Representation of interval data by weak orders yields robustness of the data fusion outcomes

S V Muravyov, L I Khudonogova and I A Marinushkina
Department of Computer-aided Measurement Systems and Metrology
National Research Tomsk Polytechnic University
Tomsk, 634050, Russia

E-mail: muravyov@tpu.ru

Abstract. It is shown in the paper that an interval data fusion procedure can be carried out by means of representation of intervals on the real line by weak order relations (or rankings) over set of discrete values belonging to these intervals. It is possible to determine a consensus ranking for collection of discrete values rankings, corresponding to initial intervals. The highest ranked value, accepted as a result of the fusion, guarantees improved accuracy and robustness of the interval data fusion procedure outputs.

1. Introduction

Description of measurement results in form of intervals with bounds defined by experimentally obtained or given beforehand uncertainty values is rather common both in theory and practice of measurement. Moreover, interval data are typical kind of information in such areas as distributed computations, data bases, systems and networks of data acquisition, etc. [1,2].

Data fusion is a process of joint processing of data on some object obtained from multiple sources aiming to acquire fuller, more objective and accurate knowledge of a characteristic under investigation than knowledge derived from a single source. List of data fusion methods usually includes mathematical statistics and probability theory, fuzzy sets theory, possibility theory, Dempster-Shafer evidence theory [3], Bayesian approach, different artificial intelligence methods [4], methods of voting and preference aggregation [5].

A simple example of data fusion utility can be taken from a classical analysis of repeated observations that is considerably older than comparatively new concept of data fusion. Indeed, from mathematical statistics point of view, single measurement result \( x \) is characterised by standard deviation \( S \), while estimation \( \bar{x} = \sum_{i=1}^{n} x_i \) of expectation in a form of arithmetic mean of \( n \) normally distributed observations has the standard deviation \( S(\bar{x}) = S / \sqrt{n} \). Therein lies the only useful effect of arranging repeated observations instead of a single one.

Procedure of interval data fusion consists in shaping an interval to be consistent with maximal number of given initial intervals (not necessary consistent among each other) and to be with maximal likelihood including a value \( x^* \) that can serve as representative of all the given intervals.

Interval data fusion procedures are applied in interlaboratory comparisons where main task is to establish a reference value that characterizes a largest subset of consistent (reliable) measurement data [6]. Other applications are prediction of fundamental constant values on the base of different measured
values [7], conformity testing [8] and enhancement of multisensor’s readings accuracy in sensor networks [9].

In this paper, it is proposed an interval data fusion procedure that represents intervals on the real line by means of weak order relations (or rankings) over set of belonging to these intervals discrete values. One may determine a consensus ranking for collection of discrete values rankings, corresponding to initial intervals. The highest ranked value, accepted as a result of the fusion, guarantees improved accuracy and robustness of the interval data fusion procedure outputs.

2. Terms and notations

2.1. Intervals

We will consider a collection of $m$ closed intervals $\{I_k\}$, $k = 1, \ldots, m$, in the real line. Each interval is characterized by lower bound $l_k$, upper bound $u_k$ and middle point $x_k$, so that $I_k = [l_k, u_k]$; $l_k < x_k$; $x_k = 0.5 \cdot (u_k + l_k)$; $l_k, u_k, x_k \in \mathbb{R}$ (see figure 1).

Frequently, the middle point $x_k$ is a result of some measurement, uncertainty interval of which is $[x_k - 0.5 \cdot (u_k - l_k), x_k + 0.5 \cdot (u_k - l_k)] = [x_k - \varepsilon_k, x_k + \varepsilon_k]$. In this case, $k$-th interval can be represented by pair $<x_k, \varepsilon_k>$.

Any two intervals do not necessary have non-empty intersection, i.e. there can be such two intervals $I_i$ and $I_j$, $j \neq k; j, k = 1, \ldots, m$, that $I_i \cap I_k \neq \emptyset$. Such the intervals we will call inconsistent.

2.2. Range of actual values and its partition

Notice that aim of interval fusion is a selection of a point $x^*$ on the real line that belongs to maximal number of intervals from $\{I_k\}$. We are going to choose the point from finite number of elements connected to the intervals, though, generally speaking, each of closed intervals in the real axis is infinite. For this purpose, let us introduce a range of actual values (RAV) $A = \{a_1, a_2, \ldots, a_n\}$, over that there exists (inherited from the real line) binary relation of linear (full) order $a_1 < a_2 < \ldots < a_n$, i.e. transitive, antisymmetric and linear binary relation.

The RAV could be shaped by means of union of the intervals $\{I_k\}$ and following partition of the union result into equal subintervals. However, existence of inconsistent intervals can produce a disjoint output of the union. In order to exclude possible breaks, in the capacity of the RAV’s lower bound $a_1$ a least lower bound for all the intervals is chosen, i.e.

$$a_1 = \min \{l_k \mid k = 1, \ldots, m\};$$

(1)

and as the upper bound $a_n$ a largest upper bound for these intervals is taken, i.e.

$$a_n = \max \{u_k \mid k = 1, \ldots, m\}. $$

(2)

To generate elements $a_2$, $a_3$, $\ldots$, $a_{n-1}$ let us partition the obtained interval $[a_1, a_n]$ into $n - 1$ equal subintervals of length

$$c = (a_n - a_1)/(n - 1).$$

(3)

The length $c$ will be called norm (or mesh) of the partition. Clear that, after partitioning, the norm is defined by the formula

$$c = |a_i - a_{i-1}|, \quad i = 2, \ldots, n,$$

(4)

and $i$-th RAV’s element is

$$a_i = a_{i-1} + c, \quad i = 2, \ldots, n.$$
Notice that choice of an appropriate number \( n - 1 \) of partition subintervals must guarantee necessary and sufficient accuracy of values \( a_i, i = 1, ..., n \), as one of them will be taken as interval fusion result \( x^* \).

### 2.3. Representation of intervals by rankings

Ranking of \( n \) elements of a set \( A = \{a_1, a_2, ..., a_n\} \) in the form \( \lambda_k = (a_1 \succ a_2 \cdots \succ a_i \succ \cdots \succ a_n) \) specifies a relation of weak order (preference relation) on the set \( A \). The preference relation \( \lambda \) is a union of two following relations: strict preference relation \( \rho \), i.e. \( a_i \succ a_j \), and equivalence relation \( \tau \), i.e. \( a_i \sim a_j \), that is \( \lambda = \rho \cup \tau \). Set of \( m \) rankings \( \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_m\} \) of \( n \) elements is called preference profile for \( m \) and \( n \) given.

Some \( k \)-th interval of \( \{I_k\}, k = 1, ..., m \), can be represented by ranking \( \lambda_k \) using the following rules at \( i,j = 1, ..., n \):

\[
\begin{align*}
    a_i \succ a_j & \text{ if } a_i \in I_k \text{ and } a_j \notin I_k; \\
    a_i \sim a_j & \text{ if } a_i, a_j \in I_k \text{ or } a_i, a_j \notin I_k; \\
    a_i < a_j & \text{ if } a_i \notin I_k \text{ and } a_j \in I_k.
\end{align*}
\]

Notice that \( k \)-th ranking consists of two equivalence classes, one of which includes elements of \( A \) belonging to the interval \( I_k \), and another includes all other elements of \( A \). At that, the former class is strictly preferred over the latter one. Hence, each ranking contains a single symbol of strict order \( \succ \) and \( n - 1 \) symbols of equivalence \( \sim \). Thus, the interval collection \( \{I_k\}, k = 1, ..., m \), can be represented in accordance with equations (6), (7) and (8) by the preference profile \( \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_m\} \).

### 3. Fusion of intervals

By interval data fusion we will understand a procedure of determination of a point \( x^* \) position in the real line that belongs to maximal number of intervals from \( \{I_k\}, k = 1, ..., m \), and that can be considered to represent all middle points \( x_k \) at minimal uncertainty \( \pm \varepsilon_k \).

Notice that fusion methods based on classic mathematical statistics work on continuous intervals. The same is, for example, procedure A [6] applied to fuse measurement results uncertainty intervals provided by laboratories when processing comparison data. In this procedure, to estimate a reference value \( x^* \) one uses a weighted mean value \( y \), and consistency of intervals is examined by criterion \( \chi^2 \) calculated for \( y \). Procedure A can be reasonably applied if the measurement results are characterized with normal probability distribution. In the cases where measurement results distribution differs from normal or unknown it is desirable to use robust methods for interval data fusion.

The property of robustness is inherent to the procedure proposed below that will be called interval fusion with preference aggregation (IF&PA). This procedure works on the set \( A \) of discrete values obtained in the way described in section 2.2 that allows to represent intervals by rankings.

Main stages of the proposed procedure IF&PA are as follows.

1. Generation of range of actual values \( A = \{a_1, a_2, ..., a_n\} \) of initial intervals \( \{I_k\}, k = 1, ..., m \).
2. Representation of the intervals by rankings and composing preference profile \( \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_m\} \).
3. Determination of value \( x^* \) as the best alternative in consensus ranking for the profile \( \Lambda \). For this aim, any algorithm can be used that implements some of voting (preference aggregation) rules, for example, simple majority, Condorcet rule, Borda count, Kemeny rule, approval voting [10], etc. Authors prefer to use the Kemeny rule at this stage [11]. The rule produces multiple consensus rankings [12] which should be transformed into a single final consensus ranking. All the consensus rankings are strict preference relations and the final consensus ranking \( \beta \) is a weak order [13].
4. Deletion of inconsistent intervals in set \( \{I_k\} \), i.e. those not including the value \( x^* \). Then a cardinality of the output set becomes to be equal to \( m' \).
5. Determination of uncertainty \( \varepsilon^* \) of the value \( x^* \) as by the formula

\[
\varepsilon^* = \min\left( \max_{k=1}^{m'} \{ l_k \leq x^* \}, \min_{k=1}^{m'} \{ u_k \geq x^* \} \right).
\]  

The stages outcomes are illustrated in figure 2 for an example collection of seven initial intervals.

4. Robustness of fusion of intervals represented by rankings

Robustness, that is independence on particular law of interval data distribution, has been examined for the procedure IF&PA (with Kemeny rule at stage 3, see Section 3) by means of numerical experimentation. There were generated \( N = 100 \) uniformly and normally distributed interval collections (individual problems) in the vicinity of a fixed nominal value \( x_{\text{nom}} = 3 \); number of intervals \( m = 15 \). An example of generated data for 15 intervals is shown in table 1.

Table 1. Generated data for individual problem #33.

| \( k \) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| \( x_k \) | 2.918 | 3.085 | 3.225 | 2.491 | 2.989 | 3.222 | 2.738 | 3.069 | 2.559 | 3.100 | 2.918 | 2.894 | 2.626 | 2.899 | 3.077 |
| \( \varepsilon_k \) | 0.190 | 0.244 | 0.289 | 0.053 | 0.213 | 0.288 | 0.133 | 0.239 | 0.075 | 0.249 | 0.191 | 0.183 | 0.097 | 0.184 | 0.242 |

We will judge a procedure's performance by deviation

\[
\xi = |x^* - x_{\text{nom}}|
\]

of value \( x^* \) from \( x_{\text{nom}} \). The deviation \( \xi \) can serve as a measure of a data fusion process robustness [14] since if, for some individual problem, values \( \xi \) at normal and uniform data distributions are close enough to each other then the corresponding procedure should be deemed as a robust one. Some fragment of the obtained unsorted numerical experimental results is shown in table 2.
In figure 3 graphs of deviations $\xi$ are presented that were obtained by means of procedure A and procedure IF&PA for normal and uniform distributions. Values $\xi$ are shown for each of 100 individual problems and sorted in ascending order. One can see from Figure 3 that discrepancy by $\xi$ between normal and uniform distributions (curves 3 and 4), in case of using IF&PA, does not mainly exceed 0.08. Thus, the procedure IF&PA is explicitly more robust than the procedure A that produced considerably larger discrepancy about 0.46 (curves 1 and 2). It follows from the presented in Figure 3 experimental data that values $x^\star$ obtained by the procedure IF&PA are essentially closer to $x_\text{nom}$ than those obtained by the procedure A.

Table 2. A fragment of numerical experimental results (unsorted).

| Problem number | Procedure IF&PA | Procedure A |
|----------------|-----------------|-------------|
|                | Normal distribution | Uniform distribution | Normal distribution | Uniform distribution |
| $x^\star$    | $e^\star$ | $\xi$ | $x^\star$ | $e^\star$ | $\xi$ | $x^\star$ | $e^\star$ | $\xi$ |
| 1             | 2.68 | 0.07 | 0.32 | 2.97 | 0.04 | 0.03 | 2.77 | 0.13 | 0.23 | 2.96 | 0.05 | 0.04 |
| 2             | 2.99 | 0.02 | 0.01 | 2.95 | 0.03 | 0.05 | 2.97 | 0.04 | 0.03 | 2.32 | 0.01 | 0.68 |
| 3             | 2.96 | 0.01 | 0.04 | 3.01 | 0.07 | 0.01 | 2.99 | 0.11 | 0.01 | 2.88 | 0.01 | 0.12 |
| 4             | 2.93 | 0.29 | 0.07 | 2.85 | 0.01 | 0.15 | 2.99 | 0.12 | 0.01 | 2.46 | 0.01 | 0.54 |
| 5             | 3.03 | 0.03 | 0.03 | 2.98 | 0.20 | 0.02 | 2.90 | 0.05 | 0.11 | 2.84 | 0.02 | 0.16 |
| 6             | 2.99 | 0.05 | 0.01 | 3.07 | 0.01 | 0.07 | 2.97 | 0.07 | 0.03 | 2.89 | 0.08 | 0.11 |
| 7             | 2.97 | 0.03 | 0.03 | 2.95 | 0.20 | 0.05 | 2.80 | 0.11 | 0.20 | 2.83 | 0.03 | 0.17 |
| 8             | 2.94 | 0.01 | 0.06 | 3.11 | 0.05 | 0.11 | 2.97 | 0.02 | 0.03 | 3.07 | 0.04 | 0.07 |
| 9             | 2.98 | 0.02 | 0.02 | 3.23 | 0.10 | 0.23 | 2.94 | 0.07 | 0.06 | 2.88 | 0.11 | 0.12 |
| ...           | ...             | ...         | ...     |
| 96            | 2.96 | 0.02 | 0.04 | 2.98 | 0.20 | 0.02 | 2.92 | 0.09 | 0.08 | 2.30 | 0.05 | 0.70 |
| 97            | 2.94 | 0.07 | 0.06 | 2.97 | 0.01 | 0.03 | 2.93 | 0.08 | 0.07 | 2.72 | 0.05 | 0.28 |
| 98            | 2.92 | 0.24 | 0.08 | 2.88 | 0.03 | 0.12 | 2.89 | 0.14 | 0.11 | 3.10 | 0.20 | 0.10 |
| 99            | 2.99 | 0.01 | 0.01 | 2.94 | 0.09 | 0.06 | 2.99 | 0.04 | 0.01 | 2.73 | 0.08 | 0.27 |
| 100           | 2.97 | 0.03 | 0.03 | 2.96 | 0.20 | 0.04 | 2.91 | 0.09 | 0.09 | 2.89 | 0.12 | 0.11 |

Figure 3. Deviations $\xi$ obtained with procedure A and procedure IF&PA for uniform and normal distributions of interval data.
5. Conclusion

It was proposed an interval data fusion procedure IF&PA on the base of representation of intervals on the real line by weak order relations (or rankings) over set of belonging to these intervals discrete values. Numerical experimental investigations have shown that using as a result of the fusion the best value in a consensus ranking found for rankings collection, corresponding to initial intervals, guarantees improved accuracy and robustness of the interval data fusion procedure outputs.

The proposed procedure starts to work on interval data obtained in ratio scale, transforms the data into multiple ordinal scale entities, determines a best entity with aggregating multidimensional preferences, and the best entity is again a point in ratio scale. We should conclude that the transition from ratio to ordinal scale facilitates revealing essential information in input interval data what, in turn, yields property of resilience (robustness and accuracy) of the procedure IF&PA.

Acknowledgment

This work was supported by the Russian Science Foundation, Project # 14-19-00926. Part of this work was performed while the senior author was at Vignan University, India. Vignan University's, particularly Prof. B. Ramamoorthy's, support is gratefully acknowledged.

References

[1] Parhami B 2007 Distributed Interval Voting with Node Failures of Various Types Proc 12th IEEE Workshop on Dependable Parallel, Distributed and Network-Centric Systems (DPDNS '07 Long Beach, California, USA) pp 1-7
[2] Dubois D and Prade H 2001 Possibility Theory in Information Fusion Data fusion and perception (Eds. G. Della Riccia, H.-J. Lenz, R. Kruse) (Berlin: Springer) 53-76
[3] Hall D L and McMullen S A H 2004 Mathematical Techniques in Multisensor Data Fusion (Norwood: Artech House Publishers)
[4] Halpern J 2003 Reasoning about Uncertainty (Cambridge: MIT Press)
[5] Parhami B 1994 Voting algorithms IEEE Trans. Reliab. 43 617–29
[6] Cox M G 2007 The evaluation of key comparison data: determining the largest consistent subset Metrologia 44 187–200
[7] Mana G, Massa E and Predescu M 2012 Model selection in the average of inconsistent data: an analysis of the measured Planck-constant values Metrologia 49 492–500
[8] Pendrill L R 2014 Using measurement uncertainty in decision-making and conformity assessment Metrologia 51 206–18
[9] Muravyov S V and Khudonogova L I 2015 Multisensor accuracy enhancement on the base of interval voting in form of preference aggregation in WSN for ecological monitoring Proc. 7th Int. Cong. on Ultra Modern Telecommunications and Control Systems (ICUMT 2015, Brno, Czech Republic, 6–8 October 2015) pp 293-297
[10] Berg D E, Norine S, Su F E, Thomas R and Wollan P 2010 Voting in agreeable societies Amer. Math. Monthly 117 27–39
[11] Muravyov S V 2013 Ordinal measurement, preference aggregation and interlaboratory comparisons Measurement 46 2927–35
[12] Muravyov S V and Marinushkina I A 2013 Intransitivity in multiple solutions of Kemeny Ranking Problem J. Phys.: Conf. Ser. 459 012006
[13] Muravyov S V 2014 Dealing with chaotic results of Kemeny ranking determination Measurement 51 328–34.
[14] Muravyov S V and Marinushkina I A 2015 A software tool for simulation of interlaboratory comparison data evaluation method using preference aggregation Proc. 21st IMEKO World Cong. (Prague, Czech Republic, 30 August–4 September 2015)