The Importance of Statistical Evidence for Focussed Bayesian Fusion

Jennifer Sander[1], Jonas Krieger[1], Jürgen Beyerer[1,2]

[1] Lehrstuhl für Interaktive Echtzeitsysteme, Institut für Anthropomatik, Karlsruher Institut für Technologie, Adenauerring 4, 76131 Karlsruhe
[2] Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB, Fraunhoferstraße 1, 76131 Karlsruhe
jennifer.sander@kit.edu
juergen.beyerer@iosb.fraunhofer.de

Abstract. Focussed Bayesian fusion reduces high computational costs caused by Bayesian fusion by restricting the range of the Properties of Interest which specify the structure of the desired information on its most task relevant part. Within this publication, it is concisely explained how Bayesian theory and the theory of statistical evidence can be combined to derive meaningful focussed Bayesian models and to rate the validity of a focussed Bayesian analysis quantitatively. Earlier results with regard to this topic will be further developed and exemplified.

Key words: information fusion; Bayesian fusion; local Bayesian fusion; statistical evidence; Likelihood ratios; probability of misleading statistical evidence.

1 Introduction

If the range Z of the Properties of Interest is large and if the involved probability distributions are not efficiently representable, the solution of Bayesian fusion tasks causes high computational costs [19]. Local Bayesian fusion approaches [5, 18–20] reduce these costs by avoiding the complete calculation of the posterior distribution. Local Bayesian fusion is inspired by criminal investigations [5]. With regard to the given fusion task, the information which has to be fused is searched for clues, i.e., for elements of Z which are better supported by the available information than others are. Then, bearing in mind the available resources, Bayesian fusion is concentrated on the identified most task relevant part U of Z.

Ignoring all elements of $Z \setminus U$ delivers a straightforward local Bayesian fusion scheme [18] which we termed focussed Bayesian fusion [6]. Usually, it is not possible to rate the validity of a focussed Bayesian model after the focussing has been done [18]. Because of this, the availableness of reliable construction rules for meaningful focussed Bayesian models is of prime importance at the research on local Bayesian fusion.

Different criteria deliver construction rules which serve this requirement: probabilistic error bounds [18], information theoretic quality indicators [19], and probability interval schemes [20]. All resulting rules deliver the same guidelines for the design of meaningful focussed Bayesian models. Additionally, by each of these rules, the validity of a focussed Bayesian analysis is ratable quantitatively with regard to a certain aspect. This publication addresses probabilistic error bounds which originate from the theory of statical evidence.

While probability in the sense of the Degree of Belief interpretation is an adequate measure for every kind of uncertainty [15, 4], Likelihood ratios provide an adequate quantitative statistical evidence measure [11, 17]. This has been formalized by Hacking as the Law of Likelihood already in 1965 [14]. In several more recent publications, questions concerning the reliability of observed statistical evidence and (in connection with that) the concept of misleading statistical evidence are discussed. Bounds for the probability of misleading statistical evidence, i.e., for the probability that misleading statistical evidence of a certain strength occurs are given, see for example [8, 16, 17]. By these bounds, the reliability of statistical evidence becomes quantifiable.

The theory of statistical evidence is not conflicting to Bayesian theory [7], rather it is implicitly an inherent part of it [12]–provided that a Degree of Belief interpretation of probability is adopted. Using the theory of statistical evidence explicitly for the task-specific design of Bayesian models seems to be self-evident. However, by default, Bayesian models are created without respect to the values which the observable quantities adopt in a given task[3].

This paper is organized as follows: Sec. 2 is a short introduction into Bayesian fusion. In Sec. 3, we demonstrate how focussed Bayesian fusion reduces the computational complexity of Bayesian fusion. Probabilistic error bounds for focussed Bayesian fusion derive themselves from the universal bound for the probability misleading statistical evidence which is applicable to every probabilistic model. In Sec. 4, we review this bound after introducing the necessary foundations from the theory of statistical evidence. The application of the concepts from Sec. 4 to focussed Bayesian fusion is done in Sec. 5 and examples for the corresponding proceeding are given in Sec. 6. These examples are much more simpler as the one to which we refer in Sec. 3 with regard to complexity reduction. We chose them intentionally because they confirm the theoretical analysis done in Sec. 5 in an easily comprehensible manner.

2 Bayesian Fusion

Let $z = (z_1, \ldots, z_N) \in Z = Z_1 \times \ldots \times Z_N$, $N \in \mathbb{N}$, denote the Properties of Interest and $d = (d_1, \ldots, d_S) \in D = D_1 \times \ldots \times D_S$, $S \in \mathbb{N}$, denote the information from several information sources. d_s stands for the contribution of information source s, $s \in \{1, \ldots, S\}$. At a given fusion task, d adopts a value according the observed information contributions while the "true" value of the Properties of Interest is unknown. At Bayesian fusion, all available information is transformed into a probabilistic representation in the sense of the Degree of

[3] See for example the discussion in [4] about the generation of a dynamic frame of discourse.

Belief interpretation. For this, all involved quantities are assumed to be random. The Bayesian theorem states how an initial Degree of Belief has to get modified to include additional knowledge in an adequate manner: it holds

$$p(z|d) \propto p(d|z)\, p(z) \ . \tag{1}$$

It is an essential advantage of Bayesian methods that–beside the contributions of the information sources which are included in the inference via the Likelihood $p(d|z)$–the posterior distribution $p(z|d)$ also reflects prior knowledge which is included in the inference via the prior distribution $p(z)$ [6].

If the information contributions are conditionally independent given z, it holds $p(d|z) = \prod_{s=1}^{S} p(d_s|z)$. In this case, it suffices at Bayesian fusion to transform each information contribution d_s individually into a source specific Likelihood $p(d_s|z)$, $s \in \{1, \ldots, S\}$. Storage costs for the saving of $p(d|z)$ get reduced. If $p(d|z)$ has to be approximated using training data, less of them will be necessary to obtain a sufficiently good approximation. The possibility to realize Bayesian fusion via a sequential fusion scheme [5] is another advantageous consequence.

3 Reduction of Computational Complexity by Focussing

The computationally complexity for the necessary operations to obtain the posterior distribution at Bayesian fusion is $\mathrm{O}(|Z|) = \mathrm{O}(\zeta^N)$, $\zeta = \sqrt[N]{\prod_{i=1}^{N} |Z_i|}$, which may be prohibitive in real world tasks. If the actual Bayesian fusion is concentrated on $U \subset Z$ with $|U| << |Z|$, the computational complexity is reduced considerably on $\mathrm{O}(|U|)$.

To clarify this practically, we refer to an example which has been given in [20]: the task of Bayesian fusion is the determination of positions, driving directions and types of cars in a scene. For this, prior knowledge from a street map, IMINT information corresponding to three images of the scene and HUMINT information are fused. The set of possible positions is discretized to approximately 1280×960 units. There are four possible driving directions ("south", "north", "west", "east") and five possible car types. At Bayesian fusion, formula (1) has to be evaluated for over $2 \cdot 10^7$ values of the Properties of Interest. In essence, we applied the concept of statistical evidence as described in Sec. 5 to find an adequate subset U of Z for focussed Bayesian fusion. Then, formula (1) has been evaluated only for these values of the Properties of Interest which are included in U. Here, $|U|$ constituted less than 15 percent of $|Z|$. I.e., for the actual Bayesian fusion, the range of the Properties of Interest has been cut substantially.

4 Statistical Evidence

If $p(d|z^*) > p(d|z^{**})$ holds for $z^*, z^{**} \in Z$, the observation d provides statistical evidence in support of z^* vis-a-vis z^{**} because d is more probable under the assumption that the "true" value of the Properties of Interest is z^* than it was z^{**}.

In general, a low value of $p(d|z^*)$ for a certain $z^* \in Z$ does not imply that d represents statistical evidence against z^*: the value of $p(d|z)$ may be low for all $z \in Z$ and d may provide significant statistical evidence in support of z^* compared with each other possible value of the Properties of Interest [3]. A quantitative measure of the statistical evidence which is provided by d can be obtained by pairwise comparisons of Likelihood values: for $z^*, z^{**} \in Z$, the Likelihood ratio $p(d|z^*)/p(d|z^{**})$ measures the strength of the statistical evidence that is provided by d in support of z^* vis-a-vis z^{**} [14, 11, 17]. Instead of communicating Likelihood ratios with respect to each possible pair of values of the Properties of Interest, the statistical evidence from d can be represented more efficiently by the use of the relative Likelihood function $r(d|z)$ which results if $p(d|z)$ gets scaled to a maximum value of one. $r(d|z)$ communicates the statistical evidence from d in support of each $z \in Z$ vis-a-vis $\arg\max_z p(d|z)$ which is the best supported hypothesis concerning the "true" value of the Properties of Interest [7, 17]. An observation d provides misleading statistical evidence of the strength $p(d|z^*)/p(d|z^{**})$ in support of z^* vis-a-vis z^{**} if $p(d|z^*)/p(d|z^{**}) > 1$ holds although the "true" value of the Properties of Interest is z^{**} [17]. Even though it can be misleading, statistical evidence is a valuable concept: useful bounds for the probability of observing misleading statistical evidence of at least a certain strength exist. For $z^*, z^{**} \in Z$, $\epsilon \in (0, 1)$, the universal bound delivers[4]

$$\sum\!\!\!\!\!\!\int_{d \in E} p(d|z^{**}) \le \epsilon \ , \quad E := \{d \in D \,|\, p(d|z^{**})/p(d|z^*) \le \epsilon\} \ , \tag{2}$$

see for example [8, 17]. I.e., under the assumption that the "true" value of the Properties of Interest is z^{**}, the probability for the observation of a value of d which delivers statistical evidence of at least the strength $1/\epsilon$ in support of z^* vis-a-vis z^{**} is bounded by ϵ. The probability of observing misleading statistical evidence is relevant at planning a statistical analysis.

5 Application at Focussed Bayesian Fusion

At focussed Bayesian fusion, Z gets restricted to U on the basis of a pre-evaluation of the observed information prior to the actual Bayesian fusion. Needless to say that the resulting focussed posterior distribution will not represent completely the knowledge about z which is provided by prior knowledge and by d: by the focussing, some information will get lost [19]. Nevertheless, the quality of the focussed posterior distribution can be sufficiently high if the restriction of Z is done in a reasonable manner. The theory of statistical evidence helps to reach this goal: Z gets restricted on the set of these values of the Properties of Interest which are at most consistent with d at a certain level which is specified by ϵ, $\epsilon \in [0, 1)$, [7] if we set

$$U := \{z \in Z \,|\, r(d|z) > \epsilon\} \ . \tag{3}$$

[4] The symbol $\sum\!\!\!\!\!\!\int$ means summation with respect to discrete and integration with respect to continuous components of d.

The choice of ϵ should be conform to the available resources. If U is defined according to (3), (2) saves as error bound for focussed Bayesian fusion: the Degree of Belief that the "true" value of the Properties of Interest is not included in U is bounded by ϵ because it can be identified with the Degree of Belief that the information contributions adopt values which lead to an ignoring of the "true" value of the Properties of Interest at focussed Bayesian fusion.

The statistical evidence which is provided by one single information contribution d_s, $s \in \{1, \dots, S\}$, is represented by $r(d_s|z)$– a quantity which does not take into account prior information and the other information contributions d_t, $t \neq s$. Hence, an isolated evaluation of the statistical evidence from each d_s is possible. This proceeding may be generally advantageous–but especially for the fusion of heterogenous information sources which have to be evaluated using extremely different kinds of expertise. Here, the analogy between criminal investigations and local Bayesian fusion becomes again obvious: a forensic expert who analyzes a specific kind of data with respect to the delinquent of one or more specific persons has to provide an analysis of the statistical evidence from this data which can subsequently be combined by an investigating detective or a court with the respective prior odds [11] and possibly additional statistical evidence. It is not the job of the forensic expert to analyze also the values of these quantities [2]. At focussed Bayesian fusion, it also makes sense to define U to consist of these values of the Properties of Interest which are at most consistent with at least one of the information contributions at a certain level, i.e.,

$$U := \{z \in Z \,|\, r(d_s|z) > \epsilon \text{ for at least one } s \in \{1, \dots, S\}\} \tag{4}$$

with an $\epsilon \in [0, 1)$ whose value should be conform to the given resources. Especially in the case of heterogenous information sources, the information contributions may be conditionally independent. As consequence, the error bound for focussed Bayesian fusion gets significantly sharpened to ϵ^S. This fact is easily provable, e.g., by an adaption of the proof of (2) which is given in [8]. If the conditional independence assumption does not hold, the bound ϵ^S is not valid. However, it makes sense to assume that generally in this case, a more optimal bound will be lesser than the poorer–also easily derivable–bound ϵ guarantees [18].

Hence, the concepts from the theory of statistical evidence give basic guidance for the determination of U and the probabilistic error bounds rate the validity of the resulting focussed Bayesian model quantitatively.

As the example in [20] shows, it is usually not necessary to adapt the size of U exactly to the available resources. However, if such an exact adaption should be done, ideally, the precise scheme for the determination of U should be task specific. We will demonstrate this theoreticly in the rest of this section and practically in Sec. 6.

If the size of U is predefined, it is wrong to say that always all S information contributions should be evaluated to determine U in an theoreticly optimal manner. The condition "$r(d_s|z) > \epsilon$ for at least one s" in (4) means that a value z of the Properties of Interest gets ignored if its relative Likelihood with respect to all evaluated information contributions is not exceeding ϵ. If ϵ is fixed, regarding only $T < S$ information contributions for the definition of U will generally lead to a smaller size of U. Because ϵ should be chosen as low as possible according to the given resources, generally, a lower value ϵ_τ should be selected for ϵ if only $T < S$ instead of S information contributions are evaluated for defining U and, as consequence, a lower error bound for focussed Bayesian fusion may result.

However, the use of $T < S$ instead of S information contributions for the definition of U is also not always favorable if the size of U is predefined. E.g., if in the case of the conditional independence of the information contributions, the error bounds are ϵ_τ^T if $T < S$ information contributions are evaluated and ϵ_σ^S if all S information contributions are evaluated with $\epsilon_\tau < \epsilon_\sigma$, it depends on the specific values of ϵ_τ, ϵ_σ, S and T if also $\epsilon_\tau^T < \epsilon_\sigma^S$ holds.

Hence, if the size of U is predefined, the best determination scheme for U–with regard to the error bounds–depends on given fusion task. As discussed in the next two paragraphs, also the costs for the retrieval of the probabilistic representations should be always considered in practice.

In principle, basing the determination of U on the evaluation of only $T < S$ information contributions saves resources: less criteria must be checked to decide if a certain value $z \in Z$ is ignorable. Knowledge corresponding to an information contribution d_s, $s \in \{1, \ldots, S\}$, which is evaluated for the determination of U has to be transformed into a full probabilistic representation in the sense that $r(d_s|z)$ has to be determined for all $z \in Z$. In contrast, for information contributions which are used solely at the subsequent actual focussed Bayesian fusion, the Likelihood has to be determined only for $z \in U$.

On the other hand, if the transformation of all kind of information into a full probabilistic representation consumes too much resources, the information contributions may also get pre-evaluated in a suboptimal manner: also by such a suboptimal pre-evaluation, a probabilistic representation in form of Likelihoods may be obtained such that the theory of statistical evidence is applicable. For example, in [20], a suboptimal pre-evaluation of all information contributions for the determination of U is combined with a subsequent actual focussed Bayesian fusion based on the actual probabilistic information representations. However, we stress that by this proceeding, the resulting error bounds which are based on probabilistic representations of lower quality, will be generally less reliable.

6 Exemplification of Focussed Bayesian Fusion

Here, focussed Bayesian fusion is applied exemplarily at naive Bayesian classification using two data sets from the UCI Machine Learning Repository [1]: Pendigits (16 attributes, 10 classes) and Letter Recognition (16 attributes, 26 classes). The aim of the studies is an easily comprehensible demonstration of the described theoretical results. It is not disclaimed that the number of classes in the data sets may be too low therefor that the application of focussed Bayesian fusion makes sense in reality and, of course, more sophisticated classifiers will outperform our results.

Strictly speaking, the final result of a Bayesian fusion task is the posterior distribution. However, using decision theoretic concepts [4], subsequent decisions can be made on the basis of $p(z|d)$. In a Bayesian classification task, an appropriate decision is choosing the Maximum a Posteriori estimate $\arg\max_z p(z|d)$ which minimizes the expected posterior loss for a zero-one loss function [10]. The naive Bayesian classifier simplifies a Bayesian classification task significantly by assuming that the attributes d_s, $s \in \{1, \ldots, S\}$, are conditionally independent given the class z. It is well known that this classifier often has a good performance although the conditional independence does not hold and although it may not be Bayes-optimal [10]: see e.g. [9, 13] and the references given therein.

Here, we implemented a simple form of the naive Bayesian classifier which has been used in [9] for its empirical evaluation: attribute values were discretized in 10 intervals of equal length, zeros in the probabilistic representations were avoided using Laplace corrections. All reported accuracies are averaged over 20 runs. For each run, the data get randomly divided into training and test data. The training data constitute 2/3 of all data. According to the calculated accuracies (Pendigits: ca. 87.81 %, Letter Recognition: ca. 70.77 %), the application of the naive Bayesian classifier is not completely beside the point for the chosen data sets and the stability of the accuracies in terms of sample standard deviations (with respect to 20 runs) is also acceptable (Pendigits: ca. 0.58 %, Letter Recognition: ca. 0.57 %).

In our first study, the decision which classes are ignored is based on the probabilistic representations used for the naive Bayesian classification. The results for Pendigits are depicted in Fig. 1. At rule 1, a class $z \in Z$ gets ignored unless it holds $r(d_s|z) > \epsilon$ for at least one $s \in \{1, \ldots, 16\}$. At rule 2, a class $z \in Z$ gets ignored unless it holds $p(d_s|z) > \epsilon$ for at least one $s \in \{1, \ldots, 16\}$. At rule 3, the ignored classes are selected randomly. At rule 1 and rule 2, different thresholds $\epsilon \in \{i/100 \,|\, i \in \{0, 1, 2, \ldots, 99\}\}$ are applied. Although rule 2 performs better than rule 3, the accuracies are conspicuously below these which are obtained by the application of rule 1 which respects the theory of statistical evidence. Even if ϵ is near to 1, rule 1 does not allow the ignoring of an unreasonable number of classes (for $\epsilon = 0.99$, ca. 33 % of the classes are ignored). The performance of rule 1 is even better than the sharper probabilistic error bound (which assumes the conditional independence of the attributes[5]) leads one to assume.

The results for Letter Recognition are comparable to these which have been obtained for Pendigits. Here, ignoring ca. 19 % of the classes according to rule 1 leads to a extremely low worsening of the accuracy of less then 1 %. For the highest considered threshold $\epsilon = 0.99$, about 55% of the classes are ignored and, by this, the accuracy gets lowered on ca. 65 %.

To demonstrate that the concepts of statistical evidence also apply if other reasonable probabilistic models are used for the determination of U than the one which is used for the actual focussed Bayesian fusion, we conducted a second study for the Pendigits data at which the attributes correspond to (x, y)-

[5] As noted in Sec. 5, it makes sense to assume that a more optimal bound lies between the lower bound ϵ and this bound, here.

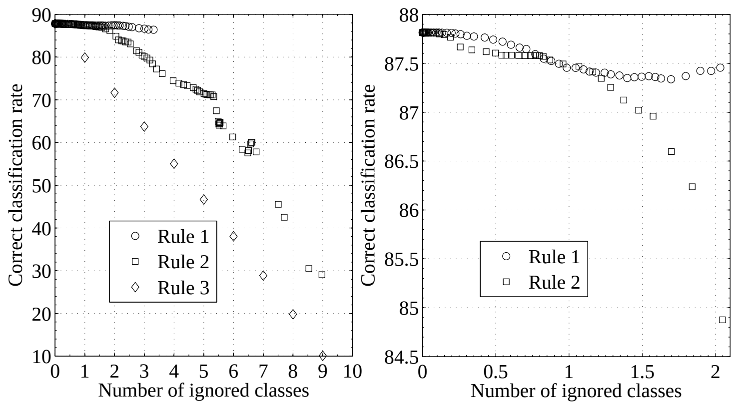

Fig. 1. Change of accuracy of naive Bayesian classification for Pendigits when classes are ignored according to rule 1, rule 2, and rule 3. For rule 1 and rule 2, each marker represents a value which corresponds to a fixed threshold (averaged over 20 runs). I.e., for theses rules, each marker relates (for a fixed threshold ϵ) the average number of classes which are ignored and the average correct classification rate over the 20 runs.

coordinates of hand-written digits. d_s corresponds to a x-coordinate if s is odd and to a y-coordinate if s is even, $s \in \{1, \ldots, 16\}$. By the combination of parts of the information which is delivered by d_1, d_5, and d_7, a new attribute $\tilde{d}_x$ which reports the kind of changes between some of the x-coordinates is created:

$$\tilde{d}_x := \begin{cases} 0, d_1 < d_5 < d_7 \\ 1, d_1 < d_5 \text{ and } d_7 \leq d_5 \\ 2, d_5 \leq d_1 \text{ and } d_5 < d_7 \\ 3, d_7 \leq d_5 \leq d_1 \end{cases} . \qquad (5)$$

For the determination of U, $\tilde{d}_x$ is evaluated on the basis of the concepts of statistical evidence: at rule 4, a class $z \in Z$ gets ignored unless it holds $r(\tilde{d}_x|z) > \epsilon$, $\epsilon \in \{i/100 \,|\, i \in \{0, 1, 2, \ldots, 99\}\}$. The probabilistic information representation which is used by rule 4 is neither equivalent to the exact Likelihoods nor to its approximations which are used at naive Bayesian classification. However, rule 4 makes sense as Fig. 2 shows. Indeed, for larger values of ϵ, also here an absurdly large number of classes is ignored and, as consequence, the accuracy stays not longer acceptable. However, it is not astonishing that rule 1 (which considers all relative Likelihood functions with respect to the original 16 attributes) outperforms rule 4 (which relates only three of them via one attribute $\tilde{d}_x$) with regard to this aspect. For a reasonable number of ignored classes, the accuracy of rule 4 may be acceptable high. In the cases in that not more than 20 % of the classes are ignored, the accuracy reached with rule 4 even exceeds the accuracy which is reachable using rule 1. Note that the newly created attribute $\tilde{d}_x$ comprises some information with regard to the conditional dependencies between d_1, d_5, and d_7.

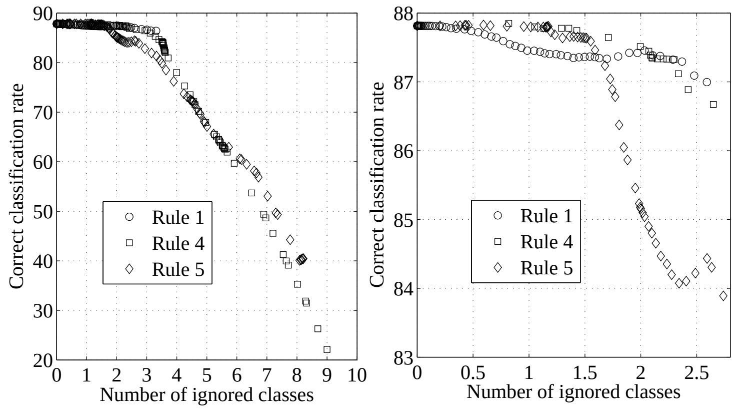

Fig. 2. Change of accuracy of naive Bayesian classification for Pendigits when classes are ignored according to rule 1, rule 4, and rule 5. Each marker represents a value which corresponds to a fixed threshold (averaged over 20 runs).

Parts of the information which is delivered by d_2, d_6, and d_8 has been used to create a new attribute $\tilde{d}_y$ analogously to $\tilde{d}_x$. When we analyzed the performance of both new attributes together, it became obvious that this is a case in that focussed Bayesian fusion based on the analysis of a smaller number of information contributions for the determination of U may be preferable if the size of U is predefined. At rule 5, a class $z \in Z$ gets ignored unless it holds $r(\tilde{d}_x|z) > \epsilon$ or $r(\tilde{d}_y|z) > \epsilon$. It becomes clear from Fig. 2 that rule 4 outperforms rule 5 if the size of U is predefined and not absurdly large. The reason of this has been identified in Sec. 5. The following exemplary numbers clarify it additionally: at rule 4, the lowest value of ϵ for that at least three classes are ignored (on average) is 0.25. Applying rule 5, the lowest value of ϵ for that at least three classes are ignored (on average) is 0.59. Even the sharper probabilistic error bound leads to the conclusion that the use of rule 4 is preferable in this situation.

7 Conclusion

The theory of statistical evidence provides concepts which are extremely valuable at the research on focussed Bayesian fusion. Their application outperforms ad hoc solutions by far. It offers rules for the exact creation of focussed Bayesian models and it provides error bounds by which the validity of a focussed Bayesian fusion is ratable. There exists no unique recipe for the exact creation of focussed Bayesian models if the size of U is predefined. However, the error bounds represent helpful mathematical tools for making this choice.

References

1. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository. `http://www.ics.uci.edu/~mlearn/MLRepository.html`. Irvine: University of California, School of Information and Computer Science (2007)
2. Aitken, C.G.G., Taroni, F.: Statistics and the Evaluation of Evidence for Forensic Scientists. Wiley, Chichester (2004)
3. Bartelborth, T.: Wofür sprechen die Daten?. Journal for General Philosophy of Science 35(1), 13–40 (2004)
4. Bernardo, J.M., Smith, A.F.M.: Bayesian Theory. Wiley, Chichester (2004)
5. Beyerer, J., Heizmann, M., Sander, J.: Fuselets – an Agent Based Architecture for Fusion of Heterogeneous Information and Data. In: Dasarathy, B.V. (ed.) Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006. Proceedings of SPIE 6242. SPIE, Bellingham (2006)
6. Beyerer, J., Heizmann, M., Sander, J., Gheta, I.: Bayesian Methods for Image Fusion. In: Stathaki, T. (ed.) Image Fusion: Algorithms and Applications. Academic Press, Amsterdam, Heidelberg (2008)
7. Blume, J.D.: Likelihood Methods for Measuring Statistical Evidence. Statistics in Medicine 21(17), 2563–2599 (2002)
8. Dempster, A.P.: The Direct Use of Likelihood for Significance Testing. Statistics and Computing 7(4), 247–252 (1997)
9. Domingos, P., Pazzani, M.: On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Mach. Learn. 29(2-3), 103–130 (1997)
10. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)
11. Edwards, A.W.F.: Likelihood. Cambridge University Press, London (1972)
12. Goodman, S.N.: Toward Evidence-Based Medical Statistics. 2: the Bayes Factor. Ann. Intern. Med. 130(12), 1005–1013 (1999)
13. Ekdahl, M., Koski, T.: Bounds for the Loss in Probability of Correct Classification under Model Based Approximation. J. Mach. Learn. Res. 7, 2449–2480 (2006)
14. Hacking, I.: Logic of Statistical Inference. Cambridge University Press, New York (1965)
15. Lindley, D.V.: The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems. Statistical Science 2(2), 3–44 (1987)
16. Royall, R.M.: On the Probability of Observing Misleading Statistical Evidence. Journal of the American Statistical Association 95(451), 760–768 (2000)
17. Royall, R.M.: Statistical Evidence: a Likelihood Paradigm. Chapman & Hall, London (1997)
18. Sander, J., Beyerer J.: Decreased Complexity and Increased Problem Specificity of Bayesian Fusion by Local Approaches. In: Proceedings of the 11th International Conference on Information Fusion, Fusion 2008, pp. 1035–1042. IEEE Press (2008)
19. Sander, J., Heizmann, M., Goussev, I., Beyerer, J.: A Local Approach for Focussed Bayesian Fusion. In: Dasarathy, B.V. (ed.) Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2009, Proceedings of SPIE 7345. SPIE, Bellingham (2009)
20. Sander, J., Heizmann, M., Goussev, I., Beyerer, J.: Global Evaluation of Focussed Bayesian Fusion. In: Braun, J.J. (ed.) Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2010, Proceedings of SPIE 7710. SPIE, Bellingham (2010)