Relationship between Subjective Degree of Similarity and Some Similarity Indices of Fuzzy Sets

Ayumi YOSHIKAWA and Takeshi NISHIMURA

Faculty of Engineering and design, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto 606 JAPAN
Phone: +81-75-724-7496 Fax: +81-75-724-7400
E-mail: ayumi@jupiter.dj.kit.ac.jp

Key Word: fuzzy distance, t-norm, verbal hedge, similarity index

Abstract: *We evaluate 16 mathematical similarity indices between fuzzy sets by comparison with the corresponding subjective degree of similarity. The results obtained under two different experimental conditions show that two of the indices concerning the distance of fuzzy sets describe the subjective degree of similarity most.*

1. INTRODUCTION

Significance of this study: In systems that process vague objects as fuzzy sets, an output expressed in fuzzy set is often converted into the corresponding verbal expression, such as *very,* in order to help their user to understand its extent [1]. The conversion is executed with mathematical similarity indices calculated from the two fuzzy sets for the output and the verbal expression. Some indices have already been proposed theoretically [2], and several of them were used for a linguistic approximation [1]. However, to convert adequately, we must select and utilize the indices corresponding to the user's subjective degree of similarity.

The previous studies: From the above viewpoint, *Zwick et al*. [3] experimentally evaluated 19 similarity indices (distance measures, in ref. 3) using six verbal probability phrases as stimuli and recommended the four measures calculated mainly from boundaries of α-cuts. Although the comprehensive indices were adopted, the indices not used in their experiment are worth testing.

Aim of this paper: A final goal of this study is to find out mathematical similarity indices corresponding to subjective judgment of similarity between vague objects. In this paper, as the beginning of this study, we aim to clarify the relationship between the subjective degree of similarity and 16 mathematical similarity indices, using the verbal expressions of *tallness* and *heaviness* with eight verbal hedges separately. First, the performance of each index is evaluated based on two correlation coefficients and discrepancy degrees. Next, the consistency in the results obtained from two experiments is discussed.

2. EXPERIMENTAL CONDITIONS AND METHODS

Procedure: Two experiments, named Exp. 1 and 2 respectively, were executed separately. Both of them, respectively, consist of the following two tasks: identifying membership functions (IMF) and similarity judgement (SJ). The procedures of the Exp. 1 and 2 are basically the same except for the stimuli used. So we mainly explain the Exp. 1 and point out only the differences in the Exp. 2 (these differences are shown in parenthesis below).

IMF task: Our subjects identified 21 (17) membership functions (MFs) for the phrases of subjective tallness (heaviness) that consist of verbal hedges and either *tall (heavy)* or *short (light)*, using the BASE method [4]. The MF obtained is a trapezoidal MF on the universe of discourse (abbr. UOD) that ranges from *perfectly short (light)* to *perfectly tall (heavy).*

SJ task: The subjects rated subjective degrees of similarity for 28 pairs of the eight phrases related to *tall* (*heavy*) employed in the IMF task, using a categorical rating scale of *dissimilar-similar* with six verbal ticks. The corresponding hedges of the phrases were used in both of the experiments. In the Exp. 1, each subject selected a set of ticks which he/she thought easy to rate from the hedges used in the IMF task. Meanwhile in the Exp. 2, all the subjects used the same set of ticks due to the limitation of the experimental design.

Subjects: Seventeen and twenty native speakers of Japanese participated as volunteer in the Exp. 1 and 2 respectively. Therefore all the experimental materials were printed and displayed in Japanese. Number of subjects submitted effective answer were 13 and 10 respectively. Three of them participated in both of the two experiments.

Mathematical similarity indices: Sixteen mathematical similarity indices employed are divided into two groups: the one concerning overlapping degree between two fuzzy sets and the other concerning distance between them.

As the former we adopted four t-norms: logical, algebraic, bounded, and drastic product set (abbr. *LP*, *AP*, *BP*, and *DP*, in order). From each t-norm, we calculated three values: the height of each product set, the relative cardinality of it, and the cardinality ratio of it to the corresponding logical sum set (abbr. *hgt-*P*, *RelC-*P*, and *CRLS-*P*, in order). Consequently, 12 indices were employed.

As the later we selected four distances between fuzzy sets. One was distance between the gravity centers of them (abbr. *Dist-gc*). The others were derived from fuzzy distance calculated in the extension principle [2]. We adopted the greatest lower boundary and the least upper boundary of unity-cut of the fuzzy distance, and the gravity center of it (abbr. *GLB-FDL1*, *LUB-FDL1*, and *Gc-FD*, in order). All the indices were defined by subtracting the distances from unity, diameter of the UOD. Note that all the distances in this paper are determined by elements along the UOD (cf. ref. 3).

For each pair of the phrases and each subject, the 16 indices were calculated. The values of them can range from zero to unity and give the highest degree of similarity at unity.

3. RESULTS AND DISCUSSION

3.1 Correlation between subjective degree of similarity and mathematical similarity indices

Two correlation coefficients: Subjective rating values of similarity obtained from the SJ task, that is, six verbal ticks, are at the level of ordinal scale, while all the mathematical similarity indices are at the level of interval scale. So we use two ways to examine the relationship between them.

The one is to utilize the *Spearman's rank correlation coefficient*: r_s. For each index and each subject, we ranked 28 pairs of the phrases in themselves based on their index values and calculated the rank correlation coefficients between the within rank order of the pairs and their subjective rating values.

The other is to employ the *Pearson's correlation coefficient*: r_p indirectly. Suppose that both the UOD of *similarity* in the category rating scale and that of *tallness* (or *heaviness*) in the IMF task are *isometric*. Then a MF for a rating category of similarity results in the MF for *tallness* (or *heaviness*) with the same hedge used for the rating category, for example, by changing the label of the UOD for *similar* from *tall* (See ***Figure 1***). This procedure works as a so-called psychological scaling of rating categories. For each index and each subject, we calculated the Pearson's correlation coefficients between the index values of the pairs and the scale values of subjective similarity estimated from the gravity centers of the MFs for the corresponding categories (See *Figure 1*).

Results: ***Table 1*** shows mean values, SDs, and minimum values of two correlation coefficients over

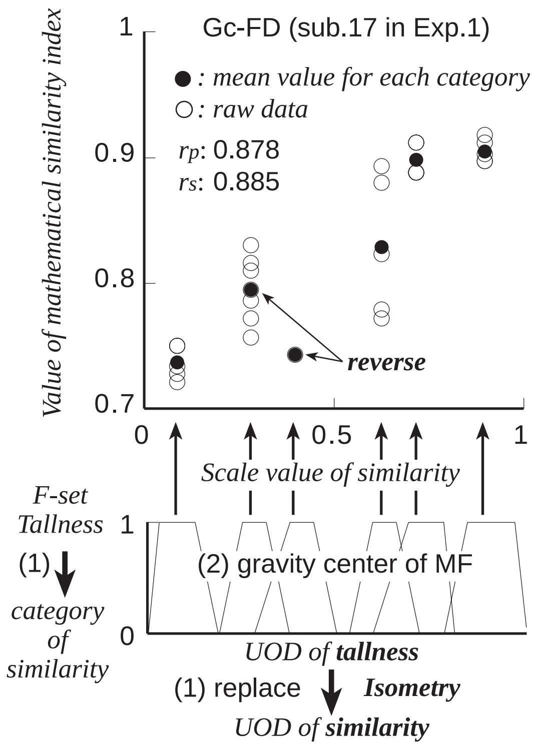

Figure 1: An example of the results (upper) and illustration of scaling the categories (lower)

Table 1: Relationship between subjective degree of similarity and 16 mathematical similarity indices

Variables	Spearman's correlation coefficient: r_s						Pearson's correlation coefficient: r_p						Discrepancy[%]			
Similarity Indices	Exp. 1			Exp. 2			Exp. 1			Exp. 2			Exp. 1		Exp. 2	
	mean	SD	min.	mean	SD	min.	mean	SD	min.	mean	SD	min.	Rev	Deg	Rev	Deg
hgt-LP	0.855	0.072	0.686	0.827	0.063	0.704	0.829	0.113	0.530	0.794	0.084	0.627	3.6	13.3	7.3	14.7
hgt-AP	0.855	0.071	0.686	0.826	0.063	0.705	0.821	0.104	0.554	0.792	0.094	0.604	4.1	13.3	9.3	14.7
hgt-BP	0.836	0.074	0.685	0.806	0.072	0.701	0.808	0.108	0.555	0.782	0.104	0.593	4.6	14.4	9.3	15.3
hgt-DP	0.836	0.074	0.685	0.806	0.072	0.701	0.808	0.108	0.555	0.782	0.104	0.593	4.6	14.4	9.3	15.3
RelC-LP	0.863	0.057	0.745	0.840	0.046	0.796	0.812	0.080	0.628	0.800	0.091	0.709	6.7	2.1	10.0	3.3
RelC-AP	0.860	0.057	0.745	0.838	0.047	0.787	0.809	0.083	0.617	0.797	0.094	0.684	6.7	2.1	10.7	3.3
RelC-BP	0.839	0.067	0.663	0.817	0.054	0.733	0.801	0.089	0.599	0.790	0.101	0.654	6.2	3.1	11.3	4.0
RelC-DP	0.832	0.052	0.655	0.801	0.063	0.682	0.784	0.096	0.589	0.779	0.115	0.595	6.2	3.1	11.3	4.0
CRLS-LP	0.879	0.052	0.752	0.843	0.049	0.795	0.800	0.083	0.588	0.783	0.096	0.637	6.2	2.1	10.7	3.3
CRLS-AP	0.879	0.053	0.752	0.842	0.051	0.791	0.803	0.083	0.591	0.786	0.100	0.622	6.2	2.1	12.0	3.3
CRLS-BP	0.859	0.066	0.670	0.820	0.064	0.714	0.801	0.086	0.585	0.784	0.106	0.602	6.7	3.1	12.0	4.0
CRLS-DP	0.853	0.067	0.670	0.810	0.069	0.673	0.794	0.089	0.590	0.782	0.118	0.558	6.7	3.1	12.7	4.0
Dist-gc	0.883	0.037	0.812	0.849	0.046	0.802	0.877	0.045	0.809	0.850	0.066	0.750	4.1	0.0	9.3	0.0
GLB-FDL1	0.854	0.062	0.687	0.832	0.056	0.737	0.802	0.113	0.498	0.776	0.073	0.666	4.6	11.8	6.7	11.3
LUB-FDL1	0.857	0.043	0.765	0.830	0.054	0.758	0.839	0.049	0.719	0.809	0.090	0.668	4.6	1.0	10.7	0.7
Gc-FD	0.883	0.037	0.817	0.849	0.054	0.796	0.856	0.054	0.776	0.828	0.068	0.714	2.6	0.0	10.0	0.0

all the subjects. First, the minimum values of them in the two experiments reveal that all the indices over all the subjects significantly correlate with the subjective degree of similarity beyond 1% point (r_s (28) = 0.448, r_p(28) = 0.479). Next, as seen from the mean values, the values of the *Dist-gc* and the *Gc-FD* are highest in all the indices. Moreover, we carry out ANOVA in two-way layout for each correlation coefficient by assigning the indices and the experiments to two main effects and the subjects to repetition. ***Table 2*** shows the F ratios for the indices, the experiments, and the interaction of them. It can be seen from *Table 2* that the differences among the indices do not significantly affect the values of two correlation coefficients. Lastly, as seen from the SDs, it can be considered that the influence of individual difference on the *Dist-gc* and the *Gc-FD* are smaller than the others. Consequently, these results suggest that the *Dist-gc* and the *Gc-FD* better correspond to the subjective degree of similarity than the others.

Table 2: Results of ANOVA

Effect	df	F ratios: r_s	r_p	Reverse	Degen
indices	15	* 1.63	1.27	0.80	## 13.75
experiments	1	## 23.71	** 4.52	## 39.49	1.06
interaction	15	0.07	0.05	0.17	0.04
residual	336				

significance ## : 0.1%, ** : 5%, * : 10%

3.2 Discrepancy between subjective rating and their estimates from indices

Importance of analysis of discrepancy: In order to evaluate the performance of mathematical similarity indices for explaining the subjective degree of similarity, we must consider also discrepancy between subjective rating of similarity and their estimates from the indices. In other words, two pairs that a subject rates at different categories of similarity must be adequately distinguished by the values of the indices. The discrepancy is divided into *degeneracy* and *reverse*. The degeneracy is the state where pairs rated at different categories are in the same value of an index, and the reverse is the state where rank order between values of an index is contrary to one between rating categories (See *Figure 1*). In these two cases, we cannot adequately predict subjective rating based on values of the index.

Results: For each index and each subject, we calculated mean values of the index within pairs rated at the same category. From these values, the relative frequencies of the discrepancy for each index, shown in *Table 1*, were summed up over all the possible paris (15) of six categories and all the subjects. As seen from *Table 1*, the differences of the reverse among the indices are relatively small. From the results of ANOVA in *Table 2*, the effect of the indices is not significant at 10% level. Theses results show that the reverse would be caused by sources independent of the indices, such as misjudgment of similarity in the SJ task.

On the other hand, *Table 1* show that the degeneracy concentrates in the indices related to the *hgt-*P* and the *GLB-FDL1*. Moreover, the degeneracy concerning the *hgt-*P* almost takes place at unity of the index values, and that concerning the *BP* and the *DP* almost does at zero. Contrary, there is no degeneracy in the indices of the *Dist-gc* and the *Gc-FD*. From *Table 2*, the effect of the indices is significant at 0.1% level. Consequently, we conclude that the *Dist-gc* and the *Gc-FD* also well describe the subjective degree of similarity in sense of the discrepancy.

3.3 Discussion

Influence of the experimental conditions: From *Table 2*, the effect of the interaction between the indices and the experiments on four variables are not significant at 10% level. Contrary, the effect of the experiments are significant beyond 5% level, except for the reverse. As seen from *Table 1*, these variables in the Exp. 2 are relatively worse than them in the Exp. 1. However, this difference can be mostly explained by the difference in selecting the ticks used in the SJ task. Consequently, we conclude that the relative relationships among 16 indices are not affected by the difference of the experimental conditions, especially the subjects. Since the isometry of the UOD holds also between tallness and heaviness, we will re-examine the influence of the attributes using the other sets of hedges and phrases.

Another formula of t-norm: From the results mentioned above, the indices concerning the overlapping degree between fuzzy sets, in which one of four indices recommended by *Zwick et al.* [3] is included, are inadequate to similarity index. One may consider improvement using the *Yegar's* or *Weber's* general formulas of t-norm, which vary continuously from LP to DP with one parameter [5]. However, all the degeneracy concerning the t-norms cause at either zero or unity of their index values. So, due to the boundary conditions of t-norm: $T(x,0) = 0$ and $T(1,x) = x$, the general t-norms cannot decrease the degeneracy either, although two correlation coefficients could be improved to some extent. Consequently, we cannot expect the improvement comparable to the cost to tune up its parameter.

4. CONCLUSION

In this paper, we have discussed the relationship between the subjective degree of similarity and the mathematical similarity indices for verbal expressions under two experimental conditions. Based on the analysis of the correlation coefficients between them and the discrepancy of their rank order, it has been clear that the two indices concerning the distance: *Dist-gc* and *Gc-FD* most correspond to the subjective degree of similarity among all of the indices adopted. Furthermore, it has been shown that the relative relationships among the indices obtained from two experiments consist with each other.

We will re-examine the relationship under the other experimental conditions in order to refine the results obtained in this paper, and compare our indices with those in ref. 3.

Acknowledgement: This study was supported by scientific research grant from the Ministry of Education Japan in 1995.

Reference

[1]Sugeno, M. & Yasukawa, T.: A Fuzzy-Logic-Based Approach to Qualitative Modeling, *IEEE Trans. On Fuzzy Systems*, **1**, 1, 7-31 (1993)

[2]Dubois, D. & Prade, H. : Fuzzy Sets and Systems: Theory and Applications, *Academic Press* (1980)

[3]Zwick, R. et al.: Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis, Int. J. Approximate Reasoning, **1**, 221-241 (1988)

[4]Yoshikawa, A et al.: Objective and subjective comparison between two methods for identifying membership functions: Computerized Fuzzy Graphic Rating Scales and Boundary Asymptotic Estimation Method, *Proc. of 11th Fuzzy System Symposium*, 405-408 (1995) *in Japanese*

[5]Mizumoto, M.:Fuzzy logic and fuzzy reasoning, *Mathematical Sciences*, 284, 10-18 (1987) *in Japanese*