Uncertain Information Fusion and Knowledge Integration: How to Take Reliability into Account

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Kittawit Autchariyapanitkul
Maejo University, kittar3@hotmail.com

Olga Kosheleva
The University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-17-02a
To appear in *Proceedings of the Joint 17th Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems*, Otsu, Japan, June 27-30, 2017.

Uncertain Information Fusion and Knowledge Integration: How to Take Reliability into Account

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA, and
Faculty of Economics, Chiang Mai University
Chiang Mai, Thailand
hunguyen@nmsu.edu

Kittawit Autchariyapanitkul
Faculty of Economics
Maejo University
Chiang Mai, Thailand
kittar3@hotmail.com

Olga Kosheleva and
Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract—**In many practical situations, we need to fuse and integrate information and knowledge from different sources – and do it under uncertainty. Most existing methods for information fusion and knowledge integration take into account uncertainty. In addition to uncertainty, we also face the problem of reliability: sensors may malfunction, experts can be wrong, etc. In this paper, we show how to take into account both uncertainty and reliability in information fusion and knowledge integration. We show this on the examples of probabilistic and fuzzy uncertainty.**

I. Information Fusion and Knowledge Integration That Takes Into Account Uncertainty and Reliability of Information Sources: Formulation of the Problem

Information fusion and knowledge integration: a brief reminder. Suppose that we are interested in an object or a system. We are therefore interested in the values of quantities $x_1, \ldots, x_n$ that characterize this object or system. The tuple of all these quantities will be denoted by $x = (x_1, \ldots, x_n)$.

A simple case is when we are interested in a simple periodic process

$$s(t) = A \cdot \sin(\omega \cdot t + \theta), \quad (1)$$

where A is the amplitude, ω is the frequency, and θ is the initial phase. This process is characterized by the above three quantities: $x_1 = A$, $x_2 = \omega$, and $x_3 = \theta$.

A more complex case is when we are interested in the trajectory of an asteroid. According to celestial mechanics, in the first approximation – when we only take into account the Sun's gravitational force – the trajectory of a celestial body is an ellipse. To describe an ellipse in a 3-D space, we need to describe the unit vector orthogonal to the ellipse's plane – which requires 2 quantities, a direction of the major axis in this plane – 1 parameters, and the value of major and minor axes – 2 more parameters, to the total of 5 quantities $x_1, \ldots, x_5$.

In many practical situations, we have several different pieces of knowledge about this object, and we need to take all this knowledge into account, so that we can get estimates for the quantities x_i that reflect all the pieces of knowledge. This is what we usually mean by information fusion and knowledge integration.

978-1-5090-4917-2/17$31.00 © 2017 IEEE

What are the pieces of information that we try to fuse? To understand how to fuse different pieces of information, let us recall what these pieces of information are.

In general, most information about the objects and systems comes from measurements. In additional to measurement results, we often also have expert information.

Sometimes, we can directly measure the corresponding quantities x_i and/or we have experts who can directly estimate the values of these quantities. However, as the above examples show, such situation are rare. For example, for a periodic signal, we do not directly measures its amplitude, frequency, or phase. What we usually measure is the value $s(t)$ of this signal at different moments of time t. Similarly, for an asteroid, what we measure is its angular position in the sky at different moments of time and when observed from telescopes located at different locations on Earth.

Let N denote the total number of available measurement and estimation results. For each j from 1 to N, let us denote the quantity estimated during the j-th estimation by y_j. These estimated quantities depend on the desired quantities $x_1, \ldots, x_n$. They also depend on the setting of the corresponding estimation; let us denote the parameters describing the setting of the j-th measurement by $a_j = (a_{j1}, \ldots, a_{js})$.

For example, for the sinusoidal wave, the only quantity that describes the setting is the time t_j of the j-th measurement, so we have $s = 1$ and $a_{j1} = t_j$. The quantity that we measure is the value $s(t_j)$ of the signal at this moment of time t_j. If we use notations x_1 for A, x_2 for ω, x_3 for θ, a_{j1} for t_j, and y_j for $s(t_j)$, then the above formula (1) for the sinusoidal signal takes the form

$$y_j = x_1 \cdot \sin(x_2 \cdot a_{j1} + x_3). \quad (2)$$

In many practical situations, the situation is even more complicated: namely, the measured quantity y_j depends not only on the quantities $x_1, \ldots, x_n$ in which we are interested, but also on some auxiliary quantities $c = (c_1, \ldots, c_m)$ that affect y_j and that are not themselves of interest to us.

For example, our observations of the periodic process are affected by the higher harmonics, with frequencies 2ω and

3ω. In this case, instead of the formula (1), we have a more complex dependence

$$s(t) = A \cdot \sin(\omega \cdot t + \theta) + A_2 \cdot \sin(2\omega \cdot t + \theta_2) + A_3 \cdot \sin(3\omega \cdot t + \theta_3), \quad (3)$$

for some unknown values A_2, A_3, θ_2, and θ_3. Here, we have $m = 4$ auxiliary quantities $c_1 = A_2$, $c_2 = A_3$, $c_3 = \theta_2$, and $c_4 = \theta_3$. By using the c-notations, the formula (3) takes the form

$$y_j = x_1 \cdot (x_2 \cdot a_{j1} + x_3) + c_1 \cdot \sin(2x_2 \cdot a_{j1} + c_3) + c_2 \cdot \sin(3x_2 \cdot a_{j1} + c_4) \quad (4)$$

In general, let us denote the dependence of y_j on x, a_j, and c by f, so we get $y_j = f(x, a_j, c)$, or, in more details:

$$y_j = f(x_1, \ldots, x_n, a_{j1}, \ldots, a_{js}, c_1, \ldots, c_m). \quad (5)$$

Information fusion and knowledge integration: formulation of the problem. In terms of the above notations, the problem of information fusion and knowledge integration takes the following form:

- we know the results $\widetilde{y}_j \approx y_j$ of measuring y_j,
- we know the settings a_j that were used in these measurements, and
- we know the function $y_j = f(x, a_j, c)$ that describes the dependence of y_j on x, a_j, and unknown auxiliary quantities c_ℓ.

Based on all this information, we want to estimate the desired quantities $x_1, \ldots, x_n$.

Need to take into account uncertainty and reliability. Measurements and estimates are never absolutely accurate; see, e.g., [7]. As a result, for each j, the measurement result $\widetilde{y}_j$ is, in general, different from the actual value y_j of the corresponding quantity: $\widetilde{y}_j \neq y_j$. This uncertainty need to be taken into account when estimating the desired quantities x_i.

We must also take into account that measurements and expert estimates are not always reliable: sometimes, they correspond not to the object of interest, but to some other object. For example, in underwater sonar measurements, when we measure the distance to an object – by the time that it takes for a signal to bounce back to us – the sensors sometimes record the signal reflected by some other object; see, e.g., [10] and references therein.

What is known and what is new in this paper. Data processing under different types of *uncertainty* is a well-studied, well-analyzed area of research. There are also many approaches to taking reliability into account; see, e.g., [8] and references therein.

Some papers take into account both uncertainty and reliability. However, to the best of our knowledge, no general algorithms are known that take into account both uncertainty and reliability. The main objective of this paper is to propose such general algorithms – and, ideally, to incorporate reliability by making the *smallest possible* changes to well-known uncertainty-related algorithms.

Structure of the paper. Our main objective is to modify the existing uncertainty-related algorithms so that they can take into account reliability as well. Because of this, in Section 2, we first describe the two main cases of uncertainty – probabilistic and fuzzy, and in Sections 3 and 4, we describe how to take both types of uncertainty into account in information fusion and knowledge integration.

Most of the methods that we describe in these sections are known. However, our description is sometimes somewhat different – since we reformulate them so as to make taking reliability into account easier.

In Section 5–7, we show how to modify methods from Sections 3 and 4 so that they also take reliability into account.

II. Two Types of Uncertainty

Two types of uncertainty. As we have mentioned, the estimates $\widetilde{y}_j$ are, in general, different from the actual (unknown) values y_j. How can we describe the corresponding inaccuracy $\Delta y_j \stackrel{\text{def}}{=} \widetilde{y}_j - y_j$?

In some cases, we know the frequency of different values of estimation inaccuracy, i.e., in precise terms, we know the probability distribution of this inaccuracy.

In other cases, all we know is the expert estimations for the size of this inaccuracy, expert estimations expressed by using imprecise ("fuzzy") words from natural language. In such cases, a reasonable idea is to use *fuzzy logic*, techniques specifically designed for handling this uncertainty [2], [4], [11].

Let us briefly describe these two types of uncertainty one by one.

Probabilistic uncertainty: examples. In some cases, we know the probability distribution for the estimation error $\Delta y_j = \widetilde{y}_j - y_j$. Each of these probability distributions can be described, e.g., by a probability density function (pdf) $\rho_j(\Delta y_j)$.

The estimate $\widetilde{y}_j$ is, in reality, never the exact number: it is usually plus minus the corresponding discretization level. For example, if a measuring instrument returns the result 0.376, this means any value from 0.3755 to 0.3765. Similarly, if an expert estimates the value as 1.1, this means any value from 1.05 to 1.15. In general, the estimate $\widetilde{y}_j$ means an interval $[\widetilde{y}_j - \delta_j, \widetilde{y}_j + \delta_j]$, for some small value δ_j. The corresponding interval for the difference $\Delta y_j = \widetilde{y}_j - y_j$ has the form

$$[(\widetilde{y}_j - y_j) - \delta_j, (\widetilde{y}_j - y_j) + \delta_j]. \quad (6)$$

Thus, we can estimate the probability P_j of this estimate by multiplying the probability density $\rho_i(\Delta y_j)$ by the width $2\delta_j$ of the corresponding interval: $P_j = \rho_j(\Delta y_j) \cdot (2\delta_j)$.

Usually, all the distributions $\rho_j(\Delta y_j)$ belong to the same family, they only differ by the values of the corresponding parameters. In precise terms, we have

$$\rho_j(\Delta y_j) = \rho(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq}) \quad (7)$$

for an appropriate function ρ and for known values of the parameters $\theta_{j1}, \ldots, \theta_{jq}$. For example, we may know that all the distributions are normal with 0 mean, and we know the standard deviations $\theta_{j1} = \sigma_j$ corresponding to different estimates. In this case, we have

$$\rho(\Delta y, \theta_{j1}) = \frac{1}{\sqrt{2\pi} \cdot \theta_{j1}} \cdot \exp\left(-\frac{(\Delta y)^2}{2\theta_{j1}^2}\right). \quad (8)$$

In more general situations, some of the parameters $\beta_i, \ldots$ of the corresponding probability distributions are unknown. For example:

- we know that the measurements come from several measuring instruments,
- we know that for each of these instruments, the distribution is Gaussian with 0 mean, but
- we do not know the standard deviations of these measuring instruments.

Alternatively:

- we may know that the estimates from several experts,
- we know that for each expert, the estimation error is normally distributed with 0 mean and unknown standard deviation, but
- we do not know the standard deviations corresponding to different experts.

Thus, we arrive at the following general description of probabilistic uncertainty.

Case of probabilistic uncertainty: general description. In general, the set $\{1, \ldots, N\}$ of all estimations is divided into several disjoint subsets S_α. The probability distribution of estimation errors Δy_j corresponding to each subset S_α are characterized, in general, by its own expression

$$\rho_\alpha(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \ldots, \beta_{\alpha t_\alpha}), \quad (9)$$

where the values $\theta_{\alpha 1}, \ldots$ are known while the values $\beta_{\alpha 1}, \ldots$ are not known.

Example. If different sets S_α correspond to different measuring instruments, with 0 mean and unknown standard deviations $\beta_{\alpha 1} = \sigma_\alpha$, then

$$\rho_\alpha(\Delta y) = \frac{1}{\sqrt{2\pi} \cdot \beta_{\alpha 1}} \cdot \exp\left(-\frac{(\Delta y)^2}{2\beta_{\alpha 1}^2}\right). \quad (10)$$

Case of fuzzy uncertainty. In the fuzzy cases, instead of probabilities, for each estimate j and for each possible value of the estimation error Δy_j, we know the degree $\mu_j(\Delta y_j)$ to which this value of the estimation error is possible. The corresponding function is known as the *membership function*.

Usually, all these membership functions belong to the same family, they only differ by the values of the corresponding parameters. In precise terms, we have

$$\mu_j(\Delta y_j) = \mu(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq}) \quad (11)$$

for an appropriate function μ and for known values of the parameters $\theta_{j1}, \ldots, \theta_{jq}$. For example, we may know that all the membership functions are triangular.

In more general situations, some of the parameters $\beta_i, \ldots$ of the corresponding membership functions are unknown. For example, we may know that the measurements come from several experts, we know that the membership functions for each of the experts is triangular with 0 maximum, but we do not know the spread of these membership functions.

In general, the set $\{1, \ldots, N\}$ of all estimations is divided into several disjoint subsets S_α. The membership functions characterizing the estimation errors Δy_j from each subset S_α are described, in general, by their own expression

$$d_j = \mu_\alpha(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \ldots, \beta_{\alpha t_\alpha}), \quad (12)$$

where the values $\theta_{\alpha 1}, \ldots$ are known while the values $\beta_{\alpha 1}, \ldots$ are not known.

III. How Probabilistic Uncertainty Is Taken into Account in Information Fusion and Knowledge Integration

General case. As we have mentioned earlier, for each estimate j, the probability P_j of having this estimate is proportional to the corresponding probability density.

Approximation errors corresponding to different measurement results are usually independent from each other. Thus, the overall probability of having all N estimates $\widetilde{y}_1, \ldots, \widetilde{y}_N$ is equal to the product of N probabilities $P_1 \cdot \ldots \cdot P_N$ and is, thus, proportional to the product L of the corresponding probability densities. This product is known as *likelihood.*

If we group together estimates corresponding to each group S_α, we get the following expression for the likelihood:

$$L = \prod_\alpha \prod_{j \in S_\alpha} \rho_\alpha(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \ldots, \beta_{\alpha t_\alpha}), \quad (13)$$

where

$$\Delta y_j = \widetilde{y}_j - f(x_1, \ldots, x_n, a_{j1}, \ldots, a_{js}, c_1, \ldots, c_m). \quad (14)$$

We need to find the desired values $x_1, \ldots, x_n$ – as well as all the remaining unknowns $c_1, \ldots, c_m, \beta_{\alpha 1}, \ldots$ A reasonable idea is to find the values for which the above probability is the largest, i.e., equivalently, the likelihood L takes the largest possible value. This idea is known as the *Maximum Likelihood Method.*

Specific case of Gaussian (normal) distributions. There are usually many different reasons for an estimation error. For example, for measurements, there is noise in each part of the measuring instrument – and all these noises contribute to the overall estimation error.

In situations when the overall estimation error is a sum of many different independent components, it is usually possible to invoke the Central Limit Theorem, according to which for large N, the distribution of the sum of N small independent random variables is close to Gaussian; see, e.g., [9]. And indeed, in many practical cases, the probability distribution of the measurement error is close to Gaussian [5], [6].

For the measurement error, it is usually safe to assume that the mean error (*bias*) is 0, since this bias can be detected if we

several times compare the results of measuring instrument with a more accurate ("standard") one, and thus, can be eliminated by simply re-scaling the measuring instrument.

It is therefore same to assume that each estimation error is normally distributed, with 0 mean and some standard deviation σ_j. The corresponding probability density function has the form

$$\rho_j(\Delta y_j) = \frac{1}{\sqrt{2\pi}\cdot\sigma_j}\cdot\exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right). \qquad (15)$$

Thus, the likelihood takes the form

$$L = \prod_{j=1}^{N}\rho_j(\Delta y_j) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\cdot\sigma_j}\cdot\exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right) =$$

$$\frac{1}{(\sqrt{2\pi})^N\cdot\prod\limits_{j=1}^{N}\sigma_j}\cdot\exp\left(-\frac{1}{2}\cdot\sum_{j=1}^{N}\frac{(\Delta y_j)^2}{\sigma_j^2}\right). \qquad (16)$$

When all the standard deviations σ_j are known, maximizing the above expression for the likelihood is equivalent to minimizing the sum in the $\exp$ part of this expression:

$$\sum_{j=1}^{N}\frac{(\Delta y_j)^2}{\sigma_j^2} = \sum_{j=1}^{N}\frac{(\widetilde{y}_j - f(x,a_j,c))^2}{\sigma_j^2}\to\min_{x,c}. \qquad (17)$$

This is the usual Least Squares approach.

In particular, when all the estimates have the same accuracy – e.g., come from using similar measuring instruments or the same expert – then $\sigma_j = \sigma$ for all j, and the above optimization problem can be further simplified, into:

$$\sum_{j=1}^{N}(\Delta y_j)^2 = \sum_{j=1}^{N}(\widetilde{y}_j - f(x,a_j,c))^2\to\min_{x,c}. \qquad (18)$$

For example, when, in addition, we only have one quantity of interest x_1 and all estimates $\widetilde{y}_j$ directly estimate this quantity, i.e., when $y_j = f(x_1) = x_1$ for all j, then the formula takes the form

$$\sum_{j=1}^{n}(\widetilde{y}_j - x_1)^2\to\min_{x_1}. \qquad (19)$$

Differentiating with respect to x_1 and equating the derivative to 0, we can conclude that the fused estimate becomes the arithmetic mean

$$x_1 = \frac{1}{N}\cdot\sum_{j=1}^{N}\widetilde{y}_j. \qquad (20)$$

In cases when we do not know the approximation errors σ_α, maximizing the likelihood L (or, equivalently, minimizing log-likelihood $-\ln(L)$) over σ_α leads to

$$\sigma_\alpha^2 = \frac{1}{N_\alpha}\cdot\sum_{j\in S_\alpha}(\widetilde{y}_j - y_j)^2, \qquad (21)$$

where N_α is the overall number of estimates j from the α-th group S_α.

Substituting these values into the expression for log-likelihood, we conclude that minimizing log-likelihood is equivalent to minimizing the sum of the logarithms of these standard deviations, i.e., minimizing the sum

$$\sum_\alpha \ln\left(\sum_{j\in N_\alpha}(\widetilde{y}_j - f(x,a_j,c))^2\right)\to\min_{x,c}. \qquad (22)$$

IV. How Fuzzy Uncertainty Is Taken into Account in Information Fusion and Knowledge Integration

Formulation of the problem. We are interested in the degree to which Δy_1 is a possible value of the first estimation error *and* Δy_2 is a possible value of the second estimation error, etc. In line with the general fuzzy techniques, to find this degree D, we apply an appropriate "and"-operation $f_\&(a,b)$ to the degrees corresponding to different values j:

$$D = f_\&(D_1, D_2, \ldots, D_\alpha, \ldots), \qquad (23)$$

where

$$D_\alpha = f_\&\{d_j : j\in S_\alpha\}. \qquad (24)$$

We need to find the desired values $x_1,\ldots,x_n$, as well as all the remaining unknowns $c_1,\ldots,c_m,\beta_{\alpha 1},\ldots$ A reasonable idea is to find the values for which the above possibility degree D is the largest.

Algorithms for solving this problem. When the "and"-operation is the algebraic product $f_\&(a,b) = a\cdot b$, then the above optimization takes the same form as for the probabilistic uncertainty, the only difference is that we have membership functions instead of the probability density functions.

In principle, however, we can have many different "and"-operations. From this viewpoint, the optimization problem corresponding to fuzzy information fusion is much more general – and thus, more complex than the Maximum Likelihood problem corresponding to probabilistic uncertainty. However, it is possible to reduce the general fuzzy case to the Maximum-Likelihood-type case of the product.

Indeed, it is known that every "and"-operation –including the widely used minimum $\min(a,b)$ – can be approximated, with any given accuracy, by an *Archimedean* "and"-operation, i.e., by an "and"-operation of the type

$$f_\&(a,b) = g^{-1}(g(a)\cdot g(b)) \qquad (25)$$

for some increasing function $g(x)$. Thus, from the practical viewpoint, we can safely assume that the actual "and"-operation is Archimedean.

For an Archimedean "and"-operation, we have

$$D = g^{-1}\left(\prod_\alpha g(D_\alpha)\right). \qquad (26)$$

Similarly, for every α, we have

$$D_\alpha = g^{-1}\left(\prod_{j\in S_\alpha} g(d_j)\right) \qquad (27)$$

and thus,

$$g(D_\alpha) = \prod_{j \in S_\alpha} g(d_j). \tag{28}$$

Substituting the formula for $g(D_\alpha)$ into the expression for D, we conclude that

$$D = g^{-1}\left(\prod_\alpha \prod_{j \in S_\alpha} g(d_j)\right). \tag{29}$$

Since the function $g(x)$ is increasing, maximizing the degree D is equivalent to maximizing the expression $g(D)$, which has a somewhat simpler form:

$$g(D) = \prod_\alpha \prod_{j \in S_\alpha} g(d_j) =$$

$$\prod_\alpha \prod_{j \in S_\alpha} g(\mu_\alpha(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \ldots, \beta_{\alpha t_\alpha})). \tag{30}$$

One can see that we arrive at the exact same expression as for the Maximum Likelihood, but with an auxiliary function $g(\mu_\alpha(\ldots))$ instead of the pdf $\rho_\alpha(\ldots)$.

V. What Do We Know About Reliability?

As we have mentioned earlier, sometimes the estimates $\widetilde{y}_j$ correspond not to the object of interest, but to some other object.

To take this into account, us recall what we know about such situations. Usually, situations when the estimate is not related to the object of interest are rare. From past experience, we can estimate how rare they can be. Thus, we can assume that for every j, we know:

- in the probabilistic case, the probability p_j that the j-th estimate is indeed related to the desired quantities, and
- in the fuzzy case, the degree of confidence q_j to which the j-th estimate is related to the desired quantity.

Let us describe how we can use this information in uncertain information fusion and knowledge integration.

VI. How to Take Reliability into Account: Probabilistic Case

How did we solve the original problem? In the above text, we had the following unknowns:

- the desired quantities $x_1, \ldots, x_n$,
- the unknown parameters $c_1, \ldots, c_m$ in the formula describing the dependence of the measurement results $\widetilde{y}_j$ on the desired quantities, and
- the parameters $\beta_{\alpha 1}, \ldots$ that describe the probability distributions of different values Δy_j of the estimation error.

To find all these parameters from observations, we used the Maximum Likelihood method.

Natural idea: use Maximum Likelihood method in case of reliability as well. If we take reliability into account, then there are other things that we do not know: e.g., we do not know which estimates are related to the desired values x and which are not. In other words, we now have more unknowns than before.

A natural idea is to again use the Maximum Likelihood approach – this time, to find *all* the unknowns: both the previous unknowns and the new unknowns.

So what are the new unknowns? If we take reliability into account, then we have following additional unknowns:

- First, for every j, we do not know whether the j-th estimate $\widetilde{y}_j$ is related to the desired quantity or not. This can be described by introducing, for each estimate j, a new binary variable z_j which is:
  - equal to 1 if this estimate is related to the desired quantities, and
  - equal to 0 if the estimate $\widetilde{y}_j$ is not related to the desired quantities $x_1, \ldots, x_n$.

  The quantities z_j are new unknowns.
- Second, for those j for which the estimate is not related to the desired quantities, we do not know what quantity y_j is being estimated. Such values y_j should also be added to the list of unknown.

Thus. we should use the Maximum Likelihood approach to estimate not only the values of the previous unknowns x, c, and β, but also the values of the new unknowns:

- the values $z_j \in \{0, 1\}$ corresponding to all estimates $j = 1, \ldots, N$, and
- the values y_j corresponding to estimates for which

$$z_j = 0.$$

Let us describe the corresponding probabilities. In situations in which we take reliability into account, it is still reasonable to assume that situations corresponding to different estimates j are independent. Thus, the overall probability – that we will maximize – is still equal to the product $P_1 \cdot \ldots \cdot P_N$ of the probabilities P_j corresponding to different estimates.

The difference from the previous case is that the expressions for the probabilities P_j are now different. In the previous case, when we fixed the values of all the unknowns x, c, and β, then we concluded that the probability P_j is proportional to the value of the pdf:

$$P_j \sim \rho_\alpha(\Delta y_j, \theta_j, \beta_\alpha), \tag{31}$$

where $\Delta y_j = \widetilde{y}_j - f(x, a_j, c)$.

In the new (general) case, once we know the values of all the unknowns, i.e., once we know the values x, c, β, z_j, and y_j for those j for which $z_j = 0$, what is the probability P_j to have the corresponding values $\widetilde{y}_j$ and z_j?

A natural assumption is that the values z_j and $\widetilde{y}_j$ are independent. (Indeed, if they were dependent, we would be able, based on the estimates $\widetilde{y}_j$, to tell whether this estimate depends on the desired quantities or not – so we would not face the situation in which we do not know it.) Thus, the probability of having the values z_j and $\widetilde{y}_j$ is equal to the product of the probability to have z_j and the probability to have $\widetilde{y}_j$.

The probability $p(z_j)$ to have z_j is easy to describe:

- the probability to have $z_j = 1$ is equal to p_j, and

- the probability to have $z_j = 0$ is equal to the remaining probability $1 - p_j$.

The probability to have a given estimate $\widetilde{y}_j$ is still proportional to $\rho_\alpha(\Delta y_j, \theta_j, \beta_\alpha)$, the only difference is that now, the expression for Δy_j is more complicated:

- when $z_j = 1$, then we still have $\Delta y_j = \widetilde{y}_j - f(x, a_j, c)$;
- when $z_j = 0$, then we have $\Delta y_j = \widetilde{y}_j - y_j$ for some value y_j.

Summarizing: the overall probability is proportional to the product $E_1 \cdot \ldots \cdot E_N$ of the following expressions E_j corresponding to different estimates j:

- when $z_j = 1$, then
$$E_j = p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha); \tag{32}$$
- when $z_j = 0$, then
$$E_j = (1 - p_j) \cdot \rho_\alpha(\widetilde{y}_j - y_j, \theta_j, \beta_\alpha). \tag{33}$$

What can we conclude from the Maximum Likelihood approach? We need to find the values of all the parameters x, c, β, z_j, and y_j that maximize the product of the above expressions.

Let us start with finding the unknown values y_j corresponding to $z_j = 0$. For each j, only the value E_j depends on y_j. Thus, the product $E_1 \cdot \ldots \cdot E_N$ is the largest if this value E_j is the largest. In its turn, this value is the largest if it corresponds to the largest value of the probability density $\rho_\alpha(\Delta y_j, \ldots)$. Its largest values is thus equal to
$$E_j = (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha). \tag{34}$$

The probability of the estimation error is usually the largest when this error is 0 and decreases when $|\Delta y_j|$ decreases. In such cases, the maximum is attained when $y = 0$ and thus,
$$E_j = (1 - p_j) \cdot \rho_\alpha(0, \theta_j, \beta_\alpha). \tag{34a}$$

Now that we have found the optimal values of y_j, let us find the optimal values of z_j. Similarly to the above case, for each j, only the value E_j depends on z_j. Thus, the product $E_1 \cdot \ldots \cdot E_N$ is the largest if this value E_j is the largest. To find out which value $z_j \in \{0, 1\}$ makes the expression E_j the largest let us compare the values of E_j corresponding to $z_j = 0$ and to $z_j = 1$:

- when $z_j = 0$, we have
$$E_j = (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha); \tag{35}$$
- when $z_j = 1$, we have
$$E_j = p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha). \tag{36}$$

The largest of these two expressions is equal to
$$E_j = \max\left((1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha),\right.$$
$$\left. p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)\right)). \tag{37}$$

To find the values of the desired parameters x, c, and β, we therefore need to maximize the product of such maxima.

So, we arrive at the following precise formulation of the problem.

Formulation of the problem in precise terms. In the general case, we know:

- the function $f(x, a, c)$ describing the dependence of the estimated quantities on the desired quantities,
- the families $\rho_\alpha(\Delta y, \theta, \beta_\alpha)$ that describe the probabilities of estimation errors Δy_j for estimates j from different groups S_α, and
- for each j, we know the probability p_j that the j-th estimate is indeed related to the desired quantities $x_1, \ldots, x_n$.

In this case, according to the Maximum Likelihood method, we should select values x, c, and β that maximize the product
$$E_1 \cdot \ldots \cdot E_N, \tag{38}$$
where, for each $j \in S_\alpha$, we have
$$E_j = \max\left((1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha),\right.$$
$$\left. p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)\right). \tag{39}$$

In particular, for probability distributions for which zero estimation error is the most probable, we have
$$E_j = \max\left((1 - p_j) \cdot \rho_\alpha(0, \theta_j, \beta_\alpha),\right.$$
$$\left. p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)\right). \tag{40}$$

Algorithm: general case. Let us assume that we already know how to solve the optimization problem corresponding to the case when all the estimates are absolutely reliable. How can we transform this algorithm into an algorithm for solving the new problem?

A natural idea is to use component-wise maximization, when we first maximize over one group of variables, then over another group, etc., until the process converges; see, e.g., [1]:

1) first, we pick $z_j = 1$ for all j and use the usual Maximum Likelihood techniques to optimize over x, c, and β;
2) once we find the corresponding values of x, c, and β, we optimize over z_j: namely, we select $z_j = 1$ if
$$p_j \cdot \rho_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \geq$$
$$(1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha); \tag{41}$$
for all other j, we select $z_j = 0$;
3) then, only taking into account the estimates j selected on the previous step, we again use the maximum Likelihood method to find new estimates for x, c, and β, and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables $x_1, \ldots, x_n$ obtained on

the next iteration are sufficiently close to the values from the previous iteration.

Algorithm: case of normal distributions. Let us consider a typical case when all the estimation error are normally distributed with 0 mean and known standard deviations σ_j.

In this case, substituting the explicit formulas for the normal pdf into the above expressions, we conclude that the second term in the expression for E_j is larger when

$$1 - p_j \le p_j \cdot \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right), \tag{42}$$

i.e., equivalently, when

$$\frac{1 - p_j}{p_j} \le \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right) \tag{43}$$

and, taking negative logarithm of both sides, when

$$\frac{(\Delta y_j)^2}{2\sigma_j^2} \le \ln\left(\frac{p_j}{1 - p_j}\right), \tag{44}$$

i.e., when

$$|\Delta y_j| \le \sigma_j \cdot \sqrt{2\ln\left(\frac{p_j}{1 - p_j}\right)} \quad (45).$$

Thus, for the case of normal distributions, the above algorithm takes the following simplified form:

1) first, we pick $z_j = 1$ for all j and use the usual Least Squares method to find the values x and c for which the sum
$$\sum_{j=1}^{N} \frac{(\widetilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2} \tag{46}$$
is the smallest possible;
2) once we find the corresponding values of x and c, we select $z_j = 1$ if
$$|\widetilde{y}_j - f(x, a_j, c)| \le \sigma_j \cdot \sqrt{2\ln\left(\frac{p_j}{1 - p_j}\right)}; \tag{47}$$
for all other j, we select $z_j = 0$;
3) then, only taking into account the estimates j selected on the previous step, we again use the Least Squares Method to find new estimates for x and c by minimizing the sum
$$\sum_{j:z_j=1} \frac{(\widetilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2}, \tag{48}$$
and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables $x_1, \ldots, x_n$ obtained on the next iteration are sufficiently close to the values from the previous iteration.

VII. How to Take Reliability into Account: Fuzzy Case

Towards the precise formulation of the problem. The original fuzzy problem has the following unknowns:

- the desired quantities $x_1, \ldots, x_n$,
- the unknown parameters $c_1, \ldots, c_m$ in the formula describing the dependence of the measurement results $\widetilde{y}_j$ on the desired quantities, and
- the parameters $\beta_{\alpha 1}, \ldots$ that characterize the membership functions describing the estimation errors Δy_j.

Now, we have to also find the two new types of unknowns:

- the values $z_j \in \{0, 1\}$ that describe whether the j-th estimate is indeed related to the desired quantities $x_1, \ldots, x_n$, and
- for estimates $\widetilde{y}_j$ which are not related to the desired quantities (i.e., for which $z_j = 0$), the actual values y_j of the physical quantities which are estimated by these estimates.

How to find all these unknowns? A natural idea is to select the values of all these unknowns for which the degree D of possibility is the largest. This degree of possibility has the form

$$D = f_\&(D_1 \ldots, D_\alpha, \ldots), \tag{49}$$

where

$$D_\alpha = f_\&\{d_j : j \in S_\alpha\}, \tag{50}$$

and d_j is the degree to which the values $\widetilde{y}_j$, z_j (and, if needed, y_j) are possible.

As we have already shown, maximizing the degree D is equivalent to maximizing the value

$$g(D) = \prod_\alpha \prod_{j \in S_\alpha} g(d_j). \tag{51}$$

When $z_j = 1$, we are interested in the degree to which $\widetilde{y}_j$ is related to the desired quantities x_i *and* the difference $\Delta y_j = \widetilde{y}_j - f(x, a_j, c)$ is possible. We know the degree q_j to which $z_j = 1$: this degree is equal to q_j. Thus,

$$d_j = f_\&(q_j, \mu_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)), \tag{52}$$

hence

$$g(d_j) = g(q_j) \cdot g(\mu_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)). \tag{53}$$

When $z_j = 0$, then we are interested in our degree of confidence that $\widetilde{y}_j$ is *not* related to the desired quantities x_i (this degree is equal to $1 - q_j$) *and* that the difference $\widetilde{y}_j - y_j$ is possible. Thus,

$$d_j = f_\&(1 - q_j, \mu_\alpha(\widetilde{y}_j - y_j, \theta_j, \beta_\alpha)), \tag{54}$$

and

$$g(d_j) = g(1 - q_j) \cdot g(\mu_\alpha(\widetilde{y}_j - y_j, \theta_j, \beta_\alpha)). \tag{55}$$

Similar to the probabilistic case, the maximum is attained:

- when for $z_j = 0$, the membership function describing the estimation error reaches its maximum, and

- when we select $z_j = 0$ or $z_j = 1$ depending on which terms is larger.

Thus, we need to maximize the product $E_1 \cdot \ldots \cdot E_N$, where, for each $j \in S_\alpha$, the expression E_j takes the form

$$E_j = \max \Bigg(g(1-q_j) \cdot \max_y g(\mu_\alpha(y, \theta_j, \beta_\alpha)), \\ g(q_j) \cdot g(\mu_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)) \Bigg) . \quad (56)$$

Comment. It should be mentioned that, in contrast to the previous case, when we did not take reliability into account, this problem is not mathematically the same as for the probabilistic case:

- there, we had the weights p_j and $1-p_j$ that add up to 1, while
- here, the weights $g(q_j)$ and $g(1-q_j)$ do not necessarily add up to 1.

However, we can still use component-wise minimization to solve the corresponding optimization problem.

Resulting algorithm. Let us assume that we know how to solve the particular case of this problem when everything is perfectly reliable – e.g., we can do it by reducing this problem to the appropriate Maximum Likelihood problem. We will call the corresponding algorithm *original*.

Then, if we take reliability into account, we should do the following:

1) first, we pick $z_j = 1$ for all j and use the original optimization method to optimize over x, c, and β;
2) once we find the corresponding values of x, c, and β, we optimize over z_j: namely, we select $z_j = 1$ if

$$g(q_j) \cdot g(\mu_\alpha(\widetilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)) \geq \\ d(1-q_j) \cdot \max_y g(\mu_\alpha(y, \theta_j, \beta_\alpha)); \quad (57)$$

   for all other j, we select $z_j = 0$;
3) then, only taking into account the estimates j selected on the previous step, we again use the original optimization method to find new estimates for x, c, and β, and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables $x_1, \ldots, x_n$ obtained on the next iteration are sufficiently close to the values from the previous iteration.

VIII. Conclusion

In many application areas, we have several different pieces of information about an object or system of interest. In such situations, it is necessary to combine these pieces of information.

In this combination, we need to take into account that the information is rarely absolutely accurate – i.e., that we have uncertainty – and that sometimes, the corresponding measurement result actually correspond to other objects – and thus, that the information is not 100% reliable.

There exist many techniques for taking uncertainty into account. In this paper, we show how these techniques can be modified so as to take reliability into account as well.

Acknowledgment

This work was supported in part by the Faculty of Economics, Chiang Mai University, by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award “UTEP and Prudential Actuarial Science Academy and Pipeline Initiative” from Prudential Foundation.

The authors are greatly thankful to the anonymous referees for valuable suggestions.

References

[1] J. C. Bezdek and R. J. Hathaway, “Convergence of alternating optimization”, *Journal Neural, Parallel, and Scientific Computations*, 2003, Vol. 11, No. 4, pp. 351–368.

[2] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[3] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, “Strict Archimedean t-Norms and t-Conorms as Universal Approximators”, *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.

[4] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[5] P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).

[6] A. I. Orlov, “How often are the observations normal?”, *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

[7] S. G. Rabinovich, *Measurement Errors and Uncertainty. Theory and Practice*, Springer Verlag, Berlin, 2005.

[8] G. L. Rogova and V. Nimier, “Reliability in information fusion: literature survey”, *Proceedings of bthe 7th International Conference on Information Fusion*, Stockholm, Sweden, June 28 – July 1, 2004, pp. 1158–1164.

[9] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[10] A. Welte, L. Jaulin, M. Ceberio, and V. Kreinovich, “Robust data processing in the presence of uncertainty and outliers: case of localization problems”, *Proceedings of the IEEE Series of Symposia in Computational Intelligence SSCI’2016*, Athens, Greece, December 6–9, 2016.

[11] L. A. Zadeh, “Fuzzy sets”, *Information and Control*, 1965, Vol. 8, pp. 338–353.