



IKER
GAZTE
NAZIOARTEKO
IKERKETA EUSKARAZ

II. IKERGAZTE

NAZIOARTEKO IKERKETA EUSKARAZ

2017ko maiatzaren 10, 11 eta 12
Iruñea, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

INGENIARITZA ETA ARKITEKTURA

Web elkarrekintzan erabilitako
gailuen detekzio automatikoa

*Iñigo Perona, Ainhoa Yera,
Olatz Arbelaiz, Javier Muguerza,
Juan Eduardo Pérez eta
Xabier Valencia*

22-29 or.

<https://dx.doi.org/10.26876/ikergazte.ii.03.03>

ANTOLATZAILEA:



ELKARLANEAN:



LAGUNTZAILEAK:



Web elkarrekintzan erabilitako gailuen detekzio automatikoa

Iñigo Perona Balda¹, Ainhoa Yera Gil¹,

Olatz Arbelaiz Gallego¹, Javier Muguerza Rivero¹,

Juan Eduardo Pérez López² eta Xabier Valencia Parafita²

¹ Euskal Herriko Unibertsitatea (UPV/EHU). Aldapa: <http://aldapa.eus/>.

² Euskal Herriko Unibertsitatea (UPV/EHU). Egokituz: <http://egokituz.eus/>.

Laburpena

Gaur egun digitalki trebeak izatearen garrantziak, behar bereziak dituzten pertsonak gailu digitalak eta aplikazioak erabiltzen prestatu beharra ekarri du, eta honekin batera webgunearekin duten elkarrekintza beraien beharretara automatikoki egokitzearen beharra ere agertu da. Lan honek azken puntuak jartzen du begia, erabiltzailearen nabigazio ezaugarri garrantzitsuak identifikatzean eta konputagailuarekin elkarrekintza burutzeko erabili den gailuan. Zehazki, erabiltzailearen web elkarrekintzatik eratorritako datuetan oinarrituz eta datu meatzaritza konputazio-prozesu osoa aurrera eramanez erabiltzaileek erabili dituzten gailuak identifikatzeko gai den sistema eraiki da. Sistema honek, nabigazio erantzun antzekoa duten gailuen multzoak %99.22-ko zehaztasun-tasaz desberdintzen ditu.

Hitz gakoak: datu meatzaritza, ikasketa automatikoa, web-irisgarritasuna, web moldakorra.

Abstract

Nowadays the importance of digital competences makes important to capacitate people with disabilities in the use of digital devices and applications, and moreover, to adapt the site interaction to their necessities automatically. This work focuses on the latter, identifying important user characteristics and the device being used to interact with the computer. Based on web user interaction data, the complete data mining process has been carried out to build a system able to identify the used device. This system has obtained %99.22 of accuracy distinguishing between groups of devices with similar navigational response.

Keywords: data mining, machine learning, accessibility, adaptive web.

1 Sarrera eta motibazioa

Azken hamarkadetakoa joera aldaketek sarearen erabilera eta bertan biltzen den informazioa izugarri ugaritzea ekarri dute. Honek webguneak baliabide garrantzitsu bilakarazi ditu, bai heuretara sartuz informazioa bilatzeko edo eta jendarteko komunikazio eta parte-hartze prozesuak aurrera eramateko. Hau dela eta, gaur egun, edonork digitalki moldatzeko trebetasuna baduela oinarritzotzat hartzen da. Testuinguru honetan, bereziki garrantzitsua da behar bereziren bat duten pertsonak gailu digitalak eta bertako aplikazioak erabiltzean prestatzea eta ohitzea, eta era berean, webgunearekin duten elkarrekintza euren beharretara egokitzea nabigazioa errazteko asmoz.

Zoritxarrez, teorikoki irisgarritasun baldintzak betetzen dituen web-diseinu batek ez du ziurtatzen behar bereziko pertsonak webgunean erraztasunez nabigatu dutenik. Testuinguru honetan, webgunea erabiltzailearen beharretara egokitzea ezinbestekoa bilakatzen da. Egokitzapen hauek erabiltzaile bakoitzari eginarazitako berariazko galdetegi baten emaitzen arabera zehatz daitezke, baina egokitzapen hauen egingarritasuna erabiltzaileek galdetegian duten parte-hartzera dago murriztua. Gainera, web aplikazio orokorretan oso erraza da erabiltzaile profil guzti-guztiak aurretik definitzean huts egitea Dillon (2001) lanean ikus daitekeen moduan.

Egokitzapenak proposatzeko beste aukera bat erabiltzailea eta webgunearen arteko elkarrekintza datuak aztertzeetik ondorioztatutakoaren arabera proposatzea da. Aukera hau orokorragoa, erabiltzailearentzat gardena eta beraz, erabiltzaile guztiei aplikagarria da.

Erabiltzaile bati beren-beregi eginiko egokitzapenek erabiltzaile horren ezaugarrietan oinarritu behar ko dute, nabigatzean dituen arazoetan etab. Testuinguru honetan, elkarrekintza gauzatzen ari den unean bertan nabigazio arazoak eta erabili den gailu mota detektatzea beharrezko lehen urratsak dira. Hone-la gero, erabiltzaileari webgunea automatikoki egokitzea posible izan dadin eta beraz, bere web-bisita esperientzia hobetu dadin.

Helburu hauek lortzeko, web meatzaritzako teknikak erabiltzeak abantaila asko ematen ditu Liu (2006) liburuan zehazten denez: erabiltzaileak bere datuak ez ditu zertan esplizituki eman behar, benetako nabigazioetatik lorturiko datu estatistikoetan oinarritzen dira (suposizio okerren probabilitatea jaisten da) eta sistema moldakorrak dira (erabiltzailearen ezaugarriak aldatzen diren heinean, biltzen diren datuak ere aldatuko dira, eta honek, interakzio eskema automatikoki aldatuko du). Erabiltzailea urritasun fisikoa, sensoriala edo kognitiboa duen pertsona denean, datu meatzaritzak da erabiltzaileari buruzko erabilera-ohituren informazioa eskuratzeko edo bere ezaugarriak deduzitzeko modu errazena eta askotan bakarra.

Lan honetan, erabiltzaileek nabigatzen duten bitartean darabilten gailuak automatikoki detektatzeko hastapeneko sistema aurkezten dugu. Sistema hau Remotest tresnan, Valencia *et al.* (2015) lanean aurkeztua, bildutako datuetan oinarritzen da, non erabiltzailearen elkarrekintza datu asko biltzen diren. Proposatutako sistemak datu meatzaritzako konputazio-prozesu osoa jarraitzen du. Lehenik, datuak bildu eta web-irisgarritasuneko adituekin eginiko bileretan ezaugarri deskribatzaile zerrenda bat definitu zen. Gero, ezaugarri erauzte eta kalkulatzeko sistema eraiki zen. Eta azkenik, ikasketa automatikoko algoritmo ezberdinak, ezaugarriak detektatzeko aukera batzuekin uztarturik, nabigatzean erabilitako gailuak ezberdintzeko gai izango zen sistema eraikitzeak erabili ziren.

Emaitzek erakusten dute Remotest tresnaz bildutako datuei datu meatzaritzako prozesu osoa aplikatzea etorkizun oneko estrategia dela automatikoki erabiltzailearen ezaugarriak detektatzeko eta gero honi beren-beregi egindako egokitzapenak proposatzeko.

Artikulu honek hurrengo atalean, datuak biltzeko erabilitako tresna eta datu-basearen eraketa deskribatuz jarraituko du. 3. atalean sailkatzaileen eraketa eta ebaluazioa aztertuko dira. Azkenik, ondorioak eta etorkizuneko lanak zerrendatuko dira 4. atalean.

2 Datuen prestaketa

2.1 Datuen bilketa

Valencia *et al.* (2015) lanean aurkezturiko Remotest plataformak ikertzaileei beharrezko funtzionalitateak eskaintzen dizkie euren web nabigazioan oinarrituriko esperimenduak definitzen laguntzeko, hala nola, laborategian bertan eginiko edo urruneko web saioak maneiatzeko aukera edo eta bildutako elkarrekintza datuak aztertzeko aukera ematen die. Honengatik plataforma honek esperimendu aukera anitz onartzen ditu. Plataforma honen arkitekturak alde biko arkitektura hibrido bat jarraitzen du non funtzio batzuk bezeroaren aldeko moduluetan eta beste batzuk zerbitzariaren aldeko moduluetan burutzen diren. Plataforma hau lau modulutan banatzen da: esperimendatzailearen modulua (EXm), parte-hartzailearen modulua (PAm), koordinatzailearen modulua (COM) eta emaitzak ikusteko modulua (RVm).

Nahiz eta RVm moduluak datu batzuk prozesatu eta ikustarazten dituen, lan honetan proposaturiko sistema eraikitzeak zuzenean elkarrekintza unitatearekin lan egin da, hau da, PAm moduluak bildutako elkarrekintza gertaerekin: kurtsore mugimenduak, tekla sakaketak, leiho korritzeak, klikak etab.

Lan honetarako datu iturburu izan den estudioan hamabost norbanakok hartu zuten parte. Hauek bost taldetan sailkatu ziren nabigatzean apuntatzeko eta klikatzeko ekintzak burutzeko erabilitako sarreragailuaren arabera: (a) hatzez sakatze zuten bi teklatu erabiltzaile (HatezTeklatu), (b) buru-adarrez sakatzen zuten bi teklatu erabiltzaile (AdarrezTeklatu), (c) trackball erabiltzaile bat (trackball), (d) lau joystick erabiltzaile (joystick) eta (e) sei sagu erabiltzaile (sagua).

Lehen lau taldeetako norbanako guztiak (a, b, c eta d) narriadura motorea zuten parte-hartzaileak ziren eta gehienek ordenagailua egunero erabiltzen zazpi urtetik gorako esperientzia zuten. Aldiz, azken norbanakoen taldeak (e) ageriko behar berezirik ez zuten erabiltzaileak bakarrik biltzen zituen eta hauek ere, saguarekin ordenagailua egunero erabiltzen zazpi urtetik gorako esperientzia zuten.

Erabilitako makinari dagokionez, 64 biteko Windows 7 bertsioa martxan zuen Dell Precision M6700 eramangarri bera erabili zen saio guztietarako. Erabiltzaileei ataza-estimuluak aurkezteko 16:10 proportziodun zabalera zuen LCD monitore gehigarri bat erabili zen, bere diagonalaren luzera 24 hazbete zelarik

eta bistaratze bereizmena 1920 x 1200 pixeletara finkatu zelarik. Makina honetan Mozilla Firefox nabigatzailea erabili zen, eta bertan, kurtsorearentzat alegiazko laguntzak inplementatuak zituzten gehigarriak instalatu ziren.

Estudioa hasi aurretik parte-hartzaileei euren kurtsorea doitzea gomendatu zitzaizen hartara adierazle honen mugimendu-portaera euren beharretara egoki zezaten. Atazak egikaritzeko a, b, c eta d taldeko norbanakoek euren sarrera-gailu pertsonalak erabili zituzten. Behar berezirik ez zuten parte-hartzaile guztiek (e taldeak) USB sagu optiko bera erabili zuten (Dell M-UVDEL1).

Esperimenturako bi webgune ezberdin aukeratu ziren ataza-estimulu gisa: behar bereziko pertsoneri zuzendutako informazioa eskaintzen duen Discapnet¹ webgunea, eta instituzio baten webgunea, Gipuzkoako Foru Aldundiaren² webgunea hain zuzen. Hirugarren bat ere erabili zen parte-hartzaileek kurtsore berriaren alegiazko areagotzeak erabiltzen ikas zezaten. Horretarako, Bidasoa-Txingudi badiari buruzko informazioa duen Bidasoa Turismo³ webgunea erabili zen. Hiru webguneek ziurtatzen zuten irisgarriak izateko WCAG 1.0 gidalerroak betetzen zituztela maila batean ala bestean: Discapnetek AA mailan, Gipuzkoak A mailan eta Bidasoak A mailan.

2.2 Datu-basearen eraketa eta ezaugarrien aukeraketa

Proposaturiko gailuak detektatzeko sistemaren diseinuak datu meatzaritzaren prozesu osoa du oinarrian. Datu-basea eratzeko web-irisgarritasuneko adituekin lehen bilerak egin ostean, Remotest-ek bildutako datuetatik erauzi beharreko ezaugarriak zehaztu ziren. Ondoren, ezaugarri hauek kalkulatu eta erauziko zituen sistema eraikiko zelarik.

Erabiltzaileak webgunearekin zuen elkarrekintza informazioa Remotest tresnaz bildu zen, eta hortik datu-basea etiketatu bat sortuko zen, gero, gainbegiraturako sailkatzaileen testuinguruan erabiliko zena. Datu-basea sortu ahal izateko esperimenduek ziharduten bitartean erabiltzaileetako bakoitzaren elkarrekintza moduak bisitatutako web-orri ezberdinetan antzekoak zirela ontzat jo zen, eta baita elkarrekintza hau nolabait erabiltzailea erabiltzen ari zen gailu motaren menpekoea zela ere. Beraz, erabilitako gailua, datu-baseko kasuak etiketatzeko erabili zen, honela bost klaseko datu-basea sortuz: hatzez teklatua, buru-adarrez teklatua, trackball, joystic eta sagua.

Sortutako datu-baseko sarrera bakoitzak erabiltzaile batek bisitatutako web-orri bakoitzean egindako elkarrekintzaren laburpen bat du. Elkarrekintza hau laburki deskribatzeko asmoz, web-irisgarritasun adituek ahalik eta ezaugarri esanguratsu gehien erauzteko ahalegina egin zuten. Zehazki 19 ezaugarri-igarle erauzi ziren 1. taulan labur definitzen direnak.

Aurrera eramandako esperimenduetatik, bost klase desorekan eta 20 ezaugarri (19 ezaugarri-igarle + klase edo helburu-ezaugarri bat) dituen datu-base bat sortu zen (ikus 2. taula). Ezaugarri guztien balioak estandarizatu ziren, horrela ezaugarriek har zitzaketan balioak balio-tarte berdinetara eramanez eta beraz, sailkatzaileak eraikitzekeo prozesuan balio-tarte ezberdinek izan zezaketan eragina minimizatu.

Hastapeneko lan honen helburua nabigatzeko erabili den gailua automatikoki detektatzea den arren, etorkizun hurbileko helburua erabiltzaileen nabigazio problemak detektatzea izango da. Hori dela eta, ezaugarriak erauztean bi helburuak izan ziren buruan. Adibidez, KurtsDistOptRat edo NoraAldBort100 bezalako ezaugarriak nahiz eta hasiera batean, intuizioz, erabilitako gailu ezberdinen artean ezberdintzeko ezaugarri esanguratsuak izan daitezkeenik ez espero, hala ere sailkatzaileak eraikitzean, erauzitako ezaugarri multzo guztiarekin lan egin zen susmo hauek ebaluatzeko asmoz.

Bestalde, web-irisgarritasuneko adituen esanak kontuan hartu ziren eta hortik, erauzitako 19 ezaugarrietatik 8 apartatu ziren erabilitako gailua detektatzeko esanguratsuak zirelakoan. Zehazki, 1. taulan 1 lehentasunaz ageri direnak. Honela, ezaugarri hauen igartze gaitasuna neurtze aldera, ezaugarri hauekin soilik eraikitako sailkatzaileak ere eraiki ziren.

3 Sailkapen sistema

Kalkulatu eta erauzitako ezaugarriak, nabigatzeko erabilitako gailuaren arabera, erabiltzaileen elkarrekintza datuak sailkatuko dituen sailkatzaileak eraikitzekeo erabili ziren. Sailkatzaileak lehenik, erauzitako

¹<http://www.discapnet.com/>

²<http://www.gipuzkoa.eus/>

³<http://www.bidasoaturismo.com/>

1 Taula: Erabiltzaile eta bisitatutako orri bakoitzerako erauzi diren ezaugarrien izena eta deskribapena. Izenaren alboan adituek emaniko lehentasun maila adierazi da parentesi artean: 1, 2 edo 3 (lehentasun handitik txikira).

Izena	Deskribapena
GertaeraKop (1)	Remotest-ek orri batean gorderiko gertaera kopurua.
TeklaBereziKop (1)	Orri batean tekla berezi, alfanumerikoak ez diren sakaketa kopurua.
GurpilKop (1)	Orri batean saguko gurpila zenbat aldiz erabili den.
GurutzMugiKop (1)	Orri batean kurtsorea bere ardatz horizontal eta bertikalen norantzetan zenbat aldiz mugitu den (gurutzean mugitu).
DiagMugiKop (1)	Orri batean kurtsorea 45, 135, 220 eta 315 graduko norantzetan zenbat aldiz mugitu den (diagonalean mugitu).
GeldikMediana (1)	Orri batean kurtsorea geldirik geratu deneko denbora-tarteen mediana.
AbiadenMediana (1)	Orri batean kurtsoreak hartzen dituen abiaduren mediana.
AzelMediana (1)	Orri batean kurtsoreak egiten dituen azelerazioen mediana.
TeklaKop (2)	Orri batean tekla sakaketa kopurua.
KutsoreDist (2)	Orri batean kurtsoreak eginiko bidearen luzera pixel kopuruan emanda.
KurtsDistOptRat (2)	Orri batean kurtsoreak eginiko bidearen eta optimoaren arteko ratioa. Bide optimoa orrian kurtsorearen hasiera eta amaiera posizioen arteko bide motzena da (bide zuzena).
NoraAldBort100 (2)	Orri batean kurtsoreak eginiko norantza aldaketa bortitzen portzentaia. Aldaketa bortitzat zer hartu den ikusteko ikus NoraAldBortKop ezaugarria.
KlikKop (3)	Remotest-ek orri batean gorderiko klik kopurua.
KorritzeKop (3)	Orri batean eginiko leiho korritze kopurua.
NoraZuz100 (3)	Orri batean kurtsoreak bere gurutzean edo diagonalean egindako mugimendu kopuruen portzentaia.
NoraAldKop (3)	Orri batean kurtsoreak eginiko norantza aldaketa kopurua.
NoraAldBortKop (3)	Orri batean kurtsoreak eginiko norantza aldaketa bortitz kopurua. Bortitzat kurtsorearen norantza erlatiboki 45 gradu edo gehiagoan aldatzen denean hartzen da.
EremuzMugiKop (3)	Orri batean kurtsoreak bere ardatz horizontal, bertikal eta diagonalek zedarritzen dituzten zortzi eremuetan batetik bestera zenbat aldiz mugitu den.
IgaroDenb (3)	Orri batean igarotako denbora.

2 Taula: Sortutako datu-basean bost klaseen distribuzioa.

Klasea	Kasu kopurua.
HatzezTeklatu	347
AdarrezTeklatu	338
Trackball	171
Joystic	584
Sagua	235
Total	1675

ezaugarrien multzo osoa erabiliz eraiki ziren. Bigarrenik, web-irisgarritasunean adituek esanguratsutat azpimarratutako ezaugarriekin eraiki ziren sailkatzaileak.

Aurrekoetz gain, García *et al.* (2015) lanean azaltzen diren ezaugarriak automatikoki detektatzeko bi algoritmo erabili ziren, horrela, ikasketa automatikoaren ikuspuntutik erabiltzea komeni ziren ezaugarri onenak zehaztuko zirelarik. Honetarako ezaugarriak detektatzeko gehien erabiltzen diren bi algoritmo aukeratu ziren: Hall (1998) laneko Correlation-based Feature Subset Selection (FSS) algoritmoa non korrelazioan oinarriturik ezaugarrien azpimultzoa aukeratzeko den eta Kohavi eta John (1997) laneko Wrapper algoritmoa non emaniko sailkatzaileako optimoak diren ezaugarrien multzoa aurkitzen den. Gure kasuan Wrapper egikaritzeko Quinlan (1993) lanean aurkeztutako J48 sailkatzaile zehaztuz aukeratu zen.

Esperimentuak Witten *et al.* (2011) liburuan aurkeztzen den Weka software ingurunean egikaritu ziren

ondoko lau oinarriko sailkatzaile erabiliz: John eta Langley (1995) lanean aurkeztutako Naive Bayes (NB), Aha *et al.* (1991) lanean aurkeztutako IBK (k bizilagun gertukoenak algoritmoaren inplementazioa), Platt (1999) lanean aurkeztutako SVM (euskarri bektoredun makina) eta Quinlan (1993) lanean aurkeztutako J48 (sailkatze zuhaitzak eraikitze C4.5 algoritmoaren inplementazio librea). Guztiak defektuzko parametroekin egikaritu zirelarik. Hauetaz gain, sailkatze zuhaitzetan (J48) oinarritutako bi meta-sailkatzaile ere eraiki ziren: Breiman (1996) lanean aurkeztutako bagging eta Freund eta Schapire (1996) lanean aurkeztutako boosting, non oinarri sailkatzailearen 25 iterazio egikaritu ziren. Sailkatzaile hauen igartze-gaitasuna neurtzeko bost iteraziodun balioztatze gurutzatua (5-fold cross-validation) metodoa erabili zen: datu-basearen %80 sailkatzailea entrenatzeko erabili zelarik eta %20 beratu testatzeko.

Esperimentazioaren ezaugarriak zehaztu ondoren, lehenik eta behin, sailkatzaile ezberdinetarako ezaugarri multzo ezberdinen zehaztasun-tasak (accuracy) eta F-puntuazioaren (F-measure) balioak ebaluatu ziren. F-puntuazioa doitasuna (precision) eta estaldura (recall) asmatze neurgailuetan oinarritzen da. Ondorengo zerrendan, erabilitako ezaugarri multzo ezberdinak zerrendatzen dira.

- **L1 ezaugarriak:** adituen arabera ezaugarri garrantzitsuenak ditu (1. taulan lehentasuna 1 dutenak).
- **L3 ezaugarriak:** erauzitako ezaugarri guztiak ditu (1. taulako guztiak: 1, 2 eta 3 lehentasuna dutenak).
- **CF ezaugarriak:** korrelazioan oinarrituriko ezaugarrien azpi-multzoa detektatzeko metodoaz aukeratuak ditu.
- **Wrapper J48:** wrapper algoritmoaz detektatutako ezaugarriak ditu. Wrapper-en barne eragiketarako J48 sailkatzailea eta bilaketa algoritmo genetikoak erabili dira.

Definitutako esperimentazio hauek, Perona *et al.* (2016) lanean argitaratu ziren 1. taulan definituriko 19 ezaugarrietarako. Bertan, hasiera batean, aipatu bost klaseekin egikaritu zen definituriko esperimentazioa, baina teklatuan oinarrituriko bi klaseak bereizteko zailtasunak zeudela ikusirik teklatuan oinarrituriko bi klaseak bildu ziren lau klaseko arazoa bilakaturik. Azken finean bietan gailu bera baita erabiltzen dena, teklatura alegia. Honela, emaitzarik onena ezaugarri guztiekin lortu zen eta 93.07 zehaztasun-tasa lortzera iritsi zen (ikus 3. taularen lehen errenkada).

Lan honen helburua emaitza horiek hobetzea da eta horretarako, alde batetik irisgarritasunean adituek diotena hartuko da kontuan, eta bestetik, ikasketa automatikoak ematen dituen aukeren bidez bilatuko da hobekuntza.

3.1 Adituen proposamenak

Arestian esan bezala, Perona *et al.* (2016) lanean lau klaserako gailu detekzio zehaztasun-tasa hobereana ezaugarri guztiak erabiliz lortu zen. Emaitza hauek 3. taulan 19 ezaugarri dagokion lerroan beha daitezke.

Teklatua erabiltzeko bi moduak bereiztean zeuden arazoak medio adituek bi proposamen egin zituzten. Bata, hiru ezaugarri berri proposatzea izan zen: JerkMediana, JerkMax eta GeldiKop. Ezaugarri hauek ondorengo zerrendan definitzen dira, bakoitzari adituek esleitu ziren lehentasuna parentesi artean adierazi delarik.

- **JerkMediana** (1): Orri batean kurtsoreak egiten dituen azelerazio aldaketen mediana.
- **JerkMax** (1): Orri batean kurtsoreak egiten dituen azelerazio aldaketen balio maximoa.
- **GeldiKop** (3): Orri batean kurtsorea zenbat aldiz geratu den geldirik.

Bestea, erabiltzaileak zerbait egiteko intentzioa erakusten duen nabigazio zatian zentratzea izan zen. Azken proposamen hau gauzatzeko normalean intentzio hau nabigazioaren azken zatian erakutsiko zela suposatu zen eta azterketa kurtsoreak eginiko bidearen azken %25ean zentratu zen.

3. taulan ikus daitezkeen moduan, lehen, 19 atributuekin, zehaztasun-tasa onena boosting kasuan lortzen zen, 93.07, eta orain, 22 atributuekin, ere bai, 93.13. Lehen bezala, zehaztasun-tasa onenak eskuarki sailkatzaile konplexuenekin edo meta-sailkatzaileekin lortu dira. Hiru ezaugarri berri ghitzearekin

3 Taula: Hautatutako sailkatzaileek lau klase bereiztean lortutako zehaztasun-tasak ezaugarrien multzo ezberdinetarako eta nabigazio zati ezberdinetarako. Sailkatzaile bakoitzeko balio onenak belztu dira.

Ezaugarri kopurua eta nabigazioaren zatia	Sailkatzaileak					
	NB	IBK	SVM	J48	Bagging	Boosting
19 ezau. - nab. %100	77.31	83.16	84.54	88.90	92.42	93.07
22 ezau. - nab. %100	80.90	83.10	85.79	89.31	91.76	93.13
22 ezau. - nab. %25	79.94	80.06	78.45	86.21	88.96	90.81

lortu den hobekuntza oso txikia da eta joera berdina ikusi da gainerako esperimentu-konfigurazioetan ere, hala nola, ezaugarriak automatikoki detektatzeko algoritmoen erabilerari dagokionez.

Bestetik, nabigazioaren azken zatitik ezaugarriak kalkulatzeko testuinguru honetan ez du hobekuntzarik ekarri. Adibidez, emaitzarik onena boosting-erako lortu da eta zehaztasun-tasa 90.81-ekoa izan da, aurreko konfigurazioetan baino baxuagoa. Hala ere azpimarratu behar da, nahiz eta moldaketa honek gailuak detektatzean hobekuntzarik ekarri ez, beste testuinguru batzuetarako egokiak izan daitezkeela: erabiltzaile mota ezberdinak definitzeko, zailtasunak eta arazoak detektatzeko etab.

3.2 Ikasketa automatikoko hobekuntzak

Egoera honetan, sistema praktikoki hobetze aldera akats ezberdinak egitearen larritasuna aztertu zen. Alegia, ez da gauza bera teklatura erabiltzen duen bat sagu erabiltzaile bezala sailkatzea edo trackball erabiltzaile bat sagu erabiltzaile gisan hartzea. Larritasun ezberdineko erroreak dira. Adibidez, teklatuaren kasuan kurtsorearen mugimendua gehiago kostatzen da eta alderantziz, joystic-trackball-sagu erabiltzaileetan kurtsorearen mugimendua biziagoa izan ohi da. Hau da, teklatura hatzez erabiltzen duen bati eginiko egokitzapenak edo buru-adarrez erabiltzen duen bati egiten zaizkionak tankerakoak izango dira, bestalde, joystic, trackball edo sagu erabiltzaileei ere egokitzapen antzekoak egingo zaizkie praktikan eragin antzekoa duten gailuak baitira. Beraz, bi klase handi edo bi meta-klase definitu ziren: teklaturakoa eta mugimenduzkoa.

Ikasketa automatikoaren ikuspuntutik, egoera honi sailkatzaileen egitura hierarkiko bat eraikiz aurre egin zitzaion, zehazki bi mailako sailkatzaileen egitura hierarkiko bat eraikiz. Lehenengo mailan bi meta-klaseak bereiziko ditu sistemak, arreta, eginiko errore larrien kopuruan jarriko delarik. Bi meta-klase hauen barnean portaera antzekoa duten gailuak bilduko dira arestian aipatu bezala. Bigarren mailan meta-klaseen barruko azpi-klaseak bereiziko dira. Maila honetan sailkatzaileek aurre egin beharreko arazoa, klase gutxiagokoa bilakatuko da, sinpleagoa beraz. Gainera, maila ezberdinetako sailkatzaile bakoitzak beharko lituzkeen ezaugarrien multzoa ezberdina izan daitekeenez, detektatzean arazoak dauden klaseak isolatuz berariazko soluzioak bila daitezke. Hau da, sailkatzaile espezializatuagoak eraiki daitezkeenez, klaseak hobeto bereizteko aukera dago. Gainera, maila honetan egiten diren akatsek ez dute lehenengo mailakoen larritasunik izango.

4 Taula: Hautatutako sailkatzaileek bi meta-klase bereiztean lortutako zehaztasun-tasak ezaugarrien multzo ezberdinetarako. Sailkatzaile bakoitzeko balio onenak belztu dira.

Ezaugarrien multzoa	Sailkatzaileak					
	NB	IBK	SVM	J48	Bagging	Boosting
L1 (8 ezau.)	93.31	96.12	94.81	97.73	98.03	98.21
L3 (19 ezau.)	90.33	96.06	96.30	98.57	99.04	99.22

Lehenik eta behin, ezaugarrien multzoak nola eragiten dion sistemaren jardunari erreparatu da 4. taulan. Ematen du adituek proposaturiko ezaugarrien multzoa (L1) erabiltzea aukerarik onena Naive Bayes eta IBk sailkatzaileen kasuan dela. Gainerako sailkatzaileek portaera hobea dute ezaugarrien multzo osoa erabiltzen denean.

4. taulan ikus daiteke sailkatzaileek bi meta-klaseak zehaztasun-tasa handiz bereiz ditzaketela. Kasu honetan ere, zehaztasun-tasa onena ezaugarri guztiak erabiltzean eta boosting kasuan lortu dira, 99.22-ko balioa hain zuzen ere. Beraz, ondoriozta daiteke sistemak errore larri gutxi egingo dituela. Erroreen jatorria aztertu asmoz sailkapen onenaren nahaste-matrizea aztertu da 5. taulan. Bertan ikus daiteke mugimenduzko gailuekin eginiko 7 orri-bisita eta teklaturakoa gailuekin eginiko 6 orri-bisita, oker sailkatu

5 Taula: Ezaugarri guztiekin osatutako datu-baseari boosting aplikatu ostean lorturiko nahaste-matrizea eta F-puntuazioa.

Honela sailkatuak \Rightarrow	a	b	F-puntuazioa
Mugimenduzkoa = a	983	7	0.993
Teklatua = b	6	679	0.991

6 Taula: Hautatutako sailkatzaileek mugimenduzko hiru klaseak bereiztean lortutako zehaztasun-tasak ezaugarrien multzo ezberdinetarako. Sailkatzaile bakoitzeko balio onenak belztu dira.

Ezaugarrien multzoa	Sailkatzaileak					
	NB	IBK	SVM	J48	Bagging	Boosting
L1 (8 ezau.)	78.38	79.70	75.76	82.53	86.16	86.16
L3 (19 ezau.)	73.33	80.10	80.10	84.24	87.17	89.09

zirela. Bi meta-klaseak gutxi-asko orekan daudenez, errore berdintsua dute, hau da, sailkatzaileek bi multzoak berdintsu bereizten dituzte.

Bigarren mailari dagokionez, kurtsoarea, gailuan nolabaiteko mugimendua eraginez mugitzen duten gailuetan zentratu da, joystick-a, trackball-a eta sagua hain zuzen ere. 6. taulan lorturiko zehaztasun-tasak ageri dira. Kasu honetan tasarik onenak, berriz ere, boosting-arekin lortu dira, 89.09. Kontuan izan behar da hemen egiten den errorearen larritasuna lehen mailan egiten denarena baino txikiagoa dela. Erroreen jatorria aztertu asmoz sailkapen onenaren nahaste-matrizea aztertu da 7. taulan. Bertan ikus daiteke, alboko F-puntuazioak iradokitzen duen moduan, proportzioan hutsegite gehien sagua eta trackball gailuak bereiztean dagoela. Hau, bi gailu hauek kurtsoarearen mugimendu murriztu-gabea baimentzen dutelako gerta daiteke, alegia, praktikan portaera antzeko xamarra dutelako. Aldiz, joystick-ean kurtsoarea aurre definituta dauden norantzetan soilik mugitzen da. Errore hau nahiko txikia eta larritasun gutxikoa denez, lan honetan, hierarkiari sagua eta trackball-a bereiziko zituen hirugarren maila bat ez zaio gehitu.

7 Taula: Ezaugarri guztiekin osatutako datu-baseari boosting aplikatu ostean lorturiko nahaste-matrizea eta F-puntuazioa.

Honela sailkatuak \Rightarrow	a	b	c	F-puntuazioa
Joystick = a	111	5	1	0.933
Sagua = b	7	38	2	0.817
Trackball = c	3	3	28	0.862

4 Ondorioak

Emaitzek, Remotest tresnak bildutako datuei datu meatzaritza prozesu osoa aplikatzea etorkizun handiko estrategia dela erakutsi dute. Web-irigarritasuneko adituek proposaturiko ezaugarriak erabiliz erabiltzaile-mota automatikoki detektatu eta beren-beregi eginiko egokitzapenak proposatzeko norantzan bidea irekitzen du. Baina erabiltzaile mota ezberdinak detektatzen hasi aurretik, lan honetan, zehazki, nabigatzeko erabilitako bost gailu mota ezberdintzen saiatu da: hatzez teklatua, buru-adarrez teklatua, joystick-a, trackball-a eta sagua.

Sailkatze arazo honi modu hierarkikoan egin zaio aurre. Lehen mailan teklatu bidezko gailuak (hatz eta buru-adarra erabiliz) eta mugimendu bidezko gailuak (joystick-a, trackball eta sagua) bereizi dira, hau da, bi klaseko arazoari egin zaio aurre. Lehen maila honetan, 99.22 zehaztasun-tasa lortzera iritsi da sistema, hau da, larritasun handiko erroreak nekez egingo du. Bigarren mailan mugimendu bidezko gailu ezberdinak bereizten saiatu da eta 89.09 zehaztasun-tasa lortu da, hierarkiaren lehen mailan baino zerbait baxuagoa baina errorearen larritasuna ere txikiagoa denez onargarria.

Etorkizun lan bezala, sistema honen online lotura garatu nahi da, sistema egiazko ingurune batean martxan jarri eta gailuak automatikoki detektatuz gailu bakoitzerako definitu diren egokitzapenak aplikatu nahi baitira.

Azkenik, lan honetan erabilitako metodologia gailuak detektatzeko egokia dela baieztatu denez, meto-

dologia bera jarraituz erabiltzaile-mota eta dituzten arazo ezberdinak automatikoki detektatzeko sistema garatzen saiatzea litzateke hurrengo urratsa.

Erreferentziak

- AHA, DAVID W., DENNIS KIBLER, eta MARC K. ALBERT. 1991. Instance-based learning algorithms. *Machine Learning* 6.37–66.
- BREIMAN, LEO. 1996. Bagging predictors. *Machine Learning* 24.123–140.
- DILLON, ANDREW. 2001. Beyond usability: process, outcome and affect in human computer interactions. *Canadian Journal of Information Science* 26.57–69.
- FREUND, YOAV, eta ROBERT E. SCHAPIRE, 1996. Experiments with a new boosting algorithm.
- GARCÍA, SALVADOR, JULIÁN LUENGO, eta FRANCISCO HERRERA. 2015. *Data Preprocessing in Data Mining*. Springer International Publishing Switzerland.
- HALL, MARK A., 1998. *Correlation-based Feature Subset Selection for Machine Learning*. Hamilton, New Zealand: University of Waikato tesia.
- JOHN, GEORGE H., eta PAT LANGLEY. 1995. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, UAI'95*, 338–345, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- KOHAVI, RON, eta GEORGE H. JOHN. 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97.273 – 324. Relevance.
- LIU, BING. 2006. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Springer-Verlag New York, Inc.
- PERONA, IÑIGO, AINHOA YERA, OLATZ ARBELAIZ, JAVIER MUGUERZA, NIKOLAOS RAGKOUSIS, MYRIAM ARRUE, J. EDUARDO PÉREZ, eta XABIER VALENCIA. 2016. Automatic device detection in web interaction. In *Actas de la XVII Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA) - VIII Simposio Teoría y Aplicaciones de Minería de Datos (TAMIDA)*, 825–834.
- PLATT, JOHN C. 1999. Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods: support vector learning*, ed. by Bernhard Schölkopf, Christopher J. C. Burges, eta Alexander J. Smola, chapter Fast Training of Support Vector Machines Using Sequential Minimal Optimization, 185–208. Cambridge, MA, USA: MIT Press.
- QUINLAN, J. ROSS. 1993. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- VALENCIA, XABIER, J. EDUARDO PÉREZ, UNAI MUÑOZ, MYRIAM ARRUE, eta JULIO ABASCAL. 2015. *Human-Computer Interaction – INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part I*, chapter Assisted Interaction Data Analysis of Web-Based User Studies, 1–19. Springer International Publishing.
- WITTEN, IAN H., EIBE FRANK, eta MARK A. HALL. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 3rd edition.

5 Eskerrak eta oharrak

Lan honek hainbat unitateren finantzaketa jaso du: Parte bat Euskal Herriko Unibertsitateko (UPV/EHU) Basque Advanced Informatics Laboratory (BAILab) ikerketa unitatearentzako funts orokorretik jaso du (erreferentzia UFI11/45). Parte bat Espainiako gobernuko ekonomia eta lehiakortasun ministerioak eta Europako eskualdeen garapenerako funtsak (ERFD) finantzatutako eGovernAbility proiektutik jaso du (erreferentzia TIN2014-52665-C2-1-R). Parte bat eusko jaurlaritzako hezkuntza saileko unibertsitateak eta ikerketa atalak sustengatzen duen ADIAN ikerketa taldetik jaso du (erreferentzia IT980-16). Bestetik, egileek narriadura fisikoa duten behar bereziko pertsonen beharrei erantzuten lan egiten duen Elkartu elkarteari zinez esker ona erakutsi nahi die, eta batik bat, estudioan parte hartu zuten parte-hartzaileei.