APPENDIX A

CORRELATION ANALYSIS FOR R_{max} IN FUZZY MEMBERSHIP FUNCTION

The similarity measurement between two instances heavily depends on the value of parameter $R_{\max}$ in the UFN reasoning model. Here, the linear correlation analysis in statistics is employed to appropriately determine the value of $R_{\max}$ in the fuzzy membership function. The analysis is a process which aims to measure the strength of the association between two sets of variables that are assumed to be linearly related.

For an instance base $\boldsymbol{U}$ with N instances, the correlation analysis in the fuzzy membership function is implemented in the following steps. The first step is to determine the degree of difference between any two instances in the base $\boldsymbol{U}$ using the function of degree difference in equation (3.12). Notably, the weights of a_m in equation (3.12) are initially set as one. Thus, a total of (C_2^N +N) resembling samples, $S_{ij}(U_{i,\mathrm{O}}, U_{j,o}, d_{ij})$, can be compiled. A resembling sample contains two instances' outputs ($U_{i,O}$ and $U_{j,O}$) and the corresponding degree of difference (d_{ij}), where the entities of $U_{i,O}$ and $U_{j,O}$ are real numbers.

Thereafter, two arrays, A_t and B_t, can be assorted from resembling samples in the case of d_{ij} less than or equal to a prescribed value, say t. The elements in A_t and B_t are the first and second items, respectively, of these resembling samples.

Next, the accumulative correlation coefficient, $Ac\_CORREL(A_t,B_t,t)$, is calculated for arrays A_t and B_t with the degree of difference less than or equal to t. Assume that for a total P resembling samples with d_{ij} less than a prescribed t, the arrays A_t and B_t can be denoted as

$$\begin{aligned} A_t &= \{a_k \mid a_k = U_{i,O} \in S_{ij}, \quad for \quad d_{ij} \le t\} = \{a_1, a_2, ..., a_p\} \\ B_t &= \{b_k \mid b_k = U_{j,O} \in S_{ij}, \quad for \quad d_{ij} \le t\} = \{b_1, b_2, ..., b_p\} \end{aligned} \tag{A.1}$$

The value of the accumulative correlation coefficient equals

$$Ac\_CORREL(A_t, B_t, t) = \frac{Cov(A_t, B_t, t)}{s_{A_t} s_{B_t}} \tag{A.2}$$

$$Cov(A_t, B_t, t) = \frac{1}{p}\sum_{k=1}^{p}(a_k - m_{A_t})(b_k - m_{B_t}) \quad s.t. \;\; d_{ij} \le t \tag{A.3}$$

where s_{A_t} and s_{B_t} are standard errors of arrays A_t and B_t; m_{A_t} and m_{B_t} are the means of A_t and B_t. The formulas expressed in equations (A.2) and (A.3) represent the relationship between the accumulative correlation coefficient to any value of t. An accumulative correlation curve can be plotted as a function of t and $Ac\_CORREL(A_t,B_t,t)$. Note that the appropriate $R_{\max}$ equals a certain value of t such that instances in the instance base $\boldsymbol{U}$ have a certain degree of correlation.

The smaller the t implies a larger accumulative correlation coefficient, indicating a strong relationship between the two arrays, e.g., the strongest correlation, t=0, between the two arrays refers to the case in which the instances in the two sets are identical and the value of $Ac\_CORREL(A_t\ B_t\ t)$ equals one. In such a case, no solution to a new instance can be generated via the UFN reasoning model except for when identical instances exist in the instance base. In order to avoid this issue here, Hung and Jan (2000) set $Ac\_CORREL(A_t,B_t,t)$ equal to 0.8 as the lower bound for similarity measurement. The value of t corresponding to this lower bound is adopted as the appropriate value of $R_{\max}$.