Integrating Domain Knowledge With Data: From Crisp To Probabilistic and Fuzzy Knowledge

Hung T. Nguyen
Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

Vladik Kreinovich
Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@cs.utep.edu

Leonid Reznik
Electrical & Electronic Engineering
Footscray Campus
Victoria Univ. of Technology
MMC Melbourne VIC 8001, Australia
leonid@cabsav.vu.edu.au

Abstract

It is well known that prior knowledge about the domain can improve (often drastically) the accuracy of the estimates of the physical quantities in comparison with the estimates which are solely based on the measurement results. In this paper, we show how known (but often under-used) methods of integrating crisp and statistical domain knowledge with data can be (naturally) extended to the case when the domain knowledge is described in fuzzy terms.

1 INTEGRATING DOMAIN KNOWLEDGE WITH DATA: FORMULATION OF THE PROBLEM

A large part of information about the world comes from measurements. However, for many complex systems, some characteristics are very difficult to measure: e.g., for a jet engine, it is difficult to measure the temperature and pressure inside the jet chamber, where the temperatures are very high; for a human body, it is difficult to measure the characteristics of the internal organs, etc. In many such cases, experts have some knowledge about the domain. It is therefore desirable to use this knowledge to improve the measurement results.

Known statistical methods can indeed drastically improve (often drastically) the accuracy of the estimates of the physical quantities in comparison with the estimates which are solely based on the measurement results. This is well known in statistical filtering where model knowledge is used to improve statistical estimates; see, e.g., (Oppenheim et al. 1999). This is also well known in static case; see, e.g., (Nguyen et al. 1989), (Casella et al. 1990), (Box et al. 1992), (Goodman et al. 1997), (Goutsias et al. 1997), (Barnett 1999). Specific applications to different static measurement situations have been described, e.g., in (Reznik et al. 1985), (Kreinovich et al. 1986), (Kreinovich et al. 1990), (Luo et al. 1989), (Kreinovich Chang 1992), (Kreinovich Quintana 1992).

In this paper, we make two points:

- first, that although the statistical methods are long established, they are still sometimes under-used;
- second, that these methods can be (naturally) extended to the case when the domain knowledge is described in fuzzy terms.

2 INTEGRATING CRISP AND STATISTICAL DOMAIN KNOWLEDGE WITH DATA

2.1 TOY EXAMPLE

As we have already mentioned, if we know the *a priori* relation between the measured quantities $x_1, \ldots, x_n$, then we can often use this relation to correct the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ and thus, to increase the accuracy of our measurements.

Let us show, on a simple example, that such an improvement is indeed possible. Let us assume that we are measuring two quantities x_1 and x_2, that both quantities are measured by a measuring instrument with zero mean and standard deviation σ, and that the results of these measurements are $\widetilde{x}_1$ and $\widetilde{x}_2$.

If we do not have any *a priori* information about the relationship between x_1 and x_2, then the only thing that we know about, e.g., x_1 is that x_1 is approximately equal to $\widetilde{x}_1$ with a standard deviation σ.

If we know that, in reality, these two quantities should have the same value, i.e., that $x_1 = x_2$, then, for each of these quantities, we have, in effect, not one but *two* independent measurements with the same accuracy σ, resulting in the values $\widetilde{x}_1$ and $\widetilde{x}_2$. It is well known that in this situation, we can get a better estimate for x_1 if we take an arithmetic average of these two measurements: $\widetilde{x} = (\widetilde{x}_1 + \widetilde{x}_1)/2$; this new estimate has a standard deviation $\sigma/\sqrt{2}$.

2.2 CASE STUDY: TELEMANIPULATORS

Let us give an example showing that, although statistical methods are long established, they are still sometimes under-used.

In many real-life mechanical tasks, it is difficult or even impossible to use humans:

- Some environments are too dangerous for a human being: for example, when we manipulate objects in space, inside a radioactive part of a nuclear reactor, in a dangerous chemical environment, or even in a potentially dangerous environment such as handling viruses that cause deadly diseases.

- In other environments, there is no danger to the human operator, but there is a significant risk of contamination of the object: e.g., in handling microchips, lunar samples, etc.

In all these cases, reasonably simple mechanical tasks can be done by an automatic mechanical hand-arm. However, there is a limit on the complexity of the tasks that automatic devices can do. For more complicated tasks, for which we cannot use a completely automated system, we must use *telemanipulators*, i.e., devices in which a mechanical hand-arm copies the movements of a human operator.

To achieve the best possible accuracy of such copying, statistical methods such as statistical inference, filtering, and planning are used to determine parameters of telemanipulators. Methods used in the control and planning of telemanipulators are a particular case of general methods used to control and plan mobile robots and manipulators; see, e.g., (Lowrance et al. 1990), (Draper et al. 1994), (Doan et al. 1996), (Givan et al. 1997), (Geffner et al. 1998), (Kaelbling et al. 1998), (Littman et al. 1998), (Trejo et al. 2000).

2.3 TELEMANIPULATORS: SUCCESSES

The main goal of the telemanipulator is to reproduce the operator's movements as accurately as possible.

A human hand is a very flexible instrument. In mechanical terms, we can say that it has many degrees of freedom: we can move and rotate the hand itself, the arm, each finger, parts of each finger, etc. Thus, to reproduce its movements accurately, the manipulator also has to have many degrees of freedom.

At present, the best of widely available hand-arm manipulators, the Utah/MIT hand, has 22 degrees of freedom. It is still slightly less than a human hand, because, e.g., it only has 4 fingers and not 5. However, it can perform many important tasks that a human hand can do.

This hand was not designed for telemanipulation only. It has many other applications: e.g., it can even twist itself into the positions that would have been impossible for a human hand.

2.4 TELEMANIPULATORS: PROBLEMS

Both in the Hollywood movies and in the self-made movies that researchers show at robotic conferences, telemanipulation works perfectly well: a robotic hand exactly reproduces the operator's movements. This is indeed happening in many application areas, but this reproduction accuracy is extremely difficult to achieve.

If we simply measure the pressure, etc., applied by the operator's arm, and send exactly proportional control signal to the electric motors that control different degrees of freedom of the robotic arm, we get a behavior that is often drastically different from what the operator did. For example, the operator's firm grip on the object may be distorted into the robotic arm dropping it, and vice versa, the operator's tender approach to a fragile object may result in a robotic arm's bumping into the actual object and damaging it.

There are three main reasons for the difference between the movements of the human and robotic hands:

- first, the sensors that measure the human hand's pressure are not 100% accurate;

- second, the motors and actuators are not perfect, and do not react precisely to the commands;

- third, the mechanical characteristics of the robotic hand itself are somewhat different from the mechanical characteristics of the operator's hand.

2.5 AS MANIPULATORS GET MORE COMPLICATED, THESE PROBLEMS GET MORE AND MORE IMPORTANT

The above inaccuracy problems can be traced even on the example of simple manipulators that have a few degrees of freedom, but for more advanced manipulators, these problems become more and more acute.

Indeed, for a manipulator, more advanced means that this manipulator has more degrees of freedom. Each degree of freedom bring its own inaccuracy, so if we have 22 degrees of freedom, then in principle, we get 22 sources of inaccuracy all leading to the huge inaccuracy of the resulting action.

Let us give a simple example.

- If we have a 3-finger manipulator, then for this manipulator to grip an object, it must place one finger below it, and two fingers above it. Due to inaccuracy, we may have a slightly distorted position, but we will still keep firmly 3 points on the object.
- For a 4-finger arm that is similar to the human arm, we need to place 3 fingers on top of the object. If, e.g., the upper surface is planar, we must have all 3 fingers on one line. Due to inaccuracy, one of these fingers may be higher than the others. As a result, this finger may not contact the object at all, and hence, the grip will not be as firm as we desired.

So inaccuracy is harmful.

2.6 NEW IDEA: VIRTUAL TOOLS

Although a human hand-arm has many degrees of freedom, we rarely use all of them in the same movement. Usually, the movement in different degrees of freedom is very much *coordinated.*

For example, if we have already firmly grasped an object, then we move the arm as a whole and, unless necessary, do not use the ability to move fingers and/or or fingertips separately.

There is a limited number of typical movements of this type, and a teleoperator can pretty well describe which of these typical movements he is applying at any given moment of time. When we get this information, we can use it to set up a similar coordination between the degrees of freedom of the robotic hand-arm.

In precise terms, a fixed movement type means that we cannot have *arbitrary* values of $x_1, \ldots, x_n$: these values must satisfy one or several restrictions (constraints). For example, if we want the arm to move as a whole, then one of these constraints may take the form $x_1 = x_2$ (or $x_1 - x_2 = 0$), where x_1 and x_2 are pressures applied by two fingers. If we want to preserve the distance between the two fingertips, then we may require something like $(x_1 - x_2)^2 + (x_3 - x_4)^2 - \text{const} = 0$. In general, we may have several restrictions of this type.

These restrictions can indeed decrease inaccuracy. For example, if the two fingers have to move in the exact same way, i.e., that $x_1 = x_2$, then, as we have mentioned in the above pedagogical example, we can decrease the standard deviation $\sqrt{2}$ times in comparison with the standard deviation of the measuring instrument.

So, the new idea is as follows: Possible hand movements are divided into several movement types, such as:

- "careful translational movements", in which the distances between all the joints stay the same, and the entire finger configuration stays parallel to its original position;
- "careful rotational movements", in which the distances between all the joints stay the same, but the entire finger configuration may turn with respect to its original position;

and others. When an operator moves his hand, he also indicates which of these movement types he is performing; the telemanipulator then uses the restrictions corresponding to this particular movement type to decrease the inaccuracy of the corresponding statistical estimates and thus, to improve the accuracy of the telemanipulator.

For each movement type, the originally flexible robotic hand acts, in effect, as a new tool that is specifically designed for this type of movement. For example, when the operator is performing what he indicates to be a translational movement, the robotic arm is restricted to only three degrees of freedom corresponding to three coordinates of the corresponding translation. With several degrees of freedom switched off, this same robotic arm acts as a simplified "virtual tool". Different movement types transform the robotic arm into different virtual tools.

The actual implementation of this "virtual tool" idea for the Utah/MIT hand is described in (Nelson et al. 1995), (Fuentes et al. 1996), (Jägersand et al. 1997), (Kreinovich et al. 1998). This implementation does lead to an improved telemanipulation performance.

3 TOWARDS INTEGRATING FUZZY DOMAIN KNOWLEDGE WITH DATA

3.1 THE NEED FOR FUZZY DOMAIN KNOWLEDGE

Often, experts can only formulate their knowledge in terms of words from natural language. For example, an expert can say that "the temperature inside the chamber is about 10,000 degrees" (or "huge"), where the word "about" does not have a precise meaning.

A natural way to formalize such a knowledge is to use *fuzzy logic*, in which each such term like "small" is described by a *membership function* $\mu_{\text{small}}(x)$ which tells, for every value x, the degree $\mu_{\text{small}}(x) \in [0, 1]$ to which this value is small. After this formalization, we get a *fuzzy knowledge domain*.

3.2 HOW CAN WE COMBINE THIS KNOWLEDGE WITH DATA?

A natural question is: *How can we combine fuzzy domain knowledge with data?*

There exist many heuristic methods of combining such data, but since we want to *guarantee* that this combination indeed improves the measurement results (and, ideally, improves them in the best possible way), we would like to have *guaranteed* methods of combining fuzzy domain knowledge with data.

3.3 OUR MAIN IDEA: LET US TREAT AN EXPERT AS YET ANOTHER MEASURING INSTRUMENT

From the viewpoint of getting new information about the world, we can consider an expert as one more measuring instrument, an instrument which brings us additional information about the actual (unknown) values of different physical quantities. From this viewpoint, the only difference between an expert and a more traditional measuring instrument is that a measuring instrument returns a *number* while an expert returns a *word* ("small", "large", etc.). Within this approach, we can determine the statistical characteristics of an expert along the same lines as we determine the statistical characteristics of the measuring instrument.

Namely, to determine the statistical characteristic of a measuring instrument, we use it to measure several quantities for which the precise value has already been measured by a (more) precise (standard) measuring instrument. Then, for each possible value x of the measured quantity, we group together all the measurements in which the actual value was x (or close to x). For all measurement results $\widetilde{x}$ from the group, we compute the measurement error $\Delta x = \widetilde{x} - x$, and then we use the standard statistical analysis to determine the (conditional) probability distribution of this error $P(\Delta x | x)$.

Similarly, we take several situations in which the desired quantity has different values and whose value has been measured by an accurate measuring instrument. For each of these situations, we can then ask an expert to provide us with a word which best describes her perception of the corresponding value x. Then, for each possible value x of the measured quantity, we group together all the measurements in which the actual value was x (or close to x). Then, for each of the words w which an expert used, we can compute the (conditional) probability $P(w|x)$ that an expert used the word w under the condition that the actual value was x.

As soon as we have this statistics, we can process the expert's estimates just like we would have processed one more measurement result. In this case, mathematical statistics guarantees that on average, combining expert estimates with the measurement results do not make things worse.

3.4 COMMENTS

1. Our idea is not as heretic as it may seem (reducing fuzzy to probabilities?):
   - On one hand, our method of determining a *probability* function $P(w|x)$ corresponding to a word w is actually very similar to the main methods of determining the *membership* functions that describe a word w; see, e.g., (Dubois et al. 1980), (Norwich et al. 1984), (Turksen 1991). So, we can view the function $P(w|x)$ as a membership function.
   - On the other hand, the very fact that fuzzy set theory can be, in principle, reformulated in more statistical terms (namely, in terms of random sets) is known for more than two decades; see, e.g., (Nguyen 1979).
2. Of course, the fact that we can *formally* describe the expert's fuzziness in probabilistic terms does not mean that this description captures all aspects of fuzziness. However, the empirical success of this idea (see references below) shows that this description can be viewed as a reasonable first approximation to fusing fuzzy and statistical data.

3.5 PEDAGOGICAL EXAMPLE

Let us consider a simple example of how this combination can be done. Let us assume that an expert estimated the value as "small", and that the corresponding probability function for "small" is $P(\text{small}|x) = \mu(x)$. Let us also assume that we have made a single measurement by a measuring instrument whose measurement error Δx is normally distributed with zero mean and known standard deviation σ, and that the measurement result is $\widetilde{x}$. The corresponding probability density will be denoted by $\rho(\widetilde{x} - x)$.

We want to use Maximum Likelihood Method and find the value x for which the value of the likelihood function $L(x)$ is the largest possible. For each x, the probability density of getting a value $\widetilde{x}$ is equal to $\rho(\widetilde{x}-x)$, the probability of an expert's estimate's "small" is equal to $\mu(x)$. It is natural to assume that the measuring instrument and an expert are statistically independent. Therefore, the likelihood $L(x)$ that, for a given x, both the measurement result $\widetilde{x}$ and the expert estimate "small" is equal to the product $L(x)=\rho(\widetilde{x}-x)\cdot\mu(x)$.

To find the maximum of this function $L(x)$, it is usually convenient to seek a minimum of its negative logarithm

$$-\log(L(x))=\frac{(\widetilde{x}-x)^2}{2\sigma^2}-\ln(\mu(x))\to\min.$$

Differentiating this expression w.r.t. x and equating the derivative to 0, we conclude that

$$x=\widetilde{x}+\sigma^2\cdot\frac{\mu'(x)}{\mu(x)},$$

where $\mu'(x)$ denotes the derivative of $\mu(x)$. Since $x\approx\widetilde{x}$, we have $\mu(x)\approx\mu(\widetilde{x})$; thus, can neglect the terms which are of second order w.r.t. Δx, and get an explicit analytical expression

$$x\approx\widetilde{x}+\sigma^2\cdot\frac{\mu'(\widetilde{x})}{\mu(\widetilde{x})}.$$

The difference between x and $\widetilde{x}$ is a correction term which is motivated by the fuzzy domain knowledge and which makes the resulting estimate more accurate (provided, of course, that the experts were correct).

3.6 THIS IDEA HAS BEEN SUCCESSFULLY USED IN PRACTICAL APPLICATIONS

Our foundations are new, but the resulting formulas (which we obtained heuristically) have actually been successfully used in practical applications; see (Kreinovich et al. 1986) and references therein.

The domain to which we applied it was testing jet engines. In this domain, expert estimates are extremely important, because an important part of this testing is trying to figure out what is going on in the high-temperature regions, and the temperatures are so high there that we cannot place any sensors. So the only available information about these regions is the experts estimates.

One of the authors (L.R.) used this fuzzy representation of uncertainty in designing software for the automated jet engines testing system IVK-12. This system was successfully used to test jet engines for aircraft and spaceships.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6). by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518, and by the National Security Agency under Grant No. MDA904-98-1-0561.

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

References

V. Barnett, *Comparative Statistical Inference*, Wiley, N.Y., 1999.

G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Wiley, N.Y., 1992.

G. Casella and R. L. Berger, *Statistical Inference*, Brooks/Cole, 1990.

A. Doan and P. Haddawy, "Sound Abstraction of Probabilistic Actions in the Constraint Mass Assignment Framework ", *Proc. UAI'96*, pp. 228–235.

D. Draper and S. Hanks, "Localized Partial Evaluation of Belief Networks", *Proc. UAI'94*.

D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, N.Y., London, 1980.

O. Fuentes and R. C. Nelson, "The Virtual Tool Approach to Dextrous Telemanipulation", *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, Minneapolis, Minnesota, April 1996.

H. Geffner and B. Bonet, "High-Level Planning and Control with Incomplete Information Using POMDPs", *Proc. AIPS-98 Workshop on Integrating Planning, Scheduling and Execution in Dynamic and Uncertain Environments*, 1998.

R. Givan, S. Leach, and T. Dean, "Bounded parameter Markov decision processes", *Proc. 4th European Conference on Planning*, Toulouse, France, 1997.

I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*, Kluwer, Dordrecht, 1997.

J. Goutsias, R. P. S. Mahler, and H. T. Nguyen (eds.), *Random Sets: Theory and Applications*, Springer-Verlag, N.Y., 1997.

M. Jägersand, O. Fuentes and R. C. Nelson, "Experimental Evaluation of Uncalibrated Visual Serving for Precision Manipulation", *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, Albuquerque, New Mexico, April 1997.

L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains", *Artificial Intelligence*, 1998, Vol. 101, pp. 99–134.

V. Kreinovich, C.-C. Chang, L. Reznik, and G. N. Solopchenko, "Inverse problems: fuzzy representation of uncertainty generates a regularisation", *Proceedings of NAFIPS'92: North American Fuzzy Information Processing Society Conference, Puerto Vallarta, Mexico, December 15–17, 1992*, NASA Johnson Space Center, Houston, TX, 1992, Vol. II, pp. 418–426.

V. Kreinovich and L. O. Fuentes, "Telemanipulation: The Virtual Tool Approach and Its Interval-Based Justification", In: G. Alefeld and R. A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.

V. Kreinovich, C. Quintana, and L. Reznik, "Gaussian membership functions are most adequate in representing uncertainty in measurements", *Proceedings of NAFIPS'92: North American Fuzzy Information Processing Society Conference, Puerto Vallarta, Mexico, December 15–17, 1992*, NASA Johnson Space Center, Houston, TX, 1992, Vol. II, pp. 618–624.

V. Kreinovich and L. K. Reznik. "Methods and models of formalising a priori information (on the example of processing measurements results)," *Analysis and Formalization of Computer Experiments, Proceedings of the Mendeleev Metrology Institute*, 1986, pp. 37–41 (in Russian).

V. Kreinovich and L. K. Reznik, "Prospects of using expert systems in intelligent measuring instruments", *Proceedings of the International Workshop on Aspects of Intelligent Measurement*, Dagomys, Sochi, USSR, 20–30 October, 1989, Moscow, 1990, Part 2, pp. 97–110 (in Russian).

M. L. Littman, J. Goldsmith, and M. Mundhenk, "The Computational Complexity of Probabilistic Planning", *JAIR*, 1998, Vol. 9, pp. 1–36.

J. D. Lowrance and D. E. Wilkins, "Plan evaluation under uncertainty," *Proc. Workshop on Innovative Approaches to Planning, Scheduling and Control*, Morgan Kaufmann, San Francisco, 1990, pp. 439–449.

R. C. Luo and M. G. Kay, "Multisensor integration and fusion in intelligent systems", *IEEE Trans. Systems, Man, Cybern.*, 1989, Vol. 19, No. 5, pp. 901–931.

R. C. Nelson, M. Jägersand and O. Fuentes, "Virtual Tools: A Framework for Simplifying Sensory-Motor Control in Complex Robotic Systems", *Proceedings of the '95 Vision for Robots Workshop*, Pittsburgh, PA. August 1995.

H. T. Nguyen, "Some mathematical tools for linguistic probabilities", *Fuzzy Sets and Systems*, 1979, Vol. 2, pp. 53–65.

H. T. Nguyen and G. S. Rogers, *Fundamentals of Mathematical Statistics*, Springer-Verlag. N.Y., 1989.

H. T. Nguyen, B. Wu, and V. Kreinovich, "On Combining Statistical and Fuzzy Techniques: Detection of Business Cycles From Uncertain Data", In: H. Mohanty and C. Baral (eds.), *Trends in Information Technology, Proceedings of the International Conference on Information Technology ICIT'99, Bhubaneswar, India, December 20–22, 1999*, Tata McGraw-Hill, New Delhi, 2000, pp. 69–74.

A. M. Norwich and I. B. Türkşen, "A model for the measurement of membership and the consequences of its empirical implementation", *Fuzzy Sets and Systems*, 1984, Vol. 12, No. 1, pp. 1–25.

A. V. Oppenheim, R. W. Shafer, and J. R. Buck, *Discrete-time Signal Processing*, Prentice Hall, Upper Saddle River, N.J., 1999.

L. Reznik and G. N. Solopchenko, "Use of a priori information on functional relations between measured quantities for improving accuracy of measurement", *Measurement*, 1985, Vol. 3, No. 3, pp. 98–106.

R. Trejo, V. Kreinovich, and C. Baral, "Towards Feasible Approach to Plan Checking Under Probabilistic Uncertainty: Interval Methods", *Proc. of the 17th National Conference on Artificial Intelligence AAAI'2000*, Austin, TX, July 30–August 3, 2000 (to appear).

I. B. Türkşen, "Measurement of membership functions and their acquisition", *Fuzzy Sets and Systems*, 1991, Vol. 50, No. 1, pp. 5–38.