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

Hung T. Nguyen\textsuperscript{1}, Kittawit Autchariyapanitkul\textsuperscript{2}, Olga Kosheleva\textsuperscript{3} and and Vladik Kreinovich\textsuperscript{3}

\textsuperscript{1}Dept. of Mathematical Sciences, New Mexico State University
Las Cruces, NM 88003, USA, and Faculty of Economics
Chiang Mai University, Thailand, hunguyen@nmsu.edu
\textsuperscript{2}Faculty of Economics
Maejo University, Chiang Mai, Thailand, kittar3@hotmail.com
\textsuperscript{3}University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, olgak@utep.edu, vladik@utep.edu
1. 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.
• *Example*: for a periodic process $s(t) = A \cdot \sin(\omega \cdot t + \theta)$, we have $x_1 = A$, $x_2 = \omega$, and $x_3 = \theta$.
• In many practical situations, we have several different pieces of knowledge about this object.
• We need to 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.
2. What Are the Pieces of Information that We Try to Fuse?

• In general, most information about the objects and systems comes from measurements.

• We often also have expert information.

• Sometimes, we can directly measure or estimate the desired values \( x_i \), but such situations are rare.

• For example, for a periodic signal, we usually measure its value \( s(t) \) at different moments of time \( t \).

• Let \( y_j \) denote the result of the \( j \)-th estimate, \( j \leq N \).

• Let \( a_j = (a_{j1}, \ldots, a_{js}) \) be parameters describing the \( j \)-th setting.

• For example, for the sinusoidal wave, \( a_{j1} = t_j \) and

\[
y_j = x_1 \cdot \sin(x_2 \cdot a_{j1} + x_3).
\]
3. Data Fusion (cont-d)

- Often, $y_j$ depends also on some auxiliary quantities $c = (c_1, \ldots, c_m)$ (of no direct interest to us):

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

- For example, our observations of the periodic process maybe affected by the higher harmonics

$$s(t) = A \cdot \sin(\omega \cdot t + \theta) + A_2 \cdot \sin(2\omega \cdot t + \theta_2)$$

- In general, we know:
  - the results $\tilde{y}_j \approx y_j$ of measuring or estimating $y_j$,
  - the settings $a_j$ and the function $y_j = f(x, a_j, c)$.

- Based on this information, we want to estimate the desired quantities $x_1, \ldots, x_n$. 
4. Need to Take into Account Uncertainty and Reliability

- Measurements and estimates are never absolutely accurate: \( \tilde{y}_j \).
- This uncertainty need to be taken into account when estimating \( x_i \).
- Also, sometimes, measurements correspond to another object, not to the object of interest.
- For example, underwater sonar sensors sometimes record reflections by another object.
- Our goal is to come up with a general methods for taking both uncertainty and reliability into account.
5. Two Types of Uncertainty $\Delta y_j \overset{\text{def}}{=} \tilde{y}_j - y_j$

- In some cases, we know the frequency of different values of estimation inaccuracy.
- 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.
- These estimations are usually expressed by using imprecise ("fuzzy") words from natural language.
- In such cases, a reasonable idea is to use fuzzy logic.
- Fuzzy logic techniques were specifically designed for handling this uncertainty.
6. Probabilistic Uncertainty: Examples

- In some cases, we know the pdf $\rho_j(\Delta y_j)$ for the estimation error $\Delta y_j = \tilde{y}_j - y_j$.

- If a measuring instrument returns the result 0.376, this means any value from 0.3755 to 0.3765.

- In general, $\tilde{y}_j$ means an interval $[\tilde{y}_j - \delta_j, \tilde{y}_j + \delta_j]$, for some small $\delta_j$.

- We can estimate the probability $P_j$ of this estimate as $P_j = \rho_j(\Delta y_j) \cdot (2\delta_j)$.

- Usually, all $\rho_j(\Delta y_j)$ belong to the same family $\rho_j(\Delta y_j) = \rho(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq})$ for known $\theta_{jk}$.

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

- Sometimes, some parameters $\beta_i, \ldots$ of the pdf are unknown: e.g., we may not know $\sigma$’s.
7. Probabilistic Uncertainty: General Description

- The set \{1, \ldots, N\} of all estimations is divided into several disjoint subsets \(S_\alpha\).
- For \(j \in S_\alpha\), the pdf of \(\Delta y_j\) is
  \[ \rho_\alpha(\Delta y_j, \theta j_1, \ldots, \theta j_q, \beta_\alpha_1, \ldots, \beta_\alpha t) \].
- Here, \(\theta_\alpha_1, \ldots\) are known, while \(\beta_\alpha_1, \ldots\) are not known.
- E.g.: different \(S_\alpha\) corr. to different measuring instruments, with 0 mean and unknown st. dev. \(\beta_\alpha_1 = \sigma_\alpha\):
  \[ \rho_\alpha(\Delta y) = \frac{1}{\sqrt{2\pi} \cdot \beta_\alpha_1} \cdot \exp \left( -\frac{(\Delta y)^2}{2\beta_\alpha_1^2} \right) . \]
8. Case of Fuzzy Uncertainty

- In the fuzzy case, we have membership functions instead of pdfs.
- The set \{1, \ldots, N\} of all estimations is divided into several disjoint subsets \(S_\alpha\).
- For \(j \in S_\alpha\), we have
  \[
  \mu_\alpha(\Delta y_j, \theta_{j_1}, \ldots, \theta_{j_q}, \beta_{\alpha_1}, \ldots, \beta_{\alpha_t}) .
  \]
- Here, \(\theta_{\alpha_1}, \ldots\) are known, while \(\beta_{\alpha_1}, \ldots\) are not known.
- E.g.: different \(S_\alpha\) corr. to different experts.
9. How Probabilistic Uncertainty Is Taken into Account in Information Fusion

- For each estimate $j$, the probability $P_j$ of having the estimate $\tilde{y}_j$ is proportional to the pdf.

- Approximation errors corresponding to different measurements are usually independent.

- So, the overall probability of having $N$ estimates $\tilde{y}_1, \ldots, \tilde{y}_N$ is proportional to

$$L = \prod_{\alpha} \prod_{j \in S_{\alpha}} \rho_{\alpha}(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_{\alpha}}, \beta_{\alpha1}, \ldots, \beta_{\alpha t_{\alpha}}),$$

where

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

- A reasonable idea is to find $x$’s, $\beta$’s, and $c$’s for which the probability $L$ is the largest.

- This idea is known as the Maximum Likelihood Method.
10. Gaussian (Normal) Case

- There are usually many different reasons for an estimation error.
- For example, for measurements, there is noise in each part of the measuring instrument.
- All these noises contribute to the overall estimation error.
- For large $N$, the distribution of the sum of $N$ small independent random variables is close to Gaussian.
- This Central Limit Theorem explains the ubiquity of Gaussian distributions.
- For measurements, bias can be detected and eliminated by re-scaling.
- It is thus reasonable to assume that the mean of $\Delta y_j$ is 0.
11. Gaussian (Normal) Case (cont-d)

- In this case, minimizing $-\ln(L)$ results in the Least Squares method:

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

- If we do not know $\sigma_\alpha$ for $j \in S_\alpha$, Maximum Likelihood leads to $\sigma_\alpha^2 = \frac{1}{N_\alpha} \sum_{j \in S_\alpha} (\tilde{y}_j - y_j)^2$, hence:

$$\sum_{\alpha} \ln \left( \sum_{j \in N_\alpha} (\tilde{y}_j - f(x, a_j, c))^2 \right) \rightarrow \min_{x,c}.$$
12. How Fuzzy Uncertainty Is Taken into Account

- We are interested in the degree $D$ to which:
  - $\Delta y_1$ is a possible value of the 1st approx. error, and
  - $\Delta y_2$ is a poss. value of the 2nd approx. error, etc.
- So, $D = f_&(D_1, D_2, \ldots, D_\alpha, \ldots)$, where
  $$D_\alpha = f_&(d_j : j \in S_\alpha).$$
- We select $\beta$’s and $c$’s for which $D$ is the largest.
- When $f_&(a, b) = a \cdot b$, we get the same problem as for probabilistic uncertainty.
- Every “and”-operation can be approximated, with any given accuracy, by an Archimedean one:
  $$f_&(a, b) = g^{-1}(g(a) \cdot g(b)).$$
- In particular, $\min(a, b)$ can be thus approximated.
13. How Fuzzy Uncertainty Is Taken into Account

- Every “and”-operation can be approximated, with any given accuracy, by an Archimedean one:
  \[ f_\&(a, b) = g^{-1}(g(a) \cdot g(b)). \]

- Thus, we can safely assume that \( f_\&(a, b) \) is Archimedean.

- Then, maximizing \( D \) is equivalent to maximizing
  \[ g(D) = \prod_{\alpha} \prod_{j \in S_{\alpha}} g(\mu_{\alpha}(\Delta y_j, \theta_{j1}, \ldots, \theta_{jq_{\alpha}}, \beta_{\alpha1}, \ldots, \beta_{\alpha t_{\alpha}})). \]

- Thus, we get the same expression as in probabilistic case, but with \( g(\mu_{\alpha}(\ldots)) \) instead of \( \rho_{\alpha}(\ldots) \).

- So, we can use the same algorithms as in the probabilistic case.
14. What Do We Know About Reliability?

- Sometimes the estimates $\tilde{y}_j$ correspond not to the object of interest, but to some other object.
- Usually, such situations are rare.
- From past experience, we can estimate how rare they can be.
- Thus, we can assume that for every $j$, we know:
  - either the probability $p_j$ that the $j$-th estimate is related to the desired quantities,
  - or the degree of confidence $q_j$ to which the $j$-th estimate is related to the desired quantity.
15. How to Take Reliability into Account: Probabilistic Case

- For every $j$, we have an additional unknown:
  - $z_j = 1$ if $j$-th estimate is related to the desired quantity,
  - $z_j = 0$ otherwise.
- When $z_j = 1$, the probability of observing $\tilde{y}_j$ is
  $$E_j = p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha).$$
- When $z_j = 0$, then $E_j = (1 - p_j) \cdot \rho_\alpha(\tilde{y}_j - y_j, \theta_j, \beta_\alpha)$.
- The values $y_j$ corr. to $z_j = 0$ are also unknown, so we find them from Maximum Likelihood:
  $$E_j = (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha).$$
16. How to Take Reliability into Account (cont-d)

- We select the largest of the two values, so
  \[ E_j = \max \left( (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha), \right. \]
  \[ \left. p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \right). \]

- We then find the values \( x, c, \) and \( \beta \) for which the product \( E_1 \cdot \ldots \cdot E_N \) is the largest.

- We already know how to solve the optimization problem corresponding to \( z_j \equiv 1. \)

- How can we transform this algorithm into an algorithm for solving the new problem?

- A natural idea is to use component-wise maximization:
  - first, we maximize over one group of variables,
  - then, over another group, etc.,
  - until the process converges.
17. Algorithm: General Case

- First, we pick \( z_j = 1 \) for all \( j \) and use Maximum Likelihood techniques to optimize over \( x, c, \) and \( \beta \).
- Once we find the corresponding values of \( x, c, \) and \( \beta \), we optimize over \( z_j \).
- Namely, we select \( z_j = 1 \) if and only if
  \[
  p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \geq (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha).
  \]
- Then, using only \( j \)'s with \( z_j = 1 \), we use Maximum Likelihood to find new estimates for \( x, c, \) and \( \beta \).
- This process continues until it converges.
18. Case of Normal Distributions

- The $E_j$-condition is: $1 - p_j \leq p_j \cdot \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right)$, i.e.,

$$|\Delta y_j| \leq \sigma_j \cdot \sqrt{2 \ln \left(\frac{p_j}{1 - p_j}\right)}.$$ So:

- Find $x$ and $c$ for which $\sum_{j=1}^{N} \frac{(\tilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2} \to \max$.

- Then, we select $z_j = 1$ if and only if

$$|\tilde{y}_j - f(x, a_j, c)| \leq \sigma_j \cdot \sqrt{2 \ln \left(\frac{p_j}{1 - p_j}\right)}.$$  

- Find $x$ and $c$ s.t. $\sum_{j: z_j = 1} \frac{(\tilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2} \to \max$.

- This process continues until it converges.
19. Taking Reliability into Account: Fuzzy Case

• When \( z_j = 1 \), the degree to which \( \tilde{y}_j \) is possible is
  \[
  D_j = f_\& (q_j, \mu_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)).
  \]
• When \( z_j = 0 \), then \( D_j = f_\& (1 - q_j, \mu_\alpha(\tilde{y}_j - y_j, \theta_j, \beta_\alpha)) \).
• For \( E_j = g(D_j) \), we thus get
  \[
  E_j = g(q_j) \cdot g(\mu_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)) \quad \text{if} \quad z_j = 1;
  \]
  \[
  E_j = g(1 - p_j) \cdot \max_y g(\mu_\alpha(y, \theta_j, \beta_\alpha)) \quad \text{if} \quad z_j = 0.
  \]
• We select the largest of the two values, so
  \[
  E_j = \max \left( q(1 - q_j) \cdot \max_y g(\mu_\alpha(y, \theta_j, \beta_\alpha)),
  
  g(q_j) \cdot g(\mu_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)) \right).
  \]
• We find the values \( x, c, \) and \( \beta \) that maximize \( g(D) = E_1 \cdot \ldots \cdot E_N \).
• We can thus use a similar algorithm.
20. Algorithm: Fuzzy Case

- First, we pick $z_j = 1$ for all $j$ and use Maximum Likelihood techniques to optimize over $x$, $c$, and $\beta$.
- Once we find the corresponding values of $x$, $c$, and $\beta$, we optimize over $z_j$.
- Namely, we select $z_j = 1$ if and only if
  \[
g(q_j) \cdot g(\mu_\alpha(\tilde{y}_j - f(x, a_j, c, \theta_j, \beta_\alpha)) \geq \]
  \[
g(1 - q_j) \cdot \max_y g(\mu_\alpha(y, \theta_j, \beta_\alpha)). \]
- Then, using only $j$’s with $z_j = 1$, we use Maximum Likelihood to find new estimates for $x$, $c$, and $\beta$.
- This process continues until it converges.
21. Conclusion

- In many application areas, we have several different pieces of information about an object 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 information is about other objects – and is, thus, 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.
22. Acknowledgments

- This work was supported in part by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721, and
- by an award from Prudential Foundation.