Towards Enhanced Creativity in Interface Design through Automated Usability Evaluation

Snehal Dhengre[1], Jayant Mathur[2], Farzaneh Oghazian[3], Xiaomei Tan[1], Christopher McComb[2]
[1]Department of Industrial and Manufacturing Engineering, [2]School of Engineering Design, Technology, and Professional Programs, [3]Department of Architecture
The Pennsylvania State University
University Park, PA 16802 USA
sud702@psu.edu, jmm8886@psu.edu, fxo45@psu.edu, xzt24@psu.edu, mccomb@psu.edu

Abstract

With an increase in the number of mobile apps making their way to users, there is a growing need for tools to support the app design process. While many tools focus on increasing the pace of development, few attempt to aide the designer in generating more creative solutions. In this work, we take creativity as the combination of novelty and utility. Particularly during development of user interfaces, assessment of utility (primarily usability) is iterative, rigorous, and time-consuming. The objective of the proposed work is to explore and evaluate the use of machine learning to predict usability measures for mobile app interfaces as a means to automate usability evaluations. Specifically, a convolutional neural network (CNN) is used to accurately (nearly 90%) predict three usability measures: regularity, complexity, and touchability. This tool automates the assessment of utility in app design, freeing up the designer to seek designs that are novel and thus creative.

Introduction

Consumers downloaded 204 billion mobile applications (apps) worldwide in 2019 (Clement 2020) and the usage of mobile apps is continuing to increase at a tremendous rate of over 10 apps used per day on average by a user (Annie 2017). Mobile app users are therefore becoming selective due to the abundance of mobile apps, many of which contain similar features and services. In order to keep up with the consumer demand for engaging mobile apps, designers must create interfaces that are creative - both novel and usable. However, utility testing during user interface (UI) development limits the time and resources designers can allocate toward novelty. This utility testing is typically done through usability evaluation. The usability of an interface is defined as "the extent to which a product can be used by specified users to achieve designated goals with effectiveness, efficiency, and satisfaction in a specified context of use" (ISO9241-11 2018).

Current methodologies to evaluate usability often entail empirical assessments of observations, but these evaluations can be costly, time-consuming, and resource intensive. These limitations hinder the creativity of the designers as significant cognitive effort must be expended to ensure utility instead of pursuing novelty. There is, therefore, an opportunity to introduce automation in usability evaluation during UI development to help designers become more creative by empowering them to spend more time and effort pursuing novelty than ensuring utility. This work therefore aims to aide in improving design creativity by helping designers expedite the quantitative evaluation step in the design cycle, thereby providing designers more time and resources to focus on novelty. The goal for the current work is to investigate the efficacy of an automated evaluation tool to predict usability metrics, a step towards developing a proof of concept that will enhance creativity in interface design.

Related Work

Research in the utilization of computational tools for automation, intelligent feedback generation, and assistive design guidance shows that designers may benefit creatively when relying on computational assistive tools (Colton et al. 2019; Colton, Powley, and Cook 2018; Karimi et al. 2019). Colton et al.(2019; 2018) explain that computational automation can "empower people in a co-creative setting" thereby allowing designers to think outside their perspective and aiding creativity. Such co-creative design tools driven by machine learning (ML) can provide innovative solutions to enhance and aid in the development of creativity and creative designs (Karimi et al. 2019; Verganti, Vendraminelli, and Iansiti 2020). Moreover, ML is gaining momentum in the field of human-computer interaction (Yang, Banovic, and Zimmerman 2018) and offers potential for innovation in user-centered design as well.

Various research studies have been conducted to enhance the process of usability evaluation. For instance, PLAIN, a java-based automatic evaluation plugin, was developed for calculating the quality of the interface from a usability perspective (Soui et al. 2017). Using a similar approach, Soui et al. (2019) developed a multi-objective automatic optimization method to detect aesthetic aspects of UI. In their study, two sets of mathematical measurements for usability, guidance and coherence, were introduced as a tool for qualitative assessment of graphical user interface (GUI) (Soui et al. 2019).

While some studies incorporated quantitative methods for usability assessment of UIs and validated the methods through empirical studies, others implemented data-driven strategies to identify mobile app issues and automate the process. A data-driven based tool, ZIPT, was developed

Proceedings of the 11th International Conference on Computational Creativity (ICCC'20)
ISBN: 978-989-54160-2-8

to measure the performance of mobile app interfaces using metrics such as completion rate, time on task, and the number of interactions performed (Deka et al. 2017b). TapShoe was developed by Swearngin and Li (2019) to predict the tappability of UI elements. Both of the aforementioned studies used crowd-sourcing for collecting data and ML models for assessing usability.

Aesthetics are strongly associated with the functionality and user satisfaction of an interface (Kurosu and Kashimura 1995; Tractinsky 1997; Hartmann, Sutcliffe, and De Angeli 2008). Norman (2002) explains that good aesthetic design has a positive impact on user experience, thereby enhancing usability of the interface. These benefits of good aesthetic design have led researchers to explore the aesthetic measures of an interface to assess usability. For instance, Ngo, Teo, and Byrne (2000) defined 14 aesthetic measures that closely aligned with designers' implicit perceptions of aesthetics (Ngo, Samsudin, and Abdullah 2000).

Accessibility for interactive elements is another important usability attribute, particularly with touch screen devices where touch target size influences user's accessibility of the interface. Smaller touch targets require a higher time to tap and often leads to frustration. Parhi, Karlson, and Bederson (2006) recommended that a minimum size of 1 cm x 1 cm is required for accurate and efficient selection of touch targets. Similarly, Microsoft recommends a target size greater than 0.9 cm x 0.9 cm for frequently used interactive elements (Microsoft 2012).

Data Extraction and Augmentation

This study uses a subset of Rico dataset (Deka et al. 2017a). The dataset contains visual, textual, structural, and interactive design properties of 72,000 screens from mobile apps. Specifically, this study was conducted with annotated images of weather apps with 5 or 6 UI components (205 screens). Data was augmented by adding gaussian blur to the extracted images to increase the diversity of the dataset. This resulted in a total of 410 annotated RGB images of size 1440 x 2560 pixels for weather apps category.

Three measures (regularity, complexity, and touchability) were used as evaluation criteria to assess the usability of the mobile apps. These measures were selected to highlight the opportunity for using ML as an automated usability evaluation tool and do not resemble a holistic usability evaluation approach. Regularity measures the consistency of organization of the UI components and spacing between UI components which is given by:

$$RM = 1 - \left(\frac{N_{av} + N_{ah} + N_{sp}}{3n}\right) \in [0, 1] \quad (1)$$

where N_{av} and N_{ah} are the number of horizontal and vertical alignment points, N_{sp} the number of distinct distances between column and row starting points, and n the number of components (Soui et al. 2017).

Complexity measure determines how easily a user can find expected information and is defined in terms of the number of components and number of alignment points on a UI. The complexity measure is given by (Soui et al. 2017):

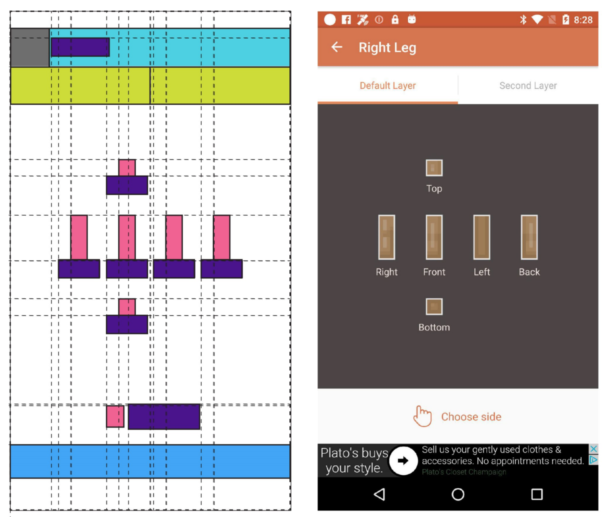
$$CM = \frac{N_{av} + N_{ah}}{(2n)} \in [0, 1] \quad (2)$$

Touchability measures the efficiency and accuracy of interaction with touch targets with regard to accessibility of UI and is defined by the authors as:

$$TM = \frac{N_c}{N_{c_{tot}}} \quad \in [0, 1] \quad (3)$$

where N_c is number of clickable components meeting the minimum touch target size requirement, and $N_{c_{tot}}$ the total number of clickable components.

To calculate these usability measures, the component class of each UI component (clickable or non-clickable), and the position of starting point and endpoint of each UI element were extracted for each mobile app image. Using these values, new index parameters were computed for each mobile app image including the number of vertical and horizontal alignment points, the number of distinct distances between row and starting points, the number of UI components, and the number of clickable components. An example of these index parameters for an image is demonstrated in Fig. 1. The usability measures, RM, CM, and TM, were then calculated using mathematical Eq. (1), Eq. (2) and Eq. (3), respectively. The minimum size for the touch object used for calculation was 0.9 cm x 0.9 cm.



Nav = 12 the numbers of vertical alignment points (number of rows).
Nah = 12 the numbers of horizontal alignment points (number of the columns).
Nsp = 16 the number of distinct distances between column and row starting points.
N = 20 the number of the components of the mobile user interface.

Figure 1: An example of index parameters for a mobile app page

Design and Implementation

Convolutional neural networks, like the model used in this paper (see Fig. 2), are a class of deep learning networks proven to be very effective for image recognition. Recently, such models have seen increased use in design (Williams et al. 2019; Raina, McComb, and Cagan 2019). The specific model in this work[1] was trained on 70% of the images from the generated dataset and tested with the remaining 30% of the images. Two layers of convolution and max pooling were applied on each mobile app screen, following which the model was forked into three branches to predict the three usability measures. Another layer of convolution and max pooling was applied for the branch predicting touchability measure for obtaining finer resolution of the input image. Sigmoidal and hyperbolic tangent activation functions were used for the output layer of RM and CM, and TM respectively, to limit the output bounds between 0 and 1. The respective activation functions ensured best fit for the model. A mean squared error loss function was used along with the Adam optimizer to compile the model.

To evaluate the performance of the model, the R-squared score was used as the performance metric, which assesses how close the predicted values are to the actual ones. The model was evaluated for the training and testing data separately by computing a coefficient of determination (R-squared score) for each of the three usability measures.

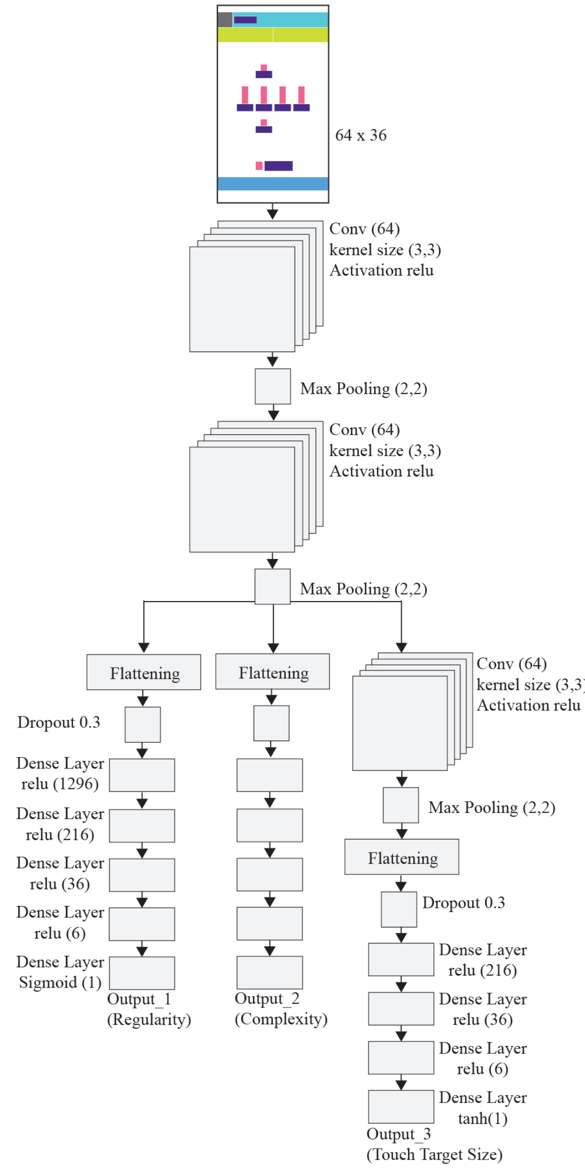


Figure 2: Architecture of ML model

Results

The R-squared score computed from the best trained model was greater than 90% for the training set and greater than 85% for the test set for all three usability measures. The corresponding R-squared score for each usability measure is shown in Table 1. These results suggest that the machine learning approach allows designers to predict close to accurate predictions to the expected usability measurement values. Prediction results of a sample mobile app screenshot is shown in Fig. 3, where low RM corresponds to fewer alignment points, high CM corresponds to difficulty in finding the required information, and low TM corresponds to fewer number of touch targets meeting minimum size criteria.

Table 1: R-squared scores of usability measures

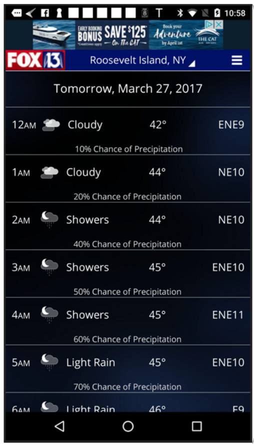
	R-squared score (%)	
Measure	Train set	Test set
Regularity	92.68	88.57
Complexity	92.17	89.11
Touchability	98.30	89.18

Conclusion

The goal for this work was to free up designers from spending time and resources on utility testing of mobile app UI designs and focus more on the novelty of designs, thereby improving the collective creative output. The proposed work therefore investigated automation of the usability evaluation of mobile app UIs. A framework for predicting usability measures for mobile apps using ML was presented as a tool for designers to conduct rapid usability evaluation. Specifically, a convolutional neural network model was developed to predict three usability measures: regularity, complexity, and touchability when given an input image of the mobile app UI. The findings indicate that ML can be used as a means of automation in the development of tools that assess the usability of mobile app UIs. There is potential to integrate such tools with GUI design platforms such as Adobe XD, Grasshopper, and Balsamiq. This can support designers by providing a quick usability assessment of interface designs, thereby allowing designers to instead focus their efforts on the novelty aspect of their designs.

The findings in the present study should be considered in light of some limitations such as small size of dataset used for training the ML model. Future work could explore validation of the tool by conducting usability study to assess its efficacy. Also, addition of mobile app screens from other

[1]https://github.com/snehal49/Usability-ML.git



Measure	Actual	Predicted
RM	0.22	0.33
CM	0.67	0.63
TM	0.00	0.07

Figure 3: Prediction results of a sample screenshot

app categories and incorporation of other usability measures could be considered to build a robust usability evaluation tool.

References

Annie, A. 2017. Spotlight on consumer app usage table of contents. http://files.appannie.com.s3.amazonaws.com/reports/1705_Report_Consumer_App_Usage_EN.pdf.

Clement, J. 2020. Number of mobile app downloads worldwide from 2016 to 2019. https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/.

Colton, S.; Pease, A.; Cook, M.; and Chen, C. 2019. The HR3 System for Automated Code Generation in Creative Settings. In *Proceedings of the Tenth International Conference on Computational Creativity*, 108–115.

Colton, S.; Powley, E.; and Cook, M. 2018. Investigating and Automating the Creative Act of Software Engineering. In *Proceedings of the Ninth International Conference on Computational Creativity*, 224–231.

Deka, B.; Huang, Z.; Franzen, C.; Hibschman, J.; Afergan, D.; Li, Y.; Nichols, J.; and Kumar, R. 2017a. Rico: A mobile app dataset for building data-driven design applications. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 845–854.

Deka, B.; Huang, Z.; Franzen, C.; Nichols, J.; Li, Y.; and Kumar, R. 2017b. Zipt: Zero-integration performance testing of mobile app designs. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* 727–736.

Hartmann, J.; Sutcliffe, A.; and De Angeli, A. 2008. Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction* 15(4).

ISO9241-11. 2018. Ergonomic requirements for office work with visual display terminals (vdts) — part 11: Guidance on usability. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en.

Karimi, P.; Maher, M. L.; Davis, N.; and Grace, K. 2019. Deep Learning in a Computational Model for Conceptual Shifts in a Co-Creative Design System. In *Proceedings of the Tenth International Conference on Computational Creativity*, 17–24.

Kurosu, M., and Kashimura, K. 1995. Apparent usability vs. inherent usability experimental analysis on the determinants of the apparent usability. In *Proceedings of the Conference on Human Factors in Computing Systems*, volume 2, 292–293.

Microsoft. 2012. UI design and interaction guide for windows phone 7 series.

Ngo, D. C. L.; Samsudin, A.; and Abdullah, R. 2000. Aesthetic measures for assessing graphic screens. *Journal of Information Science and Engineering* 16:97–116.

Ngo, D. C.; Teo, L. S.; and Byrne, J. G. 2000. Formalizing guidelines for the design of screen layouts. *Displays* 21(1):3–15.

Norman, D. 2002. Emotion & design: attractive things work better. *Interactions* 36–42.

Parhi, P.; Karlson, A. K.; and Bederson, B. B. 2006. Target size study for one-handed thumb use on small touchscreen devices. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI*, volume 159, 203–210.

Raina, A.; McComb, C.; and Cagan, J. 2019. Learning to Design From Humans: Imitating Human Designers Through Deep Learning. *Journal of Mechanical Design* 141(11).

Soui, M.; Chouchane, M.; Gasmi, I.; and Mkaouer, M. W. 2017. Plain: 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*, 127–136.

Soui, M.; Chouchane, M.; Mkaouer, M. W.; Kessentini, M.; and Ghedira, K. 2019. Assessing the quality of mobile graphical user interfaces using multi-objective optimization. *Soft Computing* (10).

Swearngin, A., and Li, Y. 2019. Modeling mobile interface tappability using crowdsourcing and deep learning. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, number Article 75, 1–11.

Tractinsky, N. 1997. Aesthetics and apparent usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 97*, 115–122.

Verganti, R.; Vendraminelli, L.; and Iansiti, M. 2020. Innovation and Design in the Age of Artificial Intelligence. *Journal of Product Innovation Management* 37(3):212–227.

Williams, G.; Meisel, N. A.; Simpson, T. W.; and McComb, C. 2019. Design Repository Effectiveness for 3D Convolutional Neural Networks: Application to Additive Manufacturing. *Journal of Mechanical Design* 141(11).

Yang, Q.; Banovic, N.; and Zimmerman, J. 2018. Mapping machine learning advances from hci research to reveal starting places for design innovation. In *CHI'18*.