Chapter 2 : Point Data Fusion

John Klein

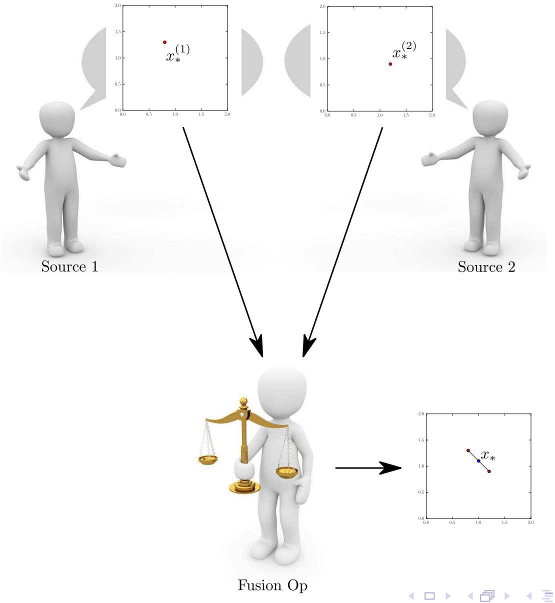
Lille1 Université - CRIStAL UMR CNRS 9189

1. Qualitative data fusion

2. Quantitative data fusion

- In this chapter, we will review techniques for **point data fusion**.
- These techniques are easy to implement and require a low computation load.
- They sometimes lack a clear theoretical justification.
- They are consequently given practical justifications by comparing performances regarding some given (application dependent) criterion.

Point data fusion setting :



Let us give a formal definition for **point data fusion problems** :

Definition

Point data fusion is a subclass of data fusion where advocacies live in the solution space : $\mathbb{X} = \mathcal{X}$.

The solutions presented in this chapter are sorted by **data nature**.

Since $\mathbb{X} = \mathcal{X}$, let us remember that sources can be viewed as estimators of x.

Chapter organization

Point data fusion for qualitative data :

- As opposed to quantitative data, qualitative data live in an abstract space with no underlying algebraic structure or topology, therefore using operators like averaging makes no sense in this context.
- The space $\mathcal{X}$ of possible values for x will be considered countable in this subsection.

Example

Suppose there are $N_s = 3$ computer vision systems scattered in a room. Each camera feeds a trained classifier which must identify the class of a moving object. The set of classes is $\{human; cat; dog\}$. Now if two of them recognize a cat while the last one recognizes a human, then averaging is impossible : $\frac{2\times\{cat\}+\{human\}}{3} = ??$.

- The most intuitive idea is then to resort to voting systems.

« *Democracy is the worst form of government except all those other forms that have been tried.* »

Sir W. Churchill, november 11th 1947, House of commons, UK.

Voting systems

Definition

A **voting system** is a selection method applicable to any data. A voting system is a made of :

- a **ballot**, which specifies the answer expected from a voter. In our notation, the ballot specifies the space $\mathbb{X}$ and the voter is an information source.
- a **tallying algorithm**, which specifies how votes are combined to produce election results. In our context, these algorithms are fusion operators.

Voting systems

- Most of the time, ballots allow for each source to select only one candidate value for x as it was the case in the classifier fusion example.
- Ballot forms :
  - In a plurality ballot, sources can only select one candidate value. We have $\mathbb{X} = \mathcal{X}$.
  - In an approval ballot, sources can select as many candidate values as they want. Observations are thus set-valued and $(\mathbb{X}, \subseteq, \cap, \cup)$ is a distributive lattice. We have $\mathbb{X} = 2^{\mathcal{X}}$.
  - In a cumulative ballot, each source has $N_v \geq 2$ possible choices but can assign several votes to the same candidate. Observation are thus score vectors. We have $\mathbb{X} = \mathbb{N}^{|\mathcal{X}|}$.
  - In a ranked ballot, each source ranks candidate values from the most likely to the least likely. Observations are ordered lists which can be exhaustive or not. These lists can be turned into vectors if one decides to assign a weight to each rank. We have $\mathbb{X} = \{1, .., |\mathcal{X}|\}^{|\mathcal{X}|}$.

Voting systems

- Ballot forms :
  - In a rated ballot, each source grades candidate values within a given range. Observations are thus score vectors. A grade can be negative, if the range allows it. A negative grade is interpreted as how much a source supports that a given value is not possible for x. Rated ballots are perhaps the most general form of ballots. We have $\mathbb{X} = \mathbb{R}^{|\mathcal{X}|}$.
- The tallying algorithm to apply depends on the ballot form.
- This section presents voting operators compatible with a point data fusion problem, *i.e.* votes with plurality ballots ($\mathbb{X} = \mathcal{X}$).
- Others ballots will be used in the forthcoming chapters.

Point data fusion for qualitative data : Let's build a histogram (reminder)

Definition

Let $\left\{x_*^{(i)}\right\}_{i=1}^{N_s}$ denote a dataset such that $\forall i,\ x_*^{(i)} \in \mathcal{X}$. The **histogram** h of the dataset $\left\{x_*^{(i)}\right\}_{i=1}^{N_s}$ is the a mapping such that :

$$\begin{aligned} h : \mathcal{X} &\longrightarrow \mathbb{N}, \\ x &\longrightarrow \sum_{i=1}^{N_s} \delta\left(x - x_*^{(i)}\right), \end{aligned} \tag{1}$$

with δ the Dirac function. In other words, $h(x)$ is the number of occurences of the value x in the dataset.

Point data fusion for qualitative data : majority voting

Definition

Let h denote the histogram computed from advocacies which are points. h_i denotes the occurences in advocacies of the value whose index is i.

The **majority operator** $\hat{\hat{x}}_{maj}$ is defined as follows :

$$\hat{\hat{x}}_{maj} : \left(x_*^{(1)}, .., x_*^{(N_s)}\right) \to \arg\max_{i \in |\mathcal{X}|} \{h_i\}, \tag{2}$$

with $|\mathcal{X}|$ the cardinality of $\mathcal{X}$.

The index i does not depend on any underlying order on $\mathcal{X}$.

Point data fusion for qualitative data : super-majority voting

Definition

(same notations)

The **supermajority operator** $\hat{\hat{x}}_{smj}$ at threshold $th \in [0;1]$ is defined as follows :

$$\hat{\hat{x}}_{smj} : \left(x_*^{(1)}, .., x_*^{(N_s)} \right) \rightarrow \begin{cases} \arg\max\limits_{i \in |\mathcal{X}|} \{h_i\} & \text{if } \max\limits_{i \in |\mathcal{X}|} \left\{ \frac{h_i}{N_s} \right\} > th \\ \emptyset \text{ otherwise} & \end{cases} . \tag{3}$$

When $th = 0.5$, this operator is often called the **absolute majority operator** $\hat{\hat{x}}_{amj}$.

The index i does not depend on any underlying order on $\mathcal{X}$.

Point data fusion for qualitative data : super-majority voting

- Supermajority operators are of course a generalization of the majority operator.
- When the target threshold *th* is not met, then the fusion system delivers no output, which is a problem. This can be seen as a consequence of conflict between sources.
- Several strategies can be thought of under such circumstances :
  - wait for more reliable data and re-try the fusion later,
  - come up with a multiple round vote. The results of the 1^{st} vote may not be sufficient to find a winner but to eliminate values with too weak occurences.

  These latter approaches are examples of iterated or hierarchical fusion systems.

Point data fusion for qualitative data : absolute majority voting

- The absolute majority operator has an interesting famous property in the binary case $\mathcal{X} = \{0;1\}$:

Theorem

Condorcet jury's theorem. *Suppose the number of sources is odd :* $\exists m \in \mathbb{N} \mid N_s = 2m+1$ *and* $m > 1$. *Suppose all sources have the same (independent) probability* θ *to choose* x *among all other possible values in* $\mathcal{X}$. *Let* p_m *denote the probability of the event* $\left\{x_*^{(amj)} = x\right\}$. *The following properties holds :*

(i) *If* $\theta > \frac{1}{2}$ *then* $p_m \xrightarrow[m \to +\infty]{} 1$.

(ii) *If* $\theta < \frac{1}{2}$ *then* $p_m \xrightarrow[m \to +\infty]{} 0$.

(iii) *If* $\theta = \frac{1}{2}$ *then* $p_m = \frac{1}{2}$.

Point data fusion for qualitative data : absolute majority voting

Sketch of proof :

- Suppose $\left\{\delta_x\left(x_*^{(i)}\right)\right\}_{i=1}^{N_s}$ is an i.i.d. sample drawn from a Bernouilli random variable Z on the probability space made of two events : $\left\{x=x_*^{(.)}\right\}$ and $\left\{x\neq x_*^{(.)}\right\}$.
- Z has a distribution which is entirely determined by θ.
- The law of large numbers implies that the proportion of sources choosing x tends to θ as m increases.
- Now if $\theta > \frac{1}{2}$, then x has a majority !

Point data fusion for qualitative data : absolute majority voting

Exercice

Suppose the hypotheses of the condorcet jury theorem hold. The same notations as in this theorem are used in this exercice.

1. Find an expression of p_m that is only depending on θ and m.
2. Find a recursive equation relating p_{m+1} and p_m. *hint* : think in terms of conditional probabilities. When do the two new source votes change the majority or not ?

Chapter organization

Point data fusion for Quantitative data :

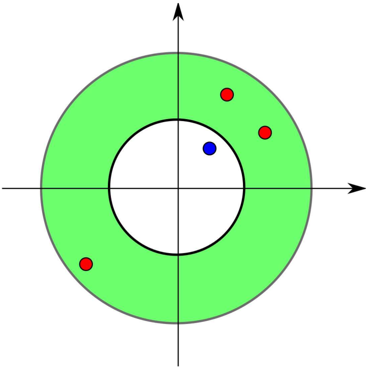
- Such data are actually numbers or vectors whose entries are numbers.
- As part of a point data fusion problem, this implies that there exists a positive interger $d \in \mathbb{N}^*$ such that $\mathbb{X} \subseteq \mathbb{R}^d$ and each advocacy $x_*^{(i)}$ is a point in $\mathbb{R}^d$.
- The main advantage of quantitative data is that we can do some calculus with them. More formally, $\mathbb{R}^d$ is an algebra and its internal or external binary operations can be used to design fusion operators.
- By definition a fusion operator is a N_s-ary internal operation : $\hat{\hat{x}} : \mathbb{X}^{N_s} \rightarrow \mathbb{X}$.

Point data fusion for Quantitative data :

- Binary operations on $\mathbb{R}^d$ are not binary operations on $\mathbb{X}$, unless $\mathbb{X}$ is a sub-algebra of $\mathbb{R}^d$. When designing a fusion operator using binary operations, there is consequently a risk that $x_* \notin \mathbb{X}$.

Example

$$\mathcal{X} = \left\{ x \in \mathbb{R}^2 \mid a < \|x\|_2 < b \right\} \subset \mathbb{R}^2$$



Point data fusion for Quantitative data :

- A similar remark concerning limit processes could be made. If the a fusion resorts to such a process, one needs $\mathbb{X}$ to be complete[1]. The only parameter whose limit is evoked in this document is N_s.
- In the sequel of this section, we will first investigate a case where there is no risk to yield an invalid fusion operator, *i.e.* when $\mathbb{X} = \mathcal{X} = \mathbb{R}^d$. Afterward, two particular cases of quantitative data will be discussed : quantized data and binary data.

1. A complete space is such that each Cauchy sequence converges to an element of the space.

Point data fusion for Quantitative data in $\mathbb{R}^d$:

- One of the most natural way to perform a data fusion in $\mathbb{R}^d$ is to compute the average of advocacies provided that the data fusion problem is simple enough to allow such an operation.
- This is tantamount to consider that x is a random variable and that each datum $x_*^{(i)}$ is a realization of that random variable.
- More generally, first order descriptive statistics are relevant tools for data fusion. We will review a few of them in this section and discuss their limitations.

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

Definition

Suppose each $x_*^{(i)}$ is a point in $\mathcal{X}$. The **arithmetic mean** or **average** $\hat{\hat{x}}_{ave}$ is defined as :

$$\hat{\hat{x}}_{ave} : \left(x_*^{(1)}, .., x_*^{(N_s)}\right) \rightarrow \frac{1}{N_s} \sum_{i=1}^{N_s} x_*^{(i)}. \tag{4}$$

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- If the $x_*^{(i)}$ are iid then $\hat{\hat{x}}$ is an unbiased estimator :

$$\hat{\hat{x}}_{ave} \underset{N_s \to +\infty}{\longrightarrow} x.$$

- F-means are a possible generalization of averaging obtained thanks to a bijective mapping $f : \mathcal{X} \longrightarrow \mathbb{R}^d$.

$$\hat{\hat{x}}_{f-mean} : \left(x_*^{(1)}, .., x_*^{(N_s)} \right) \to f^{-1} \left(\frac{1}{N_s} \sum_{i=1}^{N_s} f\left(x_*^{(i)} \right) \right). \tag{5}$$

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- F-means are potentially biased therefore they are only used in very specific situations.

- Popular f-means are :
  - the geometrical mean that is obtained when $f(x) = \log(x)$, which gives $\hat{\hat{x}}_{geo}\left(x_*^{(1)}, .., x_*^{(N_s)}\right) = \left(\prod_{i=1}^{N_s} x_*^{(i)}\right)^{\frac{1}{N_s}}$.
    The geometrical mean is useful when the processed data are normalized with respect to some reference value. This is notably the case for normalized ratios and percentages.

Example

Suppose an orange tree produced in the past years 100, 180, 210 and 300 oranges. The corresponding growth production rates are thus 1.8, 1.17 and 1.44. The arithmetic mean of these rates is 1.47 whereas the geometric mean is 1.44. Starting with 100 oranges and applying a growth rate of 1.47 gives 314.47 oranges while using 1.44 as growth rate gives exactly 300 oranges.

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- Popular f-means are :
  - the harmonic mean that is obtained when $f(x) = \frac{1}{x}$, which gives $\hat{\hat{x}}_{har}\left(x_*^{(1)}, .., x_*^{(N_s)}\right) = \frac{N_s}{\sum_{i=1}^{N_s} \frac{1}{x_*^{(i)}}}$.
  - This mean can only be used if each $x_*^{(i)} > 0$.
  - The harmonic mean processes unormalized growth rates.
  - It is also noteworthy that one always have $\hat{\hat{x}}_{har} \leq \hat{\hat{x}}_{geo} \leq \hat{\hat{x}}_{ave}$.

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- Popular f-means are : harmonic mean

Example

Suppose a car is moving by a 100m at 50km/h and by another 100m at 70km/h. The car displacement duration is 7.2 for the first move followed by 5.14s for the second move which makes 12.34s in total. We have :

- $\hat{\hat{x}}_{ave}(50; 70) = 60$. The displacement duration obtained by moving at 60km/h on 200m is 12s.
- $\hat{\hat{x}}_{geo}(50; 70) = 59.16$. The displacement duration obtained by moving at 60km/h on 200m is 12.17s.
- $\hat{\hat{x}}_{har}(50; 70) = 58.33$. The displacement duration obtained by moving at 60km/h on 200m is exactly 12.34s.

Note that if we had analyzed the same problem with constant duration and varying distance, then $\hat{\hat{x}}_{ave}$ would have been the right mean to choose. Besides, in example 3, the harmonic mean would have given 286.32 oranges so it is not appropriate in that context.

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- Another way to generalize the average is to use weights. Averaging becomes then a convex combination, or a barycenter :

$$\hat{\hat{x}}_{w-ave} : \left(x_*^{(1)}, .., x_*^{(N_s)} \right) \rightarrow \frac{\sum_{i=1}^{N_s} w_i \; x_*^{(i)}}{\sum_{i=1}^{N_s} w_i}. \tag{6}$$

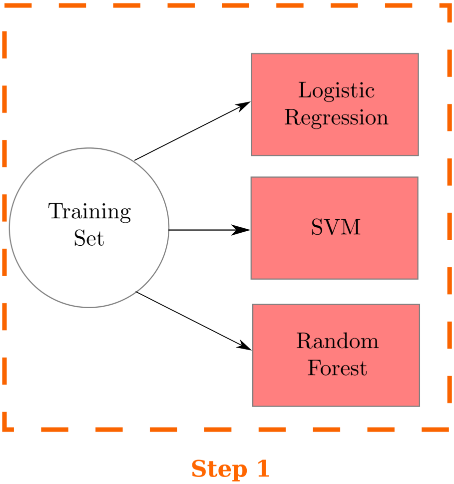
- Weights should be obtained from meta-data, which is a bit beyond the scope of this chapter.

Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- In machine learning, useful meta-data are error rates.
- Here is an example for binary classification :

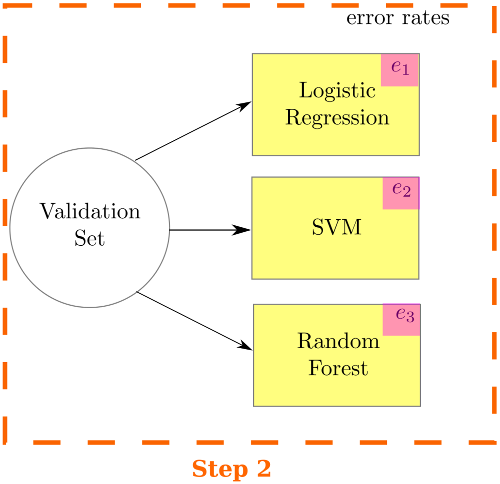
Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- In machine learning, useful meta-data are error rates.
- Here is an example for binary classification :



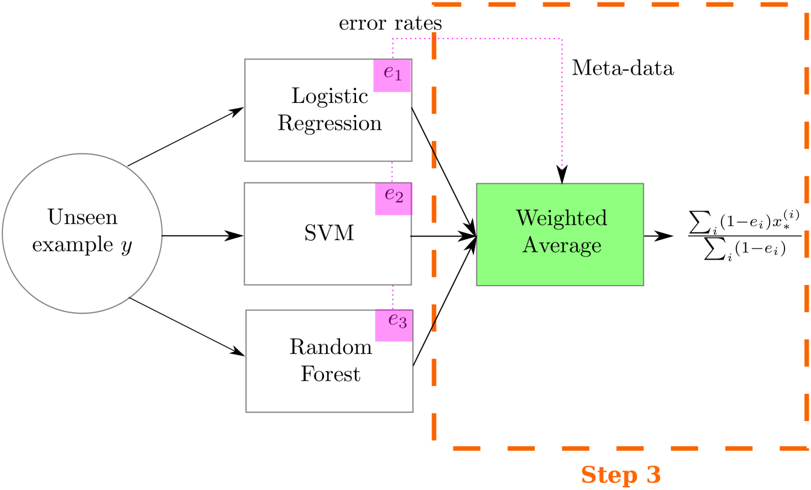
Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- In machine learning, useful meta-data are error rates.
- Here is an example for binary classification :



Point data fusion for Quantitative data in $\mathbb{R}^d$: averaging

- In machine learning, useful meta-data are error rates.
- Here is an example for binary classification :



Point data fusion for Quantitative data in $\mathbb{R}^d$: median

Definition

Let σ denote a permutation on integers from 1 to N_s such that $\forall i < j$, $x_*^{(\sigma(i))} \leq x_*^{(\sigma(j))}$. This means that the dataset $\left\{x_*^{(\sigma(i))}\right\}_{i=1}^{N_s}$ is sorted in ascending order. The **median** is defined as :

$$\hat{\hat{x}}_{med} : \left(x_*^{(1)}, .., x_*^{(N_s)}\right) \rightarrow \begin{cases} \frac{x_*^{\left(\sigma\left(\frac{N_s}{2}\right)\right)} + x_*^{\left(\sigma\left(\frac{N_s}{2}+1\right)\right)}}{2} \text{ if } N \text{ is even} \\ x_*^{\left(\sigma\left(\lceil \frac{N_s}{2} \rceil\right)\right)} \text{ if } N \text{ is odd} \end{cases}, \quad (7)$$

with $\lceil . \rceil$ the ceiling function.

Point data fusion for Quantitative data in $\mathbb{R}^d$: median

Example

Suppose your marks this semester are $\{0; 12; 12\}$, then your average mark is 8 while your median mark is 12.

- The median is known for being less sensitive to extreme values as it can be seen from the preceding example.

Point data fusion for Quantitative data in $\mathbb{R}^d$: mode

Definition

Let h denote the histogram computed from the advocacies. The **mode** is defined as :

$$\hat{\hat{x}}_{mod} = \arg\max_{x \in \mathcal{X}} \{h(x)\}. \tag{8}$$

Point data fusion for Quantitative data in $\mathbb{R}^d$: mode

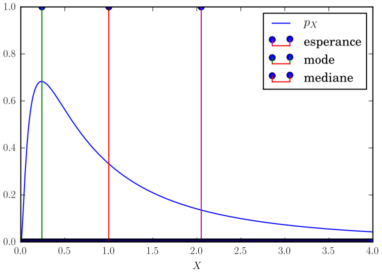
- The empirical mode can be understood as the value with biggest occurence counts and to some extent the most probable value.
- This last comment holds if the dataset $\left\{x_*^{(i)}\right\}_{i=1}^{N_s}$ is a representative sample of the underlying probability distribution. (strong similarity with the majority operator $\hat{\hat{x}}_{maj}$).

Example

Suppose your marks this semester are $\{12; 12; 14; 17; 20\}$, then your average mark is 15 while your mark mode is 12 and your median is 14.

- In practice, the mode operator is poorly efficient with non-quantized data because usually $x_*^{(i)} \neq x_*^{(j)}$ whenever $i \neq j$ which leads to an undetermination in equation (8).

Point data fusion for Quantitative data in $\mathbb{R}^d$: mode



Point data fusion for Quantized Quantitative data :

- Such data are sometimes also called discrete data.
- Quantized data means that $\mathcal{X}$ is a countable subset of $\mathbb{R}^d$.
- 2 ways to retrieve a fusion operator for quantized data from previously introduced ones :
  1. Fusion ops designed for qualitative data can be applied straightforwardly to quantized data. Indeed, the discrete possible values of x can be seen as just candidate names for which a source can vote.
  2. A second possibility is to use a fusion operator designed for data in $\mathbb{R}^d$. As previously mentionned, the only problem is that the estimated value may not belong to $\mathcal{X}$.
     Suppose had $\mathcal{X} = \mathbb{N}$ and $\left\{x_*^{(i)}\right\}_{i=1}^{N_s} = \{0; 2; 1; 0; 0\}$. The average operator gives $x_* = 0.6 \notin \mathcal{X}$.
     In this case, one usually resorts to a rounding function or a projection mapping to the desired set.

Point data fusion for Quantized Quantitative data : Rounding functions

- The most commonly used rounding function is the round to nearest function rnd :

$$\begin{aligned} \mathrm{rnd} \quad &: \quad \mathbb{R}^d \longrightarrow 2^{\mathcal{X}}, \\ &: \quad s \longrightarrow \arg\min_{x_0 \in \mathcal{X}} \left\{ d\left(s, x_0\right) \right\}. \end{aligned} \tag{9}$$

- Note that rnd is a set-valued function because ties may occur.

- For instance, suppose $\mathcal{X} = \{0; 1; 2\}$ and $\left\{ x_*^{(i)} \right\}_{i=1}^{N_s} = \{0; 2; 1; 0; 0; 0\}$, which gives $x_* = 0.5$ using $\hat{\hat{x}}_{ave}$. We have $\mathrm{rnd}\,(0.5) = \{0; 1\}$. The fusion operator is thus undecided between 0 and 1.

Point data fusion for Binary Quantitative data :

- Binary datasets are special cases of finite quantized datasets (therefore same solutions are applicable).
- In this context, the set of possible values for x is made of two values : $\mathcal{X} = \{0; 1\}$, which are sometimes referred to as $\{FALSE; TRUE\}$ instead of $\{0; 1\}$.
- Some specific fusion ops are given by logic operations such as logical AND, logical OR and so on.

Point data fusion : concluding remarks

- In the next chapters, some advanced frameworks are presented.
- These frameworks allow to formalize precisely some data fusion problems and give nice theoretical justifications.
- Each chapter is dedicated to a given nature of data to fuse :
    - partially ordered data,
    - ranked data,
    - uncertain data,
    - functional data.
- In spite of this segmentation, datasets can be at the same time ranked, uncertain and partially ordered.
- Likewise, a data fusion system can jointly rely on many of the frameworks that will be presented.

Point data fusion : concluding remarks

- The next chapters presents advanced fusion techniques for improved fusion performances.
- This improvement is naturally subject to adequate data modeling and implies an increased computation load.
- Any meticulous data fusion engineer or researcher should always ask his or herself :

  Does my advanced data fusion beat (significanty) averaging or voting ?
- Selecting a data fusion technique is a tradeoff between performance improvement and computation time.