

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

PROJEKT

Implementacija automatizacije evaluacije dizajna korisničkih sučelja

Ana Terović

Voditelj: *Igor Mekterović*

Zagreb, rujan 2023.

SADRŽAJ

1. Motivacija	1
2. Implementacija rada	2
2.1. Opis	2
2.2. Tehničke značajke	3
2.3. Upute za korištenje	6
2.4. Rezultati	7
3. Moguća proširenja	10
4. Zaključak	12
5. Literatura	13

1. Motivacija

Dizajn korisničkog sučelja (engl. *user-interface*, *UI*) je ključan za stvaranje učinkovitog i dobrog korisničkog iskustva (engl. *user-experience*, *UX*). Dobro dizajnirano sučelje može poboljšati jednostavnost korištenja i ukupnu zadovoljstvo proizvodom. Loš dizajn sučelja može dovesti do zbunjenosti, frustracije i konačno smanjenja prihvaćanja korisnika.

Dobro dizajnirano sučelje treba biti vizualno privlačno, lako za navigaciju i omogućiti korisnicima da brzo i lako obavljaju svoje zadatke. Također treba uzeti u obzir potrebe, sposobnosti i ograničenja korisnika, kao i kontekst korištenja.

U današnjem digitalnom dobu, dizajn sučelja je sve važniji jer može biti imati odlučujuću ulogu za uspjeh proizvoda. Dobro dizajnirano sučelje može poboljšati angažman korisnika, povećati lojalnost prema brendu i konačno pokrenuti poslovni uspjeh.

Kako bi se dizajnerima i developerima olakšao proces izrade korisničkih sučelja ideja je izraditi sustav koji bi omogućavao automatizaciju evaluacije dizajna korisničkih sučelja.

Tijekom prethodno napravljenom pregleda ovoga područja istaknuo se rad *Prema poboljšanoj kreativnosti u dizajnu sučelja kroz automatiziranu procjenu upotrebljivosti*, (engl. *Towards Enhanced Creativity in Interface Design through Automated Usability Evaluation*) Snehal Dhengre (2020). Rad je implementirao model strojnog učenja koji s obzirom na sliku korisničkog sučelja daje procjenu kvalitete njegovog dizajna. U sklopu ovog projekta pokušati ćemo replicirati ovaj rad te vidjeti kako ga još možemo prilagoditi i proširiti njegove funkcionalnosti.

2. Implementacija rada

U nastavku ćemo implementirati rad *Prema poboljšanoj kreativnosti u dizajnu sučelja kroz automatiziranu procjenu upotrebljivosti*, (engl. *Towards Enhanced Creativity in Interface Design through Automated Usability Evaluation*) Snehal Dhengre (2020).

2.1. Opis

Cilj rada je bio napraviti model za automatizaciju ocjene sučelja kako bi pomogli developerima i dizajnerima. Konkretno radi se o istraživanju koje se bavi ispitivanjem i evaluiranjem upotrebe strojnog učenja za predviđanje mjera upotrebljivosti sučelja mobilnih aplikacija. Kako bi donijeli ocjenu upotrebljivosti određene su tri mjere: pravilnost (engl. *regularity*), složenost (engl. *complexity*) i dodirljivost (engl. *clickability*).

Kao model strojnog učenja korištene su konvolucijske neuronske mreže kojima su predviđane prethodno spomenute tri metrike. Konvolucijska mreža se sastoji od dvije zajedničke konvolucije, nakon čega se za svaku mjeru koristi posebna konvolucijska mreža s potpuno povezanim slojem na kraju. Prilikom evaluacije izgleda sučelja modelu se predaje samo slika sučelja nakon čega on daje svoje predviđanje vrijednosti.

Tijekom treniranja modela potrebno je više podataka od samo slike te je zato korišten RICO skup podataka of Illinois koji sadrži vizualne, tekstualne, strukturalne i interaktivne opise sučelja mobilnih aplikacija. RICO skup podataka sastoji se od 77000 opisa mobilnih aplikacija, od kojih su u radu koristili podskup od njih 205. Na temelju vrijednosti iz skupa podataka računaju se vrijednosti pravilnosti, složenosti i dodirljivosti.

Rezultati na testnom skupu su: točnost 88,57%, složenost 89,11%, pristupačnost 89,18%. Ovo su impresivni rezultati ako se uspoređuju s prethodnim istraživanjima u ovom području.

2.2. Tehničke značajke

Kako bi kvantizirali upotrebljivost sučelja autori rada su odredili tri metrike koje se mogu računati iz podataka o aplikacijama. Pravilnost mjeri konzistentnost organizacije komponenti korisničkog sučelja i razmaka između komponenti korisničkog sučelja i računa se pomoću:

$$RM = 1 - \frac{N_{av} + N_{ah} + N_{sp}}{3n} \quad (2.1)$$

,

gdje su N_{av} i N_{ah} broj vodoravnih i okomitih točaka poravnanja, N_{sp} je broj različitih udaljenosti između početnih točaka stupca i retka, a n broj komponenti. Mjera složenosti određuje koliko lako korisnik može pronaći očekivane informacije i definirana je u smislu broja komponenti i broja točaka poravnanja na korisničkom sučelju. Formula za računanje složenosti:

$$CM = \frac{N_{av} + N_{ah}}{2n} \quad (2.2)$$

.

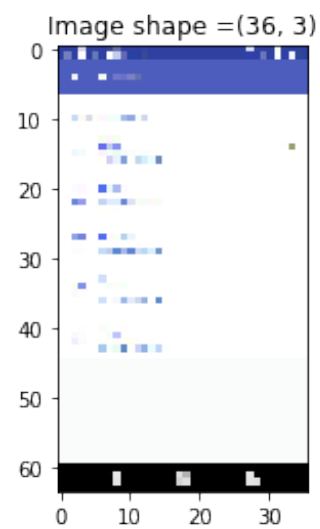
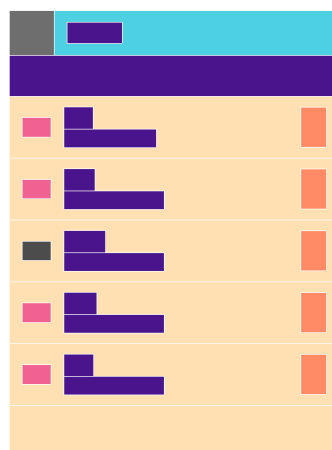
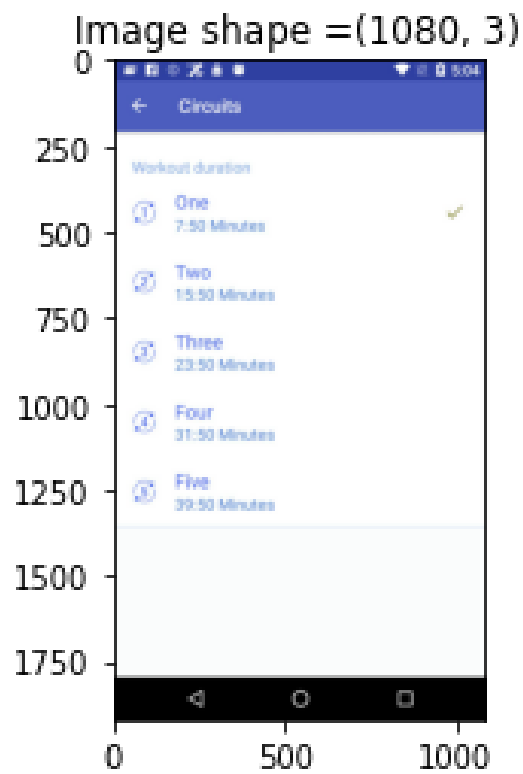
Mogućnost dodirivanja mjeri učinkovitost i točnost interakcije s ciljevima dodira s obzirom na pristupačnost korisničkog sučelja, a definiraju ju kao:

$$TM = \frac{N_c}{N_{ctot}} \quad (2.3)$$

,

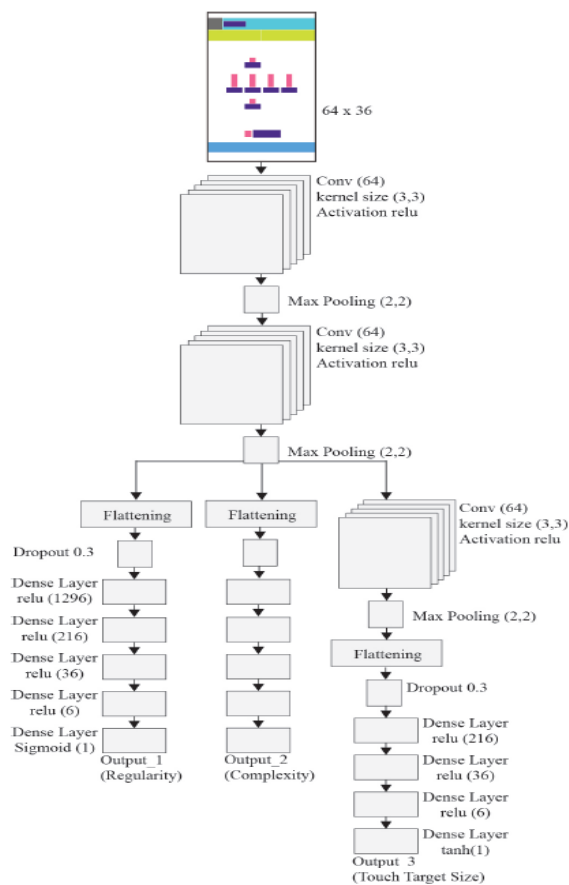
gdje je N_c broj komponenti na koje se može kliknuti i koje ispunjavaju zahtjeve za minimalnom veličinom cilja dodira, a N_{ctot} ukupan broj komponenti na koje se može kliknuti. Spomenute formule korištene su po uzoru na rad Soui i Mkaouer (2017). U kontekstu upotrebljivosti, želimo što veće vrijednosti za pravilnosti i dodirljivosti te što manju vrijednost složenosti sučelja.

Iz RICO skupa podataka korištene su .json datoteke sa semantičkim značajkama koje sadrže vrijednosti kao što su granice komponenti unutar sučelja, koje je vrste komponenta, je li na nju moguće kliknuti, itd. Uz to su bile potrebne i snimke zaslona odgovarajućih korisničkih sučelja. Slika 2.1 prikazuje primjer snimke korisničkog sučelja te slike 2.2 njegove pojednostavljene verzije koja sadrži samo obrise komponenta. Prva, originalna slika ulazi u model tijekom treniranja i testiranja dok je druga slika nastala iz podataka .json datoteke o semantičkim vrijednostima.



U procesu pripreme podataka za model strojnog učenja, na slike korisničkih sučelja je primjenjeno gaussovo zamučenje kako bi se povećala raznolikost skupa podataka te su slike reducirane na veličinu $64 \times 64 \times 3$. Rezultat ovih promjena možemo vidjeti na slici 2.3 .

Kao model strojnog učenja su korištene konvolucijske neuronske mreže zato što su dokazano vrlo dobre za probleme analize slike. Model je treniran korištenjem 70% slika iz generiranog skupa podataka i testiran s preostalih 30% slika. Korištena su dva sloja konvolucije i maksimalnog sabiranja (engl. *maxpooling*) na svakom ekranu mobilne aplikacije, a zatim je podijeljeno u tri grane kako bi predviđao tri različite mjere korisnosti. Dodatni sloj konvolucije i maksimalnog sabiranja primijenjen je za granu koja predviđa mjeru dodirljivosti kako bi se poboljšala rezolucija ulazne slike. Izlazni slojevi za različite mjere korisnosti koristili su sigmoidalne i hiperbolne tangens aktivacijske funkcije kako bi ograničili izlaz između 0 i 1. Odabrane aktivacijske funkcije pružile su najbolje podudaranje s modelom. Za funkciju gubitka se koristio MSE a za optimizaciju Adam optimizator. Slika 2.4 prikazuje arhitekturu modela.



Slika 2.4: Model strojnog učenja, Snehal Dhengre (2020)

2.3. Upute za korištenje

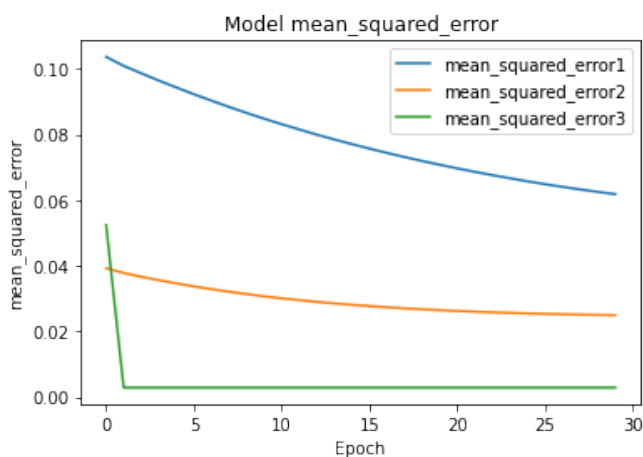
Implementacija rada dostupna je na git repozitoriju. Replikacija rezultata i testiranje su dostupni u projekt.ipynb dokumentu dok mogućnost samostalnog testiranja je moguća u testing_model.ipynb dokumentu.

Testiranje modela:

1. Otvoriti git repozitorij.
2. Iz repozitorija preuzeti datoteke:
 - model_888989.json
 - model_888989.h5
 - testing_model.ipynb.
3. Preuzeti sve potrebe biblioteke naznačene na početku testing_model.ipynb dokumenta pomoću npr. pip install.
4. Unutar testing_model.ipynb zamijeniti IMAGE_PATH varijablu sa stvarnom lokacijom slike koju želite testirati.
5. Pokrenuti datoteku testing_model.ipynb.

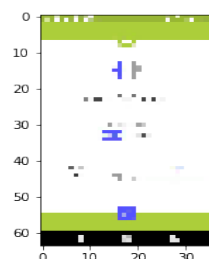
2.4. Rezultati

Rad je implementiran u cjelosti. Računanje vrijednosti za pravilnost, složenost i dodirljivost zahtjevala je malo detaljniju analizu podataka te specifikacija rada. Tijekom treniranja modela ostvarene su vrijednosti funkcije gubitka 2.5. Možemo vidjeti da je model kroz epohe konvergirao te uspio učiti tražene vrijednosti.



Slika 2.5: Vrijednosti funkcije gubitka

Model smo prvo testirali na slikama iz originalnog, RICO skupa podataka koji sadrži slike korisničkih sučelja mobilnih aplikacija. Za sliku 2.6 model je imao performanse prikazane u tablici 2.1.



Slika 2.6: Testni primjer

	stvarna vrijednost	predviđena vrijednost
pravilnost	0.061	0.102
složenost	0.727	0.738
dodirljivost	1	1

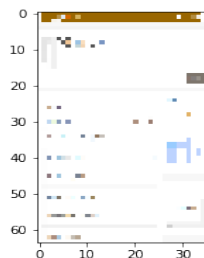
Tablica 2.1: Usporedba predviđenih i stvarnih vrijednosti modela

Gledanjem rezultata možemo vidjeti da model postiže vrlo zadovoljavajuće rezultate za slike mobilnih sučelja.

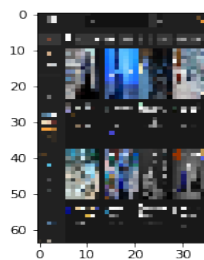
Jedna od ideja za proširenje ovoga rada je testirati može li se model generalizirati i na slike web korisničkih sučelja iako je treniran na slikama mobilnih aplikacija. Razlog za testiranje ove generalizacije je da ne postoji prikladan skup podataka za web korisnička sučelja kojim bi se mogla raditi evaluacija dizajna. Kao testne primjere koristili smo vrlo jednostavne primjere za koje se lako može predvidjeti njihove vrijednosti izlaznih metrika modela. Za sliku 2.7 možemo očekivati nisku vrijednost kompleksnosti, za sliku 2.8 očekujemo visoku vrijednost složenosti, a za sliku 2.9 visoku vrijednost kompleksnosti. Tablica 2.2 prikazuje dobivene vrijednosti zadanih sučelja. Možemo vidjeti da model nažalost ne generalizira dobro i na web korisnička sučelja.



Slika 2.7: Testni primjer korisničkog sučelja 1



Slika 2.8: Testni primjer korisničkog sučelja 2



Slika 2.9: Testni primjer korisničkog sučelja 3

	Primjer 1	Primjer 2	Primjer 3
pravilnost	0.107	0.151	0.005
složenost	0.706	0.682	0.879
dodirljivost	1	1	1

Tablica 2.2: Rezultati testiranja modela na web korisničkim sučeljima

3. Moguća proširenja

Jedna od ideja za proširenje rada je uporaba dodatnih metrika koje će nam moći reći nešto više o dizajnu sučelja osim njegove upotrebljivosti. Istraživanjem drugih radova možemo vidjeti razne implementacije procjene dizajna te bi bilo zanimljivo napraviti model koji uključuje većinu njih. Slika 3.1, napravljena u sklopu seminara, prikazuje sažeti prikaz različitih metrika upotrebljenih u drugim radovima prilikom izrade modela za procjenu dizajna sučelja.

	Vizualna složenost Boje	Kompozicija	Bijeli prostor Slike	Poveznice	Riječi	TLC	Raznolikost elemenata	Kompresija
Reinecke (2013)	+							
Aliaksei Miniukovich (2015)	+	+	+					
Thomas Schmidt (2018)	+	+						
Eleni Michailidou (2008)				+	+	+		
Aliaksei Miniukovich (2014)		+	+					
Ou Wu (2013)		+		+	+			
Maxim Bakaev (2018)	+		+	+	+		+	+
Mao (2019)		+					+	

Slika 3.1: Značajke pojedinih radova

Kako model predstavljen u ovome radu nije uspio generalizirati na web sučelja, svakako bi bilo interesatno probati napraviti model kojemu to uspijeva. Nažalost postoji velika problematiku s ovom idejom jer ne postoji skup podataka koji nam to omogućava. U zadnjih par godina izašlo je nekoliko radova, kao što je Jingwei Li (2020), koji govore o ovoj problematici te navode načine kako bi se takav skup podataka mogao napraviti. Tablica 3.2 preuzeta iz Griazev i Ramanauskait (2021), nam pokazuje razne skupove podataka web sučelja u kontekstu što sadrže. Bitno je primjetiti kako niti jedan ne sadrži sliku korisničkog sučelja što nas poprilično ograničava prilikom pokušaja automatizacije evaluacije. Sljedeći korak bi mogao biti replikacija jednog takvog rada kako bi mogli stvoriti svoj vlastiti skup podataka na kojemu bi mogli izgraditi naš model.

Datasets	Data Stored by Datasets							
	HTML	DOM	CSS	JavaScript	Images	Screenshots	Clean Text	Labeled Blocks
Cleaneval [5]	+	–	–	–	–	–	+	–
Dragnet [5,24]	+	–	–	–	–	–	+	–
GOSH [17]	–	+	+	–	+	–	–	–
Dataset-random [23]	+	+	+ ¹	+ ¹	+ ¹	–	+	+
dataset-popular [23]	+	+	+ ¹	+ ¹	+ ¹	–	+	+
wt10 g [8]	+	–	–	–	–	–	–	–
Web-KB [22]	+	–	–	–	–	–	+ ²	+ ²
L3S [24]	+	–	–	–	–	–	–	+ ³

Slika 3.2: Značajke pojedinih skupova podataka

4. Zaključak

Automatizacija evaluacije dizajna korisničkih sučelja je moguća, ali može biti teška zbog subjektivne prirode dizajna. Upravo zato pregledom već napravljenih radova susrećemo se s različitim heuristikama njegove procjene. Postoji nekoliko čimbenika koje treba uzeti u obzir prilikom ocjenjivanja dizajna web stranice, kao što su korisničko iskustvo, estetika i funkcionalnost. Korisničko iskustvo, odnosno koliko je korisnicima jednostavno i ugodno komunicirati s web stranicom, ključni je aspekt dizajna web stranice i može se ocijeniti korisničkim testiranjem i analizom metrika kao što su stopa napuštanja stranice i vrijeme na web stranici. Estetika ili vizualna privlačnost web stranice također se može ocijeniti, ali to je više subjektivno i može varirati ovisno o osobnim preferencijama. Funkcionalnost ili koliko dobro web stranica obavlja svoju namjenu također se može ocijeniti testiranjem i analizom metrika kao što su stope konverzije.

U slučaju da želimo izgraditi model na temelju strojnog učenja potreban nam je veći skup podataka kako bi mogli trenirati naš model. Ovdje se susrećemo s dodatnom problematikom manjka prikladnih skupova podataka. Za mobilna sučelja postoji prihvatljivi skup podataka ali ako se odlučimo raditi evaluaciju web korisničkih sučelja, morati ćemo sami izraditi skup podataka ili ograničiti mogućnosti našega modela.

Sve u svemu, iako se dizajn web stranice može ocijeniti, to je složen proces koji osim što zahtijeva kombinaciju objektivne i subjektivne analize također zahtijeva i resurse koji trenutno nisu dostupni.

5. Literatura

Kiril Griazev i Simona Ramanauskait. Multi-purpose dataset of webpages and its content blocks: Design and structure validation. *Applied Sciences*, 2021.

Xiangzhan Yu Jingwei Li, Chong Zhang. Webpage visual feature extraction and similarity algorithm. *CIAT 2020*, 2020.

The University of Illinois. Rico dataset. <https://interactionmining.org/rico>.

Farzaneh Oghazian Snehal Dhengre, Jayant Mathur. Towards enhanced creativity in interface design through automated usability evaluation. *Eleventh International Conference on Computational Creativity*, 2020.

M.; Gasmi I.; Soui, M.; Chouchane i M. W. Mkaouer. Plugin for predicting the usability of mobile user interface. *In Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2017.