FUZZY TREATMENT OF ANALYTICAL DATA FOR COLORIMETRY

TAKUYA SHIMOSAKA, TAKEHIKO KITAMORI, AKIRA HARATA and TSUGUO SAWADA

Department of Industrial Chemistry, Faculty of engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113, JAPAN

Abstract - We have applied fuzzy theory to non-statistical and quantitative processing of experimental errors. A fuzzy linear regression method was introduced to quantitative analysis, and an "estimation area" was derived in place of a conventional calibration curve. This theoretical treatment was applied to colorimetry and its propriety was confirmed.

Key words fuzzy, fuzzy linear regression, fuzzy number, calibration curve, error

Experimental data usually have errors caused by various sources, such as foreign noises, instrumental uncertainty and chemical procedures. These errors are one dominating factor for reliability of analytical data. Statistical and probability methods have been applied to the theory of errors, however, a quantitative treatment of errors of an individual measured value has not been used commonly. We treated these errors as an uncertainty, and have applied fuzzy theory to processing them. In the present report, the basic idea of fuzzy treatment of experimental data is proposed. It was applied to calibration curves of colorimetry. A new idea for calibration based on fuzzy linear regression (FLR) [1], called the "estimation area" was derived. The propriety of this idea was confirmed by simulations.

THEORETICAL

First, data having errors are expressed with a fuzzy number (FN). Fuzzy theory can include ambiguity in a numerical value. When the experimental values are dispersed and the reliability is numerically represented as Fig. 1, the reliability value is highest at the mean y_m . The value decreases away from the mean and it becomes zero at $y_m \pm d$. The reliability distribution, as a function of data dispersion, is called a membership function in fuzzy theory. The FN is characterized by the mean value y_m and the width d . Unlike an ordinary number (crisp number), a FN expresses not only the center value of the measured one but its error, too. As width d increases, the measured value becomes more ambiguous.

Next, a FLR method is introduced to obtain a calibration curve for the colorimetric data expressed by FNs. In the FLR theory, the calibration curve, which corresponds to the regression line $y = a_0 + a_1 x$ (a_0 and a_1 : crisp numbers) for conventional determination, is expressed as $Y = A_0 + A_1 x$ where A_0 and A_1 are FNs, and Y and x are absorbance and concentration, respectively.

In Fig. 2, the relationship between the concentration, absorbance and reliability are expressed three dimensionally. As the experimental data are FNs, the membership function for each datum is represented by the dashed line triangle of Fig. 2. The FN coefficients A_0 and A_1 are obtained by FLR with

Fig. 1 A typical example of a membership function. Measured values (y) are plotted with O.

the following restrictions.

- (1) The line segment $h=h_0$ in the dashed line triangle, which represents a certain reliability to be used in the estimation, is involved in the estimated FN value (solid line triangle).
- (2) The sum of the widths of the estimated values (the bottom lines of the triangles) becomes a minimum. Consequently, the set of triangles satisfying the FLR restriction gives the calibration area, which is indicated as an "estimation area". This estimation area corresponds to the calibration curve of a conventional least squares method for crisp number data.

Fig. 2 Relation between concentration and measured absorbance represented by FN. A membership function for each value is shown on another axis.
 $\wedge$: measured value of absorbance (FN).
 $\wedge$: estimated value of absorbance (FN).

SIMULATION

The data used for the simulation of the colorimetry were normally distributed random numbers generated by the Box-Muller method. The standard deviations were large at low concentration, and small at high concentration assuming practical data. The estimation area was calculated by the proposed theory, and the results are shown in Fig. 3.

RESULTS AND DISCUSSION

As coefficient A_0 and A_1 are FNs, the FLR results have the area as shown in Fig. 2. This area is defined as "estimation area" in the present theory, and it is a new idea for calibration area introduced by FLR. From simulation results, the estimation area was obtained as shown in Fig. 3. In the low concentration region, the width of the estimation area is wide because uncertainty of the value is high. Using this estimation area, an estimated concentration region x_a is given for a measured value y_a , although the estimated concentration does not have a region in the conventional calibration curve. The width of the estimated concentration region x_b in the higher concentration region becomes narrower as expressed in Fig. 3. This is because uncertainty of experimental values is smaller at higher concentration. This agrees well with a practical determination. With a conventional calibration curve, the estimated value is usually expressed as $x \pm 2\sigma$ (σ is standard deviation) and the confidence interval 2σ is the same value for all concentrations. Hence, the present method is considered to be more suitable for practical situations.

FLR can simultaneously process the measured value and its error, and the estimation area represents the relation between concentration, absorbance and its reliability. Therefore, propriety of introducing fuzzy theory to analytical data is shown. FLR is considered to be more suitable for processing errors of individual measured values than the conventional method.

Fig. 3 Estimation area of absorbance versus concentration.

REFERENCE

1. H. Tanaka, S. Uejima, and K. Asai, IEEE Transactions on Systems, Man and Cybernetics, SMC-12, 6, 903-907(1982)