyta, kad spalvoto triukšmo pridėjimas yra puiki priemonė dirbant su atpažinimo modelių parinkimu baigtinėms duomenų aibėms.

3. Eksperimentinis tyrimas taip pat parodė, jog kryptinis neciklinis grafas yra tinkamas klasifikatorių poroms apjungimo metodas ne tik su atraminiais vektoriais kaip klasifikatoriais poroms, bet ir su vieno sluoksnio perceptronais.

Kai kurie eksperimentai parodė, kad skirtumai tarp vieno etapo ir disertacijoje analizuojamo dviejų etapų klasifikavimo algoritmų rezultatų yra nežymūs ir statistikai nereikšmingi. Tačiau nepaisant to, dviejų etapų klasifikavimo algoritmai yra perspektyvesni dėl jų galimybės naudoti skirtingus požymius ir klasifikavimo metodus optimalius kiekvienai klasių porai. Be to, kai yra pateikiami nauji mokymo duomenys, nėra reikalo su kiekvienu duomenų vektorium iš naujo permokyti visus klasifikatorius poroms – pakanka permokyti tik $K-1$ klasifikatorių kurių vienoje iš porų yra klasė reprezentuojant naują duomenų vektorių.

Šioje disertacijoje, pasitelkiant paprastus klasifikatorius poroms (vieno sluoksnio perceptroną bei tiesinį atraminių vektorių algoritmą), buvo parodytas dviejų etapų klasifikatorių pranašumas prieš vieno etapo klasifikatorius. Tokiu būdu buvo atvertos tolesnės tyrimo galimybės pasitelkiant sudėtingesnius klasifikatorius poroms, galbūt sugebančius atrinkti optimalius požymius klasių porai ar sugebančius dar geriau parinkti optimalų konkrečios klasių poros klasifikavimo algoritmą.

Publikacijos disertacijos tema

Publikacijos, esančios periodiniuose recenzuojamuose tarptautiniuose leidiniuose:

1. Sarunas Raudys, Rimantas Kybartas, Edmundas Kazimieras Zavadskas. Multicategory Nets of Single-Layer Perceptrons: Complexity and Sample-Size Issues. IEEE Transactions on Neural Networks, Vol. 21, No. 5, p. 784 – 795, 2010
2. Rimantas Kybartas, Nurdan Akhan Baykan, Nihat Yılmaz, Sarunas Raudys. Multiclass Mineral Recognition Using Similarity Features and Ensembles of Pair-wise Classifiers. 23rd IEA-AIE Conference, 2010, Springer-Verlag, Lecture Notes in Artificial Intelligence

Publikacija kitame periodiniame leidinyje:

1. Š.Raudys, R.Kybartas. Daugelio klasių klasifikavimas naudojant vienasluoksnius perceptronus. Informacinės technologijos 2007 : konferencijos pranešimų medžiaga 2007, p. 427-430

Trumpos žinios apie autorių

Rimantas Kybartas gimė 1979 m. rugsėjo 26 d. Darbėnuose, Kretingos raj. 1997 m. baigė Palangos 3-iają (dabar „Baltijos“) vidurinę mokyklą. 1997-2003 m. studijavo Vilniaus universitete kur baigė informatikos bakalauro bei magistro studijas. 2005-2010 m. vykdė doktorantūros studijas Vilniaus universitete.

2000-2002 metais dirbo privačioje IT įmonėje programuotoju bei projektų vadovu. Nuo 2002 metų dirba Lietuvos banke, šiuo metu – vyriausiuoju programuotoju.

Nuo 2003 m. Rimantas Kybartas užsiima pedagogine veikla Viliaus universitete. Per tą laiką vadovavo kursiniams, bakalauriniams, magistriniams darbams bei vedė programų sistemų inžinerijos, informatikos, programavimo C bei objektinio programavimo C++ pratybas.

Summary

In the real world both people and industry face multi-class classification tasks, e.g. assigning characters to letters and the corresponding sounds while reading or classifying minerals classification according to their appearance. In order to overcome the shortcomings of manual classification and to make industrial multi-class classification tasks more automated, statistical methods of classification were employed. Many complex both statistical and empirical methods for solving such tasks were developed.

The multi-class classification task was defined after the introductions in Chapter 1. The overview of multi-class classifiers and their fusion methods was presented as well. It was shown that there may be many solutions for each multi-class classification task from well known classical neural networks to a complex net of multi-class classifiers. There is a lack of information or any kind of recommendations when to use one or other method. Thus there is need for such recommendations or some approach for an abstract and universal method.

Because of the cost function, the net of K single layer perceptrons sometimes fails to provide the optimal decision boundary. It was shown that two-stage multi-class classifiers based on pair-wise classifiers are a good alternative in such situations. Two voting strategies (Voting and Directed Acyclic Graph) and two probability estimating based (Hastie-Tibshirani and Wu, Lin, Weng) methods were chosen for fusing single layer perceptron and linear support vector machine as pair-wise classifiers. Due to the lack of pair-wise classifier fusion methods based on output similarities, a new method *Pair-wise Fuzzy Templates* method was presented showing that it is a good alternative to other methods when the statistical features of data classes differ significantly.

The two serious issues – small sample size and class imbalance – which are usually faced by the multi-class classifier designers were addressed in Chapter 4. Analytical experiments showed that single layer perceptron performs with optimal generalization error rate compared to standard statistical classifiers and support vector machine classifier. Therefore, SLP may absorb the addressed issues when used in a two stage classification strategy.

The experimental results presented in Chapter 5 confirmed the assumptions and analytical results of the proposed two-stage classification based on pair-wise classifiers. Since the results in some cases are not considerably better and in a few cases even a little bit worse, the main target was achieved – it was shown that the proposed two-stage strategy for multi-class classification task solving is not worse than the known standard multi-class classifiers which provides opportunities for further enhancement.

The analysis of methodical issues showed that different methods relying on different optimization criteria should be estimated carefully. It appeared that the two analyzed classifier fusion methods, H-T and WLW, are approximately the same for moderate and small sample size data in the context of classification error rate.

Usually it is difficult to analytically compare two different multi-class classifiers due to the nonlinearities of their decision functions or other reasons. Instead, experiment-based comparison is usually performed. It was also shown that the number of experiments sometimes may be critical for inexact conclusions of classifier comparison.

The fusion methods, as well as classifiers, sometimes tend to bias to training data. The analysis of classifiers' fusion correction due to their bias showed, that the method based on introduction of new pseudo-validation data is more promising than the well known analytical methods based on statistical assumptions.

The main results of the thesis were applied to complicated multi-class mineral data in Chapter 6. The results confirmed that two-stage classification strategy based on pair-wise classifiers provides not worse results than other multi-class classifiers, if properly used and leaves opportunities for further enhancement. Similarity features were employed in order to transform the data space to a more separable one. It appeared that even dimensionality-sensitive classifiers such as SLP may get considerable gain in employment of similarity features if properly fused, which considerably increases dimension when the data sample size is much larger than dimensionality. The obtained experimental results of mineral classification showed that the proposed strategy performs in generalization error which is enough in order to apply it in industry.

Defended theses of the dissertation are:

1. Two-stage pair-wise based classifiers are a good alternative to complex multi-class classification algorithms.
2. Single layer perceptrons should be used as pair-wise classifiers instead of linear support vector machines.
3. Pair-wise Fuzzy Templates method may outperform other multi-class classification algorithms when statistical parameters of pairs of classes differ considerably, while it is not recommended in symmetric deployment of classes. The introduction of a scaling parameter may improve the performance of this method in both cases.

Main conclusions of the dissertation are:

1. It was theoretically shown that the two-stage neural network based decision making procedures may outperform the single-stage ones due to the following reasons: (a) refusing the traditional K-class cost function, (b) allowing to obtain near to optimal pair-wise linear classifiers by specially organized SLP or SV training, and (c) ability to save useful discriminative information contained in the first stage classifiers by prudent fusion of the pair-wise decisions.

2. It was shown that successfully stopped pair-wise SLP based classifiers are a useful option in the first stage of decision-making due to their feature to evolve through seven different statistical classifiers and to stop at the moment when optimal statistical classifier is reached. If a prudently trained fusion rule is used, the SLPs are comparable and often outperform the linear SV classifiers in moderate dimensional situations.
3. The presented pair-wise Fuzzy Templates method is a proper classifier fusion method when statistical characteristics of class pairs are rather different.

Other results:

1. It was demonstrated that two fusion strategies, the K-L distance based Hastie-Tibshirani and the WLW method result in approximately the same performance in situations where the training sample sizes are small.
2. The repeated employment of training data to design fusion rules leads to optimistic bias and deterioration of the two-stage decision-making system. The empirical, pseudo-validation set based bias diminishing technique appeared to be more effective than the theoretical multi-dimensional Gaussian distribution model based methods (e.g. corrections for Fisher classifier).
3. Experimental results showed that directed acyclic graph is proper fusion method not only with SVM, but with SLP as well.

Literatūra

- [1] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995.
- [2] S. Haykin, *Neural Networks: A comprehensive foundation*. 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [3] S. Raudys, *Statistical and Neural Classifiers: An integrated approach to design*. Springer-Verlag, NY, 2001.
- [4] Hsu, C. W., Lin C. J., A comparison on methods for multi-class support vector machines. *IEEE Trans. on Neural Networks*, 13: 415-425, 2002.
- [5] T. Hastie and R. Tibshirani, Classification by pair-wise coupling, *The Annals of Statistics*, 26(1):451-471, 1998.
- [6] T.-F. Wu, C.-J. Lin, R. C. Weng, Probability estimates for multi-class classification by pair-wise coupling, *Journal of Machine Learning Research*, 5: 975–1005, 2004.
- [7] M. Gonen, A.G. Tanugur, and E. Alpaydm, Multiclass posterior probability support vector machines. *IEEE Tr. Neural Networks*, 19: 130–139, 2008.
- [8] J. Cid-Sueiro, J.I. Arribas, S. Urban-Munoz, and A. R. Figueiras-Vidal, Cost functions to estimate *a posteriori* probabilities in multi-class problems. *IEEE Trans. on Neural Networks*, vol. 10 (3): 645- 656, 1999.

- [9] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, “Large margin DAG’s for multi-class classification”, in: *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, vol. 12 pp. 547-553, 2000.
- [10] E. Gelenbe and K.F. Hussain, Learning in the multiple class random neural network, *IEEE Trans. Neural Networks* 13 (6): 1257–1267, 2002.
- [11] L. Kuncheva, J. C. Bezdek, M. A. Sutton, *On combining multiple classifiers by fuzzy templates*, 1998.
- [12] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*. 2nd ed. Wiley, NY, 2000.
- [13] Moshe Leshno, Vladimir Lin, Alan Pinkus and Shirnon Schocken, Multilayer Feedforward Networks With Non-Polynomial Activation Functions Can Approximate Any Continuous Function. *Journal of Neural Networks*, 6(3), 1993.
- [14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd edition. Academic Press, NY, 1990.
- [15] B. Boser, I. Guyon and V. Vapnik, A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144-152. ACM Press, 1992.
- [16] Y. S. Huang and C. Y. Suen. A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Trans. PAMI*, 17(1):90 – 94, 1995.
- [17] S. Raudys, Evolution and generalization of a single neurone. I. SLP as seven statistical classifiers, *Neural Networks*, 11: 283–96, 1998.
- [18] Z.-H. Zhou and X.-Y. Liu, Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. on Knowledge and Data Engineering*, 18(1): 63-77, 2006.
- [19] N. Japkowicz, “Learning from imbalanced data sets: a comparison of various strategies,” in *Working Notes of the AAAI’00 Workshop on Learning from Imbalanced Data Sets*, Austin, TX, pp.10–15, 2000.
- [20] Friedman, J. H., 1989. Regularized discriminant analysis. *Journal of the American Statistical Association* 84 (405), 165-175.
- [21] S. Raudys and D. Young, Results in statistical discriminant analysis: A review of the former Soviet Union literature, *Journal of Multivariate Analysis*, Vol. pp. 89, 1-35, 2004.
- [22] A. Asuncion, and D.J. Newman, *UCI Machine Learning Repository* Irvine, CA: [<http://www.ics.uci.edu/~mlern/MLRepository.html>], 2007.

- [23] Marschallinger, R., 1997. Automatic mineral classification in the macroscopic scale. *Computers & Geosciences* 23, 119–126.
- [24] Fueten, F., Mason, J., 2007. An artificial neural net assisted approach to editing edges in patrographic images collected with rotating polarizer stage. *Computers & Geosciences* 33 (9), 1176-1188.
- [25] Thompson, S., Fueten, F., Bockus, D., 2001. Mineral identification using artificial neural networks and the rotating polarizer stage. *Computers & Geosciences* 27 (9), 1081-1089.
- [26] S. Raudys, Feature over-selection. *Lect. Notes in Computer Science*, Springer. Vol. 4109, pp. 622-631, 2006.