Chapter 13

Models Based on the Dempster-Shafer Theory of Evidence

Like Bayesian approaches, the Dempster-Shafer theory of evidence aims to model and quantify uncertainty by degrees of belief. But in contrast to Bayesian approaches it permits assignment of degrees of belief to sets of hypotheses rather than to hypotheses in isolation. The underlying idea is that the process of narrowing the hypothesis set with the collection of evidence is better represented in terms of this theory than in terms of Bayesian approaches. For this reason the theory can be viewed as an alternative to Bayesian modeling in probability theory.

In this chapter we mainly consider models and implementations that are based on the classical approach to evidence theory as proposed by Shafer [Shafer 1976]. His mathematical model is essentially based on the notion of belief functions and Dempster's rule of combination. This view, however, does not rely on probability theory: in this context it is also called the transferable belief model (see e.g. [Smets 1988]). In contrast to our treatment of belief functions there are approaches that use underlying probability spaces. Some of these concepts have already been considered earlier. Other semantics as well as their relations will be discussed in the concluding remarks of this chapter.

In Sect. 13.1 the original Dempster-Shafer theory of evidence is introduced. Section 13.2 then discusses knowledge representation aspects that arise with the use of this mathematical model. This set of approaches will be split up into three sections, on

- straightforward use of belief functions,
- belief functions in hierarchical hypothesis spaces, and
- belief propagation in Markov trees.

Section 13.3 examines proposals for performing uncertain reasoning where the knowledge of experts is assumed mainly to be represented in the form of explicit expert rules. In particular, the proposals of Ginsberg and Ishizuka et al. are examined.

A characteristic feature of the models looked at in Sect. 13.4 is that the algorithms for uncertain reasoning explicitly take the fact into account that the hypothesis space may be hierarchically ordered: the reason is that Shafer's...
"mathematical theory of evidence" is based on power sets. The elements of a power set can be related together by making use of subset and superset operators. In particular, Gordon and Shortliffe's extension to MYCIN's certainty factor approach and Yen's extension to the Dempster-Shafer theory are examined.

Section 13.5 mainly considers the proposals of Shafer, Shenoy, Logan, and Mellouli. A characteristic feature of these approaches is that explicit expert rules do not appear. Instead, the knowledge on dependencies is represented by the links of a network — the underlying idea is similar to Pearl's "Bayesian networks". As a concrete implementation the system MacEvidence is introduced.

**13.1 The Mathematical Theory of Evidence**

Let $\Omega$ be a finite "frame of discernment". Then each proposition concerning the identity of the "true value" $\omega$ can be represented by a subset of $\Omega$. Sets including only one element correspond to elementary propositions. If we further assume that at most one elementary proposition is true, our task is to find just this true proposition (e.g. a hypothesis on a ship's location) by means of evidence. We generally say that a proposition is true if it contains the true elementary proposition. The Dempster-Shafer theory of evidence is based on the assumption that the partial (personal) belief in a proposition corresponding to a subset $A$ of $\Omega$ is quantified by a single number $\text{Bel}(A)$ in the unit interval. It is further assumed that these quantities satisfy the following properties:

(i) $\text{Bel}(\emptyset) = 0$, \hspace{1cm} (13.1)
(ii) $\text{Bel}(\Omega) = 1$, \hspace{1cm} (13.2)
(iii) $\text{Bel}(A_1 \cup \ldots \cup A_n) \geq \sum_{i=1}^{n} \text{Bel}(A_i) - \sum_{i,j:1 \leq i < j \leq n} \text{Bel}(A_i \cap A_j)$

$$+ - \ldots + (-1)^{n+1} \text{Bel}(A_1 \cap \ldots \cap A_n)$$ \hspace{1cm} (13.3)

for every positive integer $n$ and every collection $A_1, \ldots, A_n$ of subsets of $\Omega$.

In evidence theory it is argued that degrees of belief should satisfy the above three properties. So, for instance, the belief in the disjunction of two propositions should at least contain the sum of the belief quantities allocated to each of it and reduced by the belief allocated to both. This motivates the above condition (13.3). However, the three requirements are of course not suitable for modeling purposes since, for example, the condition (13.3) is difficult to test. So one usually starts by introducing the concept of a basic belief mass (which is also called a basic probability assignment in [Shafer 1976]). Notice the formal relation to the notion of "mass distribution" in Definition 6.5.