

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Timotej Knez

**Analiza vpliva težavnosti računalniške
igre na izmerjene vrednosti fizioloških
signalov**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM
PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Veljko Pejović

SOMENTOR: asistent Martin Gjoreski

Ljubljana, 2019

COPYRIGHT. Rezultati diplomske naloge so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavo in koriščenje rezultatov diplomske naloge je potrebno pisno privoljenje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Tematika naloge:

Cilj diplomske naloge je ugotoviti, ali obstaja povezava med fiziološkimi signali, ki jih lahko merimo s pomočjo široko dostopnih naprav mobilnega računanja, in težavnostjo igre, ki jo igralec v trenutku merjenja igra. Potrebno je implementirati celotno eksperimentalno infrastrukturo, vključno z mobilno igro, narediti eksperiment z dejanskimi uporabniki ter rezultate analizirati s pomočjo tehnik strojnega učenja.

Zahvaljujem se mentorju doc. dr. Veljku Pejoviću in somentorju Martinu Gjoreskemu za vso pomoč in številne nasvete. Zahvaljujem se tudi vsem, ki so sodelovali v raziskavi, ki je bila izvedena v okviru te diplomske naloge, posebej pa se zahvaljujem tudi družini za vso pomoč in podporo pri izvedbi naloge.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Pregled področja	3
3	Zasnova eksperimenta	5
3.1	Eksperimentalno okolje	5
3.2	Izbira igre	5
3.3	Postopek eksperimenta	6
3.4	Vprašalnik za ocenjevanje zahtevnosti	7
3.5	Osebnostni test	7
4	Izdelava programske opreme	9
4.1	Uporabljena tehnologija	9
4.2	Arhitektura Android aplikacije	11
5	Zbiranje podatkov	15
5.1	Predstavitev merjenih podatkov	15
5.2	Statistični podatki o udeležencih raziskave	16
6	Analiza podatkov	17
6.1	Uporabljene značilke	17

6.2	Uporabljene metode za analizo podatkov	21
6.3	Rezultati	23
7	Zaključek	31
	Literatura	33

Seznam uporabljenih kratic

kratica	angleško	slovensko
CA	Classification accuracy	Klasifikacijska točnost
AUC	Area under the curve	Površina pod krivuljo
RMSE	Root mean square error	Koren povprečja kvadratov napake
TLX	Task load index	Ocena obremenjenosti pri nalogi
SVM	Support vector machine	Metoda podpornih vektorjev
RF	Random forest	Naključni gozdovi
EMG	Electromyogram	Elektromiogram
EKG	Electrocardiogram	Elektrokardiogram
EEG	Electroencephalogram	Elektroencefalogram

Povzetek

Naslov: Analiza vpliva težavnosti računalniške igre na izmerjene vrednosti fizioloških signalov

Avtor: Timotej Knez

Cilj diplomske naloge je ugotoviti, ali je mogoče med igranjem igre sproti zaznavati miselni napor igralca, kar bi nam omogočilo dinamično prilagajanje težavnosti igre glede na potrebe posameznega igralca. Za določanje miselnega napora smo poskusili uporabiti fiziološke signale, ki jih lahko merijo pametne naprave. Ta ideja je danes še posebej aktualna, saj so naprave, ki omogočajo merjenje tovrstnih signalov, vse bolj pogoste. Da bi odkrili, ali je takšno zaznavanje napora mogoče, smo zasnovali eksperiment, s katerim smo preverili, ali obstaja povezava med izmerjenimi biološkimi signali ter težavnostjo igrane igre. Za potrebe izvedbe eksperimenta smo izdelali Android aplikacijo, ki uporabniku omogoča igranje igre Kača, hkrati pa s pomočjo pametne zapestnice meri fiziološke podatke. Te podatke smo kasneje obdelali s pomočjo tehnik strojnega učenja in izdelali več modelov za napovedovanje težavnosti igrane igre na podlagi izmerjenih signalov. Rezultate eksperimenta smo preverili z nekaj dodatnimi poskusi, s katerimi smo potrdili, da obstaja povezava med izmerjenimi fiziološkimi signali in težavnostjo igre.

Ključne besede: računalniške igre, dinamična težavnost, biološki signali, mobilno zaznavanje, strojno učenje, kognitivna obremenjenost.

Abstract

Title: Analyzing effect of computer game difficulty on biological signals

Author: Timotej Knez

The goal of this thesis is to determine whether it is possible to track a gamer's mental demand in real time. If possible, this would enable us to dynamically adjust the game's difficulty according to each player's unique needs. Biometric signals, measurable by smart devices, were used in order to determine mental demand. This is nowadays a popular approach as the devices able to measure such signals are getting more common. To determine whether measuring mental demand in such a way was possible an experiment was designed to verify the existence of a connection between the measured biological signals and the difficulty of the game. An Android application was designed for the needs of the experiment which allows the users to play the game Snake whilst measuring biometric data via a smart wristband. This data was later processed using machine learning algorithms to create various models of determining a game's difficulty based on the measured signals. The results of the experiment were verified with some additional experiments to confirm a link between the measured biometric signals and the difficulty of a game.

Keywords: computer game, dynamic difficulty, biological signals, mobile sensing, machine learning, cognitive load.

Poglavje 1

Uvod

Računalniške igre se v svetu razvijajo vedno hitreje. S tem postaja vse pomembnejša individualna prilagoditev uporabnikom za doseganje najboljše uporabniške izkušnje. Današnje računalniške igre običajno omogočajo prilagoditve, ki jih določa uporabnik sam, ki pa se med samo igro ne spreminjajo. V prihodnjem razvoju iger bo pomembno vlogo igralo prilagajanje igre uporabniku med samim igranjem. Izziv prihodnjega razvoja računalniških iger je tako iskanje parametrov igre, ki jih je smiselno prilagajati, kot ustreznega načina spremljanja odziva igralca. Posebno pomemben primer prilagoditev igre posamezniku je nastavljanje težavnosti igre. Večina iger ponuja nastavitve, s katerimi lahko igralec prilagodi težavnost igrane igre, kar pa bi bilo smiselno nadgraditi s sprotnim prilagajanjem težavnosti glede na trenutno stanje in zmogljivost igralca.

Za doseganje sprotnega prilagajanja se lahko uporablja podatke, ki jih pridobivamo neposredno iz spremljanja igranja, primer takšnih podatkov bi bilo število zbranih točk, doseženi nivo, uporabnikov reakcijski čas in hitrost pritiskanja gumbov, ali pa podatke, ki jih zbiramo s pomočjo dodatnih senzorjev. Naša raziskava predlaga uporabo podatkov o fizioloških signalih, pridobljenih z uporabo pametne zapestnice, ki jo nosi igralec med igro.

Ideja o zbiranju fizioloških signalov za ugotavljanje stresa, lahko tudi povezanega z miselnim naporom, ni nova. Znana je vrsta fizioloških signalov,

ki sovpadajo z nivojem stresa, kot so na primer telesna drža, smer pogleda, EMG (elektromiogram) obraznih in drugih mišic, EKG in EEG signali. Bernardi et al. [3] so prikazali povezavo med nivojem stresa in podatki, pridobljenimi iz različnih parametrov srčnega utripa. Podobno kot za srčni utrip je bila povezava prikazana tudi za druge signale, kot sta prevodnost in temperatura kože [13] [9].

Za meritev nekaterih izmed teh signalov je potrebna posebna merilna oprema, ki pa je lahko draga in za uporabnika moteča. Za praktično uporabo bi bilo koristno, če bi bilo mogoče podatke pridobivati s pomočjo enostavno dostopne večnamenske in uporabniku nemoteče opreme. Nekatero izmed fizioloških signalov je lažje meriti kot druge, zato se je za enostavno merjenje podatkov potrebno osredotočiti na lahko merljive signale, kot so srčni utrip ter prevodnost in temperatura kože. Pomembno je tudi, da so ti signali čim bolj enoznačni, ponovljivi ter da nanje vpliva čim manj stranskih dejavnikov.

Cilj tega projekta je ugotoviti, ali je mogoče uporabiti fiziološke signale, merjene s pomočjo pametne zapestnice, ki je sicer namenjena spremljanju telesne vadbe, kot dodatno orodje za določanje podatkov o miselnem naporu uporabnika, na podlagi katerih bi lahko prilagajali parametre igre. Za potrebe projekta smo uporabili pametno zapestnico Microsoft band 2, ki omogoča merjenje srčnega utripa, prevodnosti in temperature kože. Za izračun značilik, dobljenih s pametne zapestnice, je bil uporabljen program, razvit za potrebe magistrske naloge [4]. Za preverjanje naše hipoteze smo izdelali Android aplikacijo, ki omogoča igranje preproste igre na več težavnostnih stopnjah, med igranjem pa s pomočjo pametne zapestnice spremlja fiziološke podatke o igralcu. Za potrditev hipoteze je bilo potrebno izvesti meritve na čim večjem številu prostovoljcev ter pridobljene podatke analizirati.

Poglavje 2

Pregled področja

Da bi računalniške igrice naredili bolj zanimive uporabnikom, so v svetu razvili že veliko tehnik. Naš projekt se osredotoča na sprotno spreminjanje težavnosti za izboljšavo uporabniške izkušnje. Če je igra uporabniku prelahka, se bo začel dolgočasiti, če pa je pretežka, jo tudi to uporabniku naredi manj zabavno. Ker je optimalna težavnost odvisna od posameznega uporabnika in celo od posamezne situacije, je smiselno razmišljati o sprotne prilagajanju težavnosti igre. Xue et al. [15] so z uporabo dinamičnega prilagajanja težavnosti uspeli v praksi opazno izboljšati zanimivost igre, kar se je pokazalo v daljših časih in pogostejšem igranju igre. V tej raziskavi so za prilagajanje težavnosti uporabili podatke o uporabnikovem igranju. Idejo o dinamičnem prilagajanju igre lahko še razširimo tako, da za določanje zahtevnosti, ki jo občuti igralec, uporabimo dodatne podatke, pridobljene zunaj igre. V tem projektu uporabljamo zunanje fiziološke podatke, ki jih pridobimo z uporabo pametne zapestnice, ki jo nosi igralec.

V svojem članku so Gjoreski et al. [5] prikazali povezavo med fiziološkimi signali in miselnim naporom med reševanjem nalog. Za to so uporabili podatke o srčnem utripu (npr. frekvenco utripa, interval med utripi, značilnosti Poincarejevega grafa, spektralno moč vzorcev nizke in visoke frekvence). Analizirali so tudi podatke o prevodnosti in temperaturi kože. Povezavo med težavnostjo naloge in signali so preizkusili z uporabo modelov stroj-

nega učenja in tako pokazali, da se zahtevnost naloge odraža v izmerjenih fizioloških signalih.

Napovedovanje miselnega napora na podlagi fizioloških signalov so raziskovali tudi Haapalainen et al. [6], ki so poskušali ugotoviti, kateri izmed merjenih signalov največ prispevajo k uspešnemu napovedovanju miselnega napora. V testiranju so uporabili toplotno sevanje, EKG in EEG signala, srčni utrip, razširjenost zenic in prevodnost kože. Prišli so do ugotovitve, da k napovedovanju miselnega napora največ prispevajo podatki o toplotnem sevanju in EKG signalu.

Eksperiment z dinamičnim prilagajanjem težavnosti igre na podlagi fizioloških meritev so izvedli tudi Liu et al. [12] in dobili zelo dobre rezultate. Glavna razlika med njihovo in našo raziskavo je v tem, da so Liu et al. uporabili specializirane in drage naprave za merjenje fizioloških signalov, ki pa običajnim uporabnikom niso široko dostopne.

Signali, merjeni s pametno zapestnico, so bili za merjenje stresa uporabljeni tudi v magistrski nalogi [4], kar je relevantno tudi za našo nalogo, saj velik miselni napor na telo deluje tudi s stresom. Predvidevamo, da lahko s podobnimi metodami, ki so bile uspešno uporabljene za merjenje stresa, razberemo tudi nivo miselnega napora.

V raziskavi potrebujemo tudi način za merjenje subjektivne ocene napora ob reševanju naloge. Za to smo uporabili NASA-TLX test [8], ki nam pomaga s pomočjo šestih vprašanj oceniti nivo napora, ki ga je uporabnik občutil.

Poglavje 3

Zasnova eksperimenta

3.1 Eksperimentalno okolje

Za izvedbo eksperimenta smo uporabljali telefon Huawei P9, na katerem je tekel operacijski sistem Android 7.0. Telefon je bil prek 4G omrežja povezan na internet, da je lahko podatke sproti prenašal na oddaljen strežnik. Pametna zapestnica je bila na telefon povezana prek Bluetooth povezave.

3.2 Izbira igre

Za izvedbo eksperimenta smo potrebovali igro, ki ji je mogoče enostavno spreminjati težavnost in jo bodo igralci hitro razumeli ter znali igrati. Odločili smo se, da bomo uporabili igro Kača (glej sliko 4.3b), ki ustreza vsem pogojem, poleg tega pa jo je tudi dovolj enostavno implementirati. Igra poteka tako, da igralec upravlja kačo, ki se premika po zaslonu, ter na poti pobira “sadje”, s čimer nabira točke. Pri tem se kačin rep podaljšuje, kar vse bolj otežuje igro. Igralčev cilj je dobiti čim več točk s pobiranjem “sadja”, ne da bi se pri tem kača zaletela v steno ali v svoj rep. Če se kača zaleti v steno ali svoj rep, se njena dolžina zopet zmanjša na začetno, igra pa se nadaljuje do konca trajanja faze eksperimenta. Enostavna implementacija igre nam je omogočila, da smo lahko za potrebe raziskave izdelali svojo verzijo Kače, ki

smo ji lahko enostavno prilagajali težavnost, kar pa smo dosegli s pomočjo spreminjanja hitrosti premikanja kače.

3.3 Postopek eksperimenta

Eksperiment smo začeli tako, da smo prostovoljcu nadelo zapestnico, ki je merila njegove fiziološke signale, in mu razložili potek igranja igre. Nato je imel prostovoljec na voljo nekaj časa, da je igral igro v preizkusnem načinu, kjer se meritve niso beležile. Na ta način smo poskrbeli, da so prostovoljci igranje začeli z dovolj dobrim poznavanjem igre. Na ta način smo tudi rešili težavo, pri kateri je prvih nekaj meritev, ki jih izmeri zapestnica, manj natančnih, saj je bila večina nenatančnih meritev pridobljena med poskusno fazo in se je posledično zavrgla. Do manj natančnih meritev pride zato, ker zapestnica potrebuje nekaj časa, da iz signala razbere uporabnikov srčni utrip.

Temu je sledilo igranje igre, ki je potekalo na treh različnih težavnostnih stopnjah, kjer smo za mero težavnosti uporabili hitrost premikanja kače po igralnem polju. Za vsako stopnjo težavnosti je bila pri vseh osebah uporabljena enotna hitrost kače, kar je določalo dejansko težavnost igre. Vsaka stopnja igre je trajala vsaj dve minuti, potem pa se je igra prekinila, ko se je igralec naslednjič zaletel. V primeru, da se igralec do takrat še ni zaletel, se je stopnja v vsakem primeru zaključila po treh minutah igranja. Po vsaki težavnostni stopnji je prostovoljec rešil kratek vprašalnik, zasnovan za ocenjevanje občutene težavnosti igre. Tri težavnostne stopnje so si pri vsakem od prostovoljcev sledile v vrstnem redu od najlažje do najtežje ali od najtežje do najlažje. Po katerem od teh dveh vrstnih redov bo igral posamezni igralec, je bilo določeno naključno z verjetnostjo 50%. S tem smo poskusili odpraviti vpliv spreminjanja izmerjenih biometričnih podatkov zaradi časa nošenja zapestnice.

Na koncu eksperimenta so prostovoljci rešili tudi HEXACO [2] osebni test, ki ga je mogoče uporabiti za boljše razumevanje odgovorov na vprašalnike o težavnosti.

3.4 Vprašalnik za ocenjevanje zahtevnosti

Vprašalnik za ocenjevanje občutene zahtevnosti igrane igre je bil sestavljen iz osmih vprašanj. Prvi dve vprašanji sta obravnavali zahtevnost in zabavnost igre. Vprašanji sta bili dodani z namenom, da bi morda iz teh vprašanj najlažje pridobili občuteno težavnost.

Za naslednjih šest vprašanj smo uporabili Nasin TLX vprašalnik [8], ki je narejen za ocenjevanje zahtevnosti naloge. Sestavljen je iz vprašanj o naslednjih temah:

- Miselni napor
- Fizični napor
- Časovni napor
- Kvaliteta izvedbe
- Vloženi trud
- Frustracija

Na vsakega izmed vprašanj so uporabniki odgovorili z ustrezno stopnjo na Likertovi lestvici od zelo malo do zelo veliko.

Za končni izračun vrednosti testa smo uporabili surovi TLX, ki je opisan v članku [7]. To pomeni, da smo za skupno NASA-TLX oceno uporabili povprečje izbranih vrednosti kategorij težavnosti.

3.5 Osebnostni test

Za test osebnosti smo uporabili Hexaco test s 60 vprašanji [1]. Test meri človeško osebnost na šestih glavnih oseh in sicer iskrenost/ponižnost (npr. "Nikoli ne bi sprejel podkupnine, tudi če bi bila zelo visoka."), emocionalnost (npr. "Strah bi me bilo potovati v slabih vremenskih razmerah."), ekstravertnost (npr. "Raje imam dela, ki vključujejo aktivno interakcijo z drugimi

ljudmi, kot samostojno delo.”), sprejemljivost (npr. “Ko se ljudje z mano ne strinjajo, sem ponavadi kar fleksibilen glede svojih mnenj.”), vestnost (npr. “Pri delu se vedno trudim biti natančen, tudi če zaradi tega porabim več časa.”) in odprtost izkušnjam (npr. “Skrbi me mnogo manj kot druge ljudi.”)¹. Za vsako dimenzijo osebnosti test vsebuje deset vprašanj. Prosto-voljci so za vsako izmed vprašanj izbrali eno od možnosti na petstopenjski Likertovi lestvici od tega, da se močno ne strinjajo, do tega, da se močno strinjajo z dano trditvijo. Osebnostni test smo poskusili uporabiti tako, da smo njegove rezultate dodali kot vhode napovednim modelom, ki so napovedovali NASA-TLX oceno težavnosti igre. Na ta način bi lahko rezultati testa pomagali pri napovedovanju obsega ocen, ki jih je podal uporabnik. Ugotovili smo, da rezultati testa niso prispevali k izboljšanju rezultatov napovedi, zato smo jih pri nadaljnji analizi izpustili.

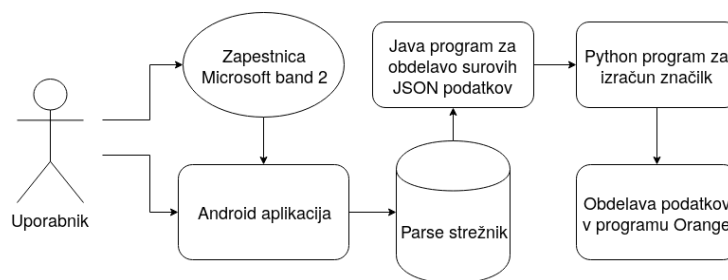
¹Prevod osebnostnih osi je povzet po magistrski nalogi [11]

Poglavje 4

Izdelava programske opreme

4.1 Uporabljena tehnologija

Za namene študije napovedovanja težavnosti igre je bil zasnovan sistem (glej sliko 4.1) z Android aplikacijo za povezavo in zbiranje podatkov s pametne zapestnice, Parse strežnika za shranjevanje podatkov, programa za pretvorbo podatkov v prijaznejši format, Python program za izračun značilik iz zbranih podatkov in Orange delovnega toka za učenje in testiranje različnih klasifikatorjev.



Slika 4.1: Diagram sistema

4.1.1 Pametna zapestnica

Za merjenje biometričnih podatkov smo uporabili komercialno dostopno pametno zapestnico Microsoft band 2. Zapestnica ima med drugimi tudi senzorje za merjenje srčnega utripa, prevodnosti kože in temperature kože, ki smo jih uporabili pri naši študiji. Omogoča enostavno in zanesljivo branje ter prenašanje podatkov s senzorjev v realnem času. Več o zajemanju podatkov je opisano v poglavju 5.1.

4.1.2 Android aplikacija

Zapestnica se prek Bluetooth povezave poveže na Android telefon, zato smo se odločili, da bo celoten eksperiment potekal s pomočjo Android aplikacije. Naloga le-te je bila, da uporabniku omogoča igranje preproste igre, medtem pa z zapestnice sproti bere podatke o uporabnikovih fizioloških signalih. Te signale tudi shranjuje na oddaljen strežnik, od koder smo jih lahko kasneje prenesli za obdelavo.

4.1.3 Oddaljen strežnik

Za shranjevanje zbranih podatkov smo uporabili strežnik z odprtokodno programsko opremo Parse, ki ga ponuja spletna storitev Back4app. Parse strežnik ponuja relacijsko podatkovno bazo, v katero smo lahko izmerjene podatke shranjevali v realnem času. Strežnik omogoča tudi preprosto integracijo z Android aplikacijo s pomočjo Android knjižnice.

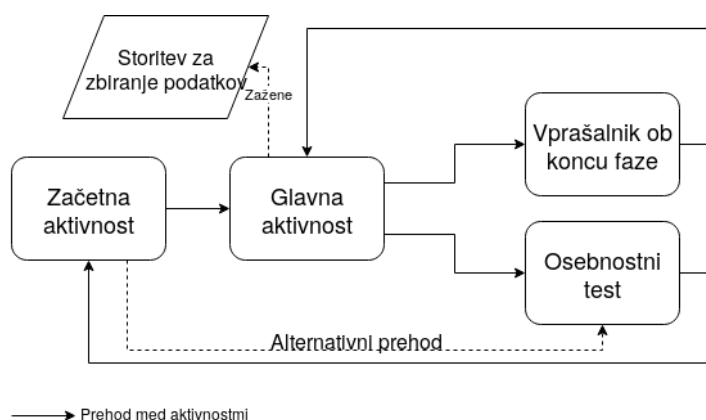
4.1.4 Python program za izračun značilk

Zbrane podatke smo najprej obdelali z uporabo Python programa, ki smo ga izdelali na podlagi programa, ki nam ga je ponudil avtor naloge [4]. S programom smo iz surovih podatkov izvlekli značilke, na podlagi katerih smo izvajali nadaljnjo analizo.

4.1.5 Orange

Za analizo značilik smo uporabili program Orange, ki je namenjen podatkovnemu rudarjenju. S pomočjo tega programa smo izdelali napovedne modele na podlagi različnih algoritmov strojnega učenja. S pomočjo napovednih modelov smo za poljubno težavnostno stopnjo napovedali težavnost in s pomočjo klasifikacijske točnosti ovrednotili povezavo med izmerjenimi podatki ter težavnostjo.

4.2 Arhitektura Android aplikacije



Slika 4.2: Diagram Android aplikacije

Android aplikacija vsebuje štiri aktivnosti (angl. Activity – Javanski razred) in eno storitev (angl. Service – Javanski razred), kot je prikazano na diagramu 4.2). Prva aktivnost se prikaže, ko uporabnik odpre aplikacijo, in vsebuje podatke o stanju povezave z zapestnico ter gumba za začetek eksperimenta in začetek osebnostnega testa (glej posnetek zaslona 4.3a). Na ta način je mogoče osebnostni test rešiti neodvisno od preostanka eksperimenta, kar je bilo koristno pri vzporednem izvajanju eksperimenta, kjer je prvi prostovoljec lahko reševal osebnostni test na enem telefonu, medtem ko je drugi igral igro na drugem. Ob pritisku na katerega koli od gumbov se uporabniku

prikaže okno, kamor mora vpisati svoj enolični identifikator, pod katerim se nato shranjujejo podatki.

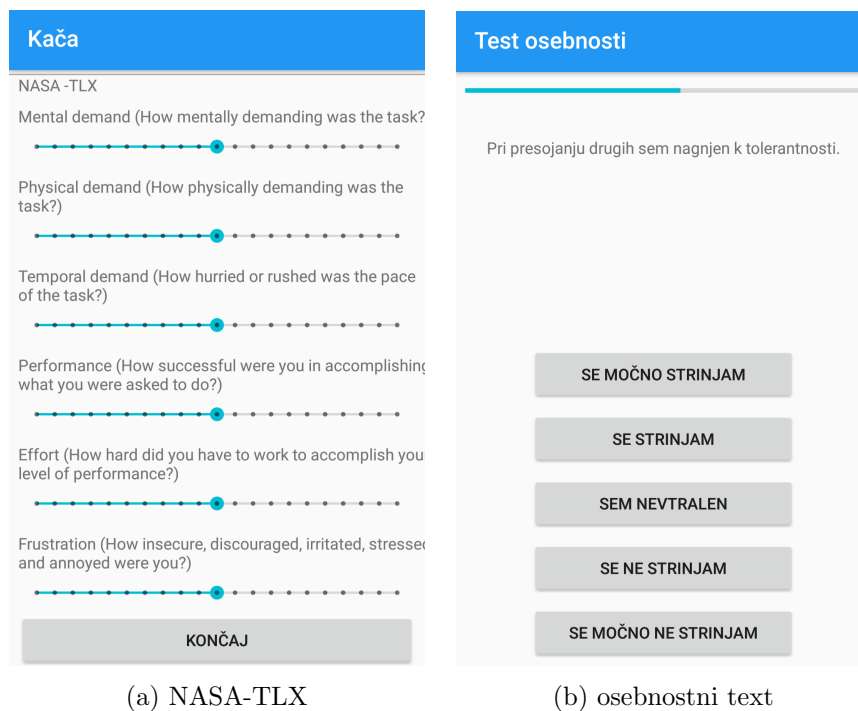


Slika 4.3: Začetna in glavna aktivnost v android aplikaciji

Druga aktivnost v aplikaciji omogoča igranje igre in koordinira vse stopnje eksperimenta (glej sliko 4.3b). Ta aktivnost tudi zažene ločeno nit, ki je odgovorna za premikanje kače glede na izbrano hitrost in glede na uporabnikove ukaze.

Poleg teh dveh aktivnosti aplikacija vsebuje tudi aktivnost, ki je odgovorna za prikaz vprašalnika o težavnosti igrane stopnje ter aktivnost za prikaz osebnostnega testa (glej sliki 4.4a in 4.4b). Na vprašanja o težavnosti uporabnik odgovarja s pomočjo drsnikov, s katerimi lahko izbere, kako močno se strinja s posamezno trditvijo, osebnostni test pa prikazuje po eno vprašanje naenkrat in uporabniku ponudi gumbe, s katerimi lahko izbere ustrezni odgovor na petstopenjski Likertovi lestvici od se močno ne strinjam do se močno

strinjam.



Slika 4.4: Zaslonski posnetek aktivnosti, ki sta odgovorni za prikaz obeh vprašalnikov.

Vzporedno s temi aktivnostmi v ozadju teče storitev za zbiranje podatkov. Ta storitev sprejema obvestila o spremembah biometričnih podatkov, ki jih pošilja zapestnica. Periodično se proži metoda, ki zbrane podatke posreduje na Parse strežnik. Storitev prav tako spremlja, koliko časa igralec že igra igro na posamezni stopnji eksperimenta in po potrebi sproži obvestilo za glavno aktivnost, da je potrebno igro prekiniti, saj je čas že potekel.

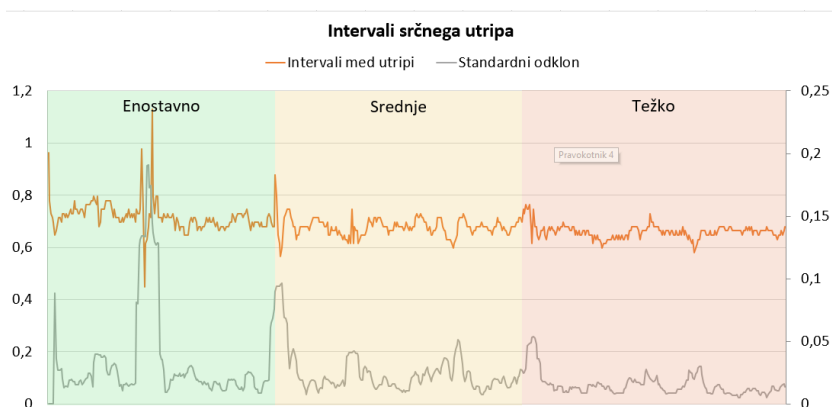
Poglavje 5

Zbiranje podatkov

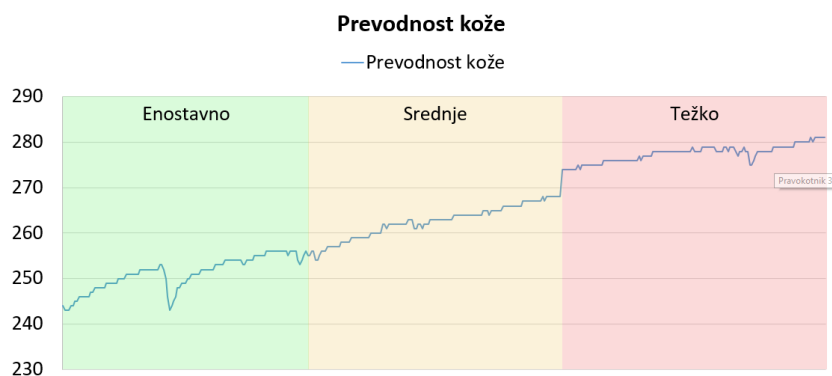
5.1 Predstavitev merjenih podatkov

Med igranjem igre smo s pomočjo pametne zapestnice merili več različnih fizioloških signalov. Prevodnost in temperaturo kože smo shranjevali enkrat na sekundo, pri čemer se je vrednost temperature zaradi omejitev senzorja posodobila samo vsakih deset sekund. Poleg tega smo ob vsakem srčnem utripu shranili tudi časovni interval od prejšnjega do trenutnega utripa. Spremljali smo tudi nekatere dodatne podatke o igrani igri. Vsakih pet sekund se je zabeležil podatek o povprečnem številu dotikov ekrana na sekundo, vsako sekundo pa se je zabeležil tudi podatek o trenutnem številu točk, ki jih je dosegel igralec. Ob vsaki meritvi so se zabeležili podatki o zaporedni številki stopnje, nastavljeni težavnosti igre za trenutno stopnjo in igralčevem identifikatorju.

Na sliki 5.1a lahko vidimo intervale med posameznimi srčnimi utripi v različnih fazah eksperimenta pri enem od uporabnikov. Prikazan je tudi standardni odklon intervalov skozi čas. Poleg tega lahko na sliki 5.1b vidimo, kako se skozi igro spreminja prevodnost kože uporabnika.



(a) Graf intervalov



(b) Graf prevodnosti

Slika 5.1: Grafa prevodnosti kože in intervalov med srčnimi utripi

5.2 Statistični podatki o udeležencih raziskave

V raziskavi je sodelovalo 22 prostovoljcev, med katerimi je izmerjena prevodnost kože enega izmed njih od skupnega povprečja odstopala za več kot 5 standardnih deviacij. Zaradi tega velikega odstopanja smo se odločili, da smo meritve te osebe izključili iz raziskave. Prostovoljci so bili stari med 11 in 50 let, kjer je bila večina stara 21 let, povprečna starost pa je bila 24 let. Med prostovoljci je bilo 14 moških in 8 žensk.

Poglavje 6

Analiza podatkov

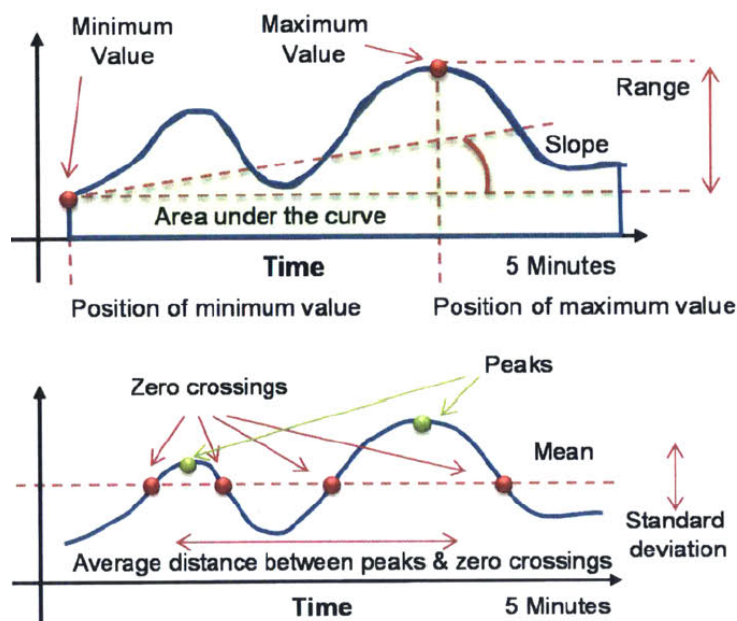
Ko smo zbrali vse potrebne podatke, jih je bilo potrebno analizirati in s tem preveriti hipotezo, da obstaja povezava med zbranimi fiziološkimi podatki in zahtevnostjo igrane igre.

6.1 Uporabljene značilke

Prvi korak pri obdelavi podatkov je izpeljava različnih značilk, osnovanih na zbranih podatkih. Pri tem smo si pomagali z delom, ki je bilo narejeno že v magistrski nalogi [4], saj smo na razpolago dobili program za ekstrakcijo značilk, ki je bil tam uporabljen. Na sliki 6.1 so predstavljene značilne točke in parametri fiziološkega signala (slika je pridobljena iz doktorske naloge [10]).

6.1.1 Značilke, osnovane na srčnem utripu

Pričakujemo, da se bosta stres in miselni napor v srčnem utripu odražala tako, da se ob prisotnosti miselne obremenitve srčni utrip zviša, posledično se povprečen interval med utripi zmanjša, zmanjša pa se tudi variabilnost intervalov med udarci srca [3, 14]. Zato so najpomembnejši podatki, ki jih želimo z izpeljanimi značilkami zajeti, frekvenca srčnih utripov in variabilnost intervalov med posameznimi utripi.



Slika 6.1: Diagram značilnih točk in parametrov fiziološkega signala iz doktorske naloge [10]

- Povprečni srčni utrip (mean_hr)
- Povprečen čas med dvema zaporednima utripoma (ibi)
- Standardna deviacija časov med zaporednimi utripi (sdnn)
- Koren povprečja kvadratov razlik med sosednimi intervali (rmsd)
- Odstotek razlik med zaporednimi intervali, ki so večji od 20 milisekund (pnn20)
- Odstotek razlik med zaporednimi intervali, ki so večji od 50 milisekund (pnn50)
- SD1 indeks Poincarjevega grafa (sd)

6.1.2 Značilke, osnovane na prevodnosti kože

Nourbakhsh et al. [13] so pokazali, da se stres in miselni napor odražata tudi v prevodnosti kože, zato smo uporabili tudi več značilk, izpeljanih iz le-te.

- Povprečje prevodnosti kože (meanGsr)
- Standardna deviacija prevodnosti kože (stdGsr)
- 25. centil prevodnosti kože (q25Gsr)
- 75. centil prevodnosti kože (q75Gsr)
- Razlika med 1. in 3. kvartilom (qdGsr)
- Vsota odvoda signala prevodnosti kože (derivGsr)
- Povprečna moč signala, ki je izračunana kot povprečje kvadratov prevodnosti (powerGsr)
- Povprečno število vrhov na sekundo (rate_peaks)
- Vsota pozitivnih vrednosti v odvodu signala deljeno s številom meritev (sum_pos_deriv)
- Delež pozitivnih vrednosti v odvodu signala (prop_pos_deriv)
- Vsota odvoda osnove signala (deriv_tonic)
- Povprečna razlika med signalom in osnovo signala (sig_tonic_difference)
- Moč signala v frekvencah 0.0-0.1 Hz (fp01)
- Moč signala v frekvencah 0.1-0.2 Hz (fp02)
- Moč signala v frekvencah 0.2-0.3 Hz (fp03)
- Moč signala v frekvencah 0.3-0.4 Hz (fp04)
- Moč signala v frekvencah 0.4-0.5 Hz (fp05)

- Sprememba amplitude signala pred največjim vrhom
(max_peak_amplitude_change_before)
- Sprememba amplitude signala po največjem vrhu
(max_peak_amplitude_change_after)
- Razmerje med spremembo signala pred in po največjem vrhu
(max_peak_change_ratio)
- Povprečna sprememba amplitude signala pred vrhovi
(avg_peak_amplitude_change_before)
- Povprečna sprememba amplitude signala po vrhovih
(avg_peak_amplitude_change_after)
- Razmerje med povprečno spremembo pred in po vrhovih
(avg_peak_change_ratio)
- Najdaljši čas dvigovanja signala pred vrhom
(max_peak_increase_time)
- Najdaljši čas spuščanja signala po vrhu
(max_peak_decrease_time)
- Najdaljše trajanje vrha (max_peak_duration)
- Razmerje med največjim trajanjem vzpona in največjim trajanjem padca
(max_peak_change_time_ratio)
- Povprečni čas dvigovanja signala pred vrhom (avg_peak_increase_time)
- Povprečni čas spuščanja signala po vrhu (avg_peak_decrease_time)
- Povprečno trajanje vrha (avg_peak_duration)
- Razmerje med povprečnim trajanjem vzpona in povprečnim trajanjem padca (avg_peak_change_time_ratio)

- Največja strmina signala pred vrhom (`max_peak_response_slope_before`)
- Največja strmina signala po vrhu (`max_peak_response_slope_after`)
- Razlika med maksimumom in minimumom signala (`signal_overall_change`)
- Razmerje med trajanjem največje spremembe in velikostjo spremembe. (`change_rate`)

6.1.3 Značilke, osnovane na temperaturi kože

Prikazano je bilo že, da se stres in miselni napor odražata tudi v temperaturi kože. Primer tega so pokazali Herborn et al.[9], ki so raziskovali vpliv stresa na temperaturo kože pri toplokrvnih živalih. Zato smo izpeljali tudi nekaj osnovnih značilk iz temperature kože.

- Povprečna vrednost temperature kože (`st_mean`)
- Standardni odklon temperature kože (`st_std`)
- Razlika med kvartili temperature kože (`st_qd`)
- Povprečna hitrost spreminjanja temperature kože (`st_diff`)

6.2 Uporabljene metode za analizo podatkov

Analiza podatkov je potekala tako, da smo najprej iz surovih podatkov izvlekli zgoraj naštetе značilke in tako za vsako fazo vsakega eksperimenta pridobili vrstico, ki je vsebovala vrednosti vsake od značilk, uporabnikove odgovore na vprašalnik ob zaključku te faze, dejansko težavnost faze igre in demografske podatke o uporabniku. Te podatke smo uvozili v program Orange, kjer smo jih lahko obdelali z različnimi tehnikami strojnega učenja.

Izkazalo se je, da dobimo mnogo boljše rezultate, če značilke pred analizo normiramo tako, da za vsakega uporabnika posebej izračunamo povprečno vrednost vsake od značilk za vse težavnostne stopnje in vrednosti značilk

delimo z izračunanimi povprečji. Na ta način nam vrednosti značilik povedo, za koliko odstopajo od uporabnikovega povprečja, in ne njihove absolutne vrednosti. To omogoča lažjo primerjavo rezultatov med osebami, saj se absolutne vrednosti med uporabniki močno razlikujejo. To pomeni, da bi bila za praktično uporabo postopka za določanje občutene težavnosti v realnem času potrebna kalibracijska doba, v kateri bi izračunali povprečne vrednosti merjenih značilik za specifičnega uporabnika.

Za analizo podatkov smo izvedli prečno preverjanje, kjer smo podatke ločili po uporabnikih in tako dobili tako imenovano *leave one person out* preverjanje. To pomeni, da so bili za vsak napovedovalni algoritem parametri nastavljeni na podlagi podatkov vseh razen enega uporabnika in testirani na izločenem uporabniku. Postopek preverjanja se nato ponavlja za podatke vsakega od uporabnikov, dokler ni bil algoritem testiran na vseh podatkih. Za mero uspešnosti smo uporabljali klasifikacijsko točnost, ki nam pove, kolikšen delež primerov je algoritem napovedal pravilno.

6.2.1 Uporabljeni algoritmi za klasifikacijo

Za napovedovanje težavnosti igre smo uporabili več različnih klasifikatorjev, ki jih ponuja program Orange.

Kot ciljno spremenljivko smo uporabili ciljno težavnost igre, ki je diskretna spremenljivka z vrednostmi lahka (0), srednja (1) in težka (2). Kot mero uspešnosti smo uporabili klasifikacijsko točnost. V nekaterih poskusih smo napovedovali tudi izračunano NASA-TLX oceno težavnosti naloge. Ta spremenljivka je zvezna, zato smo v tem primeru za mero uspešnosti uporabili koren povprečja kvadratov napak (RMSE). Pri tem so napovedni modeli dajali zelo slabe rezultate, kar je verjetno posledica tega, da so različni uporabniki zahtevnost igre ocenjevali zelo različno. Na podlagi te ugotovitve smo pri tovrstnih testih uporabniške ocene težavnosti premaknili tako, da imajo pri posameznem uporabniku povprečje 0.

Konstantni klasifikator smo uporabili kot kontrolni algoritem, da smo

lahko ocenili, kako dobri so rezultati drugih algoritmov. Ta algoritem deluje tako, da vedno napove razred, ki je bil v učnih podatkih najpogostejši.

AdaBoost algoritem deluje tako, da združi množico drevesnih klasifikatorjev v en boljši klasifikator, pri čemer prilagaja uteži, s katerimi upošteva posamezno drevo.

Random forest algoritem deluje tako, da zgradi 100 naključnih odločitvenih dreves in za vsako klasifikacijo napove razred, ki ga je vrnilo največje število dreves.

SVM (metoda podpornih vektorjev) poskusi razrede razdeliti tako, da so razdalje od meje do elementov, ki se nahajajo na pravi strani meje, čim večje, do elementov, ki se nahajajo na napačni strani meje, pa čim manjše.

Logistična regresija oceni verjetnost za vsakega od razredov na podlagi dobljenih vhodov, pri čemer poskuša minimizirati verjetnosti za napačne klasifikacije.

6.3 Rezultati

6.3.1 Skupni model napovedovanja objektivne težavnosti igre

Podatke za analizo smo pripravili tako, da smo po postopku, opisanem v poglavju 6.2, iz surovih podatkov izvlekli značilke. Napovedovalne modele smo nato testirali na rezultatih enega udeleženca naenkrat in učili na rezultatih preostalih udeležencev. Rezultati so predstavljeni v tabeli 6.1. Pri tem smo dobili najboljšo klasifikacijsko točnost z uporabo Random Forest algoritma, ki je pravilno napovedal 64% meritev.

	Večinski klas.	Random Forest	SVM	AdaBoost	Log. reg.
CA	0.333	0.642	0.494	0.519	0.506
AUC	0.500	0.815	0.636	0.639	0.660

Tabela 6.1: Rezultati testiranja napovednih modelov

Glede na to, da nam Random forest klasifikator daje najboljše rezultate, je smiselno, da si ga podrobneje ogledamo. Če pogledamo matriko napovedi, lahko vidimo, da se srednja težavnost klasificira zelo dobro, med tem ko se pri ostalih dveh težavnostih bolj pogosto zgodi, da jih klasifikator zameša.

		Napovedana vrednost			
		Lahka	Srednja	Težka	Σ
Dejanska vrednost	Lahka	18	2	7	27
	Srednja	1	23	3	27
	Težka	11	5	11	27
Σ		30	30	21	81

Slika 6.2: Matrika napovedi za Random forest algoritem

Kar nekaj izmed uporabljenih značilk ima težavo, da se močno spreminjajo v odvisnosti od časa nošenja zapestnice, kar bi lahko povzročilo, da bi se klasifikator naučil ločevati med fazami testa na osnovi časa nošenja zapestnice namesto na osnovi dejanske težavnosti naloge. Primer počasnega spreminjanja prevodnosti kože skozi čas je mogoče videti na sliki 5.1b. Ta učinek smo seveda poskusili odstraniti tako, da je polovica udeležencev različne težavnosti reševala v obratnem vrstnem redu, vendar pa se nam je ravno zaradi rezultata, da klasifikator večkrat zameša skrajni težavnostni stopnji, pojavil sum, da je težava kljub temu ukrepu še vedno prisotna.

Ta sum smo še dodatno potrdili s tem, da smo uporabljene značilke ran-

	Večinski klas.	Random Forest	SVM	AdaBoost	Log. reg.
CA	0.593	0.630	0.667	0.667	0.630
AUC	0.500	0.600	0.625	0.625	0.600

Tabela 6.2: Rezultati napovedovanja težavnosti prve faze testiranja

girali na podlagi *gain ratio*. S tem smo ugotovili, da vse najpomembnejše značilke temeljijo na prevodnosti ter temperaturi kože in so zato vsekakor odvisne od časa nošenja zapestnice.

Da bi to neželeno odvisnost odstranili, smo analizo podatkov ponovili tako, da smo pri vsakem izmed testiranj enako kot prej podatke normirali na podlagi vseh treh faz testiranja, pri klasifikaciji pa smo uporabili samo prvo fazo testiranja. Na ta način smo izločili vse učinke, ki jih ima čas nošenja zapestnice na rezultate. V tem primeru sta bila ciljna razreda samo “lahki” in “težki” in sicer je težkemu razredu pripadalo 59% podatkov.

V tem primeru vidimo, da je najboljši klasifikacijski algoritem metoda podpornih vektorjev (SVM), ki je dosegel 67% klasifikacijsko točnost (glej tabelo 6.2). Ta vrednost sicer ni veliko višja od izhodiščne vrednosti, ki znaša 59%, vendar pa kljub temu nakazuje na to, da obstaja določena povezava med izmerjenimi podatki ter težavnostjo igrane igre.

6.3.2 Skupni model napovedovanja subjektivnega občutka težavnosti igre

Analizo smo ponovili tako, da smo kot ciljno spremenljivko uporabili NASA-TLX oceno zahtevnosti naloge. Zaradi različnih obsegov ocen težavnosti, ki so jih podali uporabniki, so bile vrednosti napak pri napovedih enakovredne kontrolni napaki. Od ocen smo zato odšteli povprečno vrednost vseh ocen, ki jih je podal uporabnik (glej tabelo 6.3). Spremenljivko smo nato še diskretizirali v dva razreda, da lahko ocenjujemo klasifikacijsko točnost tako kot v prejšnjih primerih (glej tabelo 6.4). Tako je ocenjeni razred pri 92% primerov

	Kontrolni klas.	Random Forest	SVM	AdaBoost	Lin. reg.
RMSE	2.772	2.688	2.701	2.739	2.570

Tabela 6.3: Rezultati napovedovanja zvezne NASA-TLX ocene prve faze testiranja

	Večinski klas.	Random Forest	SVM	AdaBoost	Log. reg.
CA	0.519	0.593	0.556	0.593	0.630
AUC	0.500	0.688	0.650	0.600	0.850

Tabela 6.4: Rezultati napovedovanja diskretizirane NASA-TLX ocene prve faze testiranja

ustrezal predvideni težavnosti. Analizo smo nad temi razredi ponovili na isti način kot prej, s čimer smo dobili tudi podobne rezultate.

6.3.3 Personalizirani model, ki upošteva časovno odvisne značilke

Za preverjanje hipoteze smo izvedli še en eksperiment, pri katerem je prostovoljec najprej nosil zapestnico več kot eno uro, da sta se temperatura in prevodnost kože ustalili, nato pa je izvedel šest testiranj, med katerimi smo imeli daljše premore. Na ta način verjamemo, da so bili odstranjeni vsi učinki, ki sledijo iz časa nošenja zapestnice. Te meritve smo nato obdelali po istem postopku in pridobili rezultate, zapisane v tabeli 6.5.

V tem primeru sta klasifikatorja Random forest in SVM dosegla zelo podobne rezultate, ki pa niso neposredno primerljivi s prejšnjimi, saj gre tu za personaliziran model. Klasifikacijska točnost 55% je znatno večja od kontrolne klasifikacijske točnosti, ki znaša 33%. Ta rezultat nam še dodatno potrjuje povezavo med izmerjenimi signali in težavnostjo igre.

	Večinski klas.	Random Forest	SVM	AdaBoost	Log. reg.
CA	0.333	0.556	0.556	0.296	0.463
AUC	0.500	0.725	0.741	0.472	0.620

Tabela 6.5: Rezultati testiranja personaliziranih napovednih modelov

	Večinski klas.	Random Forest	SVM	AdaBoost	Log. reg.
CA	0.333	0.457	0.469	0.370	0.333
AUC	0.500	0.543	0.537	0.528	0.519

Tabela 6.6: Rezultati testiranja napovednih modelov na podlagi srčnega utripa

Poskusili smo tudi obdelati pridobljene podatke na tak način, da smo uporabili samo značilke, ki so povezane s srčnim utripom, saj te niso močno odvisne od časa nošenja zapestnice. Pri tem so bili rezultati kot po pričakovanjih precej slabši, saj je bilo število uporabljenih značilk precej manjše. V tem primeru sta nam najboljše rezultate dala algoritma Random forest in SVM (glej tabelo 6.6).

6.3.4 Analiza značilk

Ocena prispevka značilk za napovedovanje razreda

Eno od vprašanj, ki se nam postavlja, je tudi, katere značilke so za model najpomembnejše. Odgovor na to vprašanje smo poskusili najti tako, da smo značilke rangirali z algoritmi za ocenjevanje značilk. Analizo smo izvedli na podatkih dodatnega eksperimenta, opisanega v poglavju 6.3.3, saj smo se želeli čim bolj izogniti vsem učinkom na značilke, ki jih ima čas nošenja zapestnice. Za rangiranje značilk smo uporabili vrednost *Gain ratio* in dobili

		Napovedana vrednost			Σ
		Lahka	Srednja	Težka	
Dejanska vrednost	Lahka	10	7	10	27
	Srednja	3	16	8	27
	Težka	6	10	11	27
Σ		19	33	29	81

Slika 6.3: Matrika napovedi za random forest algoritem na podlagi srčnega utripa

rezultate, predstavljene v tabeli 6.7. Vidimo lahko, da sta med prvimi sedmimi značilkami dve, ki sta izpeljani iz prevodnosti kože, ena iz temperature kože, ostale štiri pa so osnovane na srčnem utripu.

Analiza značilk po vrsti signala

Vplive značilk smo preverili tudi tako, da smo primerjali klasifikacijske točnosti, ki jih doseže algoritem naključnih gozdov, če mu damo samo značilke ene vrste signala, torej samo značilke srčnega utripa, značilke prevodnosti kože ali značilke temperature kože. Rezultati so predstavljeni v tabeli 6.8.

Vidimo, da sta za klasifikacijo zelo pomembna tako srčni utrip kot prevodnost kože, temperatura kože pa ima v našem primeru manjši vpliv. To je lahko tudi posledica tega, da se temperatura kože počasi odziva na miselni napor in zato v našem relativno kratkem eksperimentu ne pride do izraza. Odziv na spremembo temperature kože je lahko počasen tudi zaradi uporabljenega merilnika temperature v zapestnici.

Rang	Značilka	Gain ratio
1	avg_peak_amplitude_change_before	0.210
2	mean_hr	0.181
3	st_mean	0.175
4	sum_pos_deriv	0.165
5	ibi	0.156
6	sdnn	0.155
7	sd	0.155

Tabela 6.7: Prvih sedem značilk rangiranih po gain ratio.

Značilke	Točnost
Vse značilke	0.556
Značilke srčnega utripa	0.519
Značilke prevodnosti kože	0.556
Značilke temperature kože	0.389

Tabela 6.8: Točnosti klasifikacije pri različnih skupinah značilk.

Poglavje 7

Zaključek

V predstavljeni študiji smo poskusili ugotoviti, ali bi bilo mogoče za ugotavljanje občutene zahtevnosti igre, ki jo igra igralec, uporabiti biološke signale, pridobljene s pomočjo pametne zapestnice. Poskušali smo potrditi hipotezo, da obstaja povezava med izmerjenimi biološkimi signali in težavnostjo igre, ki jo je igralec igral. Za potrditev hipoteze smo izvedli eksperiment, pri katerem so prostovoljci igrali preprosto računalniško igro na treh različnih težavnostnih stopnjah, pri tem pa smo spremljali njihove biološke signale in kasneje iz njih izračunali vrsto značilnk. Na podlagi teh značilnk smo z uporabo različnih algoritmov strojnega učenja izdelali modele za napovedovanje težavnosti igrane igre na podlagi opaženih značilnk.

Ugotovili smo, da smo pri klasifikaciji v tri razrede dosegli klasifikacijsko točnost okoli 64%, vendar smo odkrili, da ima na napovedi večji vpliv čas nošenja zapestnice kot sama težavnost. Analizo smo ponovili tudi tako, da smo za vsak test upoštevali le prvo stopnjo testiranja, ki je imela pri nekaterih testih težko, pri nekaterih pa lahko težavnost. Na ta način smo se izognili učinkom trajanja nošenja zapestnice. V tem primeru so klasifikatorji dosegli točnosti okoli 67%, kar je boljše kot 59%, ki jih je dosegel kontrolni klasifikator. Pri tem eksperimentu smo uporabili samo prvo od treh izmerjenih težavnostnih stopenj in s tem zmanjšali število učnih primerov, kar lahko vpliva na slabši rezultat tega eksperimenta.

Na podlagi teh rezultatov lahko potrdimo, da povezava med našimi izmerjenimi signali in težavnostjo igre res obstaja in da je težavnost mogoče spremljati z uporabo komercialno dostopne in splošno namenske opreme, ki igralcu ni moteča. Za potrditev rezultatov smo izvedli tudi manjši eksperiment, pri katerem je ena oseba najprej dalj časa samo nosila zapestnico, da so se vsi časovno odvisni signali stabilizirali, nato pa zaporedoma izvedla več testov. Pri tem so bili pridobljeni rezultati dovolj dobri, da smo z njimi lahko potrdili začetno hipotezo.

Poskusili smo ugotoviti tudi, katere izmed značilk so največ prispevale modelu za napovedovanje težavnosti igre. Ugotovili smo, da značilke, ki so bile izračunane na podlagi srčnega utripa, in tiste, ki so bile izpeljane iz prevodnosti kože, prispevajo približno enako, medtem ko značilke, ki so bile izpeljane iz temperature kože, v našem primeru ne doprinesejo veliko.

V prihodnosti želimo z različnimi metodami zmanjšati ali izničiti vpliv časovne odvisnosti merjenih signalov. Predvidevamo, da bi lahko izboljšali rezultate napovedovanja težavnostne stopnje z uporabo značilk, ki niso odvisne od prevodnosti kože, ki se je izkazala za odvisno od časa nošenja zapestnice. Za značilke, ki so povezane s prevodnostjo kože, pa bi lahko zmanjšali časovni vpliv z uporabo ustrezne normalizacije signala.

Literatura

- [1] MC Ashton and K Lee. A short measure of the major dimension of personality. *European Journal of Psychological Assessment*, 91(4):340–345, 2009.
- [2] Michael C Ashton and Kibeom Lee. The hexaco–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, 91(4):340–345, 2009.
- [3] Luciano Bernardi, Joanna Wdowczyk-Szulc, Cinzia Valenti, Stefano Castoldi, Claudio Passino, Giammario Spadacini, and Peter Sleight. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *Journal of the American College of Cardiology*, 35(6):1462–1469, 2000.
- [4] Martin Gjoreski. Continuous stress monitoring using a wrist device and a smartphone. Master’s thesis, Jozef Stefan International Postgraduate School, Slovenia, 2016.
- [5] Martin Gjoreski, Mitja Luštrek, and Veljko Pejović. My watch says i’m busy: Inferring cognitive load with low-cost wearables. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1234–1240. ACM, 2018.
- [6] Eija Haapalainen, SeungJun Kim, Jodi F Forlizzi, and Anind K Dey. Psycho-physiological measures for assessing cognitive load. In *Procee-*

- dings of the 12th ACM international conference on Ubiquitous computing*, pages 301–310. ACM, 2010.
- [7] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.
- [8] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [9] Katherine A Herborn, James L Graves, Paul Jerem, Neil P Evans, Ruedi Nager, Dominic J McCafferty, and Dorothy EF McKeegan. Skin temperature reveals the intensity of acute stress. *Physiology & behavior*, 152:225–230, 2015.
- [10] Javier Hernandez Rivera. *Towards wearable stress measurement*. PhD thesis, Massachusetts Institute of Technology, 2015.
- [11] Nikola Komes. Medsebojno zaupanje v poslovnih odnosih med slovenijo in hrvaško: magistrsko delo. Master’s thesis, Univerza v Ljubljani, Filozofska fakulteta, 2017.
- [12] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529, 2009.
- [13] Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A Calvo. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 420–423. ACM, 2012.
- [14] C Schubert, M Lambertz, RA Nelesen, W Bardwell, J-B Choi, and JE Dimsdale. Effects of stress on heart rate complexity—a comparison

between short-term and chronic stress. *Biological psychology*, 80(3):325–332, 2009.

- [15] Su Xue, Meng Wu, John Kolen, Navid Aghdaie, and Kazi A Zaman. Dynamic difficulty adjustment for maximized engagement in digital games. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 465–471. International World Wide Web Conferences Steering Committee, 2017.