

On the Normalization of Data in Chinese Meanings

Luo Yan, Wang Youqun^{*}, Wu Yong

School of Science, Chongqing University of Technology, Chong Qing, China

Email address:

wangyouqun@cqut.edu.cn (Wang Youqun), wuyong@cqut.edu.cn (Wu Yong), 353151641@qq.com (Luo Yan)

^{*}Corresponding author

To cite this article:

Luo Yan, Wang Youqun, Wu Yong. On the Normalization of Data in Chinese Meanings. *International Journal on Data Science and Technology*. Vol. 2, No. 6, 2016, pp. 72-75. doi: 10.11648/j.ijdst.20160206.13

Received: October 24, 2016; **Accepted:** November 18, 2016; **Published:** December 20, 2016

Abstract: Data normalization is important for data analysis so there are many methods of data normalization. Also because of many differences in various disciplines and fields, the concepts and methods of data normalization in literature appear vague and confusing in Chinese translations. In this paper, data normalization and data standardization are two main terms used on the basis of a detailed analysis of the situation. We recommend the "data normalization" corresponding to Chinese translations. Then theoretically, the differences between the two terms are elaborated and the intrinsic meanings and features of data normalization are studied.

Keywords: Normalization, Standardization, Non-dimensionalization, Data Processing

1. Introduction

With the rapid development of Internet, cloud computing technology, the corresponding quantity of data shows an explosive growth trend, followed by the era of big data. For each field, there is a huge amount of data generated every day and in diverse forms. So people become to focus on how to manage and use these data in a better way.

Since data exist in diverse types from different sources, wide ranges of ways to process data exist. In general, the data should be processed before application. Nowadays, the term of "data normalization" / "biaozhun hua" (Chinese meaning) is used in the domestic and foreign research reports, academic papers and non-academic articles, and it is more greatly employed than ever before.

Foreign scholars have a clear definition and explanation of data normalization; it defined the basic difference with the data standardization. In the domestic literature, and even the authoritative literature, such as Baidu Encyclopedia, the two English terms, "data standardization" or "data normalization", are translated as "shuju biao zhun hua" in Chinese translations. It has no difference even in identical use. In addition to the terminology translation, it shows that the scholars on the study of data processing also have confusing concept of these two terms. In order to facilitate international communication, our research focus on the term difference and connection

between "data normalization" and "data standardization", hoping that our research and application on data processing or other related fields can play some role in reference.

In the following, we will firstly study the literature on the related terms of data standardization and data normalization; next, study on the analysis of these two terms in Chinese to look at their difference; then summarize the features and the connotation of data normalization; finally proposed "data normalization" in the corresponding Chinese translation.

2. Term Analysis and Literature Review

In data processing, the currently frequently used term in English literature is "data normalization". In Wikipedia, the term "normalization" means any process that makes things normal and formal. There are also the implementations of standards, specifications, rules or regulations intended to acts that contain a certain range of standardization. In statistics, it means the value, or the meaning of the adjustment of the distribution. As a result, the former means standardization into a unified form, with the meaning of normalizing the existing objects to a certain range, or unifying their titles. In the technical aspect, "data normalization" means that the general data is regulated to the canonical form. "Data

normalization" also means to convert a value to the range: 0-1. The "standardization" refers to the development and implementation of technical standards process, which indicates the standards set-up, follow-up and standard evaluation process. Therefore, this term has the meaning of establishing a standard for an object and implementation this standard. For example, technical standards used in "industrial, commercial in national and international" and other aspects, have the maximum compatibility, interoperability, security, repeatability and quality advantages. "Data standardization" is to convert the data for the mean to 0 and the variance of 1, it is important to have a certain "enforcement" meaning. From the analysis of semantics, in Oxford dictionary and the British dictionary, the word formation of "normalization" and "standardization" belong to derivation. The root of normalization "normal" means "normal, formal, and standard". Its verb is "normalize", indicating to make normal, standard, or standardized behavior and processing. The root of standardization "standard" indicates the scale used in the evaluation, the verb standardize represents standards, comparative standard and a test. It also means an established standard (the higher level of norm). "Standard" is originated in the modern industrial management. From the above analysis, based on the existing literature and term application by scholars, "data normalization" translated as "shuju guifan hua", or "data guifan hua", "data standardization" translated as "shuju biao zhun hua", or "data biao zhun hua" seems more in line with research in the field context at different semantic expression "data normalization" and "data standardization" requirements [1-4].

Due to the different understanding of the semantics and the accuracy of the translation, the term application in the Chinese literature is often more complex. According to the authority of CNKI database on CNKI home page, the retrieved index to the "normalization" appeared earlier in the paper of Gu, Z. C. on helium separable wave function to the

relevant function normalizes [5]; Zhu, H. f. etc. introduced "standardization", although not explicitly given the detailed explanation in Chinese meanings. In his article, he mainly used the term for standard of China insect name [6]; "Data standardization" in the earlier paper appears in Ying, Q. R.'s paper [7], such as "normalization" in the title and keywords; then "data standardization" appeared in G, Z.'s paper [8], in English title and keywords, respectively using the "data normalization" and "data standardization". With the rise and expansion of domestic data processing research, the confusion on these terms becomes more and more serious in Chinese meanings. Today on Baihu Encyclopedia, "data standardization" explanation in English is used "normalization". From its interpretation, its meaning should be "standardization". Baidu Encyclopedia confusedly uses "standardized" with both gerunds "standardizing" and "normalizing".

With the increase of domestic academic research and international communication, many Chinese scholars have a clear understanding of "normalization" and "standardization". Therefore, using the CNKI platform, the two terms in bilingual usage advanced search, clutter and terminology used for graphical display; results are shown in table 1:

Table 1. Literature terms number correspondence tables.

	Data biao zhun hua	Data guifan hua
Data Normalization	72	45
Data Standardization	501	79

Note: search platform: CNKI (in China); retrieval methods: "advanced search", the source for "journal, doctoral, master, important meetings, international conferences, special journals"; retrieval time: September 25, 2016.

The results of the literature search will be detailed decomposition to the year, as shown below:

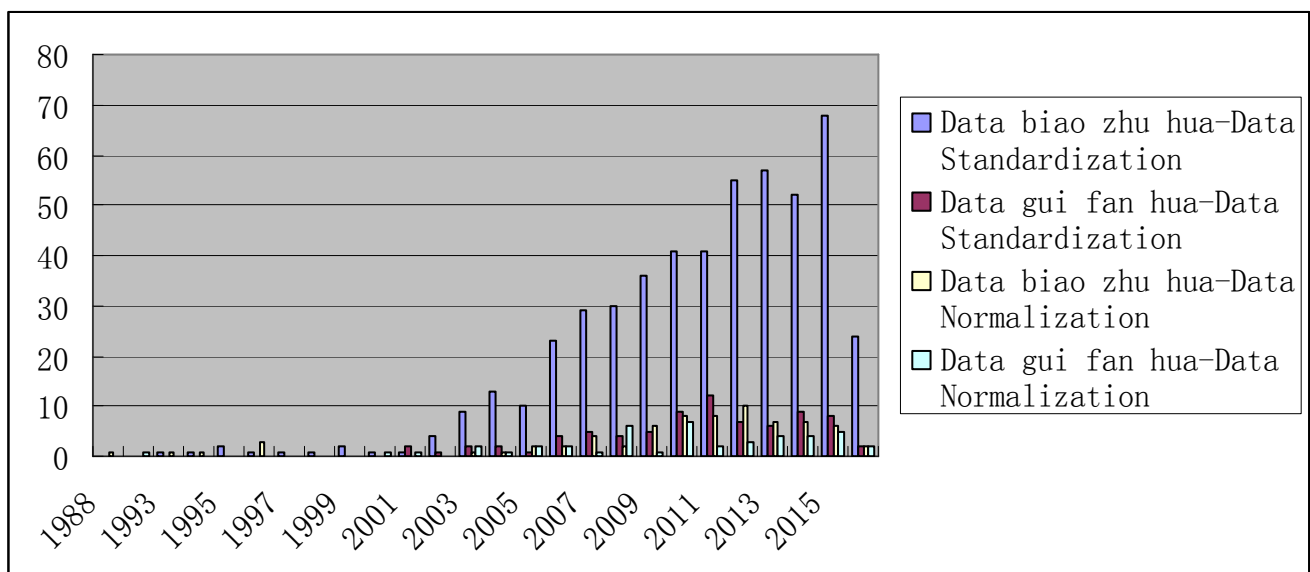


Figure 1. Term number corresponding relation between Chinese literature statistical figures.

Note: Figure 1 shows that in 2015 the data for the current data, the retrieval method is the same as in table 1.

In the library related terms usage, it shows the following problems. First, besides the obvious "data biao zhun hua-data standardization" which is a combination of the most frequently used, the other three combinations is a little visible. This means that the Chinese scholars have a big confusion about terminology; In the second place, from the number of usage cases, domestic scholars are more likely to use the term "data standardization" in their research topics, which means "data biao zhun hua" in Chinese; Thirdly, the domestic research on data processing began to change apparently from 2003.

3. Chinese Interpretation and Translation for Normalization and Standardization

According to modern Chinese dictionary, "biao zhun" is used as a noun, which means the criterion or guidelines to evaluate things, such as technical biao zhun, practice is the sole criterion for testing truth. When it is an adjective, meaning in line with guidelines and can provide a comparison norm to check for similar things, such as the biao zhun sound, biao zhun time, her pronunciation is biao zhun, etc. "Biao zhun hua" is mainly used as a verb, meaning that in order to adapt to the development of science and technology, to meet the needs for organizing production, the technical biao zhun is established for product quality control, biao zhun hua a variety of specifications, common parts, components, etc.. This is called biao zhun hua. "Guifan" is used as a noun, commonly known as agreed or stated biao zhun, such as language specification, code of ethics; when it is an adjective, it indicates guifan, such as the usage of the word but not guifan; Used as a verb, it makes normative, such as the use of new social morality to regulate people's behaviors, etc. Guifan hua also is a noun form of a verb, meaning that for certain criteria, such as language guifan hua, service industry should be guifan hua. Guifan hua is the "guifan" and a composite structure of the word "hua". According to Chinese literal meaning, it can be interpreted as "to make the guifan" and "alter, change". In the disciplines concerned, data guifan hua is the dimensionless of physics level or with chemokines, to eliminate the unit of data restrictions or heterogeneity, the original data can be converted to dimensionless or homogeneous data in order to be compared, weighted between different units or metric magnitude or nature. In the aspect of mathematics, it can be considered that guifan hua is to map the actual data values of individual indicators into a common relatively modest range, which determines a function formula between the various indicators of guifan hua [11].

From the above concept from analysis of data guifan hua, we can see that guifan hua of data should firstly go to determine whether the selected indicators/data complies with the requirements of guifan hua; then go to analyze the

characteristic of the selected indicators/data based on the principles of selection, reasonably selection of treatment methods; and finally go to analyze whether indicators/data guifan hua processing result is reasonable. Therefore, data guifan hua is to select the index data according to certain criterion (i.e. guifan hua method) for the corresponding transformation. The essence of data guifan hua is to change or remove the property and dimension of the data index, and other influencing factors.

Based on the comprehensive analysis of these two terms in Chinese and foreign languages, "biao zhun hua (standardization)" and "guifan hua (normalization)" have more complex relationship. We believe that the main difference between them is: in some sense, "guifan hua" is about common nature, technically considered lower than the level of "biao zhun hua", while "biao zhun hua" is the result of modification and abstract out of norm, having the meaning of compulsory execution. The relationship between these two has the characteristics of source and flow. "Biao zhun hua" has the characteristics of more mandatory, unique, authoritative, valuable and etc. Therefore, "guifan hua" and "biao zhun hua" should be treated differently.

In summary, we use data normalization (data guifan hua). First of all, according to the earliest use of each term on the CNKI retrieval platform, the correct meaning of the terms can be clearly distinguished; secondly, some domestic scholars have been aware of the confusion of the terms used and tried to solve by changing the terms. Secondly, some domestic scholars have been aware of the confusion of the terms used, trying to solve by changing the terms. For example, Wang X. J., Ma L. P. and etc. have tried to replace the terms in use by introducing the terms of "standardization, normalization, and uniformization" and other terms [9, 10]. In some sense, it will make people to understand the content more confused, such as "uniformization" which also has other meanings.

Therefore, according to the historical usage of relevant terms, semantic interpretation and the domestic and international literature review, the study suggests that "normalization" in data processing be translated as "guifan hua" and "standardization" as "biao zhun hua" to facilitate the normative researches, and the international communications.

4. Summary

Based on the study and analysis on the related content of normalization theory from the domestic and foreign scholars in recent years, the paper researchers have a clearer understanding of data normalization. By comparative analysis, the definition of the "data normalization" is defined. Furthermore, the difference between the terms is elaborated theoretically and proposition of "data normalization" is suggested as corresponding term in Chinese.

Due to the complex relationships of normalization and standardization, complexity of data standardization is also brought about. Coupled with the dynamic development, its

essence is also complex. We hope our research can initiate more valuable work.

Acknowledgment

The authors acknowledges the support from the National Social Science Foundation of China (No. 14BJY200)

References

- [1] Panigrahi, S., & Behera, H. S., Effect of Normalization Techniques on Univariate Time Series Forecasting using Evolutionary Higher Order Neural Network [J], International Journal of Engineering and Advanced Technology (IJEAT), 2013, 03: 208-285.
- [2] Nandakumar, A. J., & Ross, K. A., Score normalization in multimodal biometric systems [J], Pattern Recognition, 2005, 38, 2270-2285.
- [3] Shih, H. S., Shyur, H. J., Stanley Lee, E., An extension of TOPSIS for group decision making [J], Mathematical and Computer Modeling, 2007, 45, 801-813.
- [4] Chen, Y., Li, K. W., Xu, H. Y., Liu, S. F., A DEA-TOPSIS method for multiple criteria decision analysis in emergency management [J], Journal of Systems Science and Systems Engineering, 2009, 18 (04), 489-507.
- [5] Ku, T. Z., Separable wave function of the ground state of helium constructed from hylleraas wave function [J], Chinese journal of physics, 1945, 00: 21-26+79.
- [6] Chu, H. F., Kao, C. S., A Commentary on the Insects Mentioned in the Pen Ts'ao Kang Mu [J], Chinese Journal of insect, 1950, 02: 234-262.
- [7] Ying, Q. R., Normalization of Business Data [J], Shenzhen University Journal, 1987, 21: 18-23.
- [8] Zeng, G. Z., On Application of Data Normalization in cluster analysis [J], Journal of Shandong Polytechnic University, 1988, 01: 21-26.
- [9] Wang, X. J., The non-dimensional method of indexes in multi-index comprehensive evaluation [J], Population Research, 1993, 82 (4): 47-51.
- [10] Ma, L. P., Standardization of statistical data non-dimensional methods - the study of modern statistical analysis methods (three) [J], Beijing Statistics, 2000, 121: 34-35.
- [11] Wen, H. T., Ren, C. P., A New Study of Non-measurement in Evaluation of Enterprise's Performan [J], Economic issue, 2011, 06: 61-65.