

Updating the Evidence in the Dempster-Shafer Theory

Tomasz LUKASZEWSKI

*Institute of Computing Science, Poznan University of Technology
Piotrowo 3a, 60–965 Poznan, Poland
e-mail: luki@cs.put.poznan.pl*

Received: December 1998

Abstract. An algorithm for updating the evidence in the Dempster-Shafer theory is presented. The algorithm is based on an idea of indices. These indices are used to code the process of reasoning under uncertainty (the combination of evidence) using the Dempster-Shafer theory. The algorithm allows to carry out the reasoning with updating the evidence in much more efficient way than using the original Dempster-Shafer theory. Updating the evidence is necessary among others in systems that are based on changing data – adaptive and distributed reasoning systems.

Key words: reasoning under uncertainty, removing and changing pieces of evidence.

1. Introduction

The first tool developed for dealing with uncertainty was the probability theory (“the true logic for this world is the calculus of probabilities” – James Maxwell said in 1850). Initial results of applying probabilistic techniques in expert systems were promising. However, the existing systems did not scale up because of exponential numbers of probability values required in the full joint distribution (belief net algorithms were not known at that time). As a result between 1975 and 1988 a variety of alternatives were tried. The best known ones were default reasoning, rule-based approaches (certainty factor), fuzzy logic and Dempster-Shafer theory. These approaches deal with different aspects of uncertainty.

One of the most important aspects is the distinction between uncertainty and ignorance. Let us consider the following example. Given are three competing propositions (hypotheses): X , Y and Z . If we have no information we have to assert to these hypotheses some values of probability and their sum must be equal to 1 – the classical approach. The only reasonable solution is to assign to each hypothesis the same probability equal to $1/3$. If we do it we will not be able to make any distinction between this situation, when we have no knowledge, and the case when we know that all hypotheses are equally likely to happen. This distinction is possible to make using the Dempster-Shafer theory (Dempster, 1968; Klopotek, 1998; Shafer, 1976; Shafer and Logan, 1987; Shenoy and Shafer, 1986), which rejects one of the axioms of the Bayesian theory: $P(A) + P(\neg A) = 1$ for any proposition A . If we have no evidence at all, for or against A , then it is appropriate to assume that both degrees of belief for propositions A and $\neg A$ are equal to zero:

$P(A) + P(\neg A) = 0$. To model this the Dempster-Shafer theory needs two values that describe the degree of belief in A and in $\neg A$ respectively.

The main part of this theory is the algorithm of combining different pieces of evidence. It allows only to add new evidence. However, there are very often cases we would like not only add evidence but also *change* or even *remove* some pieces which were previously added. Because these processes are not defined in the original model we can not do it directly. The only possible way is to perform the whole process of combining all these pieces of evidence with the piece changed or without the one removed again. In cases it should be done regularly there is a very strong motivation for less time consuming solutions. This work presents some properties of the Dempster-Shafer theory that can be used to define these both aforementioned processes of changing and removing evidence pieces.

2. Dempster-Shafer Theory

Main Idea

Contrary to the classical probabilistic approach the Dempster-Shafer theory assigns to every proposition (hypothesis) *an interval*: [Belief, Plausibility], in which this degree of belief must lie, instead of a single degree of belief. **Belief** (usually denoted Bel) measures the strength of evidence in favour of a set of propositions. It ranges from 0 (indicating no evidence) to 1 (denoting certainty). **Plausibility** (Pl) is defined to be:

$$\text{Pl}(A) = 1 - \text{Bel}(\neg A). \quad (1)$$

It also ranges from 0 to 1 and measures the extent to which evidence in favour of $\neg A$ leaves space for belief in A . In particular, if we have certain evidence in favour of $\neg A$, then $\text{Bel}(\neg A)$ will be equal to 1 and $\text{Pl}(A)$ will be equal to 0. This tells us that the only possible value for $\text{Bel}(A)$ is also 0. In other words, Belief and Plausibility represent the lowest and the highest value of *possible* degree of belief in some proposition A .

This interval indicates not only our level of belief in some propositions but also the amount of information (evidence) we have. The width of this interval can be a good aid in deciding when we need to acquire more evidence. As evidence is accumulated this interval can be expected to shrink. As we can see, this interval approach makes it clear that we have no information when we start the process of reasoning ($\text{Bel} = 0, \text{Pl} = 1$), what is not possible in classical probabilistic approach.

Belief Function

So far, we have been talking intuitively about Believe as a measure of our belief in some proposition (hypothesis) given some evidence. Let us now define it more precisely. To do this, we need to start, just as with Bayes' theorem, with *an exhaustive universe of mutually exclusive hypotheses* – θ . It is called *the frame of discernment*. Our goal is to attach some

measure of belief to elements of θ . However, not all evidence supports single hypothesis. Often evidence supports sets of hypotheses (subsets of θ). So Dempster-Shafer theory introduces a *probability density function* (m) which is defined not just for elements of θ but for all subsets of it:

$$\begin{aligned} m : 2^\theta &\rightarrow [0, 1], \\ m[\emptyset] &= 0, \\ \sum_{A \subseteq \theta} m(A) &= 1. \end{aligned} \quad (2)$$

Having defined m , we can now define $\text{Bel}(A)$ for a set A as the sum of the values of m for A and for all of its subsets B :

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B). \quad (3)$$

Moreover we need a mechanism for performing the combination of different subsets and their probability density functions. Suppose we are given two sets X and Y (subsets of θ) and corresponding to them two probability density functions m_1 and m_2 . Let X be to which m_1 assigns a nonzero value and let Y be the corresponding set for m_2 . We define the *combination* of X and Y (m_1 and m_2) to be:

$$m_3(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) * m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) * m_2(Y)}. \quad (4)$$

EXAMPLE 1. Let us consider the following example. Given are four hypotheses: A, F, C, P so $\theta = \{A, F, C, P\}$. Let us assume that we have no information about how to choose among these hypotheses. So we start the diagnosis with $m_1(\theta) = 1$. Now suppose we acquire a piece of evidence that suggests at a level of 0.6 that the correct answer is in the set $\{A, F, C\}$. So we have $m_2(\{A, F, C\}) = 0.6$. Because the sum of m_2 for all subsets must be equal to 1, we conclude that $m_2(\theta) = 0.4$. We can compute the combination using the following table:

1st combination:

| | | |
|-------------------|--------------------------|---------------------|
| | $m_2(\{A, F, C\}) = 0.6$ | $m_2(\theta) = 0.4$ |
| $m_1(\theta) = 1$ | $m_3(\{A, F, C\}) = 0.6$ | $m_3(\theta) = 0.4$ |

The combination consists in generating all intersections of both probability density functions (m_1 and m_2) according to (4). Although it did not happen in this simple case, it is possible for the same subset to be derived in more than one way during the combination process. If that does occur, then in order to compute m for that subset, it is necessary to sum all values that are generated for the same subset (thus the summation sign in the

numerator of the combination formula 4). The denominator of this formula is equal to 1 except the situations when the intersection gives us the empty set – we will show it later.

Suppose we acquire another piece of evidence that $m_4(\{F, C, P\}) = 0.8$ and $m_4(\theta) = 0.2$. We can again compute the combination:

2nd combination:

| | | |
|--------------------------|---------------------------|---------------------------|
| | $m_4(\{F, C, P\}) = 0.8$ | $m_4(\theta) = 0.2$ |
| $m_3(\{A, F, C\}) = 0.6$ | $m_5(\{F, C\}) = 0.48$ | $m_5(\{A, F, C\}) = 0.12$ |
| $m_3(\theta) = 0.4$ | $m_5(\{F, C, P\}) = 0.32$ | $m_5(\theta) = 0.08$ |

A slightly more complex situation arises when some of the subsets generated during the combination process are empty. We assumed, that θ is exhaustive and the true value of any hypothesis must be contained in some nonempty subset of θ . If this situation happens, we must scale values of the probability density function according to (4). Let us add another piece of evidence: $m_6(\{A\}) = 0.75$ and $m_6(\theta) = 0.25$:

3rd combination:

| | | |
|---------------------------|--------------------------|----------------------------|
| | $m_6(\{A\}) = 0.75$ | $m_6(\theta) = 0.25$ |
| $m_5(\{F, C\}) = 0.48$ | $m_7(\emptyset) = 0.360$ | $m_7(\{F, C\}) = 0.120$ |
| $m_5(\{A, F, C\}) = 0.12$ | $m_7(\{A\}) = 0.090$ | $m_7(\{A, F, C\}) = 0.030$ |
| $m_5(\{F, C, P\}) = 0.32$ | $m_7(\emptyset) = 0.240$ | $m_7(\{F, C, P\}) = 0.080$ |
| $m_5(\theta) = 0.08$ | $m_7(\{A\}) = 0.060$ | $m_7(\theta) = 0.020$ |

As we can see, we have the aforementioned case, the subset $\{A\}$ is derived in more than one way. So the m value for this subset is equal to the sum of all values: $m_7(\{A\}) = 0.090 + 0.060 = 0.150$. Moreover, the degree of belief equal to 0.600 ($0.360 + 0.240 = 0.600$) is associated with \emptyset . We need to scale m -function values all remaining subsets by dividing them by 0.400 ($1 - 0.600 = 0.400$) – according to formula (4). The final result, after scaling, is the following: $m_7(\{F, C\}) = 0.300$, $m_7(\{A\}) = 0.375$, $m_7(\{A, F, C\}) = 0.075$, $m_7(\{F, C, P\}) = 0.200$, $m_7(\theta) = 0.050$.

Let us now calculate the values of Belief and Plausibility for each hypothesis:

$$\{A\} : [0.375, 0.500] \quad \{F\} : [0, 0.625] \quad \{C\} : [0, 0.625] \quad \{P\} : [0, 0.250].$$

As we can see, we have little knowledge about single hypotheses F and C . For these hypotheses intervals are relatively wide. Only for A and P we can say, that the level of belief was calculated precisely because their intervals are relatively narrow. Having this knowledge we can only say, that possible hypothesis with very high degree of belief is the set of three hypotheses: $\{A, F, C\}$:

$$\text{Bel}(\{A, F, C\}) = 0.075 + 0.300 + 0.375 = 0.750 \quad \text{Pl}(\{A, F, C\}) = 1 - 0 = 1.$$

Interval for the subset $\{A, F, C\}$ is the following: $[0.750, 1]$

To reduce the possible hypothesis to two or even one hypothesis, we need additional pieces of evidence.

Let us add another piece of evidence: $m_8(\{A, F\}) = 0.8$ and $m_8(\theta) = 0.2$:

4th combination:

| | | |
|----------------------------|-------------------------|----------------------------|
| | $m_8(\{A, F\}) = 0.75$ | $m_8(\theta) = 0.25$ |
| $m_7(\{F, C\}) = 0.300$ | $m_9(\{F\}) = 0.240$ | $m_9(\{F, C\}) = 0.060$ |
| $m_7(\{A\}) = 0.375$ | $m_9(\{A\}) = 0.300$ | $m_9(\{A\}) = 0.075$ |
| $m_7(\{A, F, C\}) = 0.075$ | $m_9(\{A, F\}) = 0.060$ | $m_9(\{A, F, C\}) = 0.015$ |
| $m_7(\{F, C, P\}) = 0.200$ | $m_9(\{A\}) = 0.160$ | $m_9(\{F, C, P\}) = 0.040$ |
| $m_7(\theta) = 0.050$ | $m_9(\{A, F\}) = 0.040$ | $m_9(\{\theta\}) = 0.010$ |

Again, some subsets are derived in more than one way. So the m -function values according to formula (4) are the following: $m_9(\{A\}) = 0.535$, $m_9(\{F\}) = 0.240$, $m_9(\{A, F\}) = 0.100$, $m_9(\{F, C\}) = 0.060$, $m_9(\{A, F, C\}) = 0.015$, $m_9(\{F, C, P\}) = 0.040$, $m_9(\{\theta\}) = 0.010$.

Let us now calculate the values of Belief and Plausibility for each hypothesis:

$$\{A\} : [0.535, 0.660] \quad \{F\} : [0.240, 0.465] \quad \{C\} : [0, 0.125] \quad \{P\} : [0, 0.050].$$

Comparing to the previous result we can say, that all intervals are narrower and possible diagnosis with very high degree of possibility consists in two hypotheses A and F :

$$\begin{aligned} \text{Bel}(\{A, F\}) &= 0.775 \\ \text{Pl}(\{A, F\}) &= 1 - \text{Bel}(\{C, P\}) = 1 - 0 = 1 \end{aligned}$$

3. Problem Formulation

The main part of the Dempster-Shafer theory is the process of combining different pieces of evidence. We can interpret this process as knowledge updating. As a result we obtain new (updated) subsets of propositions (hypotheses) with new (updated) values of the probability density function (m). This process may be performed as far as we are able to obtain new knowledge and update the result.

Let us consider the following situation. We obtained some pieces of evidence and updated subsets of hypotheses and their values of the probability density function (m). It may happen that we would like to change parameters of some previously added piece of evidence or even remove it. There are many important reasons for such actions. Firstly, after obtaining some piece of evidence and knowledge updating we may get information that this piece of evidence have to be changed because the evaluation of its parameters (sets of hypotheses or the value of the probability density function – m) was made not precisely. The correct new values are given and we have to calculate the correct result

again. Sometimes, despite the fact that the evidence parameters were estimated precisely, the surrounding world may change and we also have to change or even remove this piece of evidence. Thirdly, we may add some piece of evidence to see what happens if we do it (some kind of experiment). Then other pieces are added. After that this additional piece of evidence should be changed (to continue the experiment) or removed (to end this experiment).

Such a change of evidence parameters may happen more than once. For example it is possible in adaptive systems whose natural feature is a multiple testing of some values. Such testing may change degrees of belief in evidences, what may cause to activate a reasoning process again. Sometimes all values of belief degrees coming from this testing process are valid, but in most cases only the last value, the newest one is correct and all the previous ones must be removed (detailed discussion is presented in Lukaszewski, 1998).

We have presented only main reasons for changing and removing some pieces of knowledge. Let us concentrate on the main problem – how to change or remove a piece of evidence in the Dempster-Shafer theory. The simplest solution is to store all pieces. Each time we change parameters of some piece or remove it we have to combine *again* all pieces respectively with this piece of evidence changed or without the one cancelled. The computational complexity of this solution is rather high. Having n pieces of evidence we have to make $n - 1$ combinations even if we change (or remove) only one piece. In cases it should be done regularly there is a very strong motivation for less time consuming solutions.

4. Introduction of Indices

We will present other idea of removing and changing pieces of evidence which has much less computational complexity than the aforementioned one. Firstly, we will add some extension to the Dempster-Shafer theory, and use it to define these processes.

Let us come back to our example of evidence pieces combination that is presented partially in Fig. 1. Each combination consists in finding an intersection of “current” subsets (subsets of hypotheses) and “new” subsets. As a result we obtain “result” subsets:

$$\text{“current” subsets} \oplus \text{“new” subsets} \Rightarrow \text{“result” subsets.}$$

At the beginning we have only one subset – θ . Moreover, we assume that θ always belongs to “new” subsets. Combination of any “current” subset and θ gives us a “result” subset which is the same as the “current” one. We can say, that θ plays the same role in the combination process as “0” in addition or “1” in multiplication:

$$\text{“current” subset} \oplus \theta \Rightarrow \text{“current” subset.}$$

This is very important remark, because it is the basis for our further considerations. Of course probability density function (m) may be different for this “current” subset and

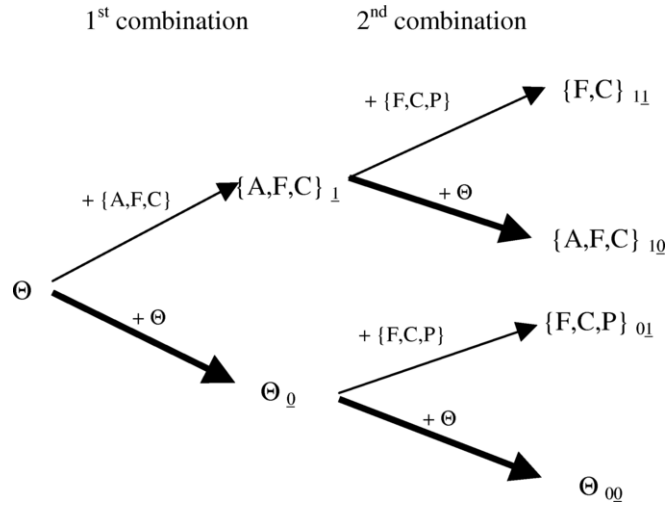


Fig. 1. The idea of the indices.

“result” subset, but the elements of these subsets are the same. The processes in which θ is a “new” subset are presented in Fig. 1 with bold arrows.

Our main extension of the Dempster-Shafer theory are *indices*. We add to every “result” subset index 0 or 1:

- *index 0* is added to “result” subsets when the “new” subset is θ . So the “result” subset is the same as the “current” one;
- *index 1* is added to “result” subsets when the “new” subset is not θ . So the “result” subset is different than the “current” one (when “current” and “new” subsets are equal the “result” subset is the same as these two ones – it is an exception).

These indices create lists which length depends on the number of combinations – see Fig. 1. The last added indices, the underlined ones, are at the end of these lists.

5. Idea of Solution

In the previous paragraph we introduced the idea of indices. We use them to present the idea of removing and changing pieces of evidence in the Dempster-Shafer theory. Having defined the removing process we can change any piece of evidence by removing the old piece of evidence and adding the correct piece of evidence.

To simplify our considerations we divide *the problem of evidence pieces removing* into two cases – without empty sets and with empty sets. For each case we will consider the same two subproblems:

A. Which subsets should be removed?

B. How to update the values of the m -function of all the not removed subsets?

Moreover, the process of evidence pieces removing is called *the combination cancelling*.

5.1. Cases without Empty Sets (\emptyset)

Let us concentrate on the first subproblem – which subsets should be removed. If the last combination should be cancelled, we can restore the state before this combination was made from a memory. We can also use our indices – subsets that *were not changed* during this combination (with index 0 at the end) should not be removed. Subsets that were created during this cancelled combination (with index 1 at the end) should be removed. If the cancelled combination is not the last one, we have to leave all subsets that are *the successors of the subsets not changed* during this combination (with index 0 at the position related to this combination). All subsets that are the successors of the subsets created during this combination (with index 1 at the position related to this combination) should be removed. We carry out these operations only for these successors that are the “result” subsets of the last combination – only for “leaves” in the combination tree.

A. Which subsets should be removed?

Remove all the subsets with index 1 at the position that is related to this cancelled combination

Let us look at our example – Fig. 2. After the second combination we have the following subsets:

$$m_5(\{F, C\}_{11}) = 0.48, m_5(\{A, F, C\}_{10}) = 0.12, \\ m_5(\{F, C, P\}_{01}) = 0.32, m_5(\theta_{00}) = 0.08.$$

We cancel the first combination. According to our rule, we have to remove all the subsets that have index 1 at the first position. We obtain the following subsets:

$$m_5(\{F, C, P\}_{01}) = 0.32, m_5(\theta_{00}) = 0.08.$$

We have found the solution for the first subproblem – which subsets should be removed. However, the second subproblem still exists – how to update the m -function values for all the not removed subsets. This cancelled combination *changed* all the m -function values of all the subsets not changed by this combination (combined with θ). This change relies on *multiplication* of all the m -function values of these subsets by the value assigned to the θ – one of the “new” subsets of this combination. This change is *propagated* to all the *successors* of these multiplied subsets. These successors are all the

subsets with index 0 at the position related to this cancelled combination. All the *not removed subsets* are the successors of these multiplied subsets. Cancelling the effect of this multiplication, we have to *divide* all the *m*-function values of all the not removed subsets by the value assigned to this θ .

B. How to update the values of the m-function of all the not removed subsets?

Divide all the m-function values of all the not removed subsets by the m-function value assigned to the “new” subset θ of the combination cancelled.

Let us look at our example. The *m*-function value assigned to the “new” subset θ in the cancelled combination is equal to 0.4. So, we have to divide all the *m*-function values of the not-removed subsets by this value. We obtain the following subsets:

$$m_5(\{F, C, P\}_{01}) = 0.8, m_5(\theta_{00}) = 0.2.$$

This is the correct result after the cancelling of the first combination. We can verify it carrying out only the second combination (omitting the first cancelled combination):

2nd combination

| | | |
|-------------------|--------------------------|---------------------|
| | $m_4(\{F, C, P\}) = 0.8$ | $m_4(\theta) = 0.2$ |
| $m_0(\theta) = 1$ | $m_5(\{F, C, P\}) = 0.8$ | $m_5(\theta) = 0.2$ |

As we can see, the results are the same.

5.2. Cases with Empty Sets (\emptyset)

In the previous section we considered only the case where each “result” subset was a non-empty set. Now we consider more general case that allows \emptyset to be created as a

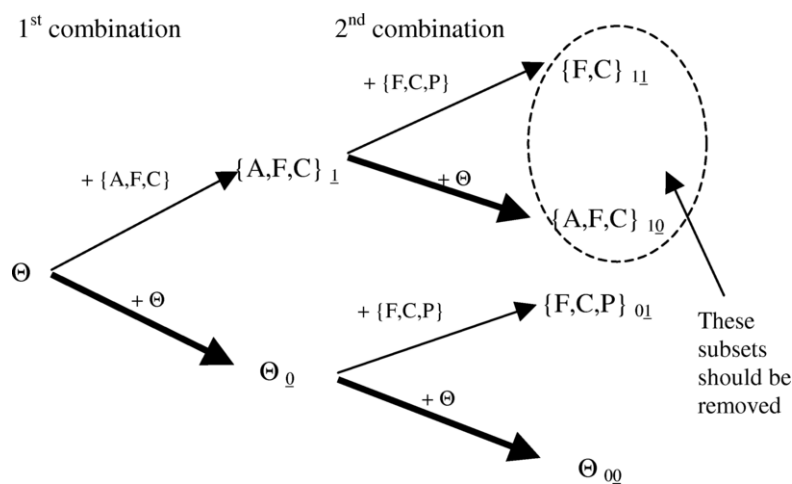


Fig. 2. The idea of the combination cancelling without empty sets.

“result” subset. Contrary to the Section 5.1, the term subsets will mean not only non-empty subsets but also empty sets.

We divide the process of the cancelling of the combination into the same two sub-problems as in the previous section. The first subproblem is solved in the same way (see Section 5.1).

A. Which subsets should be removed?

Remove all the subsets (all the non-empty subsets and all the empty sets) with index l at the position that is related to the combination cancelled.

The problem of updating the m -function values is much more difficult this time. As we remember every time an empty set is created we scale all the m -function values of all the non-empty subsets created during the combination according to formula (4). It may happen that while cancelling some combination we have to remove some empty sets. These empty sets may have been created by this combination cancelled or are the successors of some non-empty subsets created during this process cancelled. In both cases we have to remove these empty sets (Step A). These removed empty sets are used in the scaling of the m -function values of subsets (non-empty subsets and empty sets) so we have to *cancel the scaling* made by these removed empty sets. We can cancel the scaling made by these empty sets removed in the following way: *cancel the scaling made by all the empty sets* and *scale by the empty sets not removed*. Moreover, we have to remember about *cancelling the scaling made by θ* – one of the “new” subsets of this cancelled combination (see the case without \odot).

Summarising, cancelling any combination we have to carry out the following processes: removing subsets, cancelling the scaling made by θ , cancelling the scaling made by all empty sets and the scaling made by the empty sets not removed. These processes can not be carried out in an arbitrary order. The idea of the solution is the following: before we remove some empty sets we have to cancel the scaling made by them (cancel the scaling made by all the empty sets and scale by the empty sets not removed). The problem is when we can cancel the scaling made by θ ? The cancelling of the scaling made by all the empty sets should be performed in the opposite direction than the process cancelled – from the “leaves” of the tree to the combination cancelled. The scaling by the empty sets not removed should be performed in normal direction – from the combination cancelled to the “leaves” of the tree. Between these two processes we should cancel the scaling made by θ .

Let us now look closer at these processes and their description. We start our operations from the “leaves” of the tree.

- Firstly, moving from the “leaves” (the last combination) towards the combination cancelled we have to *cancel the scaling made by the empty sets*. We have to do it for each group of empty sets created by the same combination C . Each group of empty sets scales the m -function values of all the non-empty subsets created by this combination C . This scaling relies on division of all the m -function values of

these non-empty subsets by the value $1 - SUM$ where SUM is the sum of all the m -function values of all the empty sets created by combination C . This scaling is propagated to all the successors of these scaled non-empty subsets. These successors are all the subsets (the non-empty subsets and the empty sets) with any index (0 or 1) at the position related to this combination C . All the “result” subsets (all the “leaves”) are the successors of these scaled non-empty subsets. Moreover, all the empty sets created by all the combinations made later than C are also successors of these scaled non-empty subsets. Cancelling the effect of this scaling, we have to multiply all the m -function values of all the successors by the value $1 - SUM$.

- We have cancelled scaling made by empty sets created by all the combinations made later than the cancelled combination. We can now *remove subsets* created by this cancelled combination (cut some “branches” of the tree). We use the idea of indices again. All the subsets (all the non-empty subsets and all the empty sets) with index 1 at the position that is related to this cancelled combination should be removed.
- The previous process leaves all these subsets that are not changed by this cancelled combination (subsets combined with θ – one of the “new” subsets of this combination) and their successors. This cancelled combination changed all the m -function values of all these subsets. This change relies on multiplication of all the m -function values of these subsets by the value assigned to θ . This change is propagated to all the successors of these multiplied subsets. These successors are all the subsets (the non-empty subsets and the empty sets) with index 0 at the position related to the cancelled combination. All the non-empty subsets not removed (all the “leaves”) are the successors of these multiplied subsets. Moreover, all the not removed empty sets created by combinations made later than the cancelled combination are also successors of these multiplied subsets. We *cancel the scaling made by the θ* . While cancelling the effect of this scaling, we have to divide all the m -function values of all the successors of the scaled subsets by the value assigned to θ .
- In the first process we cancelled the scaling made by the empty sets. Some of these empty sets are not removed by the removing process. Moving from the cancelled combination towards the “leaves” not removed (the last combination) we have to *scale the subsets by the empty sets not removed*. We have to do it for each group of not removed empty sets created by the same combination C . Each group of empty sets scales the m -function values of all the non-empty subsets created by this combination C . This scaling relies on division of all the m -function values of these non-empty subsets by the value $1 - SUM$ where SUM is the sum of all the m -function values of all the empty sets created by combination C . This change is propagated to all the successors of these scaled subsets. These successors are all the subsets (the non-empty subsets and the empty sets) with any index (0 or 1) at the position related to combination C . All the “result” subsets not removed (all the “leaves”) are the successors of these scaled subsets. Moreover, all the empty sets

created by all the combinations made later than C are also successors of these scaled subsets. Carrying out subsets scaling, we have to divide all the m -function values of all the successors of these subsets by the value $1 - SUM$.

Summarising:

AB1. Cancel the scaling of the m -function values of non-empty subsets and empty sets:

For each group of empty sets created during the same combination C , starting from the last combination and ending with the cancelled combination:

- **find** all the successors of all the subsets scaled by this combination C – all the subsets (the non-empty subsets and the empty sets) with any index (0 or 1) at the position related to this combination;
- multiply the m -function values of these successors **by the value** $1 - SUM$ where SUM is the sum of all the m -function values of all the empty sets created by this combination.

AB2. Remove all the non-empty subsets and all the empty sets with index 1 at the position that is related to the combination cancelled.

AB3. Divide the m -function values of all the non-empty subsets and all these empty sets that have index 0 at the position related to the combination cancelled by the value assigned to the “new” subset θ of the combination cancelled.

AB4. Scale the m -function values of the non-empty subsets not removed and empty sets not removed too:

For each group of empty sets created during the same combination C , starting from the cancelled combination and ending with the last combination:

- **find** all the successors of all the subsets scaled by combination C – all the subsets (the non-empty subsets and the empty sets) with any index (0 or 1) at the position related to this combination;
- **divide** the m -function values of the successors **by the value** $1 - SUM$ where SUM is the sum of all the m -function values of all the empty sets created by this combination.

Let us now look at our example – Fig. 3. It presents the state of the problem after the third combination of our example. The m -function values of all the “result” subsets after scaling by the empty sets according to formula (4) are the following:

$$m_7(\{F, C\}) = 0.3, \quad m_7(\{A\}) = 0.225, \quad m_7(\{A, F, C\}) = 0.075, \\ m_7(\{F, C, P\}) = 0.2, \quad m_7(\{A\}) = 0.150, \quad m_7(\theta) = 0.05.$$

Let us cancel the first combination.

Step AB1 – we have to cancel scaling made by the empty sets. There is only one combination carried out after the combination cancelled that created any empty set – the

last combination. So we have to multiply all non-empty subsets by the value $1 - SUM$ where SUM is equal to 0.6. We do it only for the subsets that will not be removed. Values before multiplying:

$$m_7(\{F, C, P\}) = 0.200, m_7(\{A\}) = 0.150, m_7(\theta) = 0.050.$$

Values after multiplying:

$$m_7(\emptyset) = 0.240, m_7(\{F, C, P\}) = 0.080, m_7(\{A\}) = 0.060, m_7(\theta) = 0.020.$$

Step AB2 – we remove the non-empty subsets and the empty sets that have index 1 at the first position – Fig. 3.

Step AB3 – we have to divide all the m -function values of all the non-empty subsets and all the empty sets (they have index 0 at the first position) by the value assigned to the “new” subset θ of this cancelled combination. In this combination the m -function value assigned to θ is equal to 0.4. So, we have to divide the m -function values of all the non-empty subsets and the empty sets by this value. Values before dividing:

$$m_7(\emptyset) = 0.240, m_7(\{F, C, P\}) = 0.080, m_7(\{A\}) = 0.060, m_7(\theta) = 0.020.$$

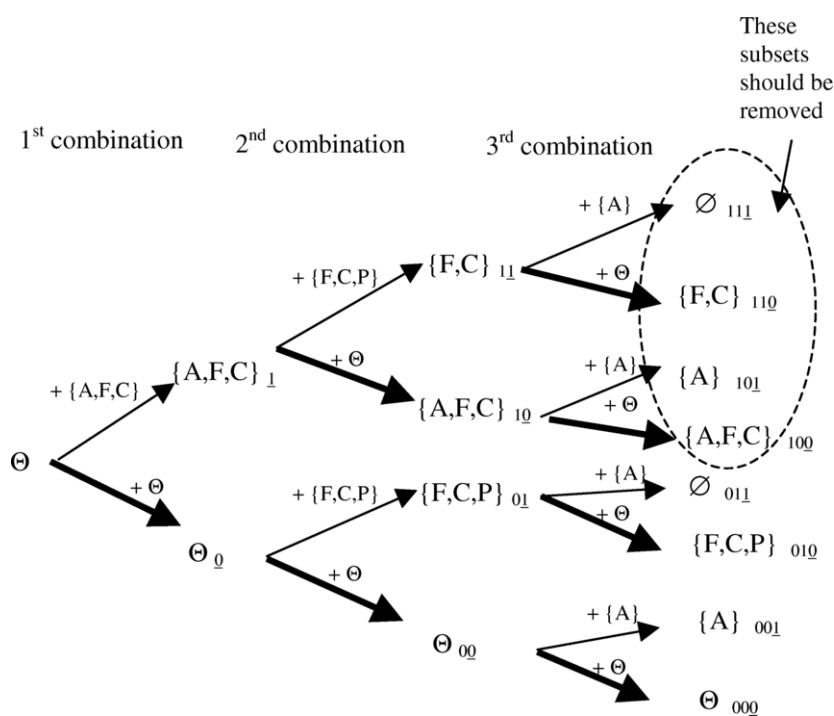


Fig. 3. The idea of the combination cancelling with empty sets.

Values after dividing:

$$m_7(\emptyset) = 0.6, m_7(\{F, C, P\}) = 0.2, m_7(\{A\}) = 0.15, m_7(\theta) = 0.05.$$

Step AB4 – There is only one combination carried out after the cancelled combination that created any empty set – the last combination. So we have to divide all the non-empty subsets with any index at the position of this combination by the value $1 - SUM$ where SUM is equal to 0.6. Values after dividing:

$$m_7(\{F, C, P\}) = 0.5, m_7(\{A\}) = 0.375, m_7(\theta) = 0.125.$$

This is the correct result. It can be verified by carrying out only the second and the third combination.

2nd combination:

| | | |
|-------------------|--------------------------|---------------------|
| | $m_4(\{F, C, P\}) = 0.8$ | $m_4(\theta) = 0.2$ |
| $m_0(\theta) = 1$ | $m_5(\{F, C, P\}) = 0.8$ | $m_5(\theta) = 0.2$ |

3rd combination:

| | | |
|--------------------------|------------------------|--------------------------|
| | $m_6(\{A\}) = 0.75$ | $m_6(\theta) = 0.25$ |
| $m_5(\{F, C, P\}) = 0.8$ | $m_7(\emptyset) = 0.6$ | $m_7(\{F, C, P\}) = 0.2$ |
| $m_5(\theta) = 0.2$ | $m_7(\{A\}) = 0.15$ | $m_7(\theta) = 0.05$ |

Values after scaling by the empty set (after dividing by 0.4):

$$m_7(\{F, C, P\}) = 0.5, m_7(\{A\}) = 0.375, m_7(\theta) = 0.125.$$

As we can see, the results are the same.

6. Conclusions

We have presented two solutions of the evidence pieces removing (the cancelling of the combination). These solutions allow removing or changing pieces of evidence in much more effective way comparing to the idea of carrying out of all combinations again except the removed or changed one. This improving of effectiveness is possible by defining the process of removing that have to be much more effective than the substitution of this process by repeatedly carrying out combinations of all not removed pieces of evidence.

These solutions base on some properties of the combination process in the Dempster-Shafer theory. It is possible to define the removing process representing all the combinations using the combination tree. This tree can be coded with indices. We proved that storing only the last result (as it is in the Dempster-Shafer theory) and indices attached to them, and all the empty sets with their indices we can cancel any combination much more effective than without this removing process.

There is very strong practical motivation for defining such the effective process of evidence pieces removing: adaptive reasoning systems, systems that allow changes in the data, experimenting with data (add data to see what happens and remove this data).

References

- Dempster, A.P. (1968). A generalisation of Bayesian inference. *J. Roy. Stat. Soc.*, **B 30**, 205–247.
- Grzymala-Busse, J. W. (1991). *Managing Uncertainty in Expert Systems*. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Kłopotek, M.A. (1998). *Metody Identyfikacji i Interpretacji Struktur Rozkładów Przekonań w Teorii Dempstera-Shafera*. Instytut Podstaw Informatyki PAN (in Polish).
- Lukaszewski, T. (1998). On reasoning under uncertainty in adaptive, distributed and multiagent systems. In *Proceedings of the Seventh International Symposium on Intelligent Information Systems – IIS'98*, Malbork-Poland, pp. 227–230.
- Rich, E. (1983). *Artificial Intelligence*, McGraw-Hill, New York.
- Russell, S.J., P. Norvig (1995). *Artificial Intelligence a Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.
- Shafer, G., R. Logan (1987). Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, **33**, 271–298.
- Shenoy, P.P., G. Shafer (1986). Propagation belief functions with local computation. *IEEE Expert*, **1**(3), 43–52.
- Yager, R.R., J. Kacprzyk, M. Ferdrizzi (1994). *Advances in the Dempster-Shafer Theory of Evidence*, J. Willey & Sons, ISBN: 0471552488.

T. Lukaszewski is an assistant in the Institute of Computing Science, Poznan University of Technology, Poland. He received his M.Sc. degree in 1994 and now he is a Ph.D. student in Computing Science at Poznan University of Technology. His research interests focus on expert systems and reasoning under uncertainty.

Informacijos atnaujinimas Dempsterio ir Šaferio teorijoje

Tomasz LUKASZEWSKI

Straipsnyje pateiktas algoritmas informacijai atnaujinti Dempsterio ir Šaferio teorijoje. Algoritmas grindžiamas indeksų naudojimu. Indeksais koduojamas Dempsterio ir Šaferio teorija pagrįstas samprotavimų neapibrėžtumo sąlygomis procesas. Jis leidžia, palyginti su klasikiniiais Dempsterio ir Šaferio teorijos metodais, atnaujinti informaciją daug efektyviau. Tai ypač svarbu kuriant sistemas, kuriose duomenys nuolat kinta, pavyzdžiui, adaptyviųjų arba išskirstytų samprotavimų sistemas.