

Determining ICT indicators using entropy theory

Narantuya Erkhembaatar*, Otgonbayar Bataa**

* Department of Communications engineering, Mongolian University of Science and Technology, Ulaanbaatar, Mongolia

** Department of Communications engineering, Mongolian University of Science and Technology, Ulaanbaatar, Mongolia

narantuya@must.edu.mn, otgonbayar_b@must.edu.mn

Abstract— A way based on so-called information entropy has been used for getting data in designing any research. In information theory, entropy is a measure of the average uncertainty of a random variable. These indicators, which are called core information and communication technology (ICT) indicators of the information society, are extracted by studying and analyzing and using content analysis, the entropy method, and the entropy weight coefficient method. The proposed methodology is developed as a reaction to the restriction of the non-existence of numerical values for some indicators in the provinces of our country. In this paper, we firstly investigate availabilities of use determining the weight based on Shannon entropy theory. Secondly, we determine the importance coefficients of the indicators, based on the stages of diffusion of ICTs and synthesize importance coefficients of indicators. Finally determining results of the core ICT indicators are presented.

Keywords—Information entropy, Entropy Weight, ICT indicators

I. INTRODUCTION

Information theory is a powerful tool in communication that provides the basis for designing efficient and reliable communication systems [1]. Today, information and communication technologies (ICT) have the potential to provide new solutions to the many challenges facing modernization societies worldwide.

In this study, definition was extracted from the statistical data of Communications and Information Technology Authority, National statistics office of Mongolia and Communication Regulatory Commission of Mongolia and in the existing data [2][3]. To this end, after extracting the indicators from the samples, those indicators that were used in more samples and where consensus was obtained from more data on these samples were given higher priority for selection as the core ones. The revised core list and two reference indicators have 46 ICT indicators. In comparison, in the 2005 list, there were 41 ICT indicators (and one reference indicator) [4] [5]. ICT today more than ever offer the potential to find new solutions to the many challenges facing the modernization of societies worldwide. The development of ICT indicators is an ongoing process and this list will be reviewed periodically. When countries get experience in ICT data collection and along with the development of policy

needs, indicators can be changed, canceled or added. There is no doubt that the rapidly developing information society requires statistics and indicators to monitor and tracking the access and use of ICT.

II. ENTROPY METHOD

A. Information entropy

The average volume of information received by a stochastic data source is determined by information entropy. The negative logarithm of pmf for value is the information entropy measuring. Random variables expected value are the information entropy, in way became the amount of information carried. The clarity of entropy used in information theory is directly analogous to the clarity used in statistical thermodynamics. In 1948 paper named "A Mathematical Theory of Communication The concept of information entropy" was submitted by Shannon [6]. Primary model of a data communication system is focused of main elements, such as source of data, a channel, and a receiver [7].

Entropy provides an absolute limitation on the shortest possible average compression length without loss of data generated by the source, and if the source entropy is less than the channel capacity of the communication channel, the data generated by the source can (at least theoretically, perhaps neglecting some practical considerations such as complexity system, necessary for data transmission, and the amount of time that may be required for data transmission). The logarithm of the probability distribution is useful as a measure of entropy, because it is additive for independent sources. For example, the entropy of a fair toss of a coin is 1 bit, and the entropy of m - tosses is m bits. Information entropy is measured in bits. The unit of the measurement depends on base of the logarithm that is applied to determine entropy. Entropy is zero when one result is determined. Entropy quantifies these considerations when the probability distribution of the original data is known.

B. Entropy Weight Coefficient Method

Shannon presented the information theory of entropy, based on the thermodynamic principle, where entropy is the degree of molecular breakage in matter for the first time. It was used as a measure of disorder, uneven distribution,

dependence degree or complexity of the system [8]. Entropy weighing is a method that consists of monitoring values of the evaluation index under objective conditions, can determine the purpose and order degree and efficiency, referring to the information entropy estimate. The weight from Entropy weighing method make ranks lists more objective [9]. This avoids the subjectivity of the weights of the various criteria, and therefore the results of the evaluation may better reflect the real situation [10].

C. Evaluation Model

To avoid subjective opinion, the model is built from measured index, where the entropy weights were determined from Shannon's theory of entropy [11]. There are the following specific steps:

1. Suppose that there are some things that need to be evaluated, their number is m . There is an object of evaluation n objects. The determining matrix of the evaluation model must be established.

$$R_{ij} = (r_{ij})_{m \times n} \quad (1)$$

Where R is the matrix, X_{ij} is the matrix elements. x is the value of the j^{th} sample of the i^{th} indicator.

2. This requires normalization. Actual operations:

$$\begin{cases} d_{ij} = x_{ij} / \max x_{ij}, \text{positive indicators} \\ d_{ij} = x_{ij} / \min x_{ij}, \text{negative indicators} \end{cases} \quad (2)$$

The new space matrix:

$$D_{ij} = (d_{ij})_{m \times n} \quad (3)$$

3. The relative weight of x_{ij} is:

$$\begin{aligned} Y_{ij} &= (y_{ij})_{m \times n} \\ y_{ij} &= d_{ij} / \sum_{i=1}^m d_{ij} \end{aligned} \quad (4)$$

4. Assume that there are estimating factors by m , estimating objects counted n , then forms an element matrix $R_{ij}=(r_{ij})_{m \times n}$. Since entropy has an extreme property [1] [12], the smaller the value of r_{ij} deviates from the equal, the more entropy there is, and the greater the uncertainty of the estimate. Shannon was submitted theory of information entropy, which is represented a measure of the average information value of a stochastic system, in his paper named "A Mathematical Theory of Communication" in 1948. Information entropy, which is based of the definition of entropy of the i^{th} indicator of the evaluation matrix Y is defined by Equation (5):

$$H_i = -\frac{1}{\ln n} \sum_{j=1}^n y_{ij} \ln y_{ij} \quad (5)$$

In addition, k -positive constant, determining $k=1/\ln(n)$, which is a normalizing positive constant for maintaining $0 \leq H_i \leq 1$.

5. The weight is calculated from the utility value of the index information defined by the EWCM. The utility value is

the largest, the largest value of the estimate. The weight of i^{th} index is determined by Equation (6):

$$w_i = \frac{1 - H_i}{\sum_{i=1}^m (1 - H_i)} \quad (6)$$

Computing results of equation (6) are used for extraction indicators.

III. WEIGHT DEFINITION BY SHANNON ENTROPY

The model built from measured index data is following specific series of steps for calculation. The evaluation criterion can be classified as an increasing or decreasing type depending on its attributes.

1. Assume there are some things need to be evaluated, their number is 48 ICT indicators. There are 22 objects of evaluation [13] and there are presented 21 provinces and the capital of Mongolia [2][3]. The determine matrix of the evaluation model is to built.

$$R_{ij} = (r_{ij})_{m \times n} \quad i = 1, \dots, 48, j = 1, \dots, 22 \quad (7)$$

2. The new space matrix is:

$$D_{ij} = (d_{ij})_{m \times n} \quad i = 1, \dots, 48, j = 1, \dots, 22 \quad (8)$$

3. The relative weight of x_{ij} is:

$$Y_{ij} = (y_{ij})_{m \times n} \quad i = 1, \dots, 48, j = 1, \dots, 22 \quad (9)$$

$$y_{ij} = d_{ij} / \sum_{i=1}^{48} d_{ij}$$

4. The information entropy is:

$$H_i = -\frac{1}{\ln 22} \sum_{j=1}^{22} y_{ij} \ln y_{ij} \quad (10)$$

5. The weight is:

$$w_i = \frac{1 - H_i}{\sum_{i=1}^{48} (1 - H_i)} \quad (11)$$

These equations calculations are used for extraction of core ICT indicators.

IV. EXTRACTION OF CORE ICT INDICATOR

A. First step

At this step, we first create a sample indicator matrix and calculate the importance coefficients of the indicators using the entropy method [14]. The main idea of making entropy very important in probabilistic and algorithmic theory of information is to convert the probabilities of events into values that can be used for their ranking [15]. Regarding the lack of

numerical values of most indicators, entropy can be a suitable method to define the importance coefficient of the indicators.

The importance coefficient of ICT indicators was computed applying the formula (12):

$$\lambda_i = \frac{w_i}{\sum_{i=1}^m (w_i)} \quad (12)$$

Table 1 shows the importance coefficients λ_i of each indicator based on entropy weight.

B. Second Step

Measuring ICT is appropriate and accurate, the cost effectiveness of ICT over time, affecting other aspects of the country, economic growth and development, technical preparation to achieve further steps forward [16].

Figure 1 shows, according to [16], at the country level of analysis there are three stages in the dissemination of ICT.

1. Stage of readiness for ICT: when technology is new to a country or region, the question of the willingness of its people to adopt a new technology is a decisive issue. Related measures should also ensure readiness in terms of business, infrastructure and the economy of the country. The types of measurements that need to be supported are primarily a search query and a preliminary assessment, as well as a comparative analysis.

2. The stage of the intensity of ICT: as acceptance increases, the intensity of acceptance, the intensity with which ICT-related activities (for example, e-commerce from business to business and e-government, initiatives, etc.) are becoming more urgent. Measuring ICT is likely to be important when ICT becomes more widespread in the country, but it has not yet reached the point where it is fully deployed and accepted, and when there is a perception of a digital divide. The type of measurement activity that is likely to be appropriate at this stage will formalize the request and analysis of the main effects.

3. This stage reflects the maturation of investments in ICT and their implementation. This is the stage when it becomes possible to assess the range of impacts that ICT can have for national economies and the business activities carried out in these countries. The most frequently observed approaches to measurement will be refinement of formal analysis, as well as retrospective analysis. Measuring the impact of ICT is likely to be of greatest interest in countries where ICTs are very well developed, but it is unclear whether the environment receives their greatest benefits. These stages of dissemination help us to measure the different needs for measurements for ICT. Based on our notion of S-shaped diffusion of ICT, we also classified the metrics for ICT in terms of the three observable stages. Most transnational differences in the digitization process are explained by the basic infrastructural conditions, which in turn lead to different intensity levels of acceptance in the digitization process. As far as the third aspect is concerned, namely the transformations in the economic and social systems, the

consequences have just occurred; therefore, an analysis of this would be inappropriate [17].

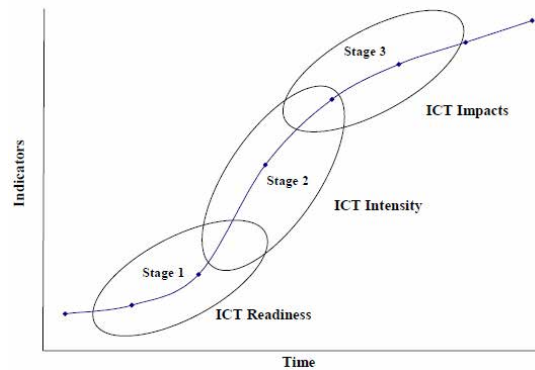


Figure 1. Different stages of ICT dissemination

Against this background, we have given the progress indicators the highest priority, and impact indicators have the lowest priority in measuring digitization. Since the exact value of the indicators can not be calculated at each stage of ICT dissemination, the use of nonparametric statistics is the most appropriate—the use of ranks instead of numerical values [18]. Nonparametric statistics suggest that the population is suitable for any parametrized distributions and is usually applied to populations that occupy a ranked order [19]. In this case, nonparametric statistics were used, because quantitative values of available indicators were not available, which would provide knowledge of their basic data distribution [20].

On the basis of the arguments just described, we assigned the highest rank to the readiness indicators and the lowest rank to the exposure level indicators, following the S-shaped diffusion of ICT. And also the importance coefficients μ_i of ICT indicators, determined by the S-shaped diffusion of ICT. Classification of indicators for the three aspects of digitization is based on the following definitions [21].

Readiness: preparation of technical, commercial and social infrastructure necessary to support digitization. The readiness indicators allow each country to compile a statistical picture of the availability status of the infrastructure needed for digital digitization.

Intensity: the use of ICT, the scope, value and use of ICT. The intensity indicators allow countries to determine who uses ICT and who does not, and identify leading sectors and applications.

Impact: added value, potentially created by ICT. Statistics are needed to assess whether ICTs are relevant in terms of efficiency and /or creating new sources of wealth.

The importance coefficients μ_i of core indicators based on S-shaped diffusion with the stages of readiness, intensity and impact are shown in Table1.

C. Third Step

In this section, λ_i and μ_i are synthesized by the following equation to determine final importance coefficients of ICT indicators [14]:

$$\rho_i = \frac{\lambda_i \cdot \mu_i}{\sum \lambda_i \cdot \mu_i} \quad i = 1, \dots, 48 \quad (13)$$

Table 1 shows the final importance coefficients ρ_i of the indicators.

TABLE1. CORE INDICATORS, IMPORTANCE STEP COEFFICIENTS

No	Index of ICT Indicators [5]	λ_i	μ_i	ρ_i
1	A1	0.01119	0.0112	0.006184
2	A2	0.03277	0.02715	0.043907
3	A3	0.02614	0.02715	0.035019
4	A4	0.02822	0.02715	0.037805
5	A5	0.02119	0.02169	0.02268
6	A6	0.03277	0.02715	0.043907
7	A7	0.02822	0.02715	0.037805
8	A10	0.01822	0.02169	0.019501
9	B1	0.02237	0.02715	0.029978
10	B2	0.02237	0.02715	0.029978
11	B3	0.02237	0.02715	0.029978
12	B4	0.02119	0.02169	0.02268
13	B5	0.02119	0.02169	0.02268
14	B6	0.02119	0.02169	0.02268
15	B7	0.0198	0.02169	0.021197
16	B8	0.0198	0.02169	0.021197
17	B9	0.0198	0.02169	0.021197
18	B10	0.01822	0.02169	0.019501
19	B11	0.01634	0.02169	0.017487
20	B12	0.01634	0.02169	0.017487
21	ED3	0.01406	0.02169	0.01505
22	ED4	0.01406	0.02169	0.01505
23	ED5	0.01406	0.02169	0.01505
24	ED6	0.01406	0.02169	0.01505
25	EDR1	0.03099	0.02715	0.041519

Each stage displays, depending to different value of importance coefficients core ICT indicators. As well different stages of readiness, intensity and impact requires different core ICT indicators.

D. Fourth Step

In this step, analysis of the 22 provinces used shows that the indicators mean number with a 95% confidence level is from 25 to 35:

$$\left(\bar{X} - \frac{s}{\sqrt{n}} z_{0.025}, \bar{X} + \frac{s}{\sqrt{n}} z_{0.025} \right) \quad (14)$$

$$= (25.26 - \frac{23.28}{\sqrt{22}} \cdot 1.96, 25.26 + \frac{23.28}{\sqrt{22}} \cdot 1.96)$$

Thus, 35 indicators with importance coefficients were selected as the core ICT indicators for our country. The 25 core ICT indicators, which have highest value of the selected

35 core indicators and importance coefficients of steps 1 and 2, and final importance coefficients are shown in the Table1. Also at this stage, two curves are shown when analyzing the core ICT indicators of countries compared to S-shaped diffusion with 3 stages.

Figure 2 shows, according to importance coefficients value of core ICT indicators based on the stages of readiness, intensity and impact, with exponential trend lines equations for developing country and developed countries [21].

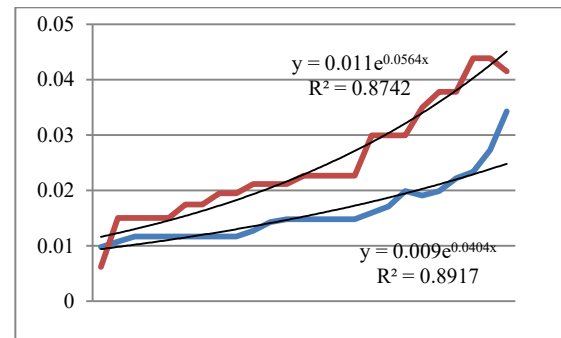


Figure2. Importance coefficients for core ICT indicators

The ICT indicators index is divided

- core indicators on infrastructure and access of ICT,
- core indicators of access to ICTs for households and individuals and their use,
- core indicators use ICT enterprises,
- core indicators in the sector ICT (production) and
- core indicators for international trade in goods ICT [4] [5].

V. CONCLUSIONS

Core ICT indicators introduce a convergence of information society indicators and digital divide models. It presents them as a standard set and those indicators are considered to be the basis for benchmarking, which are of paramount importance for information society metrics. However, not all countries are doing the same level of ICT development and they do not have well developed statistics collection and analysis systems. Thus, countries differ in terms of their needs and interest in measuring the level of digitalization, so for some of them-developing countries-measuring the level of infrastructure and access to ICT can be satisfactory, and for other developed countries they are interested in using ICT and the ensuing implications for the economic and social aspects of society [24]. These studies include determining the importance coefficients of indicators based on the method of entropy weighting coefficient and based on the stages of ICT distribution, the synthesis of the importance coefficients of indicators and determining the core ICT indicators.

This method is more efficient and practical. This paper describes the methodology for extracting the core ICT indicators of our country for 21 provinces (aimags) and the capital. In extracting the core indicators of information and communication technology, the fact that there are some

indicators of data in our country is the main limitation. However, core indicators, this method was practically feasible here. Therefore, the list of core indicators provided in the ITU index is not critical and the country should change and adjust it depending to its internal conditions [21].

The analysis on 22 selected objects shows that the mean number of core ICT indicators with 95% confidence level is between 25 and 35. Thus, the 35 indicators with the highest value of importance coefficients from the 48 ICT indicators were selected as the core ICT indicators for Mongolia. ICTs can play a significant role in the socio-economic development of a developing country [24]. The core ICT indicators curve consists of 3 different stages. Each stage is defined by different core ICT indicators. We have defined core ICT indicators for developed and developing countries.

REFERENCES

- [1] Tomas M. Cover and Joy A. Thomas, "Elements of Information Theory", Wiley-Interscience, 2006
- [2] Mongolian statistical year book, 2017, ISBN 978 - 99978 - 801 - 1 - 6
- [3] White book, Development of ICT and post, Ulaanbaatar, 2016
- [4] Organization for Economic Cooperation and Development, "Guide to measuring the information society", Partnership on Measuring ICT for Development, 2011
- [5] Core list of ICT indicators, March 2016 version, Partnership on Measuring ICT for Development, 2016
- [6] Jose C. Principe, Dongxin Xu, Qun Zhao, John W. Fisher, "Learning from Examples with Information Theoretic Criteria", 1999
- [7] Shannon C. and Weaver W., "The mathematical theory of communication", University of Illinois Press, 1949
- [8] William F. Egan, "Practical RF system design", Wiley-IEEE, 2003 ISBN 0-471-20023-9.
- [9] Zhaohong Wanga and Wei Zhan, "Dynamic Engineering Multi-criteria Decision Making Model Optimized by Entropy Weight for Evaluating Bid," International Symposium on Engineering Emergency Management 2011, Systems Engineering Procedia 5 (2012) 49 – 54
- [10] Jelena Jovanovic, Heena Shah, Aleksandar Vujovic and Zdravko Krivokapic, "Application of MCDM methods in Evaluation of Environmental Impacts," International Journal of Quality Research Volume 8, Issue 4, 517-532, 2014
- [11] Junfeng Tian, Tao Liu, Hongqiang Jiao, "Entropy Weight Coefficient Method for Evaluating Intrusion Detection Systems", International Symposium on Electronic Commerce and Security 2008, pp. 592-598, 2008
- [12] Bing He and Denghao Gao, "Variable Weight Recognition Model Of Atmospheric Quality Comprehensive Assessment And Its Application [J]," Environmental Engineering, 2001; (6): 57-58
- [13] Payam Hanafizadeh, Mohammad Reza Hanafizadeh, Mohsen Khodabakhshi, "A Methodology to Extract a New Set of Core Indicators of the Information Society" Journal of Information Technology Research, Volume 2, Issue 3, 2009, IGI Global
- [14] Saaty, T.L. "Decision making for leaders", Pittsburgh: RWS Publications. 1990
- [15] Kinsner, W., "Is entropy suitable to characterize data and signals for cognitive informatics?", International Journal of Cognitive Informatics, 2007.
- [16] Kauffman, R. J., and A. Kumar, "A critical assessment of the capabilities of five measures for ICT development", Working paper, Carlson School of Management, University of Minnesota, Minneapolis, 2005
- [17] Nicoletta Corrocher, Andrea Ordanini, "Measuring the digital divide: a framework for the analysis of cross country differences", Journal of Information Technology 17:9-19, 2002
- [18] Codoner, F. M., S. O'Dea, M. A. Fares, "Reducing the false positive rate in the non-parametric analysis of molecular coevolution", BMC Evolutionary Biology 8:106, 2008
- [19] Peter, T. Munro, and D.A. Somerton, "Maximum likelihood and non-parametric methods for estimating trawl footrope selectivity", ICES Journal of Marine Science 58:220-229, 2001
- [20] Core ICT indicators, Partnership on Measuring ICT for Development, International Telecommunication Union Place des Nations CH-1211 Geneva Switzerland, 2010
- [21] Payam Hanafizadeh, Mohammad Reza Hanafizadeh, Mohsen Khodabakhshi, "Extracting Core ICT Indicators Using Entropy Method", The Information Society 25:4, 236 – 247, 2009
- [22] Arieh Ben-Naim, "Statistical Thermodynamics Based on Information A Farewell to Entropy", World Scientific, 2008
- [23] Nicoletta Corrocher, Andrea Ordanini, "Measuring the digital divide: a framework for the analysis of cross country differences", Journal of Information Technology 17:9-19, 2002
- [24] Susan Teltscher, José Cervera-Ferri, "Producing ICT indicators in developing countries: challenges and initiatives", ESTADÍSTICA 63, 180, pp. 27-61, 2011



Narantuya Erkhembaatar graduated with a bachelor in telecommunication engineer from the Polytechnic Institute of Mongolia in 1989 and in 2002 completed a master's degree in technology at Andhra University in India. Bachelor degree thesis: "The study of electronic exchange in telecommunication network of Ulaanbaatar city". Master degree thesis: "Development of CBT on Satellite communication". Research topic for PhD degree is in information theory and distribution of information potential



Otgonbayar Bataa, in 1978 graduated from the Polytechnic Institute of Mongolia majoring in Radio communication engineer. Bachelor degree thesis: "Feasibility study of improving the efficiency of discrete information system". Master degree thesis (M.Sc) in 1995: "Some issues of speech synthesis". PhD degree thesis in 1996: "Study of Mongolian speech synthesis and applying it in telecommunication technics", in 2003, post Ph.D program thesis: "Optimal version of OFDM system frequency and time-distortion". Professor. Consulting engineer of Mongolia. Research topic: "Broadband, high speed integrated services technologies" (WiMAX, WiBro, Mobile IPTV, LTE etc).