2006 IEEE International Conference on
Multisensor Fusion and Integration for Intelligent Systems
September 3-6, 2006, Heidelberg, Germany

Fusion agents - realizing Bayesian fusion via a local approach

Jennifer Sander, Jürgen Beyerer

Abstract— **Bayesian theory delivers a powerful theoretical platform for the mathematical description and execution of fusion tasks, especially if the information delivering sources are of heterogeneous nature. However, the complexity of Bayesian fusion tasks increases exponentially with the number of sources. A new agent based architecture that is modelled on a successful operating process of the real world, namely criminalistic investigation, circumvents the high computational costs by realizing a local fusion approach. In analogy to criminalists, software agents shall be appointed to perform heterogeneous fusion tasks. In this paper, we give an overview over this concept and the potentials that are provided by it. Then, we focus on translating the proposed ideas into a formal mathematical notation.**

I. INTRODUCTION

In many real world applications, information from heterogeneous sources has to be fused for obtaining optimal results and decisions respectively. In this context, the term information refers to everything that has a potential to inform. In [7], the authors depicted that the Bayesian approach provides a very powerful fusion method that distinguishes itself by its slim calculus and its clearness. The authors also showed that the Bayesian fusion method meets the essential requirements that a feasible fusion mechanism must satisfy. By the transformation of the available knowledge, the corresponding uncertainties, and the given dependencies into Degree of Belief (DoB) distributions in the Bayesian sense, a consistent mathematical description for non-compatible kinds of information is possible. From a Bayesian point of view, a lossless information fusion is then feasible. DoBs embody the numerical degree of certainty taking account of the given knowledge. By defining DoB distributions, subjectivity may be incorporated and two subjects equipped with the same amount of information need not conclude the same DoB distribution. An established transformation method for obtaining objective DoB distributions is the maximum entropy method, see [12].

Bayesian inference offers the possibility to probabilistically draw conclusions from the given information to an underlying state of the world that is not directly observable. First of all, a probability model has to be specified that describes the properties of interests (PoI), this means the elements of the state of the real world that should be exploited, and how the given information is generated from the PoI. A necessary precondition for the fusion is that all given information refers to the same underlying "true" state of the world. In the Bayesian sense, this means that the PoI and the involved information must have a joint distribution given the selected model.

Jennifer Sander is with the Lehrstuhl für Interaktive Echtzeitsysteme, Faculty of Computer Sciences, Universität Karlsruhe (TH), Adenauerring 4, D-76131 Karlsruhe, Germany jsander@ies.uni-karlsruhe.de

Jürgen Beyerer is with the Lehrstuhl für Interaktive Echtzeitsysteme, Faculty of Computer Sciences, Universität Karlsruhe (TH), Adenauerring 4, D-76131 Karlsruhe, Germany and with the Fraunhofer IITB Institut Informations- und Datenverarbeitung, Fraunhoferstraße 1, D-76131 Karlsruhe, Germany beyerer@ittb.fraunhofer.de

II. BAYESIAN FUSION

A. Formulation of the fusion task

We assume that $S \in \mathbb{N}$ (numbered) information delivering sources are involved, and we denote by d_s the information from source number s. Then, the data vector

$$d := (d_1, ..., d_S) \in D := D_1 \times ... \times D_S$$

incorporates all available information. We assume also that we want to identify $N \in \mathbb{N}$ PoIs. These are denoted by

$$z := (z_1, ..., z_N) \in Z := Z_1 \times ... \times Z_N.$$

z is a random quantity in the sense of classical statistics in contrast to the components of d that adopt deterministic values in a special fusion task. Note that – as usual in Bayesian statistics – no difference is made between random quantities and their realizations, and consequently, the same notation for both kinds of quantities is used.

Frequently, the components of z are of different kind of nature, e.g. if the task was to detect and classify vehicles in a given aerial image, the set Z constitutes a bounded area of location and distinct vehicle types. A great benefit of the Bayesian framework is its ability to handle differently scaled variables coherently.

The aim is then to infer the "true value" of z from d.

B. Information processing via Bayesian inference

Bayesian inference is realized by the principle of inverse probability provided by the Bayesian theorem:

$$p(z|d) \propto p(d|z)p(z).$$

Up to a normalizing constant, the posterior distribution $p(z|d)$, that represents the knowledge about z after incorporating the knowledge provided by d, is equal to the product of the – not necessary normalized – likelihood function $p(d|z)$ and the prior distribution $p(z)$. The normalizing constant is the marginal density of the data vector, that is

$$p(d) = \int_{Z_1} \cdots \int_{Z_N} p(d|z_1, ..., z_N)p(z_1, ..., z_N)dz_1...dz_N.$$

Here we use that, from a measure theoretic point of view, the summation of discrete quantities is just a special case of integration and consequently, a unified integral notation can be used even if some of the PoIs have a discrete domain. dz_n represents integration with respect to the underlying dominating measure.

For a given fusion task with a fixed probability model, the likelihood principle is an immediate consequence in Bayesian statistics. It states that the information provided by the sources influences the posterior DoB distribution only via the likelihood function, and that two data vectors with $p(d^{(1)}|z) \propto p(d^{(2)}|z)$ for all $z \in Z$ cause identical inferences because they contain the same information about z.

If the nature of the information delivering sources is very different, we can assume them to be independent given the underlying truth. In this case, Bayesian fusion delivers identical results no matter if we incorporate all available information via one single inference step as described above or if the information is processed sequentially via subsequent Bayesian inference steps. The posterior at a certain stage of this sequential processing is then the prior of the next:

$$p(z|d_1,\ldots,d_{s+1}) \propto p(d_{s+1}|z)p(z|d_1,\ldots,d_s) \quad \forall s \in \{1,\ldots,S-1\}.$$

Using that $p(d_{s+1}|z) \propto p(z|d_{s+1})$ with a completely uninformative prior, we get for all $s \in \{1,\ldots,S-1\}$

$$p(z|d_1,\ldots,d_{s+1}) \propto p(z|d_{s+1})p(z|d_1,\ldots,d_s). \tag{1}$$

Additionally, the ordering in which the sources are numbered does obviously not affect the posterior DoB distribution $p(z|d_1,\ldots,d_S)$.

C. Obtaining the final fusion result

Bayesian inference is a lossless process, because its result consists of the complete posterior DoB distribution of z given the processed information. According to $p(z|d)$, the probability that a certain z originates the incidence of consequences – which became manifested in the components of d – could be "high enough" for several different values of Z.

Performing a single estimate for the "true value" of the PoI is the simplest use of the posterior DoB distribution. To this end, many techniques from decision theory that deliver various estimators can be used, e.g. the posterior expectation

$$\widehat{z}_{\mathrm{PE}} = \int_Z z p(z|d) dz$$

which minimizes the mean squared error for continuous quantities or the maximum a posteriori estimate

$$\widehat{z}_{\mathrm{MAP}} = \arg\max_{z \in Z}\{p(z|d)\}$$

that minimizes the classification error probability in the discrete case. Making these estimates is of course no longer lossless.

One sees that even for the calculation of an estimate of the "true value", a cost-intensive calculation over whole Z is necessary. The corresponding computational complexity is

$$O(|Z|) = O(\prod_{i=1}^{N}|Z_i|) \tag{2}$$

which is not feasible even for moderate dimensions.

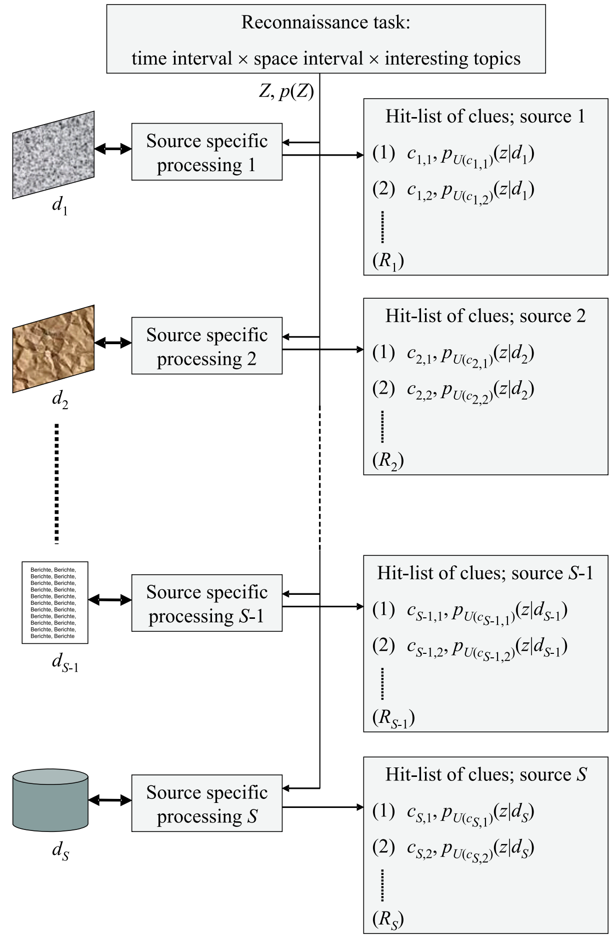

Fig. 1. The initialization process: A reconnaissance task is formalized in temporal, spatial and thematic sections that constitute the set Z. The hit-lists are ordered with falling significance of the clues which is measured by the local DoB distributions.

III. Realizing a local approach

A. The basic concept

Fusion agents, a special kind of software agents, shall perform heterogeneous fusion tasks in an analogous way as criminalists solve crimes. According to [18], software agents are regarded as entities capable of pursuing certain tasks autonomously. In an initialization step, clues – corresponding to hypotheses that should be proven – are extracted from each of the available information sources. They are given to fusion agents that try to confirm these initial suspicions by sequentially visiting the other sources. This model corresponds to a local fusion approach: every agent traces his clue c locally in a small environment $U(c) \subseteq Z$. Performing Bayesian fusion only in these environments lowers the computational complexity on $\sum_{c:\mathrm{clue}} O(|U(c)|) << O(|Z|)$.

B. Excursus: Criminalistic work

If a crime has happened, the first step done by criminalists is collecting clues and evidences. These are subsequently assigned to detectives that generate hypotheses and appendant DoBs out of it. A detective tries to strengthen his DoB in a

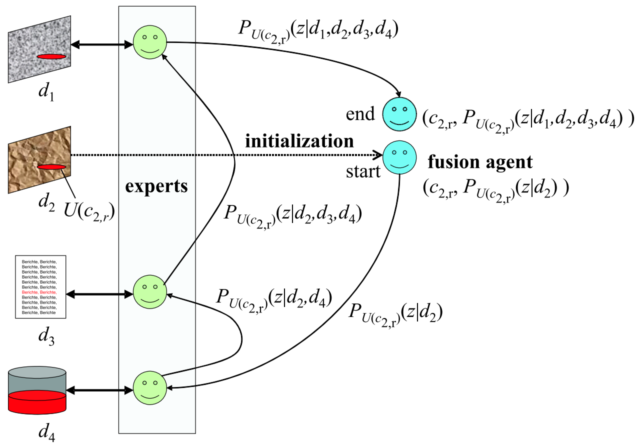

Fig. 2. A fusion agent (fuselet) is instantiated for each clue of the hit-lists generated in the initialization step. The sample depicted here shows the "work" of the fusion agent responsible for the clue $c_{2,r}$.

certain hypothesis by performing single investigation steps. Doing that, he gets confronted with heterogeneous information, e.g. testimonies, DNA traces, bullet casings, that he has to pull together. If an investigator is not capable of analyzing a certain kind of information by himself, he can assign this task to a specially trained expert. Detectives discuss together about further investigation steps and compare their results to build up a global view of the scene. Sometimes, theories have to be abandoned. A hypothesis can be considered to be true if the corresponding DoB is strong enough. In this case, the investigation can be finished. The final decision about a suspect is then not made by the investigators themselves but by a more competent jury that has to be comprehensively informed.

The most important fact concerning this process is that an investigator does not have a god's view on the whole space of interest. Instead of this, he takes up a local perspective and sees only the clues and the appendant local context.

C. An agent based architecture for the fusion of heterogeneous information sources

The basic essentials of an agent based architecture for information fusion that was first presented in [6] will be described. It is modelled on the work of detectives, that was outlined in the previous section.

1) Discovering clues: We use the maximum entropy method to determine an objective prior DoB distribution $p(z)$ on Z. This proceeding is consistent with the perception of the world which open minded criminalists should have. In the case of complete uncertainty, every element of Z should be equally conspicuous to be the true one. If prior knowledge about the "true value" of the PoI is available, this additional information is taken as constraint for the optimization problem of finding the DoB distribution with maximum uncertainty. E.g. if vehicles in a given area shall be detected, knowledge from a road map can be used to determine a non-uniformly distributed prior.

Bearing Z, $p(z)$ and the conceptual formulation in mind, source specific operators are applied to each of the information sources. The operators shall indicate suspicious points in conveniently chosen spaces $Z^s \subseteq Z$ for $s \in \{1,\dots,S\}$. For this purpose, they have to be chosen as the best possible ones: all available expertise as how to evaluate special kinds of information has to be used. E.g. if the task is to detect and classify vehicles and d_s is an aerial image, the corresponding operator shall – e.g. via matched filtering – deliver a signal $i(z|d_s)$ over an adequately chosen hybrid domain

$$Z^s = [x_1, x_{\max}] \times [y_1, y_{\max}] \times \{\text{Types of relevant vehicles}\}.$$

Theoretically, a sequential Bayesian fusion over Z would be possible by using these immediate results. However, this methodology would cause high computational costs, see (2). The essential idea is to perform Bayesian inference only in small regions of Z where a conspicuous peak, i.e. a distinct local maximum, is indicated by one of the operators. For this purpose, the peaks are detected and their arguments are stored in hit-lists of clues, see fig. 1. The lists get ordered by falling significance which is measured by local DoB distributions $p_{U(c_{s,r})}$. We can obtain the local DoB distributions from the preceding source specific processing.

2) Investigation performed by the fusion agents: For each pair $(c_{s,r}, p_{U(c_{s,r})})$, a fusion agent is instantiated. The local DoB distribution $p_{U(c_{s,r})}$ expresses the degree of confidence of the fusion agent to investigate in the right direction that is indicated by the respective clue $c_{s,r}$. The fusion agent then visits sequentially the other sources whose information is not yet incorporated in his actual local DoB. If he is not capable of accessing a certain source or of analyzing the information d_s that is delivered by it, he gets help from a specially designed expert agent that is able to perform this task. The expert agent strengthens or degrades the significance of the corresponding clue via performing a local Bayesian fusion: he interprets the actual local DoB distribution of the fusion agent as prior distribution and performs an inference step to incorporate d_s. If the nature of the sources is very different, the expert has – as indicated by (1) – just to perform a simple multiplication. The fusion agent then takes the resulting posterior that incorporates the additional information d_s as his new, i.e. updated, local DoB distribution and goes on with visiting the remaining sources. Of course, he has to keep an account of the sources that he has already visited to avoid an overweighing of certain information.

We can assume that true events will leave significant leads in more than one of the sources. Consequently, fusion agents should communicate from time to time, to find out if several of them investigate in the same direction and can be dropped. Also for realizing such a dropping, a bookkeeping of the fusion agents is highly important.

3) The final decision: If a fusion agent has finally updated his local DoB distribution by visiting all the sources, he delivers his clue together with the updated local DoB distribution to a jury. By examining all the clues that the fusion agents have investigated and the corresponding DoB distributions, the jury is then capable of deciding on the global significance of each of them.

D. The high potential of the proposed concept

As drafted in section III-A, via the proposed architecture, we can circumvent the high computational costs of Bayesian fusion methods. Additionally, the architecture is scalable according to the given computational resources: we can decide on the number of instantiated fusion agents so that, if necessary, only hot traces (which are on the first positions of the hit-lists) are investigated.
There is the possibility to integrate humans as experts agents, e.g. if for a certain source no automatic analyzing mechanism exists. It is also obvious that the described architecture could be an suitable concept if we are dealing with huge networks.

IV. The underlying local approach

A. The basic concept

The essential idea concerning the proposed architecture is that every lead should be considered only in a small environment $U(c)$ around the corresponding clue c: the investigating fusion agent possesses a local DoB that refers precisely to the lead and the corresponding local context.
Concerning quantitative properties, we define a distance measure on Z via an adequately chosen metric m on Z. E.g. if we deal with spatial properties we can chose m to be the euclidian metric. With respect to m and an $\varepsilon(c) > 0$ that may depend on the significance of c, we simply define

$$U(c) := \{z \mid m(z,c) < \varepsilon(c)\}.$$

For qualitative properties, this concept is directly applicable for ordinally scaled variables at the most. Therefore, we make another approach: if the clue c is contained in the hit-list pertaining to source number s, we define

$$U(c) := \{z \mid p(z|d_s) > T(c)\} \supset \{c\},$$

this means $U(c)$ contains exactly the elements of Z whose conditional probability exceeds a certain threshold $T(c)$. The threshold may also depend on c. In this case, we define a local DoB distribution in the following way: we specify an explicit DoB only for the elements of $U(c)$. E.g. classifying vehicles, for a good probability model, it is – without additional information – necessary to specify a conditional DoB distribution for a huge set Z of different types. However, if we know which types are highly probable and which are not, a simpler – but yet adequate – local DoB distribution can be formulated in the following way:

$$\begin{aligned} p(z = \text{type}_1|d_s) &= 0.8, \\ p(z = \text{type}_2|d_s) &= 0.15, \\ p(z \in Z \setminus U(c)|d_s) &= 0.05. \end{aligned} \tag{3}$$

with $U(c) = \{\text{type}_1, \text{type}_2\}$, e.g. type_1 = VW_Golf and type_2 = Ford_Focus. In the next sections, we will describe a formal concept for building up local DoBs for both quantitative and qualitative properties that generalizes this idea: the local point of view, that is adopted by the fusion agents, should be represented by a DoB with respect to a certain partition of the set Z. For this purpose, a closer look on the underlying probability model is necessary.

B. Describing knowledge-states via sigma-fields

Recall that a probability space is defined by a tuple $(\Omega, \mathscr{A}, P)$. $\Omega \neq \emptyset$ specifies the possible states of nature. Before defining the probability measure P, it is necessary to determine a suitable domain for it. Usually, it is not practicable to assign a probability value to all subsets of Ω. Instead, one constitutes a smaller subset of the power set $\mathscr{P}(\Omega)$ of Ω, where the probability measure is well-defined: that is the σ-field $\mathscr{A}$. A set $A \subseteq \mathscr{A}$ is called measurable because the probability measure P assigns a probability to it. Now, we review the definition of σ-fields because the concept that underlies σ-fields is essential for the understanding of the following description.
$\mathscr{A} \subseteq \mathscr{P}(\Omega)$ is a σ-field if the following holds:
(1) The empty set belongs to $\mathscr{A}$.
(2) If A belongs to $\mathscr{A}$ then so does its complement $\Omega \setminus A$.
(3) If $A_1, A_2, \ldots$ is a countable collection of sets that belong to $\mathscr{A}$ then so does the union $\bigcup_{i=1}^{\infty} A_i$.
One way to generate an adequate σ-field is to fix a collection $\mathscr{S}$ of subsets of Ω and to determine the smallest σ-field that contains all elements of $\mathscr{S}$. This σ-field is denoted by $\sigma(\mathscr{S})$. σ-fields are appropriate mathematical structures to model knowledge-states: if we have fixed a probability model that corresponds to a certain measure space $(\Omega, \mathscr{A})$, the universe of the knowledge-states on which we can build up any kind of DoB is given by the elements of $\mathscr{A}$. Let $(Z, \mathscr{Z})$ be the measure space that specifies the set of the PoIs and constitutes which knowledge-states about them can be reached. Having the definitions of $U(c)$ from the last section in mind, we will now discuss how we model the knowledge-state of a fusion agent, that mirrors his local point of view: we will assign him a private, i.e. local, measure space on which the corresponding local DoB will be defined. This concept will work for both quantitative and qualitative elements. An important aim should be holding the σ-fields in the private measure spaces as coarse as possible. To describe this we use the following definition from measure theory:
An element $B \in \mathscr{A}$ of a σ-field $\mathscr{A}$ is called an atom if no real subset $\emptyset \neq C \subset B$ is contained in $\mathscr{A}$.
If an atom B has cardinality larger than one, the probability measure does not assign probability values to real subsets of it, e.g. to single elements of Ω. For the following, we need the $\mathscr{Z}$-measurability of $U(c)$. We will assume this requirement guaranteed. Of course, the easiest possibility to define a private measure space is to use $(Z, \mathscr{Z}_1)$ with the very coarse local σ-field

$$\mathscr{Z}_1 = \{U(c), Z \setminus U(c), Z, \emptyset\}.$$

In this case, the corresponding fusion agent would just be capable of getting an idea if the "true value" of z lies in $U(c)$ or not. We would consequently lose all knowledge about the situation inside $U(c)$. To avoid this, we define a finer local σ-field in which $U(c)$ is split: we take the trace-σ-field $\mathscr{Z}_2$ of $\mathscr{Z}$ on $U(c)$, that is

$$\mathscr{Z}_2 := \{A \bigcap U(c) \mid A \in \mathscr{Z}\}.$$

Then we expand $\mathscr{Z}_2$ as crudely as possible and set

$$\mathscr{Z}^{\text{Private}} := \sigma(Z, \mathscr{Z}^{U(c)}) = \sigma(Z \backslash U(c), \mathscr{Z}^{U(c)}).$$

$\mathscr{Z}^{\text{Private}}$ is the smallest σ-field that contains both σ-fields $\mathscr{Z}_1$ and $\mathscr{Z}_2$. Just like in $\mathscr{Z}_1$, the complement $U(c)$ forms an atom of $\mathscr{Z}^{\text{Private}}$. Note that the atoms of $\mathscr{Z}^{\text{Private}}$ are influenced by the atomic structure of $\mathscr{Z}_1$ and $\mathscr{Z}_2$: every atom of $\mathscr{Z}^{\text{Private}}$ can be represented in the form $B_1 \cap B_2$ with $B_i \in \{B \mid B \text{ is an atom of } \mathscr{Z}_i \text{ or } B = Z\}$ $(i = 1,2)$.
In order to model the global point of view of the jury, it could be necessary to split $\mathscr{Z}^{\text{Private}}$ again into smaller atoms. To perform this, we can apply the maximum entropy method.

C. A note on local Bayesian fusion

Consider the local DoB distribution (3). If we use it as prior distribution, we can incorporate the information d_t ($t \in S \setminus \{s\}$) via a local Bayesian fusion step in the following way:

$$\begin{aligned}
p(z = \text{type}_1 | d_s, d_t) &\propto p(d_t | z = \text{type}_1) p(z = \text{type}_1 | d_s), \\
p(z = \text{type}_2 | d_s, d_t) &\propto p(d_t | z = \text{type}_2) p(z = \text{type}_2 | d_s), \\
p(z \in Z \setminus U(c) | d_s, d_t) &= 1 - \sum_{z \in \{\text{type}_1, \text{type}_2\}} p(z | d_s, d_t).
\end{aligned}$$

It is nontrivial to determine a local likelihood function because the function $q(z) := p(d|z)$ with a fixed d is generally non-additive. But in this case, we can perform a local Bayesian fusion task in a simple way, i.e. without determining $p(d_s | z \in Z \setminus U(c))$. This works because of the way in which we restrict $\mathscr{Z}$ to get local σ-fields $\mathscr{Z}^{\text{Private}}$: the highly probable values for the PoIs stay atomic and all other values get merged into *one* "big" atom.

D. Bayesian setups

Finding a local DoB distribution and performing local Bayesian fusion will not always be as simple as in the preceding example. Because of this, we will give an overview of a mathematical concept that fits the general case. For this, we must have a look at the whole probabilistic framework for Bayesian statistics, namely Bayesian setups:
Following [9], a general Bayesian setup[1] is defined as a product probability space, where the product measure space is given by

$$(Z \times D, \sigma(\mathscr{Z}, \mathscr{D})). \tag{4}$$

The measure space $(Z, \mathscr{Z})$ that corresponds to the PoI has been defined in section IV-B. Now we look at the measure space $(D, \mathscr{D})$, too. It consists of the set of possible data and a σ-field that indicates which knowledge-states concerning the data are possible. In (4), $\sigma(\mathscr{Z}, \mathscr{D})$ denotes the smallest σ-field containing both $\mathscr{Z}$ and $\mathscr{D}$. Finally, a complete Bayesian setup is a product probability space

$$(Z \times D, \sigma(\mathscr{Z}, \mathscr{D}), \Pi), \tag{5}$$

that is (4) equipped with a certain product probability measure.

We give each fusion agent his private Bayesian product measure space, that is $(Z \times D, \sigma(\mathscr{Z}^{\text{Private}}, \mathscr{D}))$ with $\mathscr{Z}^{\text{Private}}$ as in section IV-B – but now we do this with respect to the data, too.
To formulate private Bayesian setups completely, we need an extension of the term conditional probability that is based on the abstract conditional expectation. We will review the definitions and ideas behind this concept in the next section.

E. A review on conditional expectation

Let X be an integrable random variable on a probability space $(\Omega, \mathscr{A}, P)$ and $\mathscr{A}^{\text{Part}}$ a σ-field that is contained in $\mathscr{A}$. Generally, X is not $\mathscr{A}^{\text{Part}}$-measurable. The concept of the conditional expectation gives us a possibility to determine the "best" estimate for the value of X if only the knowledge-states provided by $\mathscr{A}^{\text{Part}}$ can be reached:
The conditional expectation $E(X|\mathscr{A}^{Part})$ of X is the unique[2] *$\mathscr{A}^{Part}$-measurable and integrable random variable that satisfies the condition*

$$\int_A E(X|\mathscr{A}^{Part}) dP = \int_A X dP \quad \text{for all } A \in \mathscr{A}^{Part}.$$

It can be shown, for a quadratic integrable random variable X on $(\Omega, \mathscr{A}, P)$, that the best approximation by an $\mathscr{A}^{\text{Part}}$-measurable quadratic integrable random variable in the sense of the least square method is given by the conditional expectation $E(X|\mathscr{A}^{\text{Part}})$. In this case, $E(X|\mathscr{A}^{\text{Part}})$ is the projection of X from $(\Omega, \mathscr{A}, P)$ into the cruder probability space that is equipped with $\mathscr{A}^{\text{Part}}$ instead of $\mathscr{A}$.
If $\mathscr{A}^{\text{Part}} = \sigma(A_1, A_2, \ldots)$ with a countable (maybe finite) partition $\{A_1, A_2, \ldots\}$ of Ω, the conditional expectation $E(X|\mathscr{A}^{\text{Part}})$ must be constant over each A_i (because of the $\mathscr{A}^{\text{Part}}$ -measurability). One gets, see [8],

$$E(X|\mathscr{A}^{\text{Part}})(\omega) = \frac{1}{P(A_i)} \int_{A_i} X dP \quad \text{for } \omega \in A_i$$

if $P(A_i) > 0$. In the case that $P(A_i) = 0$, $E(X|\mathscr{A}^{\text{Part}})$ has an arbitrarily constant value on A_i. In this sense, we can interpret $E(X|\mathscr{A}^{\text{Part}})$ as a "smoothing" of X.
Generally, $E(X|\mathscr{A}^{\text{Part}})$ represents a coarsening of X. In the extreme cases $\mathscr{A}^{\text{Part}} = \mathscr{A}$ (no coarsening), we have $E(X|\mathscr{A}^{\text{Part}}) = X$ and if $\mathscr{A}^{\text{Part}} = \{\Omega, \emptyset\}$ (maximal coarsening), $E(X|\mathscr{A}^{\text{Part}}) = E(X)$.
The conditional probability $P(A|\mathscr{A}^{\text{Part}})$ of an event $A \in \mathscr{A}$ given $\mathscr{A}^{\text{Part}} \subseteq \mathscr{A}$ is defined as $E(\mathbf{1}[A]|\mathscr{A}^{\text{Part}})$, where $\mathbf{1}[A]$ is the indicator function of A, i.e. $\mathbf{1}[A](\omega) = 1$ if $\omega \in A$ and $\mathbf{1}[A](\omega) = 0$ if $\omega \notin A$.

F. Private Bayesian setups

By decomposing the product probability Π in (5), a general concept describing Bayesian inference is obtained. By this, one recognizes how local DoBs should be structured and in which matter local Bayesian fusion is performed.
The product probability Π is a coupling $\Pi = P \otimes P(\cdot|\mathscr{Z})$ with

[1] In [9], the term Bayesian experiment is used instead of Bayesian setup.

[2] The existence and uniqueness of $E(X|\mathscr{A}^{\text{Part}})$ is proven in [8].

the following components: the set of sampling probabilities[3] for a Bayesian setup is given by the family $\{P(\cdot|z)|z \in Z\}$. The sampling probabilities represent versions of the conditional probability of Π given $\mathscr{Z}$, this means $P(\cdot|z) := P(\cdot|\mathscr{Z})(z)$. The prior probability P is the marginal probability of Π on the measure space $(Z, \mathscr{Z})$. Thus, let

$$\begin{aligned} \mathrm{SETUP}_{\mathrm{Bayes}} &= (Z \times D, \sigma(\mathscr{Z}, \mathscr{D}), \Pi) \\ \text{with } \Pi &= P \otimes P(\cdot|\mathscr{Z}) = Q \otimes P(\cdot|\mathscr{D}) \end{aligned}$$

be the Bayesian setup that corresponds to the given fusion task[4]. Q denotes the marginal probability on $(D, \mathscr{D})$.
For each of the fusion agents, a private Bayesian setup

$$\mathrm{SETUP}^{\mathrm{Private}}_{\mathrm{Bayes}} = (Z \times D, \sigma(\mathscr{Z}^{\mathrm{Private}}, \mathscr{D}), \Pi_{\sigma(\mathscr{Z}^{\mathrm{Private}}, \mathscr{D})})$$

is allocated. $\Pi_{\sigma(\mathscr{Z}^{\mathrm{Private}}, \mathscr{D})}$ denotes the restriction of Π on $\sigma(\mathscr{Z}^{\mathrm{Private}}, \mathscr{D})$. We have

$$\Pi_{\sigma(\mathscr{Z}^{\mathrm{Private}}, \mathscr{D})} = P_{\mathscr{Z}^{\mathrm{Private}}} \otimes P(\cdot|\mathscr{Z}^{\mathrm{Private}}) = Q \otimes P_{\mathscr{Z}^{\mathrm{Private}}}(\cdot|\mathscr{D}).$$

This means that the local prior of a fusion agent is exactly the marginal probability of the global prior DoB in $\mathscr{Z}^{\mathrm{Private}}$,

$$P_{\mathscr{Z}^{\mathrm{Private}}}(A) = P(A) \quad \text{for all } A \in \mathscr{Z}^{\mathrm{Private}},$$

and the same scheme holds also for the local posterior DoB. To get the local sampling probability, one has to condition the global sampling probability on $\mathscr{Z}^{\mathrm{Private}}$. To understand this, note that for σ-fields $\mathscr{A}^{\mathrm{Part1}} \subseteq \mathscr{A}^{\mathrm{Part2}} \subseteq \mathscr{A}$, the following projection result holds (see [8]):

$$\begin{aligned} E(E(X|\mathscr{A}^{\mathrm{Part1}})|\mathscr{A}^{\mathrm{Part2}}) &= E(E(X|\mathscr{A}^{\mathrm{Part2}})|\mathscr{A}^{\mathrm{Part1}}) \\ &= E(X|\mathscr{A}^{\mathrm{Part1}}). \end{aligned}$$

V. Conclusion and further work

We described an agent based architecture which has a high potential to circumvent the usually high computational costs of Bayesian fusion methods. Fusion agents shall investigate the "true" fusion result in analogy to human detectives who try to solve a crime.
Therefore, a local approach for Bayesian information fusion that uses principles and ideas from measure theory has to be realized. By restricting the σ-field of the underlying probability model in a certain way, we model the local point of view of the fusion agents. Beside a computational simplification, this concept also offers the possibility to change from a closed-world model where the set of the PoIs is fixed ex ante to an open-world model.
The proposed architecture is obviously predestinated if we deal with huge networks. It is scalable to the given resources and offers the possibility to realize man-machine-interaction in a simple way.
An important task for further research will be determining the minimum size of the environment around a clue: if we chose $U(c)$ too small, local Bayesian fusion does not deliver meaningful results. Another question is the possibility of approximating local DoB distributions approximately by simpler probability functions. A promising tool could especially be the concept of conjugate priors: local Bayesian fusion could then be performed via simple algebraic calculations.
If we deal with sources for which the dependency to other sources is unknown, additional assumptions have to be made by the fusion agents to avoid an overweighing of these sources that deliver stochastically dependent data. Additionally, obtaining clues from databases is closely related to data mining: initial clues are retrieved by searching the databases for conspicuous patterns or distinctive features. Consequently, the horizon of the initialization is broadened to yet unknown clues, and not only a priori known patterns and features could be detected and contribute to the initialization of fusion agents. Generally, in addition to the information sources that are used for the initialization, the fusion agents can be equipped to access additional sources of information, e.g. if the given ones do not provide enough evidence.

[3]We avoid here the term likelihood probability: it is just common if a function for fixed data is meant.

[4]For a Bayesian setup, the converse decomposition of Π into Q and versions of the conditional probabilities of Π given $\mathscr{D}$, the posterior probabilities, need not exist and some certain conditions are necessary to guarantee its existence. We assume these to be satisfied. In [9], such Bayesian setups are called regular.

References

[1] J.O. Berger: *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.
[2] J.O. Berger and R.L. Wolpert: *The Likelihood Principle*. Instistute of Mathematical Statistics Press, Hayward, CA, 1984.
[3] J.M. Bernardo: *Bayesian Statistics*. Encyclopedia of Life Support Systems. Probability and Statistics. Oxford, UK: UNESCO, 2003. www.uv.es/∼bernardo/teaching.html
[4] J.M. Bernardo and A. Smith: *Bayesian Theory*. Wiley, New York, 1994.
[5] M. Berthold and D.J. Hand: *Intelligent Data Analysis*. Springer, 2003.
[6] J. Beyerer, M. Heizmann, and J. Sander: *Fuselets - An Agent Based Architecture for Fusion of Information of Heterogeneous Information and Data*. In: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, Belur V. Dasarathy (ed.), Proceedings of SPIE 6242, pp 235-243, 2006.
[7] J. Beyerer, J. Sander, and S. Werling: *Fusion heterogener Informationsquellen*. Beiträge des VDI/VDE-GMA-Expertenforums "Informationsfusion in der Mess- und Sensortechnik", Eisenach, 21.-22.06.2006, Universitätsverlag Karlsruhe, 2006.
[8] P. Billingsley: *Probability and Measure*. 3. edition, Wiley, New York, 1995.
[9] J.P. Florens, M. Mouchart, and J.M. Rolin: *Elements of Bayesian statistics*. Dekker, New York, 1990.
[10] A. Gelman: *Bayesian Data Analysis*. 2. edition, Chapman and Hall, 2004.
[11] I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen: *Mathematics of Data Fusion*. Kluwer Academic Publishers Group, 1997.
[12] J.N. Kapur: *Maximum Entropy Models in Science and Engineering*. John Wiley and Sons, New York, 1993.
[13] C.F. Manski: *Learning and Decision Making when Subjective Probabilities have Subjective Domains*. The Annals of Statistics, Vol. 9, No. 1, pp 59-65, 1981.
[14] H.T. Nguyen and G.S. Rogers: *Fundamentals of Mathematical Statistics*. Springer, 1989.
[15] D. Pollard: *A user's guide to measure theoretic probability*. Cambridge University Press, 2002.
[16] M.M. Rao: *Conditional Measures and Applications*. Dekker, New York, 1993.
[17] C.P. Robert: *The Bayesian Choice*. Springer, 1994.
[18] S.J. Russell and P. Norvig: *Artificial intelligence: A Modern approach*. Prentice Hall, 2003.
[19] P.N. Tan, M. Steinbach, and V. Kumar: *Introduction to Data Mining*. Addison-Wesley, 2006.