Approximation of Belief Functions by Minimizing Euclidean Distances

Thomas Weiler and Ulrich Bodenhofer

Software Competence Center Hagenberg
A-4232 Hagenberg, Austria
e-mail: {thomas.weiler,ulrich.bodenhofer}@scch.at

Abstract. This paper addresses the approximation of belief functions by minimizing the Euclidean distance to a given belief function in the set of probability functions. The special case of Dempster-Shafer belief functions is considered in particular detail. It turns out that, in this case, an explicit solution by means of a projective transformation can be given. Furthermore, we also consider more general concepts of belief. We state that the approximation by means of minimizing the Euclidean distance, unlike other methods that are restricted to Dempster-Shafer belief, works as well. However, the projective transformation formula cannot necessarily be applied in these more general settings.

Key words: belief functions, Dempster-Shafer theory, evidence reasoning, probability theory, theory of evidence, uncertain reasoning.

1 Introduction

The relation between belief and probability plays an important role in the theory of uncertain reasoning and its applications. Belief, e.g. as in Dempster-Shafer theory, can be viewed as an extension of probability. An advantage of more general concepts of belief is that they relate to incomplete information. Probability theory, however, provides a well-established decision making theory [1,2]. In order to be able to apply probabilistic decision principles also in the presence of more general concepts of belief, therefore, different ways to transform belief functions into probability functions have been developed. So far, Dempster-Shafer belief functions have been studied most extensively in this context [5–9].

This paper is concerned with the problem how a given belief function can be approximated by minimizing the Euclidean distance in the set of probability functions, i.e. we consider the following minimization problem.

Optimization Problem (OP) For a given belief function $D : \overline{SL} \to [0,1]$ minimize the objective function

$$\sqrt{\sum_{\overline{\theta}\in\overline{SL}} \big(D(\overline{\theta}) - P(\overline{\theta})\big)^2}$$

with respect to a probability function $P : \overline{SL} \rightarrow [0,1]$, where $\overline{SL}$ denotes the Lindenbaum algebra (i.e. the Boolean algebra of well-formed formulae, where two formulae are considered as equal if all their evaluations coincide) of a finite propositional language L with n propositional variables $p_1, \ldots, p_n$.

In this paper, we first consider (OP) in relation with Dempster-Shafer belief functions. We show that (OP) can be solved explicitly by means of a projection transformation. Finally, we discuss more general concepts of belief. However, we demonstrate that the projective transformation does not necessarily give the correct result if we go beyond Dempster-Shafer belief functions.

2 Belief Functions

Definition 1. On the Lindenbaum algebra $\overline{SL}$, we define the following ordering: For all $\overline{\theta}, \overline{\psi} \in \overline{SL}$,

$$\overline{\theta} \leq \overline{\psi} \text{ if and only if } \overline{\theta} \models \overline{\psi}.$$

A formula $\overline{\alpha} \in \overline{SL}$ is called an *atom* of $\overline{SL}$ if and only if for each propositional sentence $\overline{\theta} \in \overline{SL}$ either $\overline{\alpha} \leq \overline{\theta}$ or $\overline{\alpha} \leq \neg\overline{\theta}$ holds.

Note that atoms uniquely correspond to conjunctions of the form

$$[\neg]p_1 \wedge [\neg]p_2 \wedge \cdots \wedge [\neg]p_n,$$

where the brackets indicate that each propositional variable may be prefixed with a negation or not. Therefore, $J = 2^n$ different atoms exist for a language L with n propositional variables.

Definition 2. A mapping $V : \overline{SL} \rightarrow \{0,1\}$ is called a *valuation* if and only if there exists an atom $\overline{\alpha} \in \overline{At}$ such that for all $\overline{\theta} \in \overline{SL}$,

$$V(\overline{\theta}) = \begin{cases} 1 \text{ for } \overline{\theta} \geq \overline{\alpha} \\ 0 \text{ otherwise} \end{cases}$$

where $\overline{At}$ stands for the set of atoms of $\overline{SL}$. Moreover, we denote the valuation induced by an atom $\overline{\alpha}$ with $V_{\overline{\alpha}}$.

As easy to see, a valuation is a function that assigns a truth value to each formula in $\overline{SL}$ with the particular property that the truth values assigned to the propositional variables uniquely determine the truth value of any compound formula.

Definition 3. A mapping $P : \overline{SL} \rightarrow [0,1]$ is called a *probability function* if and only if there exists a mapping $m_P : \overline{At} \rightarrow [0,1]$ satisfying

(P1) $\sum_{\overline{\alpha}\in\overline{At}} m_P(\overline{\alpha}) = 1,$
(P2) $P(\overline{\theta}) = \sum_{\overline{\alpha}\in\overline{At}} m_P(\overline{\alpha}) \cdot V_{\overline{\alpha}}(\overline{\theta})$ for all $\overline{\theta} \in \overline{SL}$.

In order to treat Dempster-Shafer belief functions in a similar way, we generalize the concept of valuations.

Definition 4. A mapping $V' : \overline{SL} \to \{0,1\}$ is called an *information function* if and only if there exists a $\overline{\theta} \in \overline{SL}\backslash\{\mathbf{0}\}$ such that for all $\overline{\psi} \in \overline{SL}$,

$$V'(\overline{\psi}) = \begin{cases} 1 \text{ for } \overline{\psi} \geq \overline{\theta} \\ 0 \text{ otherwise} \end{cases}$$

where $\mathbf{0}$ denotes the equivalence class of contradictions and $V'_{\overline{\theta}}$ stands for the information function which is generated by $\overline{\theta}$.

The crucial difference between a valuation and an information function is that an information function does not need to be generated by an atom. Information functions are closely related to the simple support functions defined by Shafer [10]. By using information functions, we are able to extend the definition of probability to Dempster-Shafer belief.

Definition 5. A mapping $D : \overline{SL} \to [0,1]$ is called a *Dempster-Shafer belief function* if and only if there exists a mapping $m_D : \overline{SL}\backslash\{\mathbf{0}\} \to [0,1]$ which satisfies

(DS1) $\sum_{\overline{\theta}>\mathbf{0}} m_D(\overline{\theta}) = 1,$
(DS2) $D(\overline{\theta}) = \sum_{\overline{\psi}>\mathbf{0}} m_D(\overline{\psi}) \cdot V'_{\overline{\psi}}(\overline{\theta})$ for all $\overline{\theta} \in \overline{SL}$.

Dempster-Shafer belief functions are, therefore, convex combinations of information functions (equivalent definitions can be found in the literature[3]). Now let us come to the most general case.

Definition 6. An arbitrary mapping $Bel : \overline{SL} \to [0,1]$ is called a *general belief function*. We denote the set of all general belief functions (with respect to a given finite propositional language) with GB.

In the following section, we define a vector space structure on the above set of belief functions. In this vector space, the optimization problem (OP) turns out to be equivalent to projecting a vector into a subspace, at least if we restrict to Dempster-Shafer belief functions.

3 The Vector Space of Belief Functions

Definition 7. The *vector space of functions from $\overline{SL}$ to* $\mathbb{R}$ is defined as

$$\mathbf{B} = (B, \oplus, \ominus, \odot, O, \|.\|, \langle .,. \rangle)$$

where B is the set of all functions from $\overline{SL}$ to $\mathbb{R}$. For all $\overline{\theta} \in \overline{SL}$, $Bel_1, Bel_2 \in B$ and $\lambda \in \mathbb{R}$, the operations $\oplus$, $\ominus$ and $\odot$ are defined by

$$\begin{aligned}(Bel_1 \oplus Bel_2)(\overline{\theta}) &= Bel_1(\overline{\theta}) + Bel_2(\overline{\theta}) \\ (Bel_1 \ominus Bel_2)(\overline{\theta}) &= Bel_1(\overline{\theta}) - Bel_2(\overline{\theta}) \\ (\lambda \odot Bel_1)(\overline{\theta}) &= \lambda \cdot Bel_1(\overline{\theta})\end{aligned}$$

O denotes the zero function $O(\overline{\theta}) = 0$. The inner product is defined by

$$\langle Bel_1, Bel_2 \rangle = \sum_{\overline{\theta} \in \overline{SL}} Bel_1(\overline{\theta}) \cdot Bel_2(\overline{\theta}).$$

The norm is uniquely given by the inner product in the usual way:

$$\|Bel_1\| = \sqrt{\langle Bel_1, Bel_1 \rangle}.$$

The precise interpretation of probability and Dempster-Shafer belief functions in $\mathbf{B}$ becomes clear by recalling the definitions of the convex and affine hull.

Definition 8. Let $\mathbf{V}$ be a vector space and $\{V_1, \ldots, V_l\}$ be a set of vectors in $\mathbf{V}$. Then

$$C(V_1, \ldots, V_l) = \{(m_1 \odot V_1) \oplus \cdots \oplus (m_l \odot V_l) \mid m_1, \ldots, m_l \geq 0 \text{ and } \sum_{i=1}^{l} m_i = 1\}$$

is called the *convex hull* of $\{V_1, \ldots, V_l\}$ and

$$A(V_1, \ldots, V_l) = \{(m_1 \odot V_1) \oplus \cdots \oplus (m_l \odot V_l) \mid \sum_{i=1}^{l} m_i = 1\}$$

defines the *affine hull* of $\{V_1, \ldots, V_l\}$.

From the definitions above, we see that the set of probability functions is obviously the convex hull of all valuations, and that the set of Dempster-Shafer belief functions is the convex hull of all information functions. The optimization problem (OP) formulated in the framework of $\mathbf{B}$ is equivalently given as: for a given belief function $D \in GB$, minimize the objective function $\|D \ominus P\|$ with respect to

$$P \in C(V_{\overline{\alpha_1}}, \ldots, V_{\overline{\alpha_J}}).$$

We proceed in the following way: we define an operator P^* on the set of Dempster-Shafer belief functions and show that $P^*_D(D) \in A(V_{\overline{\alpha_1}}, \ldots, V_{\overline{\alpha_J}})$ and, moreover, that $P^*_D(D)$ solves the modified optimization problem given next.

Simplified Optimization Problem (SOP) For a given belief function $D \in GB$, minimize the objective function $\|D \ominus P\|$ with respect to

$$P \in A(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}}).$$

The advantage of considering $A(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}})$ instead of $C(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}})$ is that every affine hull is an affine subspace. Minimizing $\|D \ominus P\|$ in a subspace of $\mathbf{B}$ means that we look for the projection of D onto the subspace. A projection has the property of the projection vector standing perpendicular on the given subspace, i.e. the inner product of the projection vector and any vector of the subspace is zero [11]. The same holds true for affine subspaces if we consider that linear subspace which is parallel to the affine subspace.

Finally, we show that even $P^*_D(D) \in C(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}})$ holds, which proves that $P^*_D(D)$ indeed solves (OP).

4 The Transformation Formula for Dempster-Shafer Belief Functions

Assume throughout this section that $D \in GB$ is a Dempster-Shafer belief function with

$$D = \sum_{\overline{\theta} > \mathbf{0}} m_D(\overline{\theta}) \cdot V'_{\overline{\theta}}.$$

Definition 9. The formula

$$P^*_D(D) = \sum_{\overline{\alpha} \in \overline{At}} m_{P^*_D(D)}(\overline{\alpha}) \cdot V_{\overline{\alpha}}$$

defines the *projective transformation* of $D : \overline{SL} \to [0,1]$, such that for all $\overline{\alpha} \in \overline{At}$,

$$m_{P^*(D)}(\overline{\alpha}) = \sum_{\overline{\theta} \geq \overline{\alpha}} 2^{1-|S_{\overline{\theta}}|} \cdot m_D(\overline{\theta}) - \sum_{\overline{\theta} > \mathbf{0}} |S_{\overline{\theta}}| \cdot 2^{1-|S_{\overline{\theta}}|-n} \cdot m_D(\overline{\theta}) + J^{-1},$$

where $S_{\overline{\theta}} = \{\overline{\alpha} \in \overline{At} \mid \overline{\alpha} \leq \overline{\theta}\}$, i.e. the set of atoms from which $\overline{\theta}$ can be inferred. Consequently, $|S_{\overline{\theta}}|$ denotes the cardinality of this set.

It can be shown that $P^*_D(D)$ maps D into $A(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}})$.

Lemma 1. $P^*_D(D) \in A(V_{\overline{\alpha_1}}, \dots, V_{\overline{\alpha_J}})$.

In order to prove that $P^*_D(D)$ solves (SOP) it is to show that the projection is perpendicular onto the plane of probability functions.

Lemma 2. $\langle V_{\overline{\alpha_i}} \ominus V_{\overline{\alpha_1}}, P^*_D(D) \ominus D \rangle = 0$ *holds for all* $i \in \{2, \ldots, J\}$.

By applying basic functional analysis it follows the required result [11].

Theorem 1. $P^*_D(D)$ *solves (SOP).*

An alternative way to prove Theorem 1 can be based on using Lagrange multipliers [9]. In order to show that $P^*_D(D)$ solves (OP), it remains to be proved that $P^*_D(D)$ is indeed contained in $C(V_{\overline{\alpha_1}}, \ldots, V_{\overline{\alpha_J}})$, i.e. that all belief values are positive.

Theorem 2. $P^*_D(D) \in C(V_{\overline{\alpha_1}}, \ldots, V_{\overline{\alpha_J}})$ *and, therefore, solves the optimization problem (OP).*

5 Generalized Projective Transformation

Alternatively to a relative frequency interpretation, it is popular to define a belief function by an expert's subjective opinion that an event might take place [3]. The motivation for this is that it is often not possible to get hold of some appropriate data to construct a belief function. Nevertheless, in more complex situations, it would be very hypothetical to assume that belief of a human being exactly matches a Dempster-Shafer belief function or a probability function. Even if an expert intends to represent her/his belief in terms of probability, in most non-trivial cases it would be far beyond her/his mental capacity to meet precisely all requirements. Nevertheless, probability functions have the advantage that they can be applied to a rational decision making process.

In fact, it has been shown by using the so-called Dutch book argument that when applying belief functions different to probability, a rational decision making process can lead to completely irrational decisions [12]. The Dutch-Book argument works when the decision maker incorrectly assumes that her/his belief is similar to probability and uses a decision making principle (e.g. maximizing expected value) that has been developed for probabilistic belief. The special case of the Dutch book argument applied to Dempster-Shafer belief is also discussed in the literature [13].

In order to address a more intuitive concept of belief and decision making, our intention is as follows. We continue with the optimization problems (OP) and (SOP) admitting arbitrary belief functions $Bel \in GB$. For this purpose, we proceed analogously to Section 4.

Definition 10. For $Bel \in GB$, the formula

$$P^*_{GB}(Bel) = \sum_{\overline{\alpha} \in \overline{At}} m_{P^*_{GB}(Bel)}(\overline{\alpha}) \cdot V_{\overline{\alpha}}$$

defines the *projective transformation* of Bel, where, for all $\overline{\alpha} \in \overline{At}$,

$$m_{P^*_{GB}(Bel)}(\overline{\alpha}) = \frac{1}{2^{J+n-2}} \left(J \sum_{\overline{\theta} \geq \overline{\alpha}} Bel(\overline{\theta}) - \sum_{\overline{\theta} \in \overline{SL}} \left| S_{\overline{\theta}} \right| Bel(\overline{\theta}) + 2^{J-2} \right).$$

Analogously to Lemma 1, it can be shown that $P^*_{GB}(Bel)$ is a member of the affine hull of valuations and, analogously to Theorem 1, we can prove that $P^*_{GB}(Bel)$ solves (SOP).

Lemma 3. $P^*_{GB}(Bel) \in A(V_{\overline{\alpha_1}}, \ldots, V_{\overline{\alpha_J}})$.

Next, it is to show that $P^*_{GB}(Bel)$ solves the (SOP). As previously mentioned, the (SOP) is meant to be re-defined in such a way that the Dempster-Shafer belief functions $D : \overline{SL} \to [0,1]$ is replaced by a belief function $Bel : \overline{SL} \to [0,1] \in GB$.

Theorem 3. *$P^*_{GB}(Bel)$ solves (SOP).*

Essentially, the projection function defined for general belief functions is exactly an extension of the projection functions for Dempster-Shafer belief. This is due to the fact that both are unique solutions of the (SOP).

Corollary 1. *Let $D : \overline{SL} \to [0,1]$ be a Dempster-Shafer belief function then $P^*_D(D) = P^*_{GB}(D)$.*

Unfortunately, as the following example demonstrates, it turns out that the generalized projective transformation $P^*_{GB}(Bel)$ does not necessarily solve the optimization task (OP).

Example 1. For a two-variable propositional language $L = \{p_1, p_2\}$, let $Bel : \overline{SL} \to [0,1]$ be given by

$$Bel(\overline{\theta}) = \begin{cases} 0 \text{ for } \overline{\theta} \geq \overline{\alpha}_1 \\ 1 \text{ otherwise,} \end{cases}$$

where $\overline{\alpha}_1 = \overline{p_1 \wedge p_2}$. For this belief function the projective transformation returns a negative value for $\overline{\alpha}_1$.

6 Concluding Remarks

In this paper, we explicitly found the probability function which minimizes the Euclidean distance to a given Dempster-Shafer belief function. This result was accomplished by re-formulating the optimization problem within the framework of the linear space of functions from $\overline{SL}$ to $\mathbb{R}$ and using methods from linear algebra. But it is not obvious how a geometric motivated transformation method for Dempster-Shafer belief can be justified.

The geometrical interpretation of distance seems to be more natural when the concept of belief is extended. For such a more general belief function, we solved the simplified optimization problem (SOP). However, it turned out the projective transformation does not always work in such settings.

Acknowledgements

The authors gratefully acknowledge partial support of the K*plus* Competence Center Program which is funded by the Austrian Government, the Province of Upper Austria and the Chamber of Commerce of Upper Austria. Many thanks also to Jeff Paris and Alena Vencovska for mathematical support and encouragement.

References

1. H. Laux, *Entscheidungstheorie*, fourth edition, (Springer, Heidelberg, 1997).
2. S. French and D. R. Insua, *Statistical Decision Theory*, first edition, (Arnold, London, 2000).
3. J. B. Paris, *The Uncertain Reasoner's Companion*, fourth edition, (Cambridge University Press, Cambridge, 1994).
4. I. Pratt, *Artificial Intelligence*, first edition, (MacMillan Press, Houndmills, 1994).
5. M. Bauer, "Approximation algorithms and decision making in the Dempster-Shafer theory of evidence — an empirical study", *Internat. J. Approx. Reason.* **17** (1997) 217–237.
6. P. Smets, "Constructing the pignistic probability function in a context of uncertainty", in *Uncertainty in Artificial Intelligence V*, eds. M. Herion, R. D. Shachter, L. N. Kanal, and J. Lemmer, (North-Holland, Amsterdam, 1990), pp. 29–39.
7. B. Tessem, "Approximations for efficient computation in the theory of evidence", *Artificial Intelligence* **61** (1993) 315–329.
8. F. Voorbraak, "A computationally efficient approximation of Dempster-Shafer theory", *Int. J. Man-Mach. Stud.* **30** (1989) 525–536.
9. T. Weiler, "Updating Towards Probability Belief", Master's thesis, University of Manchester, 1996.
10. G. Shafer, *A Mathematical Theory of Evidence*, (Princeton University Press, Princeton, NJ, 1976).
11. R. F. Curtain and A. J. Pritchard, *Functional Analysis in Modern Applied Mathematics*, (Academic Press, London, 1977).
12. B. de Finetti, *Theory of Probability*, first edition, volume I, (John Wiley & Sons, New York, 1974).
13. P. Smets, "No Dutch book can be built against the TBM even though update is not obtained by Bayes rule of conditioning", in *Proc. Workshop on Probabilistic Expert Systems*, eds. R. Scozzafava, SIS, 1993.