Data Fusion Is More Complex Than Data Processing: A Proof

Robert Alvarez, Salvador Ruiz, Martine Ceberio, and Vladik Kreinovich

Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
rjalvarez@miners.utep.edu, sruiz13@miners.utep.edu
mceberio@utep.edu, vladik@utep.edu

1. What is data processing: a brief reminder

- In many practical situations, we are interested in the value of a quantity y that is difficult – or even impossible – to measure directly.
- For example, we may be interested in tomorrow's temperature.
- Since we cannot measure this quantity directly, we can measure it indirectly. Namely:
- We find easier-to-measure quantities $x_1, \ldots, x_n$ that are related to y by a known dependence $y = f(x_1, \ldots, x_n)$.
- Then, we measure the values x_i and apply the algorithm $f(x_1, \ldots, x_n)$ to the measurement results $\widetilde{x}_i$.
- This produces an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$ for y.
- This is known as *data processing*

2. Data processing under uncertainty

- Measurements are never absolutely accurate.
- In many cases, all we know is the upper bound Δ_i on the absolute value of the measurement error $\widetilde{x}_i - x_i$.
- In such cases, after the measurement, all we know is that

$$x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i].$$

- In this case, it is desirable to find the range of all possible values of

$$y = f(x_1, \ldots, x_n).$$

- In general, computing this range is NP-hard.
- However, there are cases when computable is feasible.
- One such case is when if $f(x_1, \ldots, x_n)$ is a *Single Use Expression* (SUE), in which each variable occurs only one, (e.g., $x_1 + x_2^{x_3}$).

3. What is data fusion: a brief reminder

- To describe the state of an object, we need to know the values of the physical quantities $x_1, \ldots, x_m$ that characterize this object.
- To determine this state, we can measure all these quantities.
- Usually, the quantities are not completely independent.
- There are constraints that relate them, and these constraints can help to decrease inaccuracy.
- For example, if we know that $x_1 \in [0.9, 1.1]$, $x_2 \in [0.8, 1.0]$ and $|x_1 - x_2| \le 0.01$, then we can conclude that $x_1 \in [0.9, 1.01]$.
- This decreasing-of-inaccuracy combination of several measurement results is known as *data fusion*.

4. Problem and what we do

- Empirical evidence shows that, in general, data fusion is more time-consuming than data processing.
- In this talk, we *prove* that data fusion is indeed more complex than data processing.
- Specifically, we prove that even if all the constraints are described by SUE expressions, data fusion is still, in general, NP-hard.
- Since for SUE, data processing is feasible, this means that data fusion is indeed more complex.

5. Proof

- Let us consider the variables $x_1, \ldots, x_n, y_1, \ldots, y_n$, and y.
- Let us assume that we only measure x_i, and that the variables are related by SUE constraints $x_i = y_i$ and $y = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} y_i\right)^2$.
- Under these constraints, the range of y is equal to the range of the sample variance $\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n} \cdot \sum_{i=1}^{n} x_i\right)^2$ under interval uncertainty.
- It is known that the problem of computing this range is NP-hard.

6. References

- D. Dubois, H. Prade, and R. R. Yager, “Computation of intelligent fusion operations based on constraint fuzzy arithmetic”, *Proceedings of the 1998 IEEE International Conference on Fuzzy Systems FUZZ-IEEE’98*, Anchorage, Alaska, May 4–9, 1998, Vol. 1, pp. 767–772.
- V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing And Interval Computations*, Kluwer, Dordrecht, 1998.
- H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.

7. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).