

Denborazko serieen sailkapen goiztiarra helburu anitzeko optimizazio problema gisa aztertua

(Early time series classification analyzed as a multi-objective optimization problem)

Irati Arrieta, Usue Mori^{1*}, Alexander Mendiburu², Jose A. Lozano^{3,4}

¹ Matematika Aplikatua eta Estatistika eta Ikerkuntza Operatiboa Saila, UPV/EHU

² Konputagailuen Arkitektura eta Teknologia Saila, UPV/EHU

³ Konputazio Zientzia eta Adimen Artifiziala Saila, UPV/EHU

⁴ BCAM, Basque Center for Applied Mathematics

LABURPENA: Denborazko serieen datu meatzaritza arloko problema ohikoenetako bat da, denborazko serieen gainbegiraturatuko sailkapena. Problema honen helburua da, klaseetan banatuta dauden serie multzo batetik abiatuz, sailkatu gabeko beste serie batzuen klasea auresango duen eredu ahalik eta zehatzena eraikitzea. Problema klasiko honen hedapen gisa, kasu batzuetan, denborazko serieak denboran zehar jasotzen dira, eta ohikoa da iragarpenak ahalik eta lasterren egin nahi izatea, datuak jasotzeak dakartzan kostuak aurrezteko asmoarekin edo klaseak berandu iragartzeak ekarri ditzakeen ondorio kaltegarriak ekiditeko. Egoera honetan, denborazko serieen sailkapen goiztiarra izeneko problema agertzen da, zeinaren helburua den ahalik eta sailkatzaile zehatzena eta aldi berean iragarpen azkarrenak egingo dituen eraikitzea. Nahiko intuitiboa da pentsatzea, seriearen puntu gehiago ditugunean eskuragai, hari buruzko informazio gehiago dugula eta beraz, haren klaseari buruzko iragarpen zehatzagoak burutzea errazagoa dela. Alderantziz, seriearen klasea goiz auresan nahi badugu, informazio gutxiago izango dugu eta beraz, zailagoa izango da klasea ondo auresatea. Beraz, zehatzasuna eta azkartasuna bi helburu kontrajarriak dira. Lan honetan, denborazko serieen sailkapen goiztiarri soluzioa emango dion metodo berritzaile bat proposatzen dugu, helburu anitzeko problema gisa planteaturik. Literaturan aurretik proposatutako helburu bakarrek optimizazio problema gisa eraikitako beste eredu batekin konparatu dugu gure eredu eta ikusi dugu gureak emaitza hobek ematen dituela zenbait oinarritzko datu-basetan.

HITZ GAKOAK: Denborazko serieak, Sailkapen goiztiarra, Helburu anitzeko optimizazioa.

ABSTRACT: *One of the most prominent problems in the area of time series data mining is called supervised time series classification. The goal of this problem is to build a model that predicts the classes of new unclassified series as accurately as possible, departing from a database of time series for which the class is known. As an extension of this problem, on some occasions, the data is collected over time, and, in order to avoid costs that incur in collecting new data or negative consequences that may arise when making late predictions, the goal is to make the class predictions as early as possible. In this context, the problem denominated early classification of time series arises, whose objective is to build a classifier that is as accurate as possible, but at the same time, makes the class prediction as early as possible. It is logical to think that the more data points are made available, the more information we have about the time series and, so, it is easier to make accurate class predictions. On the contrary, if we want to make early class predictions, we will have less information and it will be more difficult to make accurate class predictions. Therefore, accuracy and earliness are two objectives which are in conflict. In this work, we propose an innovate method for early classification based on multi-objective formulation of the problem. We have compared it to a model proposed in the literature which models the problem as a single objective optimization problem and we have seen that our model provides better results on some benchmark datasets.*

KEYWORDS: *Time series, Early classification, Multi-objective optimization.*

* **Harremanetan jartzeko / Corresponding author:** Usue Mori, Konputazio Zientzia eta Adimen Artifiziala, Informatika Fakultatea, 20018, Donostia, Gipuzkoa, Euskal Herria. – usue.mori@ehu.eus – https://orcid.org/0000-0002-2057-1770.

Nola aipatu / How to cite: Arrieta, Irati; Mori, Usue; Mendiburu, Alexander; Lozano, Jose A. (2019). «Denborazko serieen sailkapen goiztiarra helburu anitzeko optimizazio problema gisa aztertua»; *Ekaia*, 36, 2019, 291-310. (https://doi.org/10.1387/ekaia.19696).

Jasoa: 14 maiatza, 2018; Onartua: 25 martxo, 2019.

ISSN 0214-9001 - eISSN 2444-3255 / © 2019 UPV/EHU



Obra hau Creative Commons Atribución 4.0 Internacional-en lizentziapean dago

1. SARRERA

Azken urteetan, datuak jaso eta gordetzeko baliabide eta ahalmenen hobekuntza dela eta, datu-base oso handiak eskura daitezke hainbat aplikazio-eremutan. Egoera honetan, datu multzo handietatik abiatuz informazio erabilgarria ateratzeko beharra eta interesa areagotu da eta honekin batera *datu-meatzaritz*a arloaren garrantzia ezinbestekoa bilakatu da [1].

Biltzen eta gordetzen diren datuetatik askok izaera tenporala dute, hau da, denboraren araberrako ordena bat dute eta, beraz, azken urteotan denborazko seriez osatutako datu-baseei buruz hitz egiten hasi gara. Datu-base hauek ezaugarri bereziak dituzte, hala nola, aldagaien denboraren araberrako korrelazioa eta baita neurketetan agertu ohi den zarata ere [2]. Hau hala, azken urteotan zientzialarien komunitatea datu-meatzaritzako ataza ohikoak datu-base mota hauetara hedatzen saiatu da, hainbat soluzio berri proposatuz [3].

Datu-meatzaritzako ataza guztietatik, sailkapen gainbegiratu da ohi-koenetako bat eta denborazko serieen datu-baseetan ere arreta handia eman zaio problema honi [4]. Zehazki, denborazko serieen sailkapeneko problema batean, serie bakoitza klase edo multzo baten partaide da. Honela, klase ezaguna duten serieen datu-base batetik abiatuz, eredu bat eraikitzea da helburua, ondoren, klase ezezagunak dituzten serie berrien klasea aurreritateko erabil dezakeguna. Xedea, beraz, eredu ahalik eta *zehatzena* eraikitzea da, klase-iragarpenak ahalik eta ondoen egingo dituen.

Batzuetan, serieen neurketa guztiak ez dauzkagu hasieratik eskuragarri, hau da, sailkatu nahi ditugun serieak denboran zehar jasotzen ditugu, eta denbora istant bakoitzean neurketa bat(zuk) gehiago eskuratzen dira. Egoera honetan, klase iragarpenak ahalik eta azkarren egitea interesgarria izan daiteke, datuak biltzeak dakartzan kostuengatik edo itxoiteak ekar ditzakeen beste ondorio kaltegarriengatik. Kasu honetan, helburua da, eredu zehatzena eta aldi berean iragarpen azkarrenak egingo dituen eraikitzea izango da. Kontrajarriak diren bi helburu izango ditugu, beraz: *zehatztasuna* eta *azkartasuna*, logikoa baita pentsatzea zenbat eta seriearen puntu gehiago eduki eskuragai orduan eta errazagoa izango dela sailkapen onak egitea, eta alderantziz. Problema honi *denborazko serieen gainbegiratu-tako sailkapen goiztiarra* deitzen zaio eta sailkapen gainbegiratu klasikoaren hedapen bat da [5].

Denborazko serieen sailkapen goiztiarra orain arte *helburu-funtzio* bakarreko optimizazio problema gisa aztertu da [6], hau da, zehatztasun balio batzuk aurrez finkatuz, iragarpenak ahalik eta azkarren egitea izan da helburu. Nahiz eta emaitza onak lortu diren, askotan zaila izaten da zehatztasun maila aurrez finkatzea, eta gainera, erabiltzaile ezberdinak zehatztasun behar ezberdinak izan ditzakete. Hau honela, lan honen helburua bi hel-

buru-funtzioak aldi berean optimizatzea da, helburu anitzeko optimizazio teknikak erabiliz. Horretarako, helburu-funtzio bakarreko eredu bat hartu dugu oinarritzat, zehazki [7] eta, honetatik abiatuz, bi helburuko problema batera hedatzeko egin beharreko aldaketak egin ditugu.

Lan hau honako atal hauetan banatzen da: 2. kapituluaren denborazko serieen sailkapen goiztiarra era formalean definituko dugu eta beste zenbait oinarritzko kontzepturi sarrera emango diogu, eta, ondoren, (lan honetan) oinarritzat hartu dugun helburu bakarreko metodoa azalduko dugu; 3. atalean proposatutako eredu eraikitzeako pausoak azalduko dira banan-banan eta, behin eraikita, ereduaren erabilera nola egin zehaztuko dugu; 4. atalean eredu ebaluatzeako egindako esperimentuak eta lortutako emaitzak azalduko ditugu eta, azkenik, 5. atalean, ateratako ondorio nagusiak laburtuko dira, etorkizunerako lan-ildoak ere zehaztuz.

2. DENBORAZKO SERIEEN SAILKAPEN GOIZTIARRA

Esan bezala, atal honetan oinarritzko definizio batzuekin eta *denborazko serieen sailkapen goiztiarra* problemaren definizioarekin hasiko gara, eta, ondoren, azalduko dugu guk proposatutako metodoak oinarritzat hartzen den helburu bakarreko metodologia.

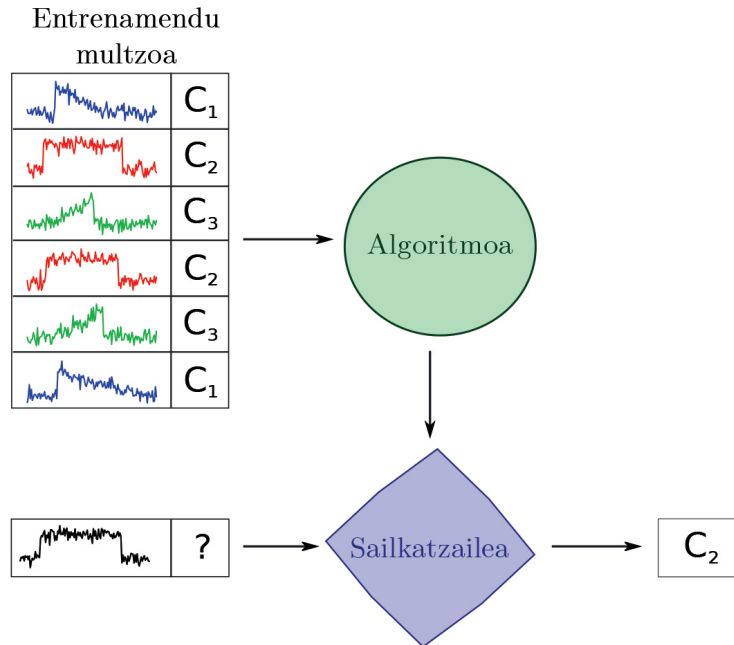
2.1. Problemaren definizioa eta oinarritzko kontzeptu batzuk

Denborazko serie bat L luzera finituko ordenatutako tupla sekuentzia bat da: $TS = \{(t_1, x_1), \dots, (t_L, x_L)\}$. x_i balioak zenbaki errealak izaten dira normalean (denborazko seriearen balioak) eta t_i aldagaiak balio positibo eta gorrakorra dira (denborak). Gainera, $TS|_l = \{(t_1, x_1), \dots, (t_l, x_l)\}$ sekuentziari TS denborazko seriearen l istanterainoko trunkaketa deritzogu, non, $l < L$.

Azken urteetan, teknologiaren aurrerapenei esker, aplikazio eremu askotan sortzen diren datuak, denboran zehar jasotzen dira. Hau dela eta, denborazko serieen datu-baseak aztertu eta hauetatik informazio erabilgarria ateratzeko beharra nagusitu da. Honelako datu-baseekin burutu ohi diren ataza ohikoen artean, *gainbegiratutako sailkapena* izeneko problema dugu, artikulu honen oinarri izango den problema [3].

Problema honetan $\mathbf{TS} = \{(TS_1, CL_1), \dots, (TS_n, CL_n)\}$ denborazko seriez osatutako entrenamendu multzo edo entrenamendu datu-basetik abiatuko gara non, CL_i, TS_i serieari dagokion klasea den. Egoera honetan, denborazko serieen sailkapen gainbegiratuaren helburua, TS_i serieak eta CL_i haien klaseen arteko erlazioa jasoko duen eredu bat eraikitzea da, sailkatu gabeko serie berrien klaseak ahalik eta zehatzen iragarriko dituen (ikus 1. irudia). Hau da, ezaguna den informazio batetik abiatuz, aurrerago eze-

zaguna izango den informazio bat aurrerateko gai izango den eredu bat doitu nahi dugu [1].



1. irudia. Denborazko serieen sailkapen gainbegiratua.

Aplikazio eremu askotan duen erabilgarritasuna dela eta, problema hau ebazteko hainbat soluzio ezberdin proposatu dira literaturan, azken urteetan denborazko serieen meatzaritzan problema klasikoenetako bat bihurtu delarik [3]. Hau hala, problema honetatik abiatuz, beste zenbait problema ere aurkeztu dira, konplexuagoak eta espezifikoagoak. Adibide gisa, lan honetan landuko dugun *denborazko serieen sailkapen goiztiarra* dugu.

Denborazko serieen gainbegiraturako sailkapen problematan bezala, denborazko serieen sailkapen goiztiarrean ere, $\mathbf{TS} = \{(TS_1, CL_1), \dots, (TS_n, CL_n)\}$ entrenamendu multzotik abiatuko gara. Aurreko kasuan bezala, denborazko serieen sailkapen goiztiarraren helburua da, TS_i serieak eta CL_i haien klaseen arteko erlazioa jasoko duen eredu bat eraikitzea. Halaber, ezberdintasuna zera da: sailkatu nahiko ditugun etiketatu gabeko serie berriak, denboran zehar jasotzen dira, hau da, denbora igaro ahala, serieko puntu berriak eskuragai jarriko dira, horrela, serie osoa jaso arte. Egoera honetan, klasea ondo aurrerateaz gain, iragarpenak ahalik eta azkarren egingo dituen eredu bat bilatzea izango da helburua, horretarako entrenamendu datu-basea erabiliz.

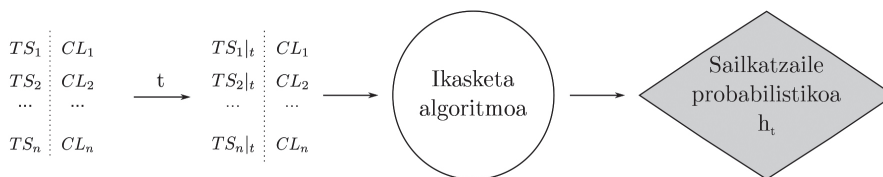
Problema hau datuak jasotzeak kostu handia duenean edo klasea berandu asmatzeak ondorio kaltegarriak dituenean agertzen zaigu. Esate baterako, ospitale batean egindako lanean [8] ikusi zuten sepsis gaixotasunaz diagnostikatutako jaio berrien bihotz taupadek emandako denborazko serieek ez-ohiko formak hartzen zituztela 24 orduko taupada azterketa batean. Bihotz-taupaden denborazko serieak monitorizatuz eta serie horiek ahalik eta azkarren sailkatuz, diagnostia jakin daiteke 24 orduak igaro baino lehen, eta horren arabera tratamendu egokiarekin has daiteke lehenbailehen. Era berean, beste hainbat aplikazio ditu problemak [5, 6, 7].

Ohartzekoa da, sailkapen gainbegiratuaren helburua zela eredu ahalik eta zehatzena lortzea. Orain, ordea, bi helburu kontrajarri dauzkagu: *zehaztasuna* eta *azkartasuna*, logikoa baita pentsatzea seriearen puntu gehiago eskuragarri dauzkagunean, informazio gehiago dugula, eta ondorioz iragarpen zehatzak egitea errazagoa dela. Helburua beraz, bi helburu horien arteko oreka lortzea izango da.

2.2. Helburu-funtzio bakarrean oinarritutako metodo bat

Eredu honen lehen osagaia *sailkatzaile probabilistiko multzo bat* izango da $\{h_t\}_{t=1}^L$, *Gauss-Prozesu*-etan [9] oinarritutako sailkatzaile multzoa, hain zuzen ere. Hauen helburua denbora istant ezberdinetan serie bat zein klasetako den iragartzea izango da, edo zehazkiago, klase bakoitzari probabilitate bat esleitzea. Beraz, $\{h_t\}$ sailkatzaileak t denboran eskuragai dagoen seriearen zatia hartuko du sarreran eta klase bakoitzari dagokion probabilitatea itzuliko du irteeran.

Sailkatzaile hauek $\mathbf{TS} = \{(TS_1, CL_1), \dots, (TS_n, CL_n)\}$ entrenamendu datu-base bat erabiliz eraikiko dira 2. irudian ikus daitekeen prozesuari jarraituz. Eraketa-prozesu honi buruzko eta ikasketa algoritmoari buruzko informazio zehatzagoa [7] artikuluan aurki daiteke.



2. irudia. h_t sailkatzaile probabilistikoen eraikuntza prozesua.

Puntu honetara iritsita, ohartu gaitzen, denbora istant bakoitzean, klase iragarpen bat eskaintzen digula metodoak beti, baina oraindik ez dugu inolako informaziorik iragarpen hauen fidagarritasunaren inguruan. Beraz, $\{h_t\}_{t=1}^k$ informazioa *gelditze-erregela* batekin konbinatuko dugu.

Erregela hauek askotariko forma eta itxurak izan ditzakete baina [7] lanean eraginkorren suertatu zen erregelak honako itxura hau zuen:

$$s_\gamma(p_1^t, p_2^t, \dots, p_k^t, t) = \begin{cases} 0, & \gamma_1 p_{1:k}^t + \gamma_2 (p_{1:k}^t - p_{2:k}^t) + \gamma_3 \frac{t}{L} + \gamma_4 < 0 \text{ bada} \\ 1, & \text{bestela} \end{cases} \quad (1)$$

non $(p_1^t, p_2^t, \dots, p_k^t)$ h_t -k t unean serie zehatz batentzat klase bakoitzerako jaulkitako probabilitateak diren. $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4) - 1$ eta 1 arteko balioak hartuko dituzten parametroak dira eta p_{1-k}^t eta p_{2-k}^t instant horretan (t) sailkatzaile probabilistikoak iragarritako lehenengo eta bigarren probabilitate handienak dira.

Honelako erregela batek 0 itzultzen badu, iragarpena egiteko informazio nahikorik ez dagoela onartzen da eta datu gehiago jaso arte itxaron beharko dugu. Aldiz, 1 ematen badu, geratu egingo gara eta klase iragarpena egingo da, iragarriko den klasea momentu horretan probabilitate handiena daukan klasea izanik.

Ohartu gaitezen, s_γ gelditze-erregelak γ parametro bektorearen menpekoak direla, hau da, $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ parametroei balio ezberdinak emanez, gelditze-erregela ezberdinak lortzen direla. Beraz, hauxe da galdera: gelditze-erregela guztietatik, zein(tzuk) dira egokienak? Galdera horri erantzuna bilatzeko, gelditze-erregela bat ebaluatzeko gai izan behar dugu, bere kalitatea balio numeriko baten bidez adieraziz. Hau hala, [6, 7] lanetan, bi kostu-funtzio definitu zituzten bi helburu hauek kuantifikatzeko. Hasteko, C_a funtzioari *zehaztasun kostua* deritzo, eta honela definituko dugu:

$$C_a(TS, \{h_1, h_2, \dots, h_L\}, s_\gamma) = \frac{1}{|TS|} \sum_{x \in TS} \mathbb{I}(C\hat{L}_x \neq CL_x) \cdot 100. \quad (2)$$

non, TS entrenatzeko serieen multzoa den, $\{h_1, h_2, \dots, h_L\}$ eraiki ditugun sailkatzaile probabilistikoen multzoa eta s_γ ebaluatu nahi dugun gelditze-erregela. CL_x eta $C\hat{L}_x$, x denborazko seriearen benetako klasea eta sailkatzaile goiztiarrak (sailkatzaile probabilistiko eta gelditze-erregelaren konbinazioak) auresandako klasea diren, hurrenez hurren. \mathbb{I} funtzioak 1 itzultzen du barruko baldintza egia bada, eta 0 bestela. Ohart(u) gaitezen, kostu hau zenbat eta txikiagoa izan, gelditze-erregela hobea (zehatzagoa) izango dela.

Aldiz, C_e *azkartasun kostua* deituko dugu, eta sailkatzaile probabilistiko multzoarekin eta gelditze-erregela horrekin egindako sailkapenak zeinen goiztiarrak diren neurtuko du:

$$C_e(TS, \{h_1, h_2, \dots, h_L\}, s_\gamma) = \frac{1}{|TS|} \sum_{x \in TS} \frac{t_x^*}{L_x} \cdot 100 \quad (3)$$

non, t_x^* , gelditze-erregelak x serieari buruzko klase iragarpena egiteko gai den lehenengo unea den (hau da, 1 itzultzen duen lehen unea), eta L_x x seriearen luzera den. Ohartu gaitzen, kostu hau zenbat eta txikiagoa izan, gelditze-erregela hobea izango dela, hau da, iragarpenak goiztiaragoak izango dira.

Ohartu, metodoaren helburua orain entrenamendu datu-basean sortutako C_a eta C_e kostuak ahalik eta txikienak dituen gelditze-erregela bilatzea dela. Ohartu, bi kostu-funtzio kontrajarri minimizatu nahi direla aldi berean. Hau da, zehaztasuna oso altua lortzeak, normalean klase iragarpenak beranduago egitea ekarriko duela, serieen informazio gehiago izan arte itxoin beharko dugulako. Hau ikusirik, [6] lanean, bi kostu-funtzio hauek bakar batean konbinatu zituzten:

$$CF(TS, \{h_1, h_2, \dots, h_L\}, s_\gamma) = \sum_{x \in X} (\alpha C_a(x, s_\gamma) + (1 - \alpha) C_e(x, s_\gamma)) \quad (4)$$

non, α bi kostuen arteko oreka moldatzeko erabiltzaileak aurrez definitu beharreko parametro bat den. Metodoaren azken pausoa, beraz, honako helburu bakarreko optimizazio-problema hau ebaztea da:

$$\min_{\gamma} CF(TS, \{h_1, h_2, \dots, h_L\}, s_\gamma) \quad (5)$$

Lan horretan, Algoritmo Genetiko bat erabiliz gauzatu zuten prozesu hau. Algoritmo hau aukeratzearen arrazoia da optimizatu nahi den funtzioaren itxura ez dela ezaguna eta propietate matematiko egokiak (adibidez, konbexutasuna) ez dituela betetzen. Ohartu, indibiduoak edo optimizazio problema honetako soluzio posibleak, 4 luzerako -1 eta 1 tarteko balio errealak hartzen dituzten γ bektoreak direla. Hala, algoritmo genetikoa bilatutako soluzio onenak, erabiliko den s_{γ^*} gelditze-erregela definituko du.

Problema ebazteko metodo honek emaitza onak erakutsi ditu [6], baina bi kostu-funtzioak helburu-funtzio bakar batean murrizteak zenbait arazo dakartza, tamalez. Hasteko, bi kostu-funtzioak eskala berdinean egon behar dute. Bestalde, α parametroa aurrez finkatu behar da, hau da, erabiltzaileak aurrez erabaki behar du zein garrantzi eman nahi dion bi kostuetako bakoitzari. Garrantzi-maila ezberdinekin probatu nahiko balu, α ezberdinekin optimizazio-algoritmoa berdefinitu beharko genuke eta problema berriro ebatzi, honek dakarren kostu konputazionalarekin, noski. Desabantaila hauei aurre egiteko, lan honetako hurrengo atalean proposatuko dugu problema bera ebaztea helburu anitzeko optimizazio teknikak erabiliz, horretarako ikuspuntu eta metodo berritzaile bat proposatuz.

3. DENBORAZKO SERIEEN SAILKAPENA HELBURU ANITZEKO EREDU BAT ERABILIZ

Atal honetan, denborazko serieen sailkapen goiztiarrerako proposatzen dugun ereduaren eraikuntza eta erabilera aurkeztuko ditugu. Ikusiko dugun moduan, eredu honen ezaugarri nagusia eta, ondorioz, lanaren ekarpen garrantzitsuen izango da lehenengo aldiz formulatzea eta ebaztea, helburu anitzeko problema bezala, sailkapen goiztiarraren problema.

3.1. Ereduaren eraikuntza

Ondorengo paragrafoetan, ereduaren eraikuntzan sakonduko dugu eta metodoaren osagai ezberdinak zeintzuk diren eta nola lortzen diren azalduko dugu banan-banan.

Hasteko, eredu honen lehen osagaiak zuzenean eratoritzen dira aurreko atalean azaldutako helburu bakarreko metodotik. Hala, bi osagarri nagusi ditu: sailkatzaile probabilistiko batzuk ($\{h_1, h_2, \dots, h_L\}$) eta gelditze-erregela sorta bat ($\{s_{\gamma_1}, s_{\gamma_2}, \dots, s_{\gamma_s}\}$).

Sailkatzaile probabilistikoak aurreko atalean azaldu moduan eraikiko dira, zuzenean oinarritzko helburu bakarreko metodoan [6] egin bezala. Gelditze-erregelaren itxura ere 1. ekuazioan erakutsitakoa izango da, hauek eman dituzten emaitza onak ikusirik aldaketarik ez egitea erabaki baitugu. Gainera, s_γ gelditze-erregelak, γ parametroen guztiz menpe daudela esan dugu jada aurreko atalean, eta definitu ditugu gelditze-erregela ezberdinak ebaluatzeko prozedura bat 2. eta 3. ekuazioetan oinarrituz.

Lan honetan proposatutako metodoak, azken pausoan erakutsiko du helburu bakarreko oinarritzko metodoarekiko ezberdintasuna. Azken pauso honetan, C_a eta C_e kostu-funtzio hauek minimizatuko dituen γ bektorea lortzea zen helburua, hau da, gelditze-erregela optimoa(k) lortzea. Helburu-bakarreko metodoan, bi kostu-funtzioak bakar batean konbinatzea proposatzen zen (ikus 4. eta 6. ekuazioak) eta aipatu dugu, honek zenbait desabantaila sortzen dituela.

Hau guztia ikusirik, guk, lan honetan, problema, helburu-funtzio bakarreko problema gisa modelatu beharrean, bi helburu-funtzioko problema bezala ebatziko dugu. Zehazki honako optimizazio problema hau ebatzi nahiko dugu:

$$\min_{\gamma} (C_e(TS, \{h_1, h_2, \dots, h_L\}, s_{\gamma}), C_a(TS, \{h_1, h_2, \dots, h_L\}, s_{\gamma})) \quad (6)$$

Optimizazio problema batean optimizatzeko helburu-funtzio bat baino gehiago dauzkagunean, soluzio optimoen bilaketari *helburu anitzeko optimizazioa* deritzo.

Ohartu gaitzen, helburu-funtzio anitzeko optimizazioan, soluzioak elkarren artean konparatzea ez dela batere erraza. Noiz da soluzio bat beste bat baino hobea? Soluzio bat beste bat baino hobea al da helburu batean hobea bada eta bestean ez?

Galdera hauei erantzuna emateko sortu zen Pareto irizpidea. Helburu anitzeko optimizazio-problema batean, x_1 eta x_2 bi soluzio posible izanik, x_1 soluzioak x_2 soluzioa *dominatzen* duela esaten da honako bi baldintza hauek betetzen baditu: helburu guztietarako, $f(x_1)$ ez da $f(x_2)$ baino okerragoa eta helburu baterako behintzat, $f(x_1) < f(x_2)$ baino hobea da. Gainera, demagun soluzio posibleen espazioan ez dela existitzen beste soluziorik x_1 soluzioa dominatzen duenik. Orduan, x_1 soluzio ez dominatua edo *Pareto-soluzioa* dela esaten da. Soluzio ez dominatu guztien multzoari edo Pareto-soluzio guztiek sortzen duten multzoari *soluzio ez-dominatuen frontea* deituko diogu eta $\mathcal{F}1$ idatziko dugu.

Beraz, soluzio ez dominatuen frontea, helburu anitzeko problema-
ren soluzio onenek osatzen dute, eta gure helburua izango da 6. ekuazioan adierazitako optimizazio problemaren soluzio ez dominatuen frontea bilatzea, hau da, zehaztasun eta azkartasun kostu ez-dominatuak dituzten gelditze-erregelen multzoa aurkitzea. Multzo hau, hasieran aipatu moduan, $\{s_{\gamma_1}, s_{\gamma_2}, \dots, s_{\gamma_s}\}$ bezala adieraziko dugu eta metodoaren bigarren osagaia izango da.

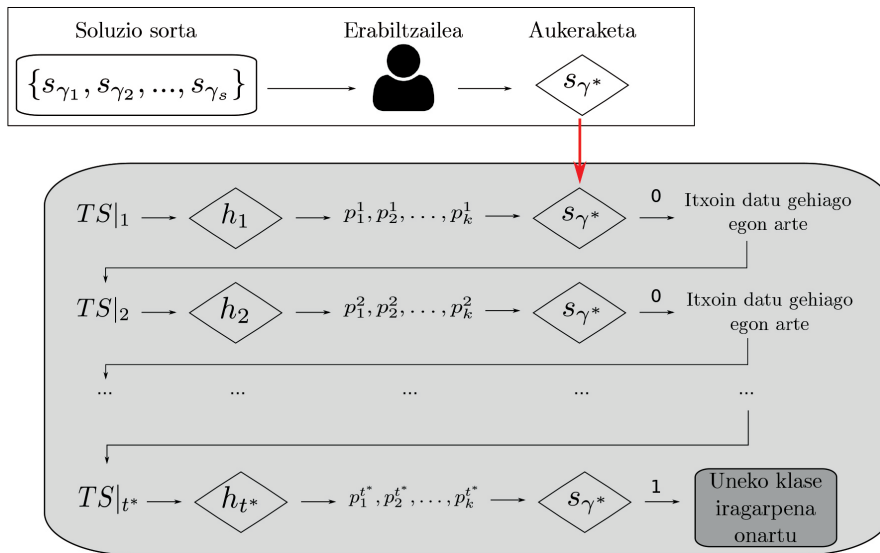
Honelako helburu anitzeko optimizazio-problema ebazteko eta soluzio ez dominatuen frontea bilatzeko hainbat metodo ezberdin existitzen dira, baina guk, aurreko kasuan bezala, algoritmo genetiko bat erabiliko dugu gure kostu-funtzioen itxura ezezaguna delako. Zehazki, helburu anitzeko optimizazio-problema ebazteko algoritmo genetiko ezagunenetakoa erabiliko dugu, hain zuzen NSGA-II algoritmoa [11].

Gogoan izan, metodo honen bitartez zuzenean soluzio multzo bat lortuko dugula, hau da $\{\gamma_1, \gamma_2, \dots, \gamma_s\}$ bektore sorta bat eta ondorioz, $\{s_{\gamma_1}, s_{\gamma_2}, \dots, s_{\gamma_s}\}$ gelditze-erregela sorta bat. Gelditze-erregela bakoitzak bi kostuen arteko oreka ezberdina eskainiko digu eta hau izango da helburu bakarreko metodoarekiko ezberdintasun eta abantaila nabariena: erabiltzaileak, gelditze-erregela ezberdinek emandako emaitzak ikusita, gehien komeni zaiona hautatzeko aukera izango du, zehaztasun eta azkartasun beharren arabera. Metodo honekin, jada ez da beharrezkoa izango alde aurretik azkartasun eta zehaztasunari eman beharrekiko garrantzia zehaztea, metodoak berak bi helburu hauek modu ezberdinetan orekatzen dituen soluzio sorta bat emango baitigu algoritmoaren exekuzio bakarrean. Gainera, optimizazio problemaren kostu-funtzio bakoitza bere aldetik agertzeak, halakoak eskala ez-berdinetan egotea baimentzen du, helburu-bakarreko metodoaren bigarren desabantaila nagusia ekidinez.

3.2. Ereduaren erabilera

Behin ereduak eraiki dugula, hau nola erabili azaldu behar dugu. Ikusi dugun moduan, helburu bakarreko eta helburu anitzeko ereduak emandako soluzioak oso ezberdinak dira (gelditze-erregela bat batean eta gelditze-erregela multzo bat, bestean), beraz erabilera ere apur bat ezberdina izango da.

Helburu bakarreko metodoak, aukeratutako α baliorako soluzio bakar bat itzuliko du. Beraz gelditze-erregela hau sailkatzaile probabilistikoekin batera zuzenean erabili dezakegu sailkapen goiztiarrak egiteko. Serie zati berri bat iristen zaigun bakoitzean, instant horri dagokion sailkatzaile probabilistikoari emango diogu sarreratzat delako serie zati hori, eta momentu horretan klase bakoitzari dagozkion probabilitateak lortuko ditugu. Ondoren, probabilitate hauek gelditze-erregelan sartuko ditugu 0 edo 1 emaitza lortuz. 0 emaitza lortuz gero, iragarpena oraindik fidagarria ez dela esan nahiko du eta datu gehiagoren zain geratuko gara; aldiz, 1 balioa lortuz gero momentuko klase-iragarpenarekin geratuko gara.



3. irudia. Helburu anitzeko ereduaren erabilera.

Bestalde, bi helburu-funtzioko metodoan, gelditze-erregela multzo bat lortuko dugu: Pareto-Soluzioen multzoa. 3. irudian ikus daitekeen bezala, erabiltzaileak gelditze-erregela hauetatik bat aukeratuko du, haren zehaztasun- eta azkartasun-beharrei erantzungo diena. Horretarako, entrenamendu datu-basean lortutako azkartasun eta zehaztasun kostuak erabiliko ditu

erreferentzia gisa. Azkenik, aukeratutako soluzio hau aplikatuko da, 3. irudiko koadro grisean erakutsitako pausoei jarraituz. Ohartu gaitezen, pauso hauek helburu bakarreko ereduaren erabileraren pausoak direla.

4. ESPERIMENTUAK

Kapitulu honetan, gure eredia ebaluatzeko egin ditugun probak azalduko ditugu. Horretarako, lehenengo, erabili ditugun datuak aurkeztuko ditugu. Ondoren, egindako parametroen aukeraketa azalduko dugu eta azkenik, lortu ditugun emaitzak ikusiko ditugu.

4.1. Datuak

Gure eredia ebaluatzeko, denborazko serieen UCR datu-base artxiboko [15] 8 datu-base erabili ditugu. Artxibo honetan denborazko serieen sailkapenen metodoak ebalua tzeko hainbat datu-base erreal eta sintetiko daude eskuragai eta arlo honetako oinarritzko erreferentzia bilakatu da azken urteotan.

1. taula. Erabilitako datu-baseen ezaugarri batzuk.

Datu basea	Entrenamendu serie kopurua	Test serie kopurua	Serieen luzeera	Klase kopurua
CBF	30	900	129	3
Coffee	28	28	287	2
ECG200	100	100	97	2
FaceFour	24	88	351	4
fish	175	175	464	7
Gun Point	50	150	151	2
ItalyPowerDemand	67	1.029	25	2
Trace	100	100	276	4

Artxiboan dauden datu-base guztietatik aukeraketa bat egin dugu, zehazki entrenamendu multzo txikia duten datu-baseetan zentratuz. Honela, esperimentazioaren kostu konputazionala mugatu dugu. Azpimarra dezagun, datu basearen tamainak bi aspektutan izan dezakeela eragina: h_t sailkatzaileen eraikuntzan eta C_a eta C_e kostuen kalkuluan. Bi aspektu hauek berdin-berdinak dira bai helburu anitzeko eta bai helburu bakarreko metodoetan, eta, beraz, laginaren tamaina mugatu arren, bi metodoen konpara-

ketak bidezkoa izaten jarraitzen du. Gainera, UCR artxiboko datu-baseak ehundaka artikulua zientifikotan erabili dira gaur arte eta ikusi da guztiek ikasketa prozesuak aurrera eramateko nahikoa handiak direla eta orokorrean, tamainak baino, beste faktore batzuek (klase kopurua, serieen itxura eta luzera, klaseen arteko ezberdintasuna) dutela eragin handiena lortutako emaitzetan. Aspektu horietan, aukeratutako datu-baseek ezaugarri ezberdinak dauzkate. 1. taulan laburbiltzen dira datu bakoitzaren ezaugarri batzuk.

Nabarmendu dezagun ezen artxibo honetako datu-base bakoitzak ez duela zerikusirik besteekin, ezaugarri ezberdinak dauzkala eta problema guztiz ezberdin bati dagokiola. Adibide gisa, ECG200 datu baseak 200 bihotz-taupaden serieak jasotzen ditu eta bi klaseak taupada normalak eta miokardioko infartua pairatzen dutenak banatzen ditu; edo ItalyPowerDemand datu baseak Italiako hilabeteko luzerako elektrizitate kontsumoaren serieak jasotzen ditu, lehenengo klasekoak urritik martxorako (udazkena-uda) kontsumoak direlarik eta bigarren klasekoak apiriletik irailera artekoak (udaberri-uda). Datu-base bakoitzaren azalpen zehatza, serieen irudiak eta klaseen esanahiak UCR artxiboan bertan edo [16] helbidean aurkitu daitezke.

4.2. Parametroen aukeraketa

Esperimentazioan gure metodoa exekutatzear gain, helburu-funtzio batenko ereduarekin konparatu nahi izan ditugu emaitzak, beraz, hori ere doitu eta exekutatu dugu.

Helburu bakarreko metodoan, jatorrizko lanean egindako esperimentazioa errepikatu dugu: R-ko *GA* [12] paketea erabili dugu algoritmo genetikoak inplementatzeko eta aldatu dugun parametro bakarra iterazio kopurua izan da, *maxiter* = 10. Algoritmo honek *linealki eskalatutako Fitness-aren araberako aukeraketa* erabiltzen du aukeraketarako; *aritmetika lokaleko gurutzaketa*, gurutzaketarako; eta *ausazko mutazio ez-uniformea*, mutaziorako [10, 13, 14]. Beste parametroak defektuz dauden bezala utzi ditugu: *pcrossover* = 0.8 (gurutzaketa probabilitatea) eta *pmutation* = 0.1 (mutazio probabilitatea). Gainera, metodo honen esperimentazioa ($\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$) 9 pisu balioetarako burutu dugu, guztira datu-base bakoitzerako 9 soluzio (gelditze-erregela) lortuz.

Bi kostu-funtzioko eredia optimizatzeko, R-ko *mco* paketeko *nsga2* [17] algoritmoa erabili dugu. Algoritmo genetiko honetan, aukeraketa gauzatzeko *binakako lehiaketa bidezko aukeraketa* erabiltzen da, gurutzaketarako *gurutzaketa binario simulatua* eta mutaziorako *mutazio polinomiala* [10, 13, 14]. Parametro gehienak, defektuz dauden bezala utzi ditugu: *cprob* = 0.7 (gurutzaketa probabilitatea), *cdist* = 5 (n-ren balioa), *mprob* = 0.2 (mutazio probabilitatea) eta *mdist* = 10 (η_m -ren balioa). Bestalde, aurrekoan bezala, aldatu dugun parametro bakarra *popsiz*e parame-

troa izan da, defektuz $popsize = 100$ (populazioaren tamaina) da, baina, guk, $popsize = 32$ definitu dugu, kodea azkarrago exekutatu ahal izateko.

4.3. Ebaluazio metodoa

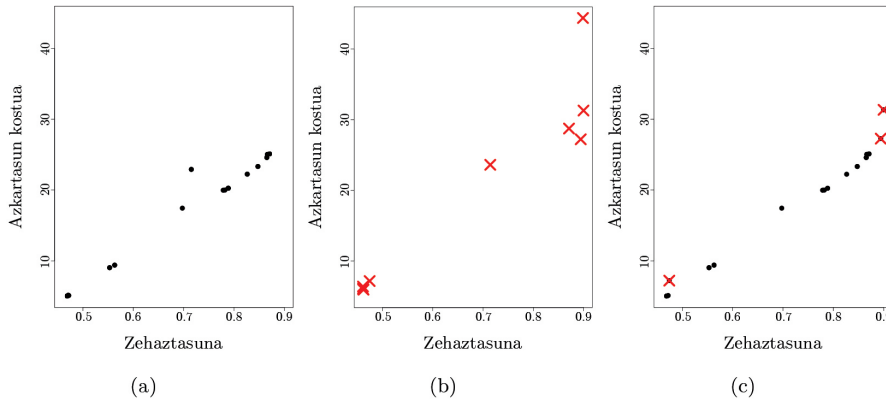
UCR artxiboan datuak jada bi multzotan banatuta hartu ditugu UCR artxibotik [15], alde batetik entrenamendu multzoa eta bestetik test multzoa. Aurrerago, beste eredu en emaitzekin konparatu ahal izateko, horrela mantendu ditugu. Beraz, entrenamendu multzoa erabili dugu eredu a doitzeko eta test multzoa, berriz, eredu a ebaluatzeo. Ebaluazio neurriak, *zehaztasuna* (2. ekuaziotik abiatuz kalkulata) eta *azkartasun* kostua (3. ekuazioa) izan dira.

4.4. Emaitzak

Esan bezala, datu-base bakoitzarekin helburu-funtzio bateko eredu a eta bi helburu-funtzioko eredu a doitu ditugu eta emaitzak konparatu ditugu. Ereduak modu errazean konparatzeko, datu bakoitzarentzat grafiko mota bera egin dugu: lehenengo, bi helburu-funtziorekin lortutako soluzioak irudikatu ditugu X ardatzean zehaztasuna eta Y ardatzean azkartasun-kostua irudikatuz; Ondoren, helburu-funtzio bakarreko ereduarekin $\alpha \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ balioetarako lortutako soluzioen kostuak irudikatu ditugu era berean; azkenik, konturatu beharra dago, datu desberdinekin eraiki eta ebaluatzen ari garenez eredu a, entrenamendu multzorako ez-dominatuak diren soluzio guztiek ez dutela zertan test multzorako ez-dominatuak izan, bestela esanda, Pareto-frontearen hurbilpen bat lortzen dugu test multzorako. Horregatik, badaude dominatuak diren soluzio batzuk ere. Hori dela eta, hirugarren grafiko bat irudikatuko dugu helburu bakarreko ereduarekin eta bi helburutako ereduarekin lortutako soluzio guztien artean, ez-dominatuak diren soluzioak bakarrik irudikatuz.

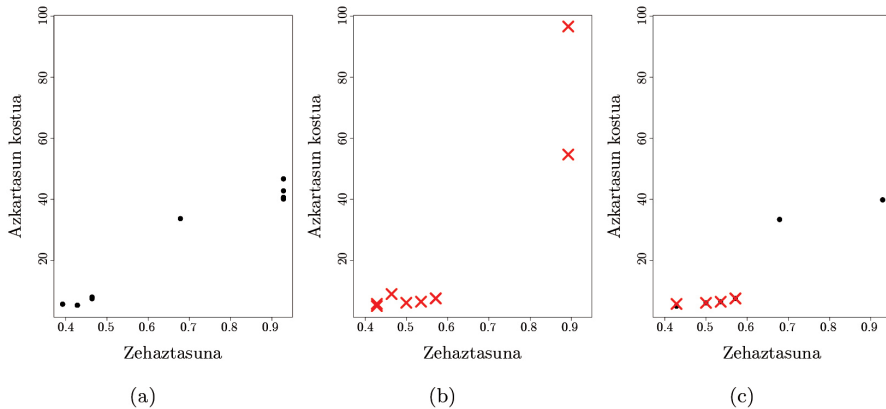
CBF datu-baserako aurretik aipatu ditugun hiru grafikoak 4. irudian ikus ditzakegu adibidez: (a) irudian gure helburu anitzeko ereduarekin lortutako emaitzak irudikatu ditugu. Esan bezala, badaude dominatuak diren emaitzak ere. (b) irudian helburu bakarreko ereduarekin, α bakoitzerako lortutako 9 soluzioak irudikatu ditugu. (c) irudian, bi eredu ekin lortutako soluzio guztien artean ez-dominatuak direnak bakarrik irudikatu ditugu. Gurutzeak, helburu bakarreko ereduarekin lortutako soluzioak dira eta ikusten da nola, α desberdinekin paretoa hurbiltzen saiatzen garen, α handitzen den heinean *zehaztasunari* pisu handiagoa emanez eta *azkartasunari* txikiagoa. Horren adierazgarri da, zenbat eta α handiagoa izan, hainbat eta *zehaztasun* handiagoa eta *azkartasun* kostu handiagoa dutela soluzioek. Gainera, irudiei erreparatu, argi dago bi helburu-funtziotako ereduak Paretoaren errepresentazio zabalagoa ematen digula, exekuzio bakar batean, helburu-funtzio bakarrekoan 9 eginda baino. Ikus daiteke nola 9 soluzioeta-

tik 3 bakarrik agertzen diren Pareto-soluzioen artean eta gainera, nola hiru-
rek garrantzia handiagoa ematen dioten helburu bati besteari baino. Soluzio
hauek gehienetan baztertu egiten dira, aplikazio eremu gehienetan interes-
garrin suertatzen diren soluzioak oreka bilatzen dutenak dira eta.



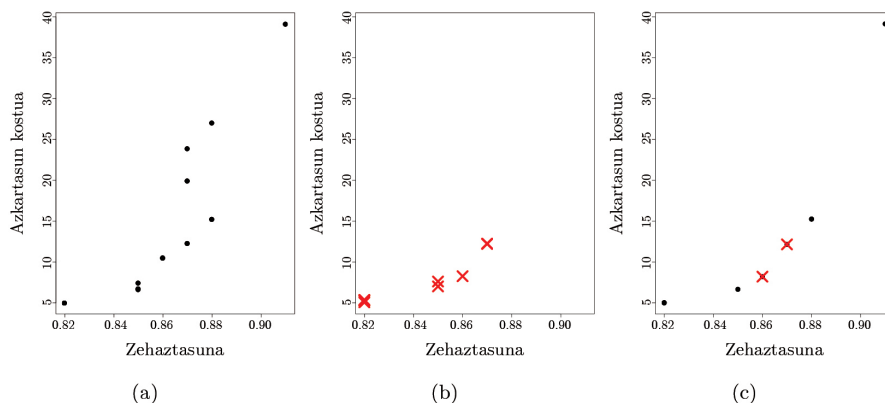
4. irudia. CBF datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto soluzioak bakarrik.

Coffee datu-basearen kasuan, 5. irudian ikus ditzakegu eredu bakoitzaren soluzioak eta Pareto-frontea. Kasu honetan, bi helburuko ereduaren soluzio desberdinak gutxi dira eta ondorioz, soluzio gutxi dira test multzorako ez-dominatuak direnak ere. Hala ere, bi helburuko metodoarekin lortutako emaitzak dira orekatuenak kasu honetan ere.

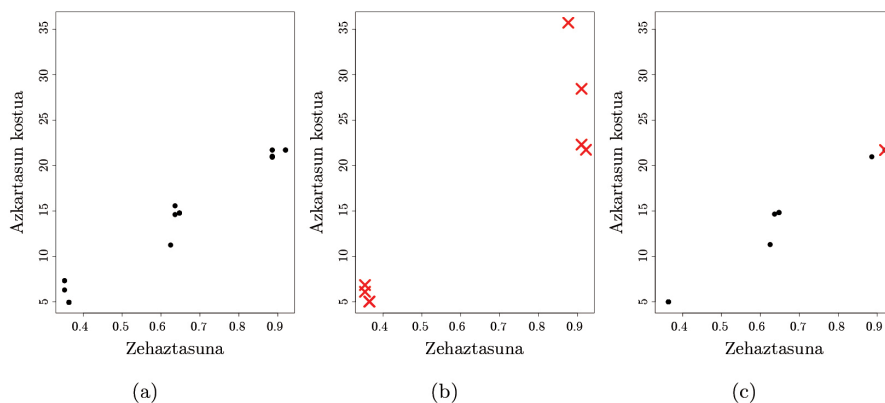


5. irudia. Coffee datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak.

ECG200 datu-basearen kasuan, 6. irudian ikus dezakegu nola helburu bakarreko ereduak bi helburuen arteko oreka bat daukaten soluzio asko lortzen dituen, nahiz eta bi helburukoak gehiago lortu. Hala ere, Pareto-soluzioak irudikatzerakoan helburu bakarreko ereduak bi soluzio besterik ez dira agertzen, besteak bi helburuko soluzio batek dominatzen baititu.



6. irudia. *ECG200* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak bakarrik.

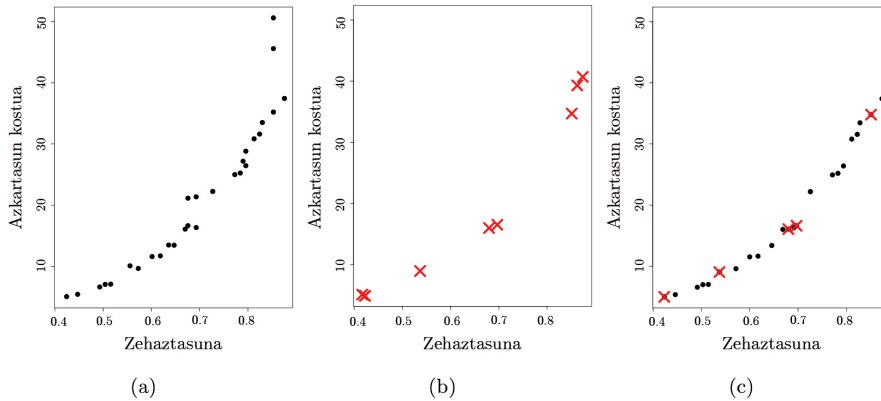


7. irudia. *FaceFour* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak bakarrik.

FaceFour datu-basearen kasuan, 7. irudiari erreparatu, argi geratzen da bi helburuko ereduak bi helburuen arteko soluzioen arteko oreka bat lor-

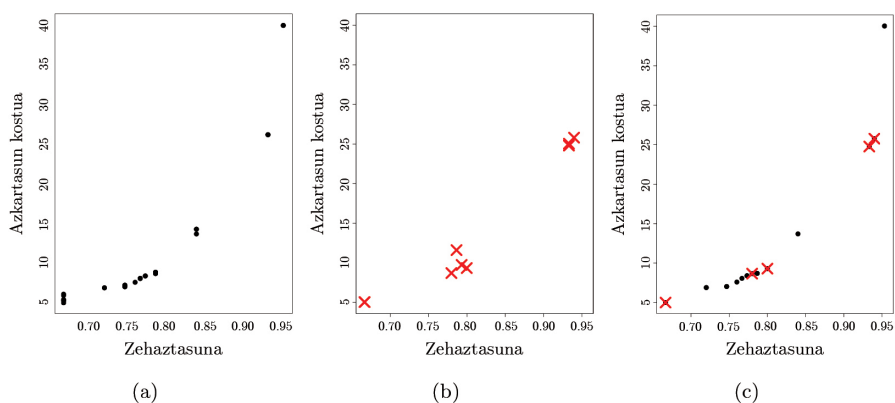
tzen duten soluzioak bilatzen dituela eta helburu bakarrekoak, aldiz, helburu bati edo besteari ematen diola pisu gehiago. Honen adierazgarri da helburu bateko soluzio gehienek zehaztasun altua baina azkartasun kostu altua dauzkatela, edo alderantziz. Bi helburuko soluzio asko ordea, erdian daude. Gainera, Pareto-soluzioak irudikatzerakoan ikus daiteke helburu bakarreko ereduak lortutako soluzio bakarra agertzen dela, eta, gainera, soluzio hori izan ohi dela azkartasun-kostu handiena daukana, nahiz eta soluzio horrek izan zehaztasun handiena.

Fish datu-baserako grafikoak aztertuz (8. irudian), kasu honetan ikus dezakegu helburu bakarreko funtzioak bi helburuen arteko oreka bilatzen duten soluzioak badauzkala, baina hala ere, gehienek helburu bati edo besteari ematen diote garrantzia. Bi helburuko ereduak aldiz, exekuzio bakar batean askoz soluzio gehiago bilatzen ditu eta gehienek bi helburuen arteko oreka bat ematen digute, interesatzen zaigun bezala.



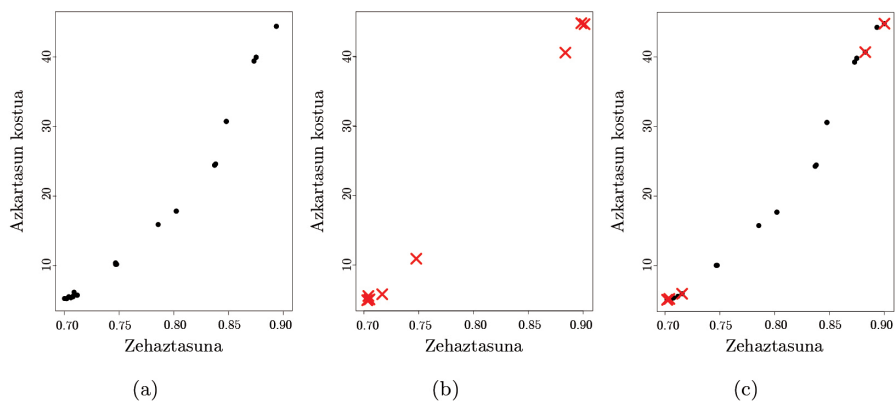
8. irudia. *Fish* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak bakarrik.

Gun Point datu-basearen kasuan, 9. irudian erreparatuz, ikusten da bi eredutan badaudela muturretako soluzioak, hau da, helburu bati edo besteari garrantzia handiagoa ematen dioten soluzioak, baina baita bien arteko oreka bat lortzen dutenak ere. Hala ere Pareto-soluzioak irudikatzerakoan, oreka hori daukaten helburu bakarreko eredutik lortutako bi soluzio bakarrik aurkitzen ditugu. Helburu bakarreko beste soluzioek, helburu bati edo besteari ematen diote garrantzia. Bi helburuko ereduaren kasuan ordea, hainbat soluzio daude bi helburuen arteko oreka hori ematen digutenak.



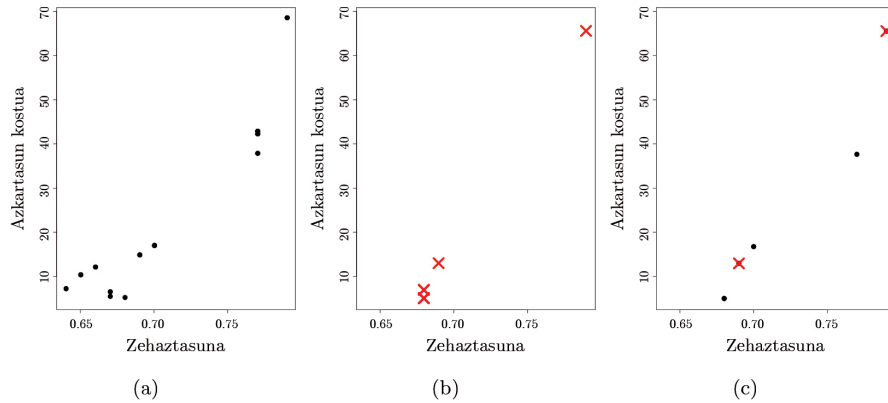
9. irudia. *Gun Point* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak bakarrik.

ItalyPowerDemand datu-basearen kasuan (ikus 10. irudia), helburu bakarreko ereduarekin lortutako soluzio guztiak muturrekoak dira, hau da, edo zehaztasun oso altua baina azkartasun kostua ere oso altua dute edo azkartasun kostu txikia baina zehaztasuna ere halako daukate. Bi helburuko ereduarekin lortutako soluzioak ordea, azkartasun kostu txikitik eta zehaztasun txikitik hasiz, azkartasun kostu handietara baina zehaztasun handietara mugitzen dira, horrela, erdiko soluzio asko lortuz. Hau da, bi helburuen arteko oreka daukaten soluzioak emanez. Pareto-soluzioak irudikatzerakoan, ikusten dugu soluzio gehienak bi helburuko ereduarekin lortuak direla.



10. irudia. *ItalyPowerDemand* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto-soluzioak bakarrik.

Trace datu-basearen kasuan (11. irudian), ikus dezakegu nola hasieran bi helburuko ereduarekin lortutako soluzio asko test multzoan ez diren Pareto-soluzioak. Gauza bera gertatzen da helburu bakarreko ereduarekin lortutako soluzioekin ere. Hau, test eta entrenamendu multzoak oso desberdinak direlako izan daiteke. Pareto-soluzioak irudikatzerakoan, bost soluzio desberdin dauzkagu: horietatik hiru, bi helburuko ereduarekin lortuak; eta bi, helburu bakarreko ereduarekin lortuak.



11. irudia. *Trace* datu-basearen emaitzak, (a) irudian helburu anitzeko ereduarekin lortuak, (b) irudian helburu bakarreko ereduarekin lortuak eta (c) irudian soluzio guztietatik Pareto soluzioak bakarrik.

5. ONDORIOAK ETA ETORKIZUNERAKO LANA

Orokorrean, emaitzak ikusita, argi ikusten da gure helburu anitzeko ereduaren abantaila nagusia: bi helburu-funtzioko ereduak exekuzio batean soluzio multzo bat ematen du, eta helburu bakarrekoak exekuzio bakoitzeko soluzio bakarra ematen digu. Honek, eredu eraikitzerako garaian denbora asko aurrezteko baimentzen du.

Gainera, gehienetan helburu-funtzio bakarreko ereduak soluzioek garrantzia handiagoa ematen diote helburu bati besteari baino eta oso zaila da oreka bilatzen duen α balio egoki bat bilatzea, balio hori datu-base bakoitzean ezberdina izanik.

Bi helburu-funtzioko ereduaren kasuan aldiz, paretoaren hurbilpen askoz ere hobea lortzen dugu exekuzio bakar batean eta α bezalako parametrorik gabe. Gainera, esperimuntuan egiaztatu dugu, egoera gehienetan, bi helburu-funtzioko metodotik lortutako soluzioek helburu-funtzio bakar bateko soluzio gehienak dominatzen dituztela, eta horrek erakusten du optimizazio-metodo egokiagoa dela.

Laburbilduz, ikusi dugu bi helburu-funtzioko ereduak, exekuzio bakar batean, Paretoaren hurbilpen hobe ematen digula eredu bakarreko ereduak baino. Beraz, eraginkorragoa da bi helburu-funtzioak aldi berean optimizatzea biak funtzio batean nahastu beharrean.

Egindako esperimentazioan ereduak emaitza onak eman dituela ikusi dugunez, esperimentuarekin jarraitzea da etorkizuneko lana. Honela, algoritmo genetikoaren iterazio gehiago, errepikapenak zorizkotasunaren efektua ekiditeko, populazio handiagoak eta parametro konbinaketa gehiagorekin probatu beharko genuke emaitzak sendagoak izateko. Gainera, UCR artxiboko beste datu-baseekin ere esperimentua egin beharko genuke, lehen aipatutako aspektu (klase kopurua, serieen itxura eta luzera, klaseen arteko ezberdintasuna) ezberdin gehiago kontuan hartzean, ateratako ondorioak eta emaitzak mantentzen direla ziurtatzeko.

6. BIBLIOGRAFIA

- [1] WITTEN, I. H., and FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. 2005. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [2] CHUNG FU, T.. 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, **24:1**, 164-181.
- [3] ESLING, P., and AGON, C. 2012. Time-series data mining. *ACM Computing Surveys*, **25:1**, 1-34.
- [4] WANG, X., MUEEN, A., DING, H., TRAJCEVSKI, G., SCHEUERMANN, P., and KEOGH, E. 2013. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, **26:2**, 275-309.
- [5] XING, Z., PEI, J., and YU, P. S.. 2011. Early classification on time series. *Knowledge and Information Systems*, **31:1**, 105-127.
- [6] MORI, U., MENDIBURU, A., DASGUPTA, S., and LOZANO, J. A. 2018. Early Classification of Time Series by Simultaneously Optimizing the Accuracy and Earliness. *IEEE Transactions on Neural Networks and Learning Systems*, **In press**, 1-10.
- [7] MORI, U., MENDIBURU, A., DASGUPTA, S., and LOZANO, J. A. 2015. Early classification of time series from a cost minimization point of view. In *NIPS time series workshop*.
- [8] GRIFFIN, M., and MOORMAN, J. 2001. Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics* **107:1**, 97-104.
- [9] RASMUSSEN, C. E. &., and WILLIAMS, C. K. I. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.

- [10] EIBEN, A. E., and SMITH, J. E. 2003. *Introduction to Evolutionary Computing*. Springer Verlag.
- [11] DEB, K., PRATAP, A., AGARWAL, S., and MEYARIVAN, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, **6:2**, 182-197.
- [12] SCRUCICA, L. 2013. GA: A package for genetic algorithms in R. *Journal of Statistical Software*, **53:4**.
- [13] K.C, T., E.F., K., and T.H., L. 2002. Evolutionary algorithms for multiobjective optimization: Performance assessments and comparisons. *Artificial Intelligence Review* .
- [14] DEB, K., and DEB, D. 2014. Analysing mutation schemes for real-parameter genetic algorithms. *Int. J. Artif. Intell. Soft Comput.*, **4:1**, 1-28.
- [15] CHEN, Y., KEOGH, E., HU, B., BEGUM, N., BAGNALL, A., MUEEN, A., and BATISTA, G. The UCR time series classification archive, July 2015.
- [16] BAGNALL, A., LINES J., VICKERS W. and KEOGH E., The UEA & UCR Time Series Classification Repository, www.timeseriesclassification.com
- [17] MERSMANN, O. 2014. *mco: Multiple Criteria Optimization Algorithms and Related Functions*. R package version 1.0-15.1.