Digestion Efficiency of Texts and Images in Information Transfer

Hiroaki Koike [1] and Teruaki Hayashi [1]

[1] *The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, Japan*

Abstract

The proliferation of information and communication technologies has augmented and diversified information sources. However, an increase in the volume and selection of information does not necessarily promote understanding [1]. In addition, conventional evaluations of information transfer have focused only on the arrival of information at receivers. They should adequately contemplate the recipients' comprehension of the data post-acquisition [2]. In this study, we propose the concept of "information digestion", which refers to receivers' adequate understanding of the acquired information. We proposed an evaluation model of information digestibility using hierarchical factor analysis and extracted factors that constitute digestibility using four types of media.

Keywords

Information digestion, Information transfer, Hierarchical factor analysis

1. Introduction

The receiver comprehends the content and intent of the information, with no misunderstandings or omissions. This can be distinguished from the arrival of certain information to the receiver. A series of information transfer flows of information digestion consisting of information, the sender, and the receiver is called an information transfer system. The efficiency of information digestion (ρ) is represented by Eq. (1), where $\omega_{\text{group}}(i)$ refers to the fraction of variation in the data that can be explained by the $No.i$ group factor F_i alone, and the evaluation of each group factor is denoted by $\text{ev}(F_i)$.

$$\rho = \sum_i \frac{\omega_{\text{group}}(i) \cdot \text{ev}(F_i)}{\sum_{j=0}^{n} \omega_{\text{group}}(j)} \qquad (1)$$

Applying the extracted group factors by media to Eq. (1), we obtain the respective quantitative evaluation equations, that is, Eq. (2) is obtained for the following four media: A, online news articles; B, online advertisements; C, online shopping; and D, articles/reports. The evaluation of F_i , each group factor in the equation, is calculated based on the subjective evaluation by several people of the degree to which the observed variable that constitutes each factor has the property that it represents.

$$\rho(A) = \text{ev}(Intimacy) \cdot 0.272$$
$$+\text{ev}(Unexploredness) \cdot 0.360$$
$$+\text{ev}(Simplicity) \cdot 0.378$$
$$\rho(B) = \text{ev}(Inclusiveness) \cdot 0.384$$
$$+\text{ev}(Simplicity) \cdot 0.312$$
$$+\text{ev}(Accessibility) \cdot 0.304 \qquad (2)$$
$$\rho(C) = \text{ev}(Intimacy) \cdot 0.279$$
$$+\text{ev}(Simplicity) \cdot 0.400$$
$$+\text{ev}(Unexploredness) \cdot 0.321$$
$$\rho(D) = \text{ev}(Accountability) \cdot 0.557$$
$$+\text{ev}(Simplicity) \cdot 0.443$$

2. Experiments

In Experiment 1, to devise a quantitative evaluation model of information digestibility, we conducted a hierarchical factor analysis [3] on the evaluation data (A, B, and C:400 cases; D:100 cases), which consisted of 22 observed variables

---

at three levels on each side and six levels in total, to extract factors and structures of information digestibility. We used a twin-factor model [4] with two layers: a general factor (g) that affects all the observed variables and a group factor (F_i) that affects clusters of specific observed variables. Equations (1) and (2) are based on the interpretation results of each extracted factor and $\omega_{\text{group}}(i)$.

To directly observe and evaluate information digestion, in Experiment 2, we collected free response data for information groups with different types (text, image, and complex) and amounts of information, and evaluated the digestion rate on a 5-axis, 4-point scale. The text information group (four types) consisted of different numbers of characters. The image information group (four types) consisted of differences in the plurality of subjects and complexity of backgrounds. The composite information group (eight types) consists of combinations of these types of information. We compared the digestibility of the information that demonstrated significant differences ($p < 0.05$) in the Scheffé test with the results of the evaluation.

In Experiment 3, two examples were considered for each medium and processed based on the model. In the process, we referred to the constructs and applied each group factor. We attempted to verify the criterion-related validity of the proposed model by analyzing the evaluation data of relative digestibility before and after processing.

3. Results and Discussions

Experiment 1 yielded A = 3, B = 3, C = 3, and D = 2 group factors. The estimated model explained over 90% of the variation in data. In Experiment 2, no significant differences were found among the different groups but differences were found within and among some of the textual/complex information groups (see Figure 1). Experiment 3 yielded a 60% approval rate for adopting the model in B, C, and D. However, in A, there was no significant difference between the two groups. However, for A, 56.3% of the respondents supported the model after processing one type of information, and 87.6% supported the model before processing the other type of information.
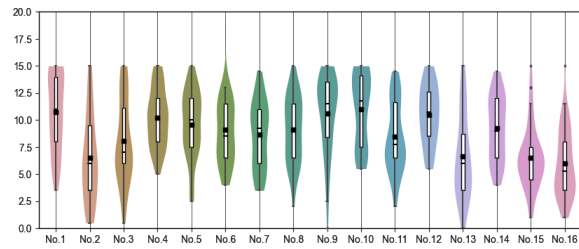

Figure 1: Distribution of scores in Experiment 2

The group factors for A were *Intimacy*, *Unexploredness*, and *Simplicity*, regarding the results of interpreting factors for each medium - *Intimacy* indicating excess prior knowledge, *Unexploredness* signifying no prior knowledge or preconceptions, and *Simplicity* indicating the concentration of main points and few non-main elements. The general factor was interpreted as the *Richness* (amount of information). Although *Intimacy* and *unexploredness* are opposite factors, they are not contradictory; rather, they indicate that the possession of incomplete knowledge is the most inefficient approach to digest knowledge. From the results of Experiment 2, it is clear that the most inefficient digestion occurs when the ratio of known elements is approximately 25%. From $\omega_{\text{group}}(i)$, it is found that being intimate is preferable to *Unexploredness* for A. Experiment 3 solely supports A, as the amount of information on the subject is limited pre-processing, and few factors impede digestion. The pre-processed version was more highly evaluated because the additional information was compressed or reduced by applying *Simplicity*. The utility of the factors may vary in situations of limited information, such as those in Experiment 2, where the digestibility of composite data was altered.

The group factors for B were *Inclusiveness* (i.e., the richness of information items, types, and supplementary information), *Simplicity*, and *Accessibility* (ease of understanding the main points). The general factor is interpreted as *low digestion cost* (thought and time effort required to understand the information). In B, it is desirable that the conclusions, main points, and explanations are made in various ways and concisely.

The combination of group factors for C was generally similar to that for A, but there was a difference in $\omega_{\text{group}}(i)$. Moreover, the general factor was *Satisfiability* (i.e., more niche needs should be met). C should contain various information but should be organized individually.

Finally, in D, *Accountability* (validity and understandability of the background leading to the conclusion) and *Simplicity* were extracted. *Simplicity* was the only factor extracted from all media. The factor of *Readability* implies that D requires a clear conclusion, novelty, and legibility with no lag in argument and background development. D had the largest model fit, which means that the factor has a high generality as a criterion for digestibility.

From the results of Experiment 2, it was found that for text alone, digestibility reaches the lowest at approximately 50 characters. Comparisons with the group of textual information revealed that the effect of the increase in the amount of information on the digestibility of textual information differed between the cases of textual information alone and textual information combined with images. Although no significant difference was found among the image information groups, and no difference in digestibility was due to the plurality of subjects or the complexity of the background, the increase in the amount of information in the image information has a greater effect on digestibility than that in the text information in the composite information.

Based on the above, considering the differences in the information handled in Experiments 1 and 2 from the viewpoint of the information transmission system, it may be possible to analyze the information across media and types by focusing on three points: purpose of information transmission, content ratio of information types, and common recognition of purpose within the information transmission system.

4. Acknowledgements

5. References

[1] M. Shalini & D. Stokols: Psychological and Health Outcomes of Perceived Information Overload, Env. and Behavior, 44(6), 2012.

[2] K.S. McCarthy & D.S. McNamara: The Multidimensional Knowledge in Text Comprehension framework, Educ. Psychologist, 56(3), pp.196-214, 2021.

[3] J.C. Westland: Structural Equation Models, Stud. Syst. Decis. Control 22.5:152, 2015.

[4] Flora, David B.: Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. Advances in Methods and Practices in Psychological Science 3.4: 484-501. 2020.