Analysis and Prediction of University Websites Perceptions by Different User Groups

Maxim A. Bakaev[1], Tatiana A. Laricheva[1], Sebastian Heil[2], Martin Gaedke[2]
[1]Novosibirsk State Technical University, Novosibirsk, Russia
[2]Technische Universität Chemnitz, Chemnitz, Germany

Abstract **– Users' subjective impressions of websites are actively being researched, as they are formed very early in the interaction and affect the whole subsequent user experience. The subjective visual complexity is known to affect both website users' cognitive load and overall affective impressions. In our work we construct artificial neural network model to predict users' perceptions of selected universities' websites without the actual users. To train the network, we joined the two kinds of data that we specially collected. The visual complexity-related metrics of websites were extracted from web interface screenshots by our dedicated visual analyzer software implementing the "human-computer vision" approach. The assessments of the homepages orderliness and overall complexity were collected from 61 human annotators of different ages and cultural/national groups. We were able to achieve moderate subjective perceptions prediction accuracy, with relative error of 73.0%. The analysis of the factors' importance suggests that the proposed index of visual-spatial complexity had considerably higher importance than the baseline frequency-based entropy measure for images. The results of our work can aid in controlling the visual complexity perception in website visitors, ultimately contributing to better web usability.**

Index Terms **– Web interfaces, user satisfaction, computer vision, neural networks, visual complexity.**

I. Introduction

THE MODERN research in Human-Computer Interaction (HCI) recognize that user needs go beyond just the objective usability metrics, but that rather integral user experience must be considered. Thus, emotional dimensions, such as the interface attractiveness, are increasingly being considered in research works. This splendor is of significant interest to web design practicians as well, since it is capable of "masking" usability disadvantages. The aesthetics and visual attractiveness factors are perceived very early in the interaction and affect the whole subsequent user experience. Indeed, research results imply that users make up their impression of a website even quicker than the brain can perceive it on a conscious level [1]. A positive first impression shapes the subjective perception of objective negative aspects in the interaction, which can be diminished or even ignored. On the contrary, even objectively useful and usable website can cause overall negative user experience, if the first impression was unfavorable.

In his influential book, D. Norman claimed that the design aesthetics can be more important in terms of user preferences than the traditionally esteemed functionality [2]. However, it was also found that highly attractive websites with low usability had high user satisfaction, but low usability ratings [3]. Thus, what is perceived beautiful is not necessarily perceived usable, and both these dimensions must be attended to by designers. Moreover, aesthetic perceptions are more personal and more variable within a user group. Universal design guidelines are hardly applicable here, and studying the perceptions generally requires gathering large amounts of data for analysis, accounting for both individual and group differences, such as gender, age, cultural background, etc.

Such research undertakings also need to consider influential visual properties of websites, objectively measured [4]. Particularly, visual complexity is promising, since it's known to substantially affect attractiveness – e.g. in [5] it was found that complexity expressed as the number of compositional elements significantly influenced interfaces' aesthetic appeal ratings. The results of other research works imply that users appreciate websites that fit into the range of moderate complexity [6].

Besides the number of elements, images, or the amount of text, the visual complexity measurements in interfaces need to take into account visual presentation characteristics. For instance, high number of saturated colors can lead to increased colorfulness causing visual "overload" in the perceiving users. The visual complexity has been also measured as the share of screen space under a certain element type – text, imagery, etc. – determined by textures [7], where the images are determined based on the distinct edges [8]. It was also reported that website complexity is affected by the webpage structural characteristics, such as symmetry, balance, etc.

Here one finds the notable difference between the techniques for measuring websites' visual complexity versus assessing their usability based on website metrics. The latter are mostly evaluated based on the website code (HTML/CSS) analysis. They assess the website metrics affecting the three classic usability dimensions – effectiveness, efficiency and satisfaction – or find usability problems and risks [9]. However, code-based approaches have limited for visual complexity measuring, since the same webpage code can be rendered differently in different browsers, screen sizes, mobile devices – as modern adaptive designs alter the layout and even the number of displayed elements. Hence, visual complexity objective measurement, which is still largely seen as desirable [10], has to employ different approaches, such as computer vision-based ones.

978-1-5386-7054-5/18/$31.00 ©2018 IEEE

II. Problem Definition

In our paper, we employed specially developed software, the website visual analyzer, to collect selected metrics related to visual complexity of webpages. These metrics were used in training the artificial neural network (ANN) user behavior model that we constructed to predict the users' subjective perception of the webpages orderliness and complexity. The training data were obtained in experimental sessions, where 61 target users interacted with 21 websites of German and Russian universities. The model can be applied to predict users' subjective perceptions of websites from the considered domain without involvement of the actual users. We believe this approach has the potential to aid web designers produce websites attaining higher levels of the users' subjective appreciation.

In Section 2, we describe the visual analyzer's architecture and the data it produced for the ANN model. There we also provide some background on the neural networks in general. Section 3 is dedicated to the training data collection in the experimental session and the ANN model's construction, training and quality evaluation. We also use the analysis of the factors' importance to provide recommendations for the future data collection sessions. In Conclusions, we summarize our findings and outline prospects for further work.

III. Methods and Tools

A. The Visual Analyzer

Arguably, webpage screenshot analysis is the most feasible way to obtain its visual complexity metrics. The visual analyzer that we developed is actually capable of performing various intelligent user interface (UI) analysis tasks (see in [11]), and in the current work we applied it to measure the visual complexity.

The analyzer's architecture consists of two parts: the Analyzer Frontend and the Visual Analysis backend. The Analyzer Frontend is a Web Application that communicates with the Visual Analysis Backend, implemented as Web Service, via an HTTP Interface. Our visual page analysis algorithm is based on [12]: it takes a screenshot of a user interface as input and tries to identify the UI elements of which the interface is formed. Higher-level structures can be identified through analysis of the visual hierarchy of the interface, using closeness, alignment, containment etc.

During the preprocessing stage, the screenshot is transformed into black-and-white image and upscaling is performed with the respective OpenCV's library function. The recognition of interface elements in the screenshot is based on detection of rectangular objects, trained detectors (implemented in dlib) for identifying the elements' types, and optical character recognition for texts (combination of OpenCV's close-edge detection and Tesseract). Ultimately, the identified UI elements and structures are represented as JSON in the HTTP API:


```
{
  "elements":
   [
    {
    "height": 35, "positionX": 10, "positionY": 50,
    "text": "Name", "type": "label", "width": 40
    },
    {
    "height": 35, "positionX": 70, "positionY": 50,
    "text": "Enter name", "width": 250,
    "type": "labeled dropdown or textfield"
    },
    {
    "height": 20, "positionX": 130, "positionY": 100,
    "text": "OK", "type": "button", "width": 50
    },
   ],
  "page": { "height": 1024, "width": 1280}
}
```


More detailed description of the analyzer's algorithm can be found in [11], where we also demonstrate fair validity of most of the data output by the analyzer.

The analyzer frontend aggregates the visual analysis results with the results from DOM analysis and calculates several metrics for the UI. The following metrics related to visual complexity were used in the current research:

1. The number of all identified UI elements: *Num_Elem*. Note that this number results from visual analysis and is agnostic to programming styles and HTML elements invisible to users, in contrast to elements in DOM analysis.
2. The number of different elements types identified by the analyzer: *Num_Type*.
3. *Whitespace*, which is the area calculated as the full area of the screenshot minus the areas under all identified interface elements (text, graphics, etc.).

B. ANNs for Assessing Web Interaction

Artificial neural networks were designed to represent structure and capabilities of the human brain, in a way copying its organization [13]. The architectural components of ANNs are computational units resembling the neurons in the brain, and the network is formed from one or several neuron layers whose interconnections have synaptic weights. Each neuron in the network is engaged in the computations for the network training process, the results of which are stored as the synaptic weights' values. After the training, ANNs can be tested on real data, to classify, predict, support decision-making, etc.

The available dataset is usually partitioned into:

- The training set, used in the network training process, where the weights are generally corrected based on the error backpropagation method.
- The testing set, used to prevent the model overtraining. The data from this set are not used in the weights' correction, and can be used to stop training when the error on the validation dataset increases, as this is a sign of overfitting to the training dataset.
- The holdout (validation) set is used after the training is finished, to provide unbiased evaluation of the resulting models' quality (fit to the data).

The quality of the model is measured as the percentage of incorrect predictions (for categorical output neurons) or as the relative error that is calculated as sum-of-squares relative to the mean model (the "null" hypothesis).

The collection and preparation of the training data is a crucial step, as ANNs can deliver reasonably accurate classifications and predictions in many tasks, although the model setup can take a lot of time. Neural networks are considered a potent tool for predicting human behavior, including subjective impressions [14], provided that enough data are available and appropriate model structure and training algorithms are employed. Generally, modifications in the input data, as well as alterations in the number of hidden neuron layers and the number of neurons significantly affect the predictions accuracy. Among certain disadvantages of neural networks are the facts that they are "black boxes", rarely allowing meaningful interpretation of the model, and that they require lots of diverse data for training [15].

C. *The Metrics for the ANN Model*

For the sake of our task, i.e. predicting the users' visual complexity perceptions, the output neurons were subjective complexity assessments in ordinal scale (1 – "completely disagree", 7 – "completely agree"). We used the following statements, expressed in the terms matching the users' thesaurus:

1. "The elements in the webpage are well-ordered" (the *Compress* dependent variable).
2. "The webpage appears very complex" (the *Complex* dependent variable).

The input neurons had to reflect the context of use and thus incorporated the website metrics and the users' characteristics:

a) User-based factors (the selection of the factors is based on general knowledge of important user attributes):
   1. *Age*,
   2. *Gender*,
   3. National/cultural group (*Nationality*).

b) Website-based factors (the metrics provided by the visual analyzer):
   4. The number of interface elements (*Num_Elem*),
   5. The number of different types of elements (*Num_Type*),
   6. The entropy value provided by Matlab's entropy(I) function that returns a scalar value E reflecting the entropy of grayscale image I (*Entropy*),
   7. Share of whitespace in the web interface area (*Whitespace*),
   8. Index of visual complexity (*Index_Compl*).

The index of visual complexity was calculated based on the results of our previous research work on interface complexity [16] as the following:

$$Index\_Compl = \\ = Num\_Elem * \log_2(Num\_Type) * \log_2(JPEG) \quad (1),$$

where *JPEG* is the size (in bytes) of the webpage screenshot compressed using JPEG-100 algorithm (per ISO/IEC 10918-1:1994), to represent the spatial orderliness of the web interface.

In the next Section we describe the construction and training of the ANN model with the structure outlined above.

IV. The Predictive ANN Model

A. *The Experimental Setup*

In the training data collection session, we employed screenshots of 21 operational websites from the fixed domain, *Career and Education*: 11 websites of German universities and 10 websites of Russian ones (for both groups, English versions of the websites were used) with varying levels of visual complexity. More detailed description of the employed websites can be found in the related technical report [17].

In total, the evaluations were collected from 61 participants (Age ranged from 19 to 72, mean=27.61, SD=8.19), whose demographics are presented in Table I.

TABLE I
THE PARTICIPANTS' DATA

Factor		The number of participants	The share in the group
Gender	Male	29	47.5%
	Female	32	52.5%
Language used	Russian	41	67.2%
	English	11	18.0%
	German	9	14.8%
Cultural group	Russian	40	65.6%
	German	10	16.4%
	Argentinean	3	5.0%
	Iranian	1	1.6%
	Bulgarian	1	1.6%
	Other	6	9.8%

The survey for the data collection was implemented in LimeSurvey software that the participants accessed via a web browser. Some of them worked in university computer rooms, while the others used their own computer equipment, so on overall diverse range of equipment and screen resolutions was represented. Each of the subjects was asked to assess the 21 website screenshots (presented sequentially in random order) per the set of 7 subjective scales that included the *Compress* and *Complex* evaluations.

B. *Descriptive Statistics*

In total, 2562 assessments were collected for the two subjective scales. Of them, 2440 (95.2%) were considered valid; one website (#14) was removed from the analysis due to technical problem with the screenshot. The values for the scales and the qualitative factors used in the ANN model are presented in Table II. Mean and standard deviations for the Compress and Complex variables that are of ordinal scale are provided just for reference.

TABLE II
THE DESCRIPTIVE STATISTICS

Role in the ANN model	Name	Mean (SD)
Input neurons	*Num_Elem*	63.15 (23.78)
	Num_Type	5.75 (1.20)
	Entropy	4.18 (1.00)
	Whitespace (area)	2658365 (963047)
	Index_Compl	3279 (1472)
	Age	27.61 (8.19)
Output neurons	*Compress*	4.30 (1.30)
	Complex	3.59 (1.44)

In the subjects' assessments, negative correlation (ordinal scale's Kendall's tau-b) between *Compress* and *Complex* was highly significant, $p<0.001$, $\tau=-0.336$. *Compress* assessments also had significant correlation with website's *Whitespace* ($p=0.001$, $\tau=0.072$) and *Entropy* ($p<0.001$, $\tau=0.091$). *Complex* assessments had significant correlations with *Num_Elem* ($p<0.001$, $\tau=0.116$), *Whitespace* ($p<0.001$, $\tau=0.075$), and *Index_Compl* ($p<0.001$, $\tau=0.148$) – the latter having the highest value.

In the website's visual complexity metrics extracted by the visual analyzer, significant correlations were found for *Whitespace*: with *Num_Elem* ($p=0.011$, $\rho=0.553$), *Num_Type* ($p=0.019$, $\rho=0.519$), and *Index_Compl* ($p=0.008$, $\rho=0.577$). Somehow unexpectedly, the baseline *Entropy* factor had no significant correlations with any of the other website's metrics.

C. The Model Implementation

We employed the Multilayer Perceptron to construct the model, since it is considered a robust and reliable structure for predictive ANNs [18]. There was one hidden layer with 4 neurons; the input neurons were the 8 factors, the output neurons were the 2 subjective perception scales (see Tab. II). Optimization algorithm – gradient descent, activation functions – hyperbolic tangent for the hidden layer and identity – for output layer. The partition of the data was 70% for training, 20% – testing, 10% – holdout sets (1220 assessments in total).

The resulting overall relative error was relatively high, 73.0% (71.0% for *Compress* and 75.0% for *Complex*). Thus we have to conclude that although the constructed ANN model can be used for predicting users' perceptions of web interfaces' visual complexity parameters, it still needs further refinement. The results of the factors' importance analysis are presented in Table III and Fig. 1. The index of visual-spatial complexity (*Index_Compl*) that we proposed in one of our previous works [16] has shown the highest importance, while the standard frequency-based entropy measure for images had the lowest importance in the model.

TABLE III
THE IMPORTANCE OF THE FACTORS IN THE MODEL

Factor	Importance	Normalized importance
Index_Compl	0.297	100.0%
Num_Elem	0.207	69.8%
Age	0.120	40.3%
Nationality	0.105	35.5%
Whitespace	0.101	34.2%
Num_Type	0.096	32.4%
Gender	0.052	17.5%
Entropy	0.021	7.2%

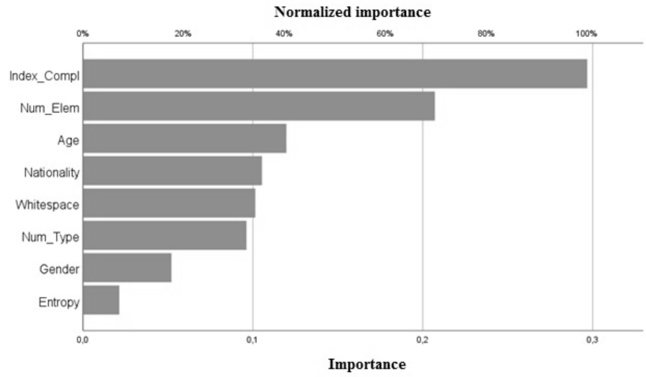

Fig. 1. The results of the factors' importance analysis in the ANN model.

V. CONCLUSION

The study of users' subjective impressions is actively performed by researchers and is of considerable interest for practicing web designers. Particularly, perception of visual complexity affects both web interface user's cognitive load and overall affective impression from visiting a website. Being able to predict the visual complexity perception dimensions using trained user behavior models can provide significant boost in enhancing user experience and satisfaction.

In our work we relied on trained artificial neural network model to predict users' perceptions of selected universities' websites without the actual users. The visual complexity-related metrics of websites were extracted from their homepage screenshots by our dedicated visual analyzer [11]. The assessments of the homepages orderliness and overall complexity were collected from 61 human annotators of different ages and cultural/national groups. We used the collected data to train the ANN model and so far were able to achieve moderate prediction accuracy, with relative error of 73.0%. We plan to further refine the model through more in-depth interface elements-related factors evaluation and consideration of their spatial allocation.

This prospect is supported by the analysis of the factors' importance that suggested that the index of visual-spatial complexity that we proposed previously in [16] had the highest importance. Next in importance was the number of identified elements, while the correlations analysis highlighted the influence of whitespace – both these findings are in line with the existing research on interfaces' visual complexity and web design practice. In the user group-related factors, the importance of age and nationality were considerably higher than the one of gender, which should be considered in the sampling for future training data collection sessions. The standard frequency-based entropy measure for images had shown the lowest importance in the model, which calls for more advanced metrics to represent visual complexity in web interfaces.

ACKNOWLEDGEMENT

The reported study was funded by RFBR according to the research project No. 16-37-60060 mol_a_dk.

REFERENCES

[1] Lindgaard G. et al. Attention web designers: You have 50 milliseconds to make a good first impression! // Behaviour & information technology. – 2006. – V. 25. – No. 2. – pp. 115-126.
[2] Fishwick M. Emotional design: Why we love (or hate) everyday things // The Journal of American Culture. – 2004. – V. 27. – No. 2. – pp. 234-234.
[3] Lindgaard G., Dudek C. What is this evasive beast we call user satisfaction? // Interacting with computers. – 2003. – V. 15. – No. 3. – pp. 429-452.
[4] Reinecke K. et al. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. – ACM, 2013. – pp. 2049-2058.
[5] Bauerly M., Liu Y. Effects of symmetry and number of compositional elements on interface and design aesthetics // Intl. Journal of Human–Computer Interaction. – 2008. – V. 24. – No. 3. – pp. 275-287.
[6] Geissler G. L., Zinkhan G. M., Watson R. T. The influence of home page complexity on consumer attention, attitudes, and purchase intent //Journal of Advertising. – 2006. – V. 35. – No. 2. – pp. 69-80.
[7] Ivory M. Y., Sinha R. R., Hearst M. A. Empirically validated web page design metrics // Proceedings of the SIGCHI conference on Human factors in computing systems. – ACM, 2001. – pp. 53-60.
[8] Bucy E. P. et al. Formal features of cyberspace: Relationships between web page complexity and site traffic // Journal of the Association for Information Science and Technology. – 1999. – V. 50. – No. 13. – pp. 1246.
[9] Grigera, J., Garrido, A., Rivero, J.M. Rossi, G. Automatic detection of usability smells in web applications. International Journal of Human-Computer Studies, 2017, V. 97, pp. 129–148.
[10] Wu, O., Hu, W., Shi, L. Measuring the visual complexities of web pages. ACM Transactions on the Web (TWEB), 2013, V. 7, No.1, p.1.
[11] Bakaev, M., Heil, S., Khvorostov, V., Gaedke, M. HCI Vision for automated analysis and mining of web user interfaces. // Lecture Notes in Computer Science, 10845: Web Engineering (ICWE 2018). – in print.
[12] Kong J. et al. Web Interface Interpretation Using Graph Grammars. IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.), 2012, V. 42, No. 4, pp. 590–602.
[13] Haykin, S., Neural Networks, A Comprehensive Foundation, 696 pp., Macmillan, Indianapolis, Indiana, 1994.
[14] Bakaev M., Khvorostov V., Laricheva T. Assessing Subjective Quality of Web Interaction with Neural Network as Context of Use Model // International Conference on Digital Transformation and Global Society. – Springer, Cham, 2017. – pp. 185-195.
[15] Yao X. Evolving artificial neural networks // Proceedings of the IEEE. – 1999. – V. 87. – No. 9. – pp. 1423-1447.
[16] Bakaev M., Razumnikova O. Opredeleine slozhnosti zadach dlya zritelno-prostranstvennoi pamyati i propustkoi spospobnosti cheloveka-operatora (Defining complexity for visual-spatial memory tasks and human operator's throughput). Upravlenie bol'shimi sistemami = Large-Scale Systems Control, 70. – 2017. – pp. 25–57 (2017). – In Russian.
[17] Bakaev M., Gaedke M., Heil S. Kansei engineering experimental research with university websites.// Technical Report. – 2016.
[18] Oztekin A. et al. A machine learning-based usability evaluation method for eLearning systems // Decision Support Systems. – 2013. – V. 56. – pp. 63-73.

Maxim Bakaev – PhD in Technical Sciences, works as Senior Researcher in Automated Control Systems department of Novosibirsk State Technical University (NSTU), Russia.
His research interests include Human-Computer Interaction, Web Design, User Behavior Models, and Machine Learning. He serves as reviewer for several international journals, including Large-Scale Control Systems, Algorithms, Advances in Systems Science and Applications, etc.

Email: maxis81@gmail.com

Tatiana Laricheva – Master student in Applied Informatics, Automated Control Systems department, Faculty of Automation and Computer Engineering of NSTU. Her research interests include Usability Engineering, Web Interfaces, and Artificial Neural Networks. She also served as researcher in "Modeling interaction in human-computer systems with interface information complexity project" (RFBR grant 16-37-00184 mol_a).

Email: laricheva-t-a@mail.ru

Sebastian Heil – PhD student in Distributed and Self-organizing Systems Professorship of Chemnitz Technical University, Germany. His research interests include Web Engineering, Software Migration, Crowdsourcing and Collaborative Development, and Linked Open Data. He is a member of Institute of Systems and Technologies of Information, Control and Communication.

Email: sebastian.heil@informatik.tu-chemnitz.de

Martin Gaedke – Full Professor, chair of Distributed and Self-organizing Systems group and Vice Dean of the Department of Computer Science, Chemnitz University of Technology.
His research focuses on Web Engineering, Service Engineering, and trustworthy Service-based Systems Engineering to address needs of future distributed hyper-connected collaborative work environments. He serves as invited expert for W3C's WebID and Social Web and served as reviewer for FP7 and ERC and as member of FP7 FAST Advisory Expert Group and NMC's Horizon.JISC Advisory Board.

Email: martin.gaedke@informatik.tu-chemnitz.de