Research on the Fuzziness in the Design of Big Data Visualization

Tian Lei, Qiumeng Zhu(✉), Nan Ni, and Xin He

School of Mechanical Science and Engineering,
Huazhong University of Science and Technology, Wuhan, China
qiumeng_zhu@hust.edu.cn

Abstract. In consecution to use and process information immediately, the relationship among a huge number of information is necessary to be read and understand. Information visualization as an effective method to optimize this process, using the charts to help people comprehend and process information intuitively and quickly. The accuracy of the information in the visualization chart is based on the readability and integrity of the information transition, once the chart does not meet this requirement, the accuracy of the information will be greatly reduced, and even may be misunderstood or cannot obtain the problem of information.

This paper will analyze and deduce the causes of ambiguous in the information visualization from the aspects of ambiguity definition and fuzziness experimental research. To solve this problem, the investigation collects 30 samples based on five complex information visualization charts, we will use infographic as the research object to explore the impact of fuzziness on the user in the visualization process and explore the causes and mechanisms of this effect by quantitative experiments.

Keywords: Information visualization · Fuzziness

1 Introduction

In the time of information explosion, people need to quickly handle the mass of information and understand the complex relevance among them. Information visualization as an effective method to optimize this process, using the charts to help people comprehend and process information intuitively and quickly. There is no doubt that the information visualization will promote the information resource development and utilization. If users can quickly and accurately comprehend all the information contained in the infographic, then the method of visualizing these information is effective and accurate. Otherwise, if users misunderstand the information or cannot get all the information content of the problem, so this appearance means that the method of visualizing the information remains fuzziness.

2 Literature Review

In contemporary times, there are many problems with the process of visualization due to huge and complex data. For example, improper or unreasonable key information extraction may result in chaotic and unsystematic problems in the information

S. Yamamoto and H. Mori (Eds.): HIMI 2018, LNCS 10904, pp. 70–77, 2018.
https://doi.org/10.1007/978-3-319-92043-6_6

visualization; inconsistent or irregular characterization of information types may result in that users are unable to quickly and accurately understand the information in a short time. When the origin comprehensible information becomes incomprehensible, the fuzziness of the visualization occurs. In the Webster's Dictionary, the "understanding" means that "the ability to distinguish the nature of things, or to grasp the inherent behavior of things or results" [1]. If users cannot distinguish the nature of things or cannot understand the internal relations, the thing like "incomprehension", "misunderstand" will occur. This is called information fuzziness.

Fuzziness is a very common topic. The problem of fuzziness exits in many fields. For example, in the linguistics, a kind of words like "many" "little" "possible" "probably" is full of fuzziness. However, there is no common acceptable definition until now, fuzziness is an objective attribute without a certain boundary in the amount. In the Oxford Advanced Learner's Dictionary, fuzziness has two meanings: one is the quality of being not clear in shape or sound and the other one is the fact of being confused and not expressed clearly [2]. Nevertheless, this does not affect scholars' enthusiasm for the study of ambiguity. For instance, scholars use mathematical methods to study the fuzziness - fuzzy mathematics. Its essence is to discretize the associated problem, in consecution to study and deal with the phenomenon of fuzziness [3]; scholars also use psychological methods to study the fuzziness - fuzzy psychology. That is, in a case of uncertain decision-making, subconscious psychology as the main tone, to make a subconscious psychological choice [4].

In the branches of design science, fuzziness is a widespread phenomenon, because the design itself is full of numerous emotional information. On the base of the emotional information, how to propose a more user-friendly design, or to evaluate these designs, is a very important part of the design process. For example, KY Chanb et al. used the method of intelligent fuzzy regression to study the nonlinear fuzzy relation between emotional response and design variables, and they used it as a tool to obtain more reliable information about customer demand [5]; Wang established an evaluation model in an inaccurate, non-deterministic product design environment, which based on the integrity of the user preference information [6]. These studies are interesting, but they are fragmented and not systematic. In the design of information visualization, there are similar problems in theory and practice.

In this paper, we will use infographic as the research object to explore the impact of fuzziness on the user in the visualization process and explore the causes and mechanisms of this effect by quantitative experiments.

3 Investigation of Fuzziness

The aim of this investigation is to understand the mechanism of the generation about fuzziness in the cognize space during the process of information visualization, and to find a method of avoiding the fuzziness.

The research is divided into two part:

(1) The evaluate to the subjects' cognition of infographic;
(2) The experimenter evaluate the subjects' cognition degree.

3.1 The Experiment About the Subjects' Cognition on Infographic

For information visualization of simple data, such as line chart, pie chart, it is less chance that the fuzziness happens. Since the data is simple, the relationship of data is simple, and the way of information visualization is simple. But for more complex data, namely the data itself is complex, since the relationship is complicated, the kinds of graphics, symbols needed are complex, it is easy to produce cognitive fuzziness generally. According to the relation of the data in this study, the main methods of complex information visualization are divided into four categories, shown in Table 1.

Table 1. The classification of infographic

Relationship type	Members
Logic relationship	Venn diagram
	Clustermap
Hierarchal relationship	Tree diagram
	Circle packing
	Sunburst
	Mosaic plot/cube tree map
Circulation relationship	Chord diagram
	Network diagram
	Sankey diagram
	Flow chart
Date comparison relationship	Bullet chart
	Scatter plot matrix
	Radar chart
	Boxplot

Logical relationship charts and contrast data chart are learned in the Chinese education, based on the degree of complexity, five methods of information visualization are selected from other kinds of charts in the hierarchical and circulation relation. They are chord diagram method, Sankey diagram method, circle packing method, sunburst method and tree diagram method. The independent variable is the five degree of the information visualization degree. The control variable of this study is the content of the information and the color of the chart. The dependent variable as the degree of familiarity, the degree of complexity, the degree of intact, the degree of consecution, the degree of access to information, the degree of color influence, the degree of language influence, the degree of readable. The corresponding problem are shown in Table 2. While the material of experiment are shown in Fig. 1.

The experiment subjects are thirty university students aged between 18 and 25, 15 subjects are men and 15 subjects are women, among them, 25 subjects are undergraduate students and 5 are postgraduate students. The major of 10 subjects are associated with design engineering and rest 20 subjects' major are not related to design engineering.

Table 2. Experimental questions

No.	Questions
Q1	Are you familiar with this kind of charts?
Q2	Do you think the chart is complicated?
Q3	Can you understand the complete information expressed in this chart?
Q4	Do you think the chart is organized?
Q5	Do you think the chart is useful for getting data?
Q6	How much do you think the color of the chart affecting reading?
Q7	How much do you think the language has caused you to read?
Q8	How was your reading experience on the chart?

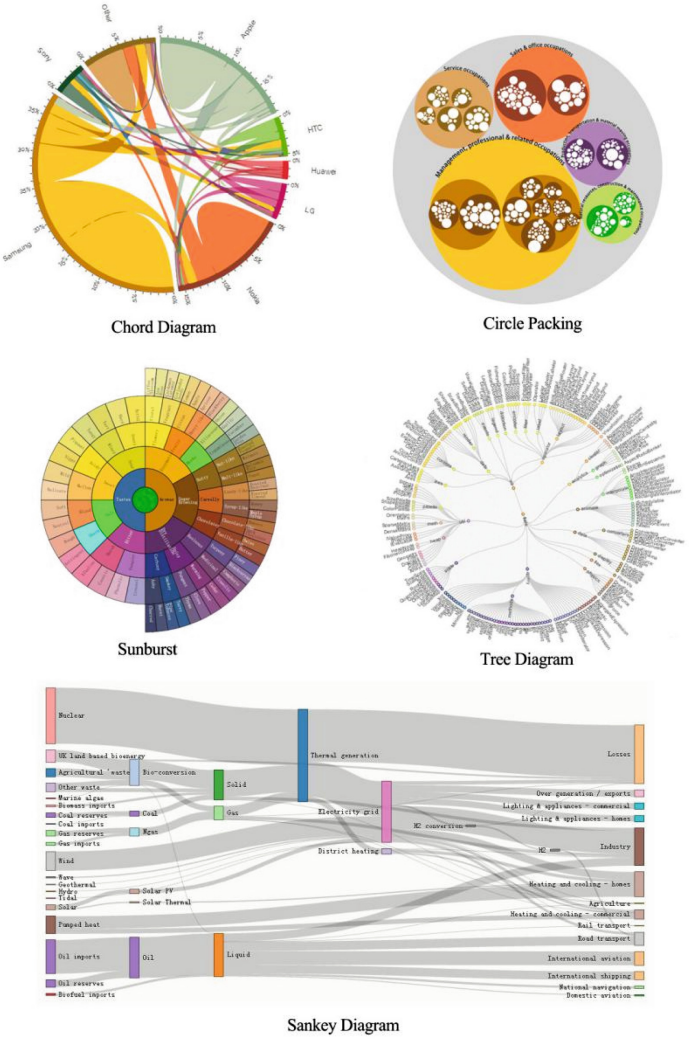

Fig. 1. Experimental materials (Chord Diagram, Circle Packing, Sunburst, Tree Diagram, Sankey Diagram)

The experiment steps are divided into two parts:

(1) The subjects read the prepared experiment material.
(2) The subjects mark the experiment material based on the five points scale method.

3.2 The Evaluate to the Cognition Degree of Subjects

The experimenter interview each subjects after the questionnaire finished. Then according to the experimental material of the entire information content, the experimenter judge the subjects' understanding of the experimental materials and mark the degree of subjects' understanding by using five points scale method.

4 Analysis and Deduction

The investigation result reveals:

(1) The method of information visualization does have an impact on the subjects' accurate understanding of the information content.

We conducted an ANOVA analysis of the data obtained from the three subjects "readability of information graphics", "degree of information obtained from information graphics" and "mastery of information integrity", Table 3 shows the results.

Table 3 shows that the average variance of subjects' scores on "the degree of readability of the information graphics" is 15.001, and the probability value of the corresponding null hypothesis "the subjects do not have any difference on the readability of the information graphics" is 0.000, less than the critical value of $\alpha = 0.01$, which shows that the null hypothesis is not established, that is, the degree of readability of these information graphics does exists the differences.

The results of the two ANOVA analyzes, "the degree of information obtained from the information graphics" and "the degree of mastery of information integrity" are similar to the results of the analysis of "the degree of readability of the information graphics". And the respective probability values of the null hypothesis are less than the critical value $\alpha = 0.01$, indicating that their respective null hypothesis does not hold. Which means that the two cases "subjects in different types of information obtained different levels of information", and "subjects have different degrees of mastery of information integrity" are exists indeed.

Table 3. The ANOVA analysis of the "degree of subjects' readable of infographic", "degree of subjects' information obtained from infographic" and "mastery of degree of experimenters' information integrity"

	Sum of squares	df	Mean square	F	Sig.
The degree of readability of infographic	60.004	4	15.001	20.959	0.000
The degree of access to information from infographic	18.624	4	4.656	4.405	0.002
The degree of mastery of infographic	22.131	4	5.533	8.267	0.000

Based on the results of ANOVA analysis of the three data types, concluded that same information through different visualization methods, the users have cognitive differences in terms of readability, comprehension, and integrity.

(2) The better visualization chart organized, the less ambiguous of the subjects incorrect interpretation on the chart of the ambiguous.

The bivariate correlation analysis are applied on the scores of "the degree of organization of information graph" and the scores of "the degree of readability of information graph", "the degree of information obtained from information graph" and "the degree of mastery of information integrity", finishing results are shown in Table 4.

From Table 4, it shows that there are significant correlation between "the degree of organization of information graph" and "the degree of readability of information graph", "the degree of mastery of information integrity" on the level of $\alpha = 0.01$, which corresponds to the correlation coefficient of 0.619 and 0.431 respectively. This shows that the higher degree of organization of the information graph, the more conducive to the participants to read the information, the subjects are also easier to think they have completely understand the information in the graph.

The null hypothesis that the correlation exists between "the degree of the organization of information graph" and "the degree of information obtained from the information graph" is 0.37, which is greater than the critical value of $\alpha = 0.05$, indicating that there is no correlation between them. That is, the degree of participants get information from the infographic has no relevance to the degree of the organization of graph. This is a bit different from our usual perception. In most cases, we think that the well-organized information may be more beneficial to users to obtain information. The unusual result may be related to the user's cognitive style, which users may extract information according to their own mode of thinking or focus, rather than read all the information step by step like a scanner.

Table 4. The partial correlations are on the subjects' marks on "degree of subjects' readable of infographic", "degree of subjects' information obtained from infographic" and "subjects' mastery degree of information integrity"

		The degree of readability of infographic	The degree of information obtained from infographic	The degree of mastery of infographic
The degree of organization of infographic	Pearson correlation	0.619	−0.075	0.431
	Sig. (2-tailed)	0.000	0.370	0.000

(3) The subjects are more familiar to the theme of the information, the less likely they have ambiguity in understanding.

We conducted a bivariate correlation analysis of "subjects' familiarity with the information involved in the infographic" and "the degree of mastery of information

integrity". Table 5 shows the results that there is a significant correlation between "subjects' familiarity with the information involved in the infographic" and "the degree of mastery of information integrity" at the level of $\alpha = 0.01$, the corresponding correlation coefficient is 0.601. This result is consistent with common experience. For example, from the time-share charts of stocks, we can see that subjects with relevant experience can get the meanings of curves, bars and other representations in the graph in a relatively short period of time. And the whole meanings can also be obtained in a relatively short period of time. This is because subjects are more familiar with the information content, it is easier to understand the infographic.

Table 5. The correlation analysis on "degree of subjectes' familiarity to involved infographic information" and "the degree of mastery of information integrity"

		The degree of mastery of infographic
The degree of familiarity with the information involved in infographic	Pearson correlation	0.601
	Sig. (2-tailed)	0.000

(4) The color collocation of the infographic and reading disorder caused by the language are not correlated with whether the subjects grasp of the information completely.

We tested participants' scores on the influence of "chart color on reading" and "the chart language", respectively. Bivariate correlation is conducted with those two scores with the score of "subject's degree of mastery of information integrity", the finishing results in Table 6. From Table 6, we can see that, the probability values of the null hypothesis that "the influence of chart color on reading" and "the reading disorder caused by language of chart" are related to "subjects' degree of mastery of information integrity" are greater than the critical value of $\alpha = 0.05$. It indicates that the fuzziness of reading caused by the color of the chart or graphic language does not correlate with completely mastered the information.

This conclusion seems inconsistent with our daily experience. For example, when we use a pie chart to represent the consumption of a physical quantity, we might give the consumed part a black color and the remaining part a white color. However, because subjects have different semantic relations to black and white in their daily life, some people may think that white is consumed, black persists, while others are opposite. In this case, it is difficult to judge whether all users have an accurate understanding of this information.

Therefore, we suppose that the relationship between the color of the infographic and the reading disorder caused by the language and the subjects' complete mastery of the information may also be influenced by the user's background knowledge and educational level. If participants have a high level of education or a background in design, they may self-correct some of the information in the cognitive system to reduce the ambiguity of the individual's understanding of the infographic. In this experiment,

one third of the participants had a design knowledge background and all of the participants were undergraduate students from Huazhong University of Science and Technology (China Top 10).

Table 6. The correlation analysis is done to "the reading impact degree of color of chart" and "the reading barrier caused by language of chart" with "subjects' mastery degree of information integrity" respectively.

		The degree of mastery of infographic
The reading impact degree of color of chart	Pearson correlation	−0.111
	Sig. (2-tailed)	0.176
The reading barrier caused by language of chart	Pearson correlation	0.059
	Sig. (2-tailed)	0.475

5 Conclusion

Based on the experimental study and analysis of the differences in user perception space under different visualization methods, this research finds out the method of information visualization, the information graph's own consecution degree, information content and the subjects' familiarity is the main cause of ambiguity in information visualization. And the language, color and collocation of the information graph do not affect the subjects producing ambiguity of information content in sensation space.

Acknowledgments. We would like to express my gratitude to all those who helped me during the writing of this thesis. We gratefully acknowledge the help of the 30 subjects who participated in the experiment, who has offered us valuable data in the academic studies.

We also owe a special debt of gratitude to all the companions in lab who have always been giving us valuable suggestions and critiques which are of help and importance in making the thesis a reality.

Last but not the least, my gratitude also extends to my family who have been assisting, supporting and caring for me all of my life.

References

1. Merriam-Webster's Learner's Dictionary. http://www.learnersdictionary.com/
2. Oxford Advanced Learner's Dictionary. http://www.oxfordlearnersdictionaries.com/
3. Haoliang, Z.: Basic theory and application of fuzzy mathematics. Mine Constr. Technol. (Z1), 70–80+96 (1994)
4. Tongtao, Z.: Fuzzy language and fuzzy psychology. J. Xiamen Liniversity **4**, 140–146 (1987)
5. Chanb, K.Y., Kwonga, C.K., Dillonb, T.S., Funga, K.Y.: An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness. J. Eng. Des. **22** (8), 523–542 (2011)
6. Wang, J.: Int. J. Prod. Res. **35**(4), 995–1010 (1997). 16 p.