JOURNAL OF THE KOREAN SOCIETY OF AGRICULTURAL ENGINEERS, MAY. 2007, pp. 69-75

한국농공학회논문집
제49권 제3호

Development of Interactional Information Measuring Technique Using Information Gathering and Utilizing System in Analysis of Correlations between Rural Amenities

Lee, Je-Myung* · Jung, Nam-Su · Lee, Jeong-Jae***,†**

Abstract

Defining the correlations between rural amenities, while difficult, is important for the adequate and efficient development of these amenities. In this research, an Interactional Information Measuring Technique(IIMT) using Information Gathering and Utilizing System(IGUS) was developed to objectively analyze the correlations between abstract ideas. In order to analyze correlations between rural amenities, this model used the Korean Dusan World Encyclopedia as IGUS and the relative interactional information was measured. The correlations between rural amenities were analyzed objectively using IIMT with the results satisfying the basic conditions of interactional information suggested in this research.

Keywords : Rural amenity, Interactional Information, Relative Interactional Information, Correlation Measure, Information Gathering and Utilizing System(IGUS)

I. Introduction

Rural amenities are attractive factors as a means of rural development, with the result that there are many attempts to improve the rural living environment and economy through the development of rural amenities. When developing rural amenities, it is necessary to consider the correlation between rural amenities, because amenities are complexly related to each other. The profit of rural development can be increased by considering the correlation between rural amenities and the opposite can be true. However, it is difficult to analyze the correlations of rural amenities objectively since rural amenities are abstract ideas. Qualitative Analysis Methods using surveys(ex. Analytic Hierarchy Process; AHP) can be used as a means of correlation analysis. However, although the survey conditions are the same, the results given by Qualitative Analysis Methods can differ, making it difficult to reproduce the same results these methods.

There has been much research about the objective analysis of correlations using statistical methods to overcome the difficulties of using Qualitative Analysis Methods.

After the research of Church & Hanks(1990),

* Seoul National University graduate school
** Kongju National University Bio-Industry Engineering
*** Seoul National University Rural System Engineering
† Corresponding author. Tel.: +82-2-880-4581
Fax: +82-2-873-2087
E-mail address: ljj@snu.ac.kr

many research areas, such as information retrieval and natural language processing, have proceeded to measure the correlation between words by applying mutual information(Lee, 2003). Traditional methods of mutual information estimate mutual information by the calculation of the concurrent probability of words within the retrieval data sets, resulting in different results if the retrieval data sets are different(Kim, 1994; Kang, 1997; Heo et al., 2006). Traditional methods calculate the concurrent probability of words based on the spelling of words and not the meaning, leading to measurement errors produced by ambiguous words(Heo et al., 2006).

In this research, Interactional Information Measuring Technique(IIMT) using Information Gathering and Utilizing System(IGUS) was developed to analyze the correlations between abstract ideas objectively. IIMT was applied to the measurement of the correlations between rural amenities for objective analysis.

II. Information and Mutual Information

1. Information

Shannon(1948) researched methods to measure information uncertainty as a function of probabilities. If $p(x)$ means an appearance probability of event x, and x is an event of random variable X, then average self information $H(X)$ can be defined as eq. (1) and is called the entropy(Shannon, 1948).

$$H(X) = -\sum_{x} p(x)\log_2 p(x),\ x \in X \quad \text{(Shannon, 1948)} \quad (1)$$

where $H(X)$ is the entropy of set X and $p(x)$ is the appearance probability of element x in set X. Entropy $H(X)$ represents the average number of times to select each event in random variable X, and it also means the number of bits which are necessary for the translation of random variable X as information (Jung, 2006).

After Shannon's research, Gell-Mann & Lloyd (1996) researched methods to measure the worth of an entity by adapting Shannon's information theory, and Jung(2003) presented Information Measure Technique(IMT) using Information Gathering and Utilizing System(IGUS) to estimate the information of natural language. These methods can estimate the total quantity of one entity as information. However they cannot measure the correlation between two entities.

2. Mutual Information

Mutual Information(MI) refers to the information that two random variables hold in common. Mutual Information is defined as eq. (2) by the ratio of the concurrence probability $p(x, y)$ to the appearance probability $p(x)$, $p(y)$ (Fano, 1961).

$$MI(X, Y) = \sum_{x}\sum_{y} p(x,y)\log_2 \frac{p(x,y)}{p(x)p(y)},$$
$$x \in X \text{ and } y \in Y \quad \text{(Fano, 1961)} \quad (2)$$

where $MI(X, Y)$ is the mutual information, $p(x)$ is the appearance probability of element x, and $p(x, y)$ is the concurrent probability of elements x and y. The relation between information and mutual information is illustrated in Fig. 1, and mutual information has to satisfy three basic requirements illustrated at eq. (3)~(5).

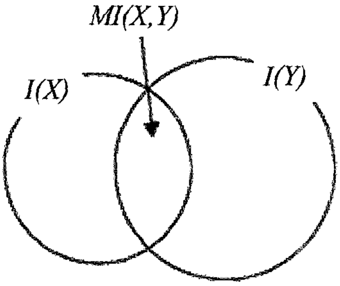

Fig. 1 Diagram of Mutual Information

$$0 \leq MI(X,Y) \leq min[H(X), H(Y)] \quad (3)$$

$$MI(X,Y) = MI(Y,X) \quad (4)$$

$$MI(X,X) = H(X) \quad (5)$$

where $MI(X,Y)$ is the mutual information and $H(X)$ is the entropy of set X. The boundary of mutual information at eq. (3) is $0 \leq MI(X,Y) \leq min[H(X), H(Y)]$ and is determined by the size of the entropies $H(X)$, $H(Y)$. To solve this problem, Lee (2003) suggested the concept of Relative Mutual Information(RMI) and presented eq. (6) to normalize the mutual information.

$$RMI(X,Y) = \frac{MI(X,Y)}{\sqrt{H(X)\,H(Y)}} \quad \text{(Lee, 2003)} \quad (6)$$

where $RMI(X, Y)$ is the relative mutual information, and $H(X)$ is the entropy of set X. Mutual information was used to analyze the property of Korean syllables (Lee & Oh, 1989) and it was also used to analyze the collocation of language (Church & Hanks, 1990). Likewise, mutual information was widely applied to information retrieval and analysis of words correlations(Kim, 1994; Kang, 1997; Lee, 2003). Traditional methods of identifying mutual information estimate mutual information by the calculation of concurrent probability of words within the retrieval data sets. However, although the measured entities are same, these methods present different results if the retrieval data sets are different. Furthermore, by using traditional methods, measurement errors are produced due to the ambiguity of some words.

In this research, the concept of mutual information was adapted to develop the Interactional Information Measuring Technique(IIMT) using Information Gathering and Utilizing System(IGUS) to analyze the correlations between abstract ideas objectively. To eliminate the effects caused by the size of entropy, the concept of the relative mutual information illustrated in eq. (6) was applied to the relative interactional information in eq. (10).

III. Development of an Interactional Information Measure Technique (IIMT)

The IIMT presented in this research the interactional information based on the concept of mutual information is calculated by using the data gathered from a particular boundary, which is defined by IGUS. The IIMT differs from previous mutual information methods as IIMT estimates the correlation between words by using the data of words defined in IGUS.

1. Development of an Interactional Information Equation

In this research eq. (7) was suggested as an equation of interactional information for IIMT which uses the data gathered from IGUS.

$$II(X,Y) = \sqrt{\sum_y p_X(y)\log_2\frac{1}{p_X(y)} \times \sum_x p_Y(x)\log_2\frac{1}{p_Y(x)}} = \sqrt{\sum_i p_X(i)\log_2\frac{1}{p_X(i)} \times \sum_i p_Y(i)\log_2\frac{1}{p_Y(i)}} \quad (7)$$

where $x \in X$, $y \in Y$ and $i \in (X \cap Y)$, $II(X, Y)$ is the interactional information, and $p_X(y)$ is the appearance probability of elements y in random variable X. x is an element of random variable X, y is an element of random variable Y, $p_X(y)$ is the appearance probability of y in X and $p_Y(x)$ is the appearance probability of x in Y. $p_X(y)$ and $p_Y(x)$ are defined as eq. (8)~(10). By using interactional information equation presented in this research, the interactional information between oneself $II(X,X)$ becomes entropy $H(X)$, so eq. (7) satisfies the basic requirements.

$$p_X(y) = \frac{N_X(y)}{\sum_x N_X(x)} = \frac{N_X(y)}{N(X)}, \ x \in X \text{ and } y \in Y \tag{8}$$

$$p_X(x) = \frac{N_X(x)}{\sum_x N_X(x)} = \frac{N_X(x)}{N(X)} = p(x), \ x \in X \tag{9}$$

where $N_Y(x)$ is the number of element x in set Y, and $N(X)$ is the number of total element in set X. Relative interactional information equation eq. (10), which is given by adaptation eq. (6) in eq. (7), was presented to solve the problem of the size of the entropies $H(X)$, $H(Y)$ determining the boundary of interactional information.

$$RII(X, Y) = \frac{II(X,Y)}{\sqrt{H(X)\,H(Y)}} = \frac{\sqrt{\sum_i p_X(i)\log_2\frac{1}{p_X(i)} \times \sum_i p_Y(i)\log_2\frac{1}{p_Y(i)}}}{\sqrt{\sum_x p_X(x)\log_2\frac{1}{p_X(x)} \times \sum_y p_Y(y)\log_2\frac{1}{p_Y(y)}}} \tag{10}$$

where $RII(X,Y)$ is the relative interactional information, and $H(X)$ is the entropy of set X.

$$0 \le RII(X,Y) \le 1 \tag{11}$$

$$RII(X,Y) = RII(Y,X) \tag{12}$$

$$RII(X,X) = 1 \tag{13}$$

The relative interactional information defined in this research satisfies the basic requirements of eq. (11)~(13). In this research, eq. (10) was used to measure the relative interactional information, and X is the rural amenity and x is the word that explains X in the IGUS.

2. Interactional Information Measure Technique (IIMT)

The IIMT model based on relative interactional information is illustrated in Fig. 2. Items to measure for correlation were selected and data on the items was gathered from the IGUS. Na et al. (2005) revealed that the character of IGUS has an effect on the result of information measure, so an appropriate IGUS should be chosen depending on the object of research. Next, nouns are extracted from the data gathered from IGUS. $p_X(y)$ and $p_Y(x)$ are calculated and the relative interactional information is estimated. If relative interactional information is close to '1' this is denotes a high correlation, while if it is close to '0' this is indicates low correlation.

In this research, the relative interactional information was measured using an IGUS, which provides the same information regardless of observer, to estimate the same results regardless of researcher. The relative interactional information between the data of items was measured. To solve the problem of the boundary of interactional information being determined by the size of the entropies the interactional information was normalized as relative interactional information.

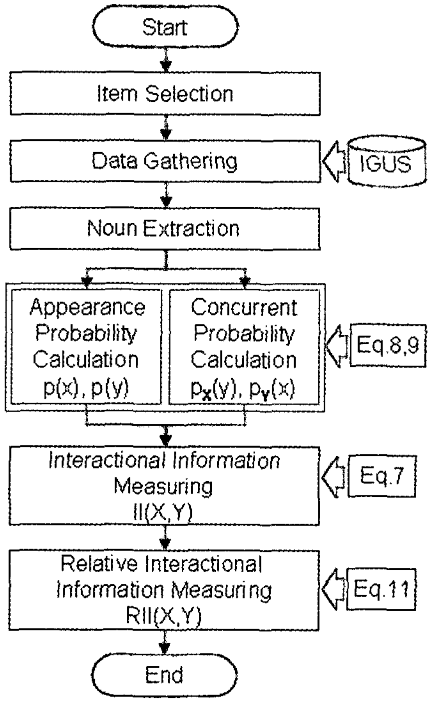

Fig. 2 Flow Chart of IIMT

Table 1 Category of rural amenity resources

Rural amenity resources	Description
Water resources	river, reservoir, lake, spring, well, dam etc.
Geography resources	marsh, habitat, rocks, cliff, peak, waterfall etc.
Plant resources	forest, plant community, protected trees, old and large trees etc.
Animal resources	mammal, birds, fishes, insect species, reptile, invertebrate etc.
Environment resources	natural environment, noise, pollution etc.
Tradition resources	tradition structure, tradition house, religious place, traditional street, symbol or great-man of rural community, legend etc.
Landscape resources	landscape of residential · agricultural area, landscape of mountain · river and forest etc.
Community resources	rural-urban exchange, cultural activities, community activities, publicity activities etc.

Data) Choi et al., 2006

IV. Application of IIMT

In this research, IIMT was applied to rural amenities to analyze correlations between water resources, geography resources, plant resources, animal resources, environment resources, tradition resources, landscape resources and community resources which were categorized in Choi et al. (2006). The category of rural amenity resources is illustrated in Table 1.

1. Selection of Information Gathering and Utilizing System (IGUS)

Information was gathered in an IGUS. However, as previously mentioned, care must be taken in the selection of the IGUS due to the influence that the IGUS has on the result of the information measure (Na et al., 2005). If an IGUS provides the same information regardless of the observer then the same results can be estimated regardless of the researcher. An encyclopedia was selected as an IGUS since it contains an enormous quantity of information and provides the same information regardless of the observer. Among the encyclopedias, the Korean Doosan World Encyclopedia was utilized as the IGUS because of the ease of gathering information from this resource.

2. Results

Information regarding eight categories of rural amenity resources was gathered from the Korean Dusan World Encyclopedia and the nouns were extracted. Relative interactional information was measured by using extracted nouns and eq. (7)~(10), and the result is shown in Table 2. The relative interactional information between itself shows the maximum value '1.0', the boundary is from '0.0' to '1.0' and it satisfies $RII(X,\ Y)$ =

$RI(Y, X)$. So the relative mutual information measured by IIMT satisfied the three basic requirements eq. (11)~(13).

The result was analyzed on the assumption that relative interactional information which is close to '1' denotes high correlation and that which is close to '0' means a low correlation. Water resources show a high relation with plant resources(0.131) and landscape resources(0.127), indicating that water resources such as rivers, streams and lakes have a large effect on the plant and landscape resources such as mountains and forests. Animal resources also have a strong relation with plant resources (0.366), showing that a wide variety of plants make for a more varied animal kingdom. Community has a high relation with tradition resources(0.194), indicating that traditional culture have a great influence on rural communities. Landscape resources has a high relation with water resources(0.127), geography resources(0.115) and plant resources (0.102), showing that a good landscape consists of plentiful water resources, plant resources and beautiful geography resources.

The results shown in this research were not compared with the results of other research because there is no sufficient research about the correlations between rural amenities. However, the correlations between rural amenities were analyzed objectively using IIMT.

Table 2 Relative interactional information between rural amenity resources

Category	water	geo-graphy	plant	animal	environ-ment	tradition	land-scape	com-munity
water	1.000	0.066	0.131	0.073	0.086	0.088	0.127	0.085
geography		1.000	0.067	0.097	0.099	0.009	0.115	0.081
plant			1.000	0.366	0.049	0.115	0.102	0.080
animal				1.000	0.073	0.065	0.060	0.058
environment					1.000	0.055	0.052	0.095
tradition		symmetry				1.000	0.089	0.194
landscape							1.000	0.084
community								1.000

V. Conclusion

Rural amenities are attractive factors as a means of rural development, with the result that there are many attempts to improve the living environment and economy of rural areas through the development of rural amenities. When developing rural amenities, it is necessary to consider the correlation between each amenity resources because amenities are complexly related to each other. However, difficulties arise in the objective analysis of the correlations of rural amenities since rural amenities are abstract ideas. Qualitative Analysis Methods using surveys (ex. Analytic Hierarchy Process; AHP) can be used as a means of correlation analysis. However, although the survey conditions are the same, the results given by Qualitative Analysis Methods can differ, making it difficult to reproduce the same results. Traditional methods of mutual information estimate mutual information by the calculation of the concurrent probability of words within the retrieval data sets, resulting in different results if the retrieval data sets are different. And traditional methods calculate the concurrent probability of words based on the spelling of words and not the meaning, leading to measurement errors produced by ambiguous words.

In this research, Interactional Information Measuring Technique(IIMT) using Information Gathering and Utilizing System(IGUS) was developed to analyze the correlations between abstract ideas objectively. Interactional information equation was developed to consider the information gathered from IGUS and relative interactional information was suggested. IIMT was applied to measurement of the correlations between rural amenity resources, for objective analysis.

The interactional information measured by IIMT satisfied the basic requirements suggested in this research, and the correlations between rural amenity resources were analyzed objectively by applying IIMT. The conclusion is that IIMT can be utilized in cases where it is difficult to apply qualitative analysis methods or where verification is needed. In development of rural amenities, it will be possible to consider the correlation between rural amenities by using IIMT and it will help more efficient, more synthetic development of rural community.

This research was supported by 'Application and Management Techniques on Rural Amenity Resources Databases' (Rural Resources Development Institute)

References

1. Choi, J.Y., J.J. Lee, D.G. Lee, H.J. Kim, Y.C. Han, Y.G. Oh, J.M. Lee, E.J. Yoon, W.H. Nam, J.M. Lee, H.J. Lee, T.G. Kim & S.H. Kwub, 2006, Application and Management Techniques on Rural Amenity Resources Databases, Rural Resources Development Institute, Korea
2. Church, K.W. & P. Hanks, 1990, Word Associatioin Norms, Mutual Information, and Lexicography, Computational Linguistics, 16(1), pp.22-29.
3. Fano, R.M., 1961, Transmission of Information: A Statistical Theory of Communications, M.I.T. Press., New York.
4. Gell_Mann, Murray & Seth Lloyd, 1996, Information Measures, Effective Complexity, and Total Information, Complexity, John Wiley & Sons, Inc., pp.44-52.
5. Heo, J., H.C. Seo & M.G. Jang, 2006, Homonym Disambiguation based on Mutual Information and Sense-Tagged Compound Noun Dictionary, Korean Information Science Society, 33(12), pp.1073-1089.
6. Jung, N.S. & J.J. Lee, 2006, A Cultural Property Priority Assessment Using Information Measure Technique, Journal of The Korean Society of Agricultural Engineers, 48(1), pp.41-48.
7. Jung, N.S., 2003, Development of an information measure technique for rural facilities location and amenities assessment, Seoul National University, Korea.
8. Kang, H.K., 1997, Two-level Document Ranking Methods Using Mutual Information in Natural Language Information Retrieval, Korea Advanced Institute of Science and Technology, Korea.
9. Kim, P.G., 1994, Automatic Indexing of Compound Words Based on Mutual Information for the Korean Information Retrieval, Seoul National University, Korea, pp.36-73.
10. Lee, J.H. & S.H. Oh, 1989, Entropy and Average Mutual Information for a 'Choseong', a 'Jungseong', and a 'Jongseong' of a Korean Syllable, The Institute of Electronics Engineers of Korea, 26(9), pp.1299-1307.
11. Lee, J.Y., 2003, A Study on Relative Mutual Information Coefficients, Korean Society for Library and Information Science, 37(4), pp. 177-198.
12. Na, J.Y., N.S. Jung & J.J. Lee, 2005, Development of a Ranking Model of Evaluation Indexes in Public Works, Journal of The Korean Society of Agricultural Engineers, 47(2), pp.25-32.
13. Shannon, C.E., 1948, A Mathematical Theory of Communication, The Bell System Technical Journal, 27, pp.379-423.