

Evidence combination for incremental decision-making processes

Ghita Berrada¹, Maurice van Keulen¹ and Ander de Keijzer²

¹Faculty EEMCS, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
Email: {g.berrada,m.vankeulen}@utwente.nl

²University of Applied Sciences Utrecht, PO Box 13102, 3507 LC Utrecht, The Netherlands
Email: ander.dekeijzer@hu.nl

Abstract. The establishment of a medical diagnosis is an incremental process highly fraught with uncertainty. At each step of this painstaking process, it may be beneficial to be able to quantify the uncertainty linked to the diagnosis and steadily update the uncertainty estimation using available sources of information, for example user feedback, as they become available. Using the example of medical data in general and EEG data in particular, we show what types of evidence can affect discrete variables such as a medical diagnosis and build a simple and computationally efficient evidence combination model based on the Dempster-Shafer theory.

Keywords: uncertain databases; incremental decision-making processes; evidence combination; user feedback; Dempster-Shafer evidence theory

1. Introduction

Reaching an accurate diagnosis as soon as possible is key to treating patients' ailments effectively. The case of teenager Rory Staunton who died of sepsis a few days after having been diagnosed with a benign flu at the ER and sent back home illustrates how critical it is to reach a timely accurate diagnosis (Dwyer; 2012b).

Though quite extreme, Rory Staunton's case is not an isolated case of misdiagnosis and is just one particularly striking example of the many errors of diagnosis that occur in the healthcare system. In fact, the prevalence of misdiagnoses is estimated to be up to 15% in most areas of medicine (Eta S. Berner; 2008) and a study of physician-reported diagnosis errors (Schiff et al.; 2009) finds that 28% of the misdiagnoses are major (i.e resulting in death, permanent disability, or near life-threatening event) and 41% moderate (i.e resulting in short-term morbidity, increased length of stay, higher level of care or invasive procedure). Even common

conditions such as pneumonia, asthma or breast cancer are routinely being misdiagnosed especially when the presenting symptoms are atypical (Kostopoulou et al.; 2008; Singh et al.; 2013). Missed diagnoses alone account for 40,000 to 80,000 preventable deaths annually in the US (Leape et al.; 2002).

Not only is reaching a correct diagnosis quite a challenge but the process leading to a reliable diagnosis is often rather lengthy as it may involve many patient consultations and referrals to other clinicians as well as various tests and scans. The study of physician-reported errors of diagnosis cited earlier (Schiff et al.; 2009) also finds that 32% of the cases are due to clinician assessment errors. This figure coupled with the misdiagnosis prevalence figure suggests that the misdiagnosis problem is far from being an individual clinician's problem but rather a systemic problem. Or in the words of (Leape; 2000):

“Errors are rarely due to personal failings, inadequacies, and carelessness. Rather, they result from defects in the design and conditions of medical work that lead careful, competent, caring physicians and nurses to make mistakes that are often no different from the simple mistakes people make every day, but which can have devastating consequences for patients. Errors result from faulty systems not from faulty people, so it is the systems that must be fixed.”

And since “to err is human”, systems must be designed in such a way as to make errors hard to commit or to quote the Institute of Medicine landmark report on medical errors (Kohn et al.; 2000), “Human beings, in all lines of work, make errors. Errors can be prevented by designing systems that make it hard for people to do the wrong thing and easy for people to do the right thing”.

As such, instead of focusing on assigning blame to physicians/nurses, which does little to fix systemic problems and only ensures that preventable errors are made again and again, it would be more beneficial to try and identify the factors that contribute to making it difficult to reach a correct diagnosis in a timely fashion or that lead to erroneous/delayed diagnoses. Some of these factors include the following:

1. *Only finite resources can be allocated to the diagnosis process.* Even with the best of intentions, a doctor can only devote a limited amount of time and energy to each patient under his/her care. Furthermore, to decrease costs and minimize patient discomfort, the number of tests performed to reach a diagnosis needs to be kept as low as possible. There is also only a fixed (small) number of specialists and doctors are encouraged to make as few referrals as possible. And obviously, even with the best will in the world, doctors, being human, have only a limited amount of memory and knowledge to draw on to make diagnoses.
2. *The diagnosis process is highly dependent on the accuracy of the initial diagnosis hypothesis.*
The patient is at best an unreliable source of information: he/she may give vague information or omit crucial clues that he/she feels are not significant. Moreover, patient history, which may shed a different light on some non-specific presenting symptoms, is usually fragmented and scattered across different healthcare institutions that don't necessarily share information between themselves. Therefore, the first patient consultation only provides incomplete and highly noisy information on which the clinician needs to rely to form his/her initial hypothesis and order the relevant tests and/or referrals required to unearth further relevant diagnostic clues and evidence.
3. *Finding the right clues and evidence for a diagnosis is comparable to searching for a needle in a haystack.* Patient consultations/referrals and medical tests generate a huge amount of data that may or may not contain the needed

diagnostic clues (depending on whether the right hypotheses were tested for) and is mostly irrelevant for the diagnosis task at hand. There is at the same time too much and too little data available.

4. *Medical knowledge is fragmented.* Due to the sheer amount of medical knowledge accumulated through time, no single clinician can know all there is to know inevitably leading to a spread of expertise and knowledge between clinicians.
5. *The diagnosis process is fragmented.* The patient often has to consult several doctors and undergo several tests. This is a direct consequence of the fragmentation of knowledge and expertise driven by the massive amount of medical knowledge available.

As a result of these factors, the potential of communication breakdowns between healthcare agents and crucial information being lost in the process increases as does the likelihood of clinicians falling back on potentially harmful cognitive biases.

Rory Staunton's case (Dwyer; 2012a,b) is a case in point of how a *breakdown in communication between healthcare agents* can result in erroneous diagnosis and inadequate care. In Rory's case, because critical blood tests' results had not been communicated to the clinicians in charge and important observations by the pediatrician had gone missing from the charts, each of the parties involved in Rory's care had only access to fragments of information on his condition, each of which could be construed to result from something other than sepsis. Taken in conjunction, all of Rory's symptoms and tests pointed clearly to sepsis but the flu diagnosis was not outlandish given the information available to the first ER practitioner at the time of diagnosis. This case perfectly exemplifies the situation described in the tale of the blind men and the elephant:¹ While the conclusions of the blind men might have been right individually, taken as a whole, they missed the target completely.

Rory's case also illustrates another source of diagnostic failure: *cognitive biases* (Segal; 2007; Groopman; 2007; Croskerry; 2003). Two biases were in play in Rory's case: *representativeness bias* and *premature closure*. The representativeness bias is the tendency for a clinician to look for prototypical manifestations of a condition, thus rejecting the possibility of a particular condition if the presenting symptoms are atypical or if the patient is not part of the stereotypical population in which the condition occurs. In Rory's case, the possibility of sepsis was not considered because sepsis rarely occurs in teenagers. Premature closure is the tendency for a clinician to decide on a diagnosis to the exclusion of others too soon in the process and before it has been fully verified by tests for example. In Rory's case, there was no indication that the attending clinicians had considered another diagnosis than flu. Cognitive biases are essentially reasoning shortcuts and heuristics that come into play when doctors try to cope with time and resources constraints. Cognitive biases are necessary and time-saving but may result in wrong, missed or delayed diagnoses.

In addition to the representativeness and premature closure biases, a few more

¹ The story, which has several versions (see http://en.wikipedia.org/wiki/Blind_men_and_an_elephant), basically goes like this: some blind men or men in a dark room touch different parts of an elephant trying to figure out what they are touching. Depending on which part they touch (trunk, leg ,etc.), they come to completely different conclusions.

biases may become problematic if applied indiscriminately: zebra retreat, availability and confirmation biases and diagnosis momentum. A clinician usually follows the well-known maxim "If you hear hoofbeats, think horses, not *zebras*"², i.e. a clinician tends to only consider the most common diagnoses that fit the symptoms exhibited by the patient. Failing to consider a zebra even when likely based on the clinical findings is called zebra retreat. Taken as a whole, zebras are not so uncommon: 8% of the US population (ie about 25 million) are estimated to be affected by one of the approximately 7000 zebras.

The *confirmation bias* can be especially harmful when associated with premature closure: it is the tendency for a clinician to look for the evidence, even not present, that supports his/her preferred diagnosis and dismiss the existing evidence that disproves it. The confirmation bias can cause the clinician and patient to go on a wild-goose chase and delay the diagnosis especially if it intervenes while forming the initial diagnosis hypothesis since the whole process hinges on that initial hypothesis.

The *availability bias* and *diagnosis momentum* may be consequences of the fragmentation of expertise. The availability bias is the tendency of a clinician to reach for the most easily recalled diagnosis that fits the clinical findings, whether the clinician recalls that diagnosis because he has more expertise on it or because he has recently encountered it. Diagnosis momentum is the fact for a diagnosis to stick in particular because it keeps being passed on by all the agents and intermediaries involved in the diagnosis process. Diagnosis momentum also makes reaching a correct diagnosis during the initial patient consultation critical.

We contend that, to obviate or at least mitigate the aforementioned factors leading to misdiagnosis, different forms of computer-support could be used to assist clinicians in their decision-making task. One form of computer-support is *(semi-)automatic interpretation* of tests and scans, such as the semi-automated EEG interpretation performed by (Lodder; 2014; Cloostermans et al.; 2009). A different form of computer-support, which is the focus of this paper, is *evidence combination*. We view medical diagnosis primarily as an incremental process where at each point in time, there is an intermediary diagnosis based on 'what is known so far': symptoms and clinical evidence from consultations, tests/measurements/scans, interpretations thereof by experts, second opinions/feedback of experts on other experts' interpretations/conclusions, etc. Each interpretation, opinion, or feedback is a piece of evidence that is combined to produce a well-weighted intermediary diagnosis.

1.1. Contribution

The model presented in this paper

1. provides a combined diagnosis constructed from all known evidence and opinions known so far at a point in time,
2. is based on Dempster-Shafer theory,
3. allows the inclusion of evidence that stems from the processing of historic

² Zebra is the medical slang for rare or surprising diagnosis. For examples of zebras, see the Medical Mysteries column in the Washington Post: http://www.washingtonpost.com/linksets/medical-mysteries/2010/07/06/ABELr7D_linkset.html.

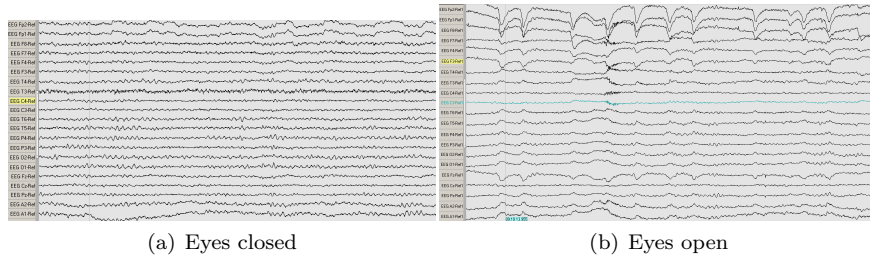


Fig. 1. Normal EEGs in different contexts

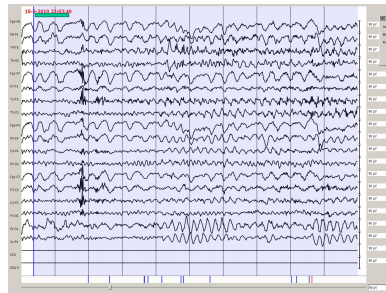


Fig. 2. EEG of a toothbrush artifact

- evidence found in electronic patient records and usually not or insufficiently considered with the help of computers,
4. is open to including the outputs of computer-based interpretation tools as evidence,
 5. allows a clinician to take into account more alternatives so as to notify him/her of rare diseases becoming sufficiently likely to warrant consideration,
 6. can incorporate meta-evidence, i.e., feedback from one clinician on the diagnosis of another, and
 7. protects him/her against ill-advised cognitive biases (Segal; 2007; Groopman; 2007; Croskerry; 2003; Graber et al.; 2002)

1.2. Examples

To illustrate the potential of computer-support through evidence combination with these properties, consider the following two examples.

Example 1: Toothbrush case

A suspicious sequence is detected in the EEG recording of an ICU patient (see Figure 2). Several clinicians debate but they can't agree on a diagnosis based on this sequence: opinions are split between epilepsy and artifact. A few clinicians (2%) think it is something else, i.e., unknown. Subsequently, the video recorded simultaneously with the suspicious EEG sequence is reviewed. It shows without a doubt that the sequence is actually an artifact due to the patient brushing his teeth. Figure 3 shows a timeline of events for the toothbrush case.

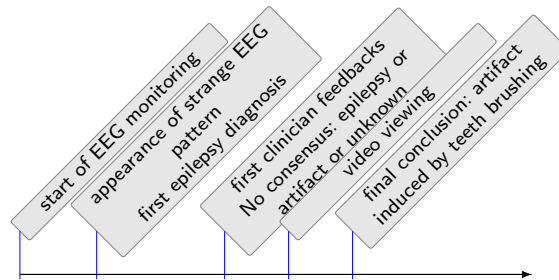


Fig. 3. Chronology of events for the toothbrush case

Example 2: Hemochromatosis case

This example is a real case reported in a Washington Post article in the Medical Mysteries section (see (Boodman; 2011)).

After an initial set of seemingly unrelated complaints and symptoms (blurry and rapidly decreasing vision, increased sleepiness, fatigue, high blood sugar level leading to a diabetes type I diagnosis), the patient lands in the ER with symptoms such as severe confusion and disorientation, internal bleeding and liver cirrhosis. Tests rule out the possibility of an infection or of hepatitis C and the ER doctors conclude that the symptoms exhibited by the patient result from a combination of diabetes type I and severe alcoholism (some symptoms being seen as signs of alcohol withdrawal). However, both the patient and his family deny the alcoholism especially since he hadn't drunk any alcohol in the two weeks before the ER visit as a result of fatigue. Moreover, some tests, undisclosed by the ER personnel to the patient at that point, show an alcohol level of 0g/L and extremely high blood iron levels.

Unconvinced by the diagnosis given at the ER, the patient, with the help of a pathologist friend, researches possible explanations for the complaints that landed him in the ER. Hemochromatosis, a disease found while perusing a medical textbook,³ appears a very likely possibility to him and after some tests (level of iron in the blood and genetic test), the hemochromatosis diagnosis is definitively confirmed. Figure 4 shows a timeline of events for the hemochromatosis case.

These examples illustrate several things. First, medical tests and scans interpretation such as, for instance, EEG interpretation, are *inherently uncertain*. As (Niedermeyer and Silva; 1999) puts it: “there is an element of science and an element of art in a good EEG interpretation” (p.167). The uncertainty of the interpretation can stem from the massive amount of data whose perusal is required to make a conclusion. For instance, the interpretation of a routine 20-minute EEG, usually done visually by a trained neurologist, requires the perusal of large amounts of data⁴ that contain age-dependent, context-dependent and

³ genetic disease that causes the body to absorb and store excessive amounts iron, resulting in organ damage

⁴ at least 109 A4 sheets of paper (1 sec of EEG being represented by at least 25mm on paper taken in landscape format) following the guidelines of the American Clinical Neurophysiology Society (American Clinical Neurophysiology Society; 2006).

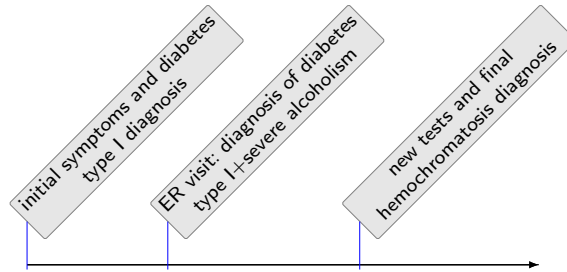


Fig. 4. Chronology of events for the hemochromatosis case

non-specific patterns. Uncertainty can also stem from the fact data patterns have no standard definition making the interpretation process not reproducible. For instance, a study by (Webber et al.; 1993) showed that, even when done by one single clinician at two different points in time, markings on EEG recordings for patterns such as epileptiform discharges may not be identical.

Computer-based interpretation tools may help make the data interpretation process more reproducible. Crucially, interpretation tools, while producing uncertain results, provide a quantification of the result uncertainty as opposed to visual inspection by humans for which the uncertainty estimation is tricky. This means that it can effectively be included in the evidence combination. (Semi)-automated interpretation tools could also provide cognitive aids and exploitable markings to clinicians, thus reducing their workload and their reliance on memory and freeing up enough time for them to focus properly on hard cases only. This would not only make the diagnosis process faster but also less prone to errors due to mistaken cognitive biases (Croskerry; 2003). And while not substitutes for clinician input, computer-based interpretation tools show good accuracy: a study trying to predict cancer outcomes based on applying machine-learning algorithms on electronic administrative records finds such algorithms at least as accurate as a panel of 5 expert clinicians (Gupta et al.; 2014).

Second, crucial clues or evidence may be ‘hidden’ in the vast amounts of available data, eg the blood iron levels in the hemochromatosis case. As computers are able to process a larger amount of data than humans possibly can and faster than they do, computer-based data pre-processing may assist the clinician in *uncovering clues and evidence hidden in mountains of data* even in cases when the clinician does not even suspect that there is something to find and point to zebras when they become likely due to multiple clues occurring together. For example, in the hemochromatosis case, software matching clinical findings with possible diagnoses may have highlighted the conjunction of high iron blood levels, no alcohol in the blood, diabetes and cirrhosis and pointed out that the hemochromatosis diagnosis became likely with those findings and should be explored. This would effectively result in forcing clinicians to consider several alternatives thus reducing the incidence and impact of premature closure. Note that this does require the data be available in digital format, which is not as straightforward as it may seem. According to (Manyika et al.; 2011), 30% of data in medical records, laboratory, and surgery reports, is not digitized. And 90% of the data generated

by healthcare providers is discarded, for example, almost all video recordings from surgery.

Third, often *unexpected circumstances may mislead* a clinician during the decision making process. Even when known with hindsight, it is hard to correctly trace back the process and properly reconsider the then known evidence. For example, it may transpire, after several EEGs have been recorded, that a set of electrodes used for one or more EEGs were faulty, which would mean that the diagnoses in which these EEGs were involved may need to be reconsidered. By storing data provenance, i.e., derivation lineage that represent which diagnosis is derived from what evidence taken from what data, the clinicians can be supported with batch-wise reconsideration of their diagnoses. And clinicians may only need to be notified if the unexpected event changes their diagnosis significantly.

Fourth, each step in the diagnosis process and all evidence leading to it should be *accessible for review*. In our example, the initial epilepsy diagnosis could be reviewed by several clinicians and modified significantly because of the video evidence. As such, a review of the EEG and its accompanying video helped override a faulty initial conclusion resulting from incomplete evidence and get to the correct conclusion. Computer-based evidence combination can assist the clinicians in properly incorporating new evidence as well as meta-evidence (feedback) in the overall diagnosis thus insuring that available data is used to make a decision.

The remainder of this paper is organized as follows. The medical diagnosis examples of Section 1.2 serve as running examples for further supporting our claims for evidence combination as well as for categorizing evidence into types (Section 2). We then give some background on the Dempster-Shafer theory underlying our model in Section 3. In Section 4, we explain how to represent evidence and the uncertainty inherent to it using the Dempster-Shafer framework. We then proceed to present our evidence combination model (Section 5) and validate it analytically (Section 7). We show how our model can be used in practice by working out the running examples as well as an important theoretical example from literature in detail (Section 6). Finally, Section 8 discusses how the model can be implemented by storing evidence in a probabilistic database which naturally supports uncertainty in data and maintaining lineage.

2. Categorization of evidence types for evidence combination

The examples given in Section 1.2 highlight the fact that the diagnosis process is an *incremental process*. New evidence modifies the state of knowledge: every step in Figures 3 and 4 introduces one or more pieces of evidence. By evidence, we mean the interpretation of a lab result or any other medical test or scan but not the raw data itself. What we call *evidence* is the new diagnosis that the clinician forms based on raw medical data (eg lab results or scans). The number of pieces of evidence is generally limited since clinicians tend to and need to focus on a limited set of alternatives. Note that computer-based interpretations of medical tests (eg lab tests and scans) and/or presenting symptoms are also considered to be evidence since they are an interpretation of raw medical data.

We distinguish three main types of evidence:

1. Evidence on already considered alternatives.
2. Evidence that introduces a new alternative.
3. Meta-evidence: Evidence on the reliability of other pieces of evidence.

The first type of evidence assigns a likelihood either to one or more existing alternatives or to part of one or more existing alternatives (e.g., new evidence supporting epilepsy while previously only evidence supporting both epilepsy and artifact existed). Special cases of this type of evidence include corroboration and rejection, i.e., positive or negative feedback from one clinician on the diagnosis of another. In our model, these are represented by a likelihood of 1 and 0, respectively, assigned to one particular alternative.

The second type of evidence occurs when new evidence, that may or may not support one or more previous conclusions, also points to a diagnosis hypothesis not previously considered. An example of this type of evidence occurs in the hemochromatosis case when hemochromatosis becomes a possible diagnosis aside from the combination of diabetes and alcoholism diagnosed at the ER.

An example of the third type of evidence is the genetic test and blood test in the hemochromatosis case: the tests invalidated the ER conclusions, i.e., reduced their reliability. Another example is the video evidence, in the toothbrush case, that confirmed the suspicion of artifact in the toothbrush case and lead to the rejection of the epilepsy diagnosis.

Evidence has several characteristics, on top of having a type (as explained earlier in the section). These characteristics can be summarized as follows:

- Evidence is uncertain and depends, for example, on the reliability of its source. We therefore attribute a confidence score c to each piece of evidence to quantify its reliability. If no knowledge on source reliability is available, one needs to assume, by default, that all sources of evidence are equally reliable.
- It is crucial that a concrete record of the dependencies between evidences (i.e., evidence provenance) be kept to ensure that pieces of evidence are properly combined or re-considered at any time. The reason for this is that, while evidence obviously appears in certain order during the diagnosis process, the evidence that arises at a certain point in time may refer to other specific pieces of evidence that arose earlier, for instance, in case of corroboration or meta-evidence.

3. A brief introduction to the Dempster-Shafer model

The Dempster-Shafer theory is a mathematical theory of evidence and can be viewed, in a finite discrete space, as a generalization of the traditional probability theory where probabilities are assigned to sets and not to mutually exclusive singletons. So, whereas, in traditional probability theory, evidence has to be associated with only one event, the Dempster-Shafer theory makes it possible to assign the evidence to a set of events. The Dempster-Shafer theory is therefore useful when evidence is available for a set of possible events and not for each possible event within the set and collapses to traditional probability theory in the case where evidence is available for all possible events within a set.

Let $\Theta = \theta_1, \dots, \theta_N$ be a finite set of possible hypotheses. Θ is called the *frame of*

discernment in the Dempster-Shafer theory. Note that, according to the Dempster-Shafer theory, each element $\theta_i \in \Theta$ where $i = 1, \dots, N$ doesn't have to be a singleton. For example, in the case of a clinician defining possible diagnosis with non-mutually exclusive diseases (for example migraine M , sinusitis S and labyrinthitis L), the frame of discernment Θ could be defined as: $\Theta = \{\emptyset, M, S, L, MS, ML, SL, MSL\} = 2^{\{M,S,L\}}$. As can also be seen in the previous example, the frame of discernment is usually defined as an exclusive⁵ and exhaustive non-empty⁶ set of possible alternatives.

The Dempster-Shafer theory defines three important functions on this frame of discernment: the basic probability assignment also called mass function (otherwise known as m), the belief function (denoted Bel) and the plausibility function (denoted as pl).

The *mass function* (or *basic probability assignment*) m is a function that assigns a value in $[0, 1]$ to every subset \mathcal{A} of Θ (such that $\cup_{\mathcal{A} \subseteq \Theta} \mathcal{A} = \Theta$) and satisfies:

$$m(\emptyset) = 0$$

$$\sum_{\mathcal{A} \subseteq \Theta} m(\mathcal{A}) = 1$$

(This means, in practice, that the sum of all basic probability assignments of subsets of the frame of discernment Θ of whom some may overlap may be different from 1 and actually higher than 1. It simply means that some evidence is counted more than once). The basic probability assignment defined for a subset \mathcal{A} , $m(\mathcal{A})$ is actually the degree of belief that the variable of interest falls within interval \mathcal{A} . However, $m(\mathcal{A})$ gives no indication as to the degree of belief that the variable of interest falls within any of the subintervals of interval \mathcal{A} . Additional evidence is needed for that.

The *belief* and *plausibility* functions can be interpreted respectively as the lower and upper bounds for probabilities, with the actual probability associated with the considered subset of 2^Θ in between the belief and plausibility values for that subset. The *belief* function Bel assigns a value in $[0, 1]$ to every non-empty subset \mathcal{B} of Θ and is defined as:

$$Bel(\mathcal{B}) = \sum_{\mathcal{A} \subseteq \mathcal{B}} m(\mathcal{A})$$

The *plausibility* function Pl assigns a value in $[0, 1]$ to all sets \mathcal{A} whose intersection with the set of interest \mathcal{B} is not empty:

$$Pl(\mathcal{B}) = \sum_{\mathcal{A} \cap \mathcal{B} \neq \emptyset} m(\mathcal{A})$$

Both the *belief* and *plausibility* functions are non-additive, which means that the sum of all belief values associated with values in 2^Θ is not required to be equal to 1 and similarly for plausibility. Furthermore, the *mass function* m can be defined using the *belief* function with:

$$m(\mathcal{B}) = \sum_{\mathcal{A} \subseteq \mathcal{B}} (-1)^{|B-A|} Bel(\mathcal{A})$$

⁵ in the previous example, it would mean that only one of the elements of Θ is the true diagnosis

⁶ it contains at least the empty set \emptyset

where $|B - A|$ is the cardinality of the difference between sets A and B. And we can derive *plausibility* from *belief* with: $Pl(\mathcal{B}) = 1 - Bel(\bar{\mathcal{B}})$ where $\bar{\mathcal{B}}$ is the complement of set \mathcal{B} . If $Bel(\mathcal{B}) = Pl(\mathcal{B}) = m(\mathcal{B})$, then we have defined a probability in the classical sense of the term.

An underlying assumption in the Bayesian theory is the existence of an ultimate refinement, that is “a frame of discernment so fine that it encompasses all possible distinctions and admits no further refinement” (Shafer; 1976, p. 119). In other words, the Bayesian theory supposes that all possible worlds are known and defined. While such an ultimate refinement would be conceptually convenient, it is also unrealistic as, in most real world applications, existing possible worlds for the system are discovered as we go and more evidence is gathered. In contrast, the Dempster-Shafer theory allows for ignorance, does away with the ultimate refinement hypothesis and instead defines frame refinements and coarsenings. According Shafer (Shafer; 1976, p.120), a frame of discernment is defined as “a set of possibilities that one does recognize on the basis of knowledge that one does have — or at least on the basis of distinctions that one actually draws and assumptions that one actually makes”. In other words, the frame of discernment reflects the state of knowledge at a given point in time so it is quite normal, in practice, to begin by defining a coarse frame of discernment and then refine it (that is split sets defined in the initial frame of discernment into finer subsets) as more knowledge is accumulated. The existence of such a possibility of refinement is what will allow us to perform user feedback that actually adds new alternatives (see Section 5.2 for more details). For a more detailed and formal definition of frames of discernment, frame compatibility and frame refinements and coarsenings, we refer to chapter 6 in (Shafer; 1976).

The Dempster-Shafer theory also provides means to combine evidence obtained from multiple sources, that provide different assessments for the frame of discernment and are supposed independent from each other. One such combination rule is the Dempster combination rule, which can be defined by (given two sources denoted 1 and 2):

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K} \quad \text{when } A \neq \emptyset$$

$$\text{where } K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$$

$$m_{12}(\emptyset) = 0$$

The denominator in Dempster’s rule is a normalization factor and has the effect of completely ignoring conflict and attributing all probability masses associated with conflict to the null set.

According (Zadeh; 1984), this omission of conflict may lead to some counterintuitive results as in the following example (and hereafter referred to as Zadeh’s example). A patient is seen by two physicians for troubling neurological symptoms. The first physician gives a diagnosis of meningitis with an associated probability of 0.99 while admitting the possibility of a brain tumour with an associated probability of 0.01. The second physician believes the patient has a concussion with a probability of 0.99 or a brain tumour with a probability of 0.01. If we use the Dempster’s combination rule with the available data, we get

$m(\{\text{brain tumour}\}) = Bel(\{\text{brain tumour}\}) = 1$. This result would imply that the most likely diagnosis is actually the one that both physicians find extremely unlikely.

Furthermore, the Dempster combination as well combination rules derived from it, such as Yager’s combination rule and Zhang’s center combination rule to name a few, suppose that all sources of evidence are equi-reliable. In our application, we suppose this is not the case and that a reliability score w_i — between 0 and 1 — is associated with each user giving feedback (how this reliability score is determined is beyond the scope of this paper, see Section 5.3). One way of taking into account the difference of reliability between sources of evidence would be to use the mixing (or averaging) rule described in (Senz and Ferson; 2002):

$$m_{1\dots n}(A) = \frac{1}{W} \sum_{i=1}^n w_i m_i(A) \quad \text{where } W = \sum_{i=1}^n w_i$$

where n is the number of sources, w_i the weight associated with the i -th source and m_i the mass function associated with the i -th source. For more details on the Dempster-Shafer theory and the evidence combination rules, see (Shafer; 1976), (Senz and Ferson; 2002), (Salicone; 2007) and (Užga-Rebrovs and Kuļešova; 2008).

4. Representation of uncertain evidence

Ideally, the uncertainty surrounding evidence is precisely known, but in practice it is often incomplete, coarsely known or completely missing. Therefore, evidence needs to be represented under various circumstances:

1. *Exact evidence likelihood values available.* For example, “the EEG of the patient points to epilepsy with confidence 0.8 and to an artifact with confidence 0.2.”
2. *Missing likelihood values* For example, “the EEG of the patient shows an epileptic seizure or an artifact.”
3. *Imprecise likelihood values.* For example, “the EEG of the patient shows epilepsy with a confidence of at least 0.7.”
4. *Coarse likelihood values.* For example, “the EEG of the patient shows epilepsy or an artifact with confidence 0.8 or a normal pattern with confidence 0.2”

Of particular interest for our application, are the cases where evidence likelihood values are exactly known and where likelihood values are missing (by far the most common case in our application). We represent piece of evidence i and its associated uncertainty with (a) a mass function m_i assigning likelihood values to several alternatives, combined with (b) a weight w_i representing the reliability of the evidence (relative to other pieces of evidence). For example,

$$m_i(S) = \begin{cases} 0.8 & \text{if } S = \{\text{epilepsy}\} \\ 0.2 & \text{if } S = \{\text{artifact}\} \\ 0 & \text{otherwise} \end{cases}$$

represents evidence that the EEG points to epilepsy with confidence 0.8 and to an artifact with confidence 0.2. This is a case where exact likelihood values are available: all mass is comprised in singletons, i.e., sets containing a single label. If no likelihood values are known, the mass can be assigned to a set containing multiple labels. For example,

$$m_i(S) = \begin{cases} 1 & \text{if } S = \{\text{epilepsy, artifact}\} \\ 0 & \text{otherwise} \end{cases}$$

represents evidence that the EEG points to epilepsy or an artifact.

Note that in the verbal expression of such evidence, one often does not mention the possibility that it could be entirely something else. We make this explicit in our model by introducing the explicit label ‘other’. This label represents all other diseases or conclusions not considered (yet). This allows a likelihood to be assigned to this label. For example,

$$m_i(S) = \begin{cases} 0.8 & \text{if } S = \{\text{epilepsy, artifact}\} \\ 0.2 & \text{if } S = \{\text{other}\} \end{cases}$$

represents the evidence that the EEG points to epilepsy or an artifact with confidence 0.8, but that one keeps open the possibility that it could be entirely something else with a confidence of 0.2. This conclusion can, for example, be drawn from circumstances where one estimates that the reliability of the sources is not perfect, but 80%. Note that with the inclusion of the explicit label *other*, there is no need for an ‘otherwise’; the mass function m_i representing a piece of evidence is always complete.

In the sequel, we will consistently use the term *label* and symbol a for a single interpretation, such as *epilepsy* or *artifact*, and the term *alternative* and symbol A for a set of labels, such as $\{\text{epilepsy, artifact}\}$. We denote with \mathcal{L} the set of all considered labels; $\mathcal{L} = \{\text{epilepsy, artifact, other}\}$ in the example above. Therefore, the frame of discernment is $\mathcal{F} = 2^{\mathcal{L}}$. In the example above,

$$\mathcal{F} = \{\emptyset, \{\text{epilepsy}\}, \{\text{artifact}\}, \{\text{other}\}, \{\text{epilepsy, artifact}\}, \{\text{epilepsy, other}\}, \{\text{artifact, other}\}, \{\text{epilepsy, artifact, other}\}\}$$

5. Evidence combination model

5.1. Core of the model: the mixing rule

In Section 3, we introduced a basic probability assignments’ combination rule, called the mixing (or averaging) rule defined as:

$$(\forall A \in \mathcal{F}) \quad m_{1\dots n}(A) = \frac{1}{W} \sum_{i=1}^n w_i m_i(A) \quad \text{with} \quad W = \sum_{i=1}^n w_i$$

where n is the number of evidences, w_i the weight associated with the i -th piece of evidence, W the normalization factor being the sum of all weights, and m_i the mass function associated with the i -th piece of evidence.

We assume that a database actually contains both the individual pieces of evidence with all associated information as well as an aggregation of the evidence

obtained from l previous sources. So, if we want to combine a new piece of evidence m_{l+1} with the l previous evidences, the total number of sources of evidence combined is $n = l + 1$. Also, the introduction of a new weight w_{l+1} updates the normalization factor $W' = W + w_{l+1}$. We apply the mixing rule as follows:

$$\begin{aligned}
(\forall A \in \mathcal{F}) \quad m_{1\dots n}(A) &= \frac{1}{W'} \sum_{i=1}^n w_i m_i(A) \\
&= \frac{1}{W'} \left(\sum_{i=1}^l w_i m_i(A) + w_{l+1} m_{l+1}(A) \right) \quad \text{since } n = l + 1 \\
&= \frac{1}{W'} \left(W \frac{1}{W} \sum_{i=1}^l w_i m_i(A) + w_{l+1} m_{l+1}(A) \right) \\
&= \frac{W}{W'} m_{db}(A) + \frac{w_u}{W'} m_u(A)
\end{aligned}$$

where $m_{db}(A)$ is the basic probability assignment associated with alternative A in the database, $m_u(A)$ is the basic probability assignment associated with alternative A by the user providing the new evidence and w_u the weight representing the reliability of this evidence.

Observe here that m_n representing the new combined diagnosis of all $n = l + 1$ evidences, can be calculated incrementally in terms of m_{db} , m_u , and w_u . Also, defined in this way, the combination rule is trivially associative and commutative as well as idempotent.

5.2. Basic operations of the model

We model all types of evidence with three atomic operations. We present the third atomic operation in two separate cases: in practice there are two different types of evidence that can be handled with one atomic operation, i.e., 3a and 3b are formally the same operation.

- 1 Adding a (weighted) basic probability assignment m_u with weight w_u due to a new piece of evidence.
- 2 Updating the weights associated with one or more previously given evidences $J \subseteq [1..n]$.
- 3a Refining the frame of discernment by splitting a known label a into multiple more refined ones.
- 3b Refining the frame of discernment by adding a new label for something previously not considered.

The notations used in this section can be found in Table 1. Section 5.2.5 describes how to determine which atomic operation to use.

As explained earlier, we assume that the database also contains the mass function m_{old} representing the combined diagnosis of all n previous evidences. Through the atomic operations formulas derived below, we aim to recalculate a new combined mass function incrementally, i.e., to define m_{new} in terms of m_{old} and the basic probability assignment $m_u(A)$ associated with the new evidence.

Notation	Meaning
a	existing label, e.g., epilepsy .
A	existing alternative, e.g., { epilepsy , artifact }.
n	number of evidence sources prior to new evidence being added.
m_u	new evidence represented as a basic probability assignment. $\text{dom}(m_u) = \mathcal{F}$
m_{old}	stored combined basic probability assignment derived from m_1, \dots, m_n prior to new evidence
m_{new}	combined basic probability assignment after taking new evidence into account
w_u	weight associated with m_u
W	normalization factor prior to new evidence being added, $W = \sum_{i=1}^n w_i$.
W'	normalization factor after taking new evidence into account $W' = W + w_u$.
$J \subseteq [1..n]$	set of indices corresponding to evidence sources for which the weights must be updated because of new evidence
w'_j	new weight due to new evidence ($j \in J$)
\mathcal{L}	set of all considered labels
\mathcal{F}	frame of discernment, i.e., set of all considered alternatives, $\mathcal{F} = 2^{\mathcal{L}}$

Table 1. List of notations

5.2.1. Adding a (weighted) basic probability assignment

This operation is used when evidence is added without any change in the weights associated with previous evidence sources. The number of sources with the addition of new evidence becomes $n + 1$. The new normalization factor is $W' = W + w_u$. Applying the mixing rule gives us:

$$\begin{aligned}
 (\forall A \in \mathcal{F}) \quad m_{new}(A) &= \frac{1}{W'} \left(\sum_{i=1}^n w_i m_i(A) + w_u m_u(A) \right) \\
 &= \frac{W}{W'} m_{old}(A) + \frac{w_u}{W'} m_u(A)
 \end{aligned}$$

5.2.2. Updating weights

This operation is used when new evidence leads to updating one or more weights w_j associated with previously given evidence with a new weight w'_j (e.g., decreasing the weight associated with a previous evidence because the source of evidence has been discovered to be less reliable than previously thought, or altogether canceling a piece of evidence with setting $w'_j = 0$).

The new normalization factor can be defined as

$$\begin{aligned}
 W' &= \sum_{j \notin J} w_j + \sum_{j \in J} w'_j \\
 &= W - \sum_{j \in J} w_j + \sum_{j \in J} w'_j \\
 &= W + \sum_{j \in J} (w'_j - w_j)
 \end{aligned}$$

We denote the latter *normalization correction* term with $W_{\Delta} = \sum_{j \in W} (w'_j - w_j)$.

According to the mixing rule, we have:

$$\begin{aligned} (\forall A \in \mathcal{F}) \quad m_{old}(A) &= \frac{1}{W} \sum_{i=1}^n w_i m_i(A) \\ &= \frac{1}{W} \sum_{j \notin J} w_j m_j(A) + \frac{1}{W} \sum_{j \in J} w_j m_j(A) \end{aligned}$$

The updated basic probability assignment for alternative A is obtained by using the mixing rule as follows:

$$\begin{aligned} (\forall A \in \mathcal{F}) \\ m_{new}(A) &= \frac{1}{W'} \left(\sum_{j \notin J} w_j m_j(A) + \sum_{j \in J} w'_j m_j(A) \right) \\ W' m_{new}(A) &= \sum_{j \notin J} w_j m_j(A) + \sum_{j \in J} w'_j m_j(A) \\ &= \sum_{j \notin J} w_j m_j(A) + \sum_{j \in J} w'_j m_j(A) + \sum_{j \in J} w_j m_j(A) - \sum_{j \in J} w_j m_j(A) \\ &= \sum_{j \notin J} w_j m_j(A) + \sum_{j \in J} w_j m_j(A) + \sum_{j \in J} w'_j m_j(A) - \sum_{j \in J} w_j m_j(A) \\ &= W m_{old}(A) + \sum_{j \in J} (w'_j - w_j) m_j(A) \\ m_{new}(A) &= \frac{W}{W'} m_{old}(A) + \frac{1}{W'} \sum_{j \in J} (w'_j - w_j) m_j(A) \end{aligned}$$

Updating the weights basically consists of canceling the terms in which the weights to be updated appear and then adding the newly weighted basic probability assignments. A full incremental calculation is not possible in this case, but one needs to revisit all evidences in the database for which the weight is updated. Usually, this remains rather limited.

5.2.3. Refining the frame of discernment: splitting a label

The label *epilepsy* actually represents a set of *epileptic syndromes* that differ by the specific features that are present. For example, benign rolandic epilepsy, childhood absence epilepsy, and juvenile myoclonic epilepsy are all particular cases of epilepsy. Suppose that in a diagnostic process, there have only been pieces of evidence where a confidence is assigned to alternatives that include the label *epilepsy*, but that a new piece of evidence points to, say, childhood absence epilepsy, or juvenile myoclonic epilepsy, how can we properly represent this new evidence and combine it with the existing ones?

Talking about measurement results, (Salicone; 2007, p.38) says:

When measurement results are considered, the basic probability number $m(A)$ can be interpreted as the degree of belief that the measurement result falls within interval A ; but $m(A)$ does not provide any further evidence in support to the belief that the measurement result

belongs to any of the various subintervals of A . This means that, if there is some additional evidence supporting the claim that the measurement result falls within a subinterval of A , say $B \subset A$, it must be expressed by another value $m(B)$.

The labels used in our model are very similar to concepts in description logic (Nardia and Brachman; 2002). A classic example of a description logic definition is $\text{Person} \equiv \text{Male} \sqcup \text{Female}$. It defines the concept of a person to be equivalent to either a male or a female. Important to note, is that this definition also states that the union of all possible males in the past, present, and future, and all possible females in the past, present, and future, will *exactly* give one the set of all possible persons in the past, present, and future. In other words, this definition truly *refines* the concept Person into two sub-concepts (it doesn't state that Male and Female are disjoint though).

We also apply this technique of refining concepts to our labels. Here we call it *splitting a label*. We may define

$$\begin{aligned} \text{epilepsy} &\equiv \text{benign rolandic epilepsy} \sqcup \text{childhood absence epilepsy} \\ &\sqcup \text{juvenile myoclonic epilepsy} \sqcup \text{other epileptic syndromes} \end{aligned}$$

Note that the inclusion of a label $\text{other epileptic syndromes}$ is necessary, because otherwise the equivalence doesn't hold. For brevity, we use the shorthands bre , cae , jme and oes in the sequel.

We can use this equivalence of concepts for refining our frame of discernment in a non-interfering manner: by replacing epilepsy with the four sub labels. Formally,

$$\begin{aligned} \mathcal{L}' &= (\mathcal{L} \setminus \{\text{epilepsy}\}) \cup \{\text{bre}, \text{cae}, \text{jme}, \text{oes}\} \\ \mathcal{F}' &= 2^{\mathcal{L}'} \end{aligned}$$

Furthermore, we need to adapt all existing pieces of evidence to the new frame of discernment. This is done by similarly replacing epilepsy in all pieces of evidence, i.e., whenever a mass function contains an alternative $A = \{\text{epilepsy}, \dots\}$, we replace A by $\{\text{bre}, \text{cae}, \text{jme}, \text{oes}, \dots\}$. Assigned confidences and weights remain unchanged.

Observe that the old frame of discernment \mathcal{F} is compatible with the new one \mathcal{F}' , because it is a proper refinement. Analogously, the thus constructed mass functions are proper refinements as well.

Typically, this atomic operation is triggered by the occurrence of evidence for a sub-label of an existing label. The operation of splitting the label hence is something to be carried out before the actual adding of the new evidence. It in a sense makes the frame of discernment and all existing evidence compatible with the refined nature of the new evidence. The new evidence is subsequently added using the atomic operation of Section 5.2.1.

Generically speaking, the atomic operation of splitting a label a into sub-labels a_1, \dots, a_m is defined with the following steps:

1. Define the equivalence $a \equiv a_1 \sqcup \dots \sqcup a_m$.
2. Let $(\forall A \in \mathcal{F}) \text{refine}(A) = \begin{cases} (A \setminus \{a\}) \cup \{a_1, \dots, a_m\} & \text{if } a \in A \\ A & \text{otherwise} \end{cases}$
3. Refine the frame of discernment $\mathcal{L}' = \text{refine}(\mathcal{L})$; $\mathcal{F}' = 2^{\mathcal{L}'}$.
4. For every $i \in [1..n]$, we define a refined mass function $m' = \text{refine}(m)$ as

$$(\forall A' \in \mathcal{F}') \quad \begin{cases} m'(A') = m(A) & \text{if } \exists A \in \text{dom}(m) : A' = \text{refine}(A) \\ 0 & \text{otherwise} \end{cases}$$

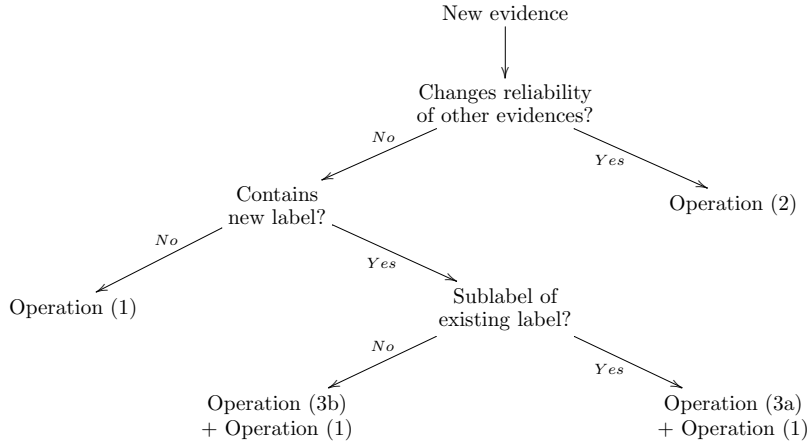


Fig. 5. Decision tree for combining atomic operations to handle all types of evidence

(Note that we have overloaded *refine* to work both on alternatives as well as mass functions).

5.2.4. Refining the frame of discernment: adding a new label

When a clinician makes a diagnosis, (s)he not only makes a diagnosis, but effectively excludes all other possible diagnoses. In this diagnosis, (s)he implicitly assigns a zero confidence to alternative $\{\text{other}\}$. The existence of the **other** label makes the frame of discernment as well as all mass functions exhaustive.

It is the existence of the **other** label, however, that makes it possible to apparently “expand” the frame of discernment and add a new label a . For example, adding the hemochromatosis alternative after a diagnosis of diabetes combined with alcoholism has been reached in the hemochromatosis example. Because any new unknown label is already included in the **other** label, we can split it into a and a new **other'** by defining $\text{other} \equiv a \sqcup \text{other}'$ and applying the atomic operation of Section 5.2.3. Therefore, adding a new label is only a special case of refining the frame of discernment by splitting a label.

5.2.5. Deciding on which atomic operations to use

In Section 5.2, we introduced the atomic operations that are used to model the addition of new evidence. We here provide a decision tree that illustrates how the atomic operations need to be combined to handle all types of evidence.

5.3. Deciding on a weighting method

The setting of weights is not the purpose of this paper. However, some ways to set the weights include defining rules to set weights (e.g., “clinician A is twice as reliable as clinician B” or “video-based evidence supersedes EEG-based evidence”) or deducing the weights by, for example, evaluating experts or sources of evidence through a set of calibrating questions for which the answer is known.

The mixing rule is a generalization of the averaging of probability distributions ((Sentz and Ferson; 2002)) also known as linear opinion pool. The linear opinion pool is widely used as a way to combine expert opinions in a probabilistic framework and several ways to set the weights in the linear opinion pool have been studied ((Cooke and Goossens; 2008; Ouchi; 2004; O’Hagan et al.; 2006; Rougier et al.; 2013)). Similar strategies can be applied to set the weights for the mixing rule as well. One such strategy is the performance-based Cooke “classical” method . Cooke argues that using equal weights for all experts leads to a suboptimal solution as it doesn’t evaluate the quality of each expert’s opinion. Cooke suggests assigning the weight base on the performance of experts on an elicitation exercise based on “seeding variables”, the “seeding variables” being quantities from the same area as the uncertain quantity of interest for which the true value is known to the one administering the exercise and not the experts. The experts may be asked to choose the probability bin in which they think the “seeding variable” they are given falls. Two scores are deduced from the experts performance: a calibration score which is the likelihood that the expert’s answer corresponds to the known answer and an information/informativeness score that measures how concentrated the distribution given by expert is. Those two scores are then combined into a weight assigned to the expert. An expert that is “highly reliable” scores high on both calibration and informativeness.

Another way of determining appropriate weights is through data mining. At the end of a diagnostic and treatment process, the correct diagnosis is known. All evidence given in the process can then be evaluated based on its degree of correctness. By accumulating these evaluations for pieces of evidence given by a certain expert or a certain evidence source, one can determine an appropriate reliability score. Over time, one could determine a set of weights that is based on how accurate experts and sources actually are on average.

5.4. Rationale for using Dempster-Shafer framework instead of Bayesian framework

The Bayesian theory is a special case of the Dempster-Shafer evidence theory according to (Shafer; 1976, Chp.1), with the Bayesian belief functions a subset of the Demspster-Shafer belief functions. The Dempster-Shafer theory is shown in (Hoffman and Murphy; 1993) to be more suited in cases of missing priors and ignorance. (Shafer; 1976) tries to show through an example (Example 1.6, chapter 1, pages 23–24) that applying the Bayesian theory to cases of complete ignorance could lead to counter-intuitive results. In the example given by Shafer, the question is to know whether or not there is life around the star Sirius. And though some scientists may have evidence on this question, Shafer takes the point of view of the majority of people who profess complete ignorance on the subject and that

$$Bel(A) = \begin{cases} 0 & \text{if } A \neq \Theta \\ 1 & \text{if } A = \Theta \end{cases}$$

where $\Theta = \{\theta_1, \theta_2\}$, with θ_1 denoting the possibility that there is life on Sirius and θ_2 denoting the possibility that there is no life on Sirius.

He then considers a more refined set of possibilities $\Omega = \zeta_1, \zeta_2, \zeta_3$ where ζ_1 is the possibility that there is life on Sirius, ζ_2 the possibility that there are planets

around Sirius but no life, and ζ_3 the possibility that there are not even planets around Sirius. The original frame of discernment Θ and the refined set Ω are related in that

$$\theta_1 = \zeta_1 \quad \text{and} \quad \theta_2 = \{\zeta_2, \zeta_3\}$$

which means that

$$Bel(A) = \begin{cases} 0 & \text{if } A \neq \Omega \\ 1 & \text{if } A = \Omega \end{cases}$$

So translating complete ignorance in the Dempster-Shafer framework is straightforward. Shafer goes on to try and show that it is difficult to specify consistent degrees of belief over Θ and Ω in the Bayesian framework when representing complete ignorance. Complete ignorance on Θ may be represented by $Bel(\{\theta_1\}) = Bel(\{\theta_2\}) = \frac{1}{2}$. On Ω , however, according to him, complete ignorance would mean that $Bel(\{\zeta_1\}) + Bel(\{\zeta_2\}) + Bel(\{\zeta_3\}) = 1$, hence $Bel(\{\zeta_1\}) = Bel(\{\zeta_2\}) = Bel(\{\zeta_3\}) = \frac{1}{3}$.

This yields

$$Bel(\{\zeta_1\}) = \frac{1}{3} \quad \text{and} \quad Bel(\{\zeta_2, \zeta_3\}) = \frac{2}{3}$$

These results are inconsistent with the ones found on Θ since $\{\theta_1\}$ and $\{\zeta_1\}$ have the same meaning as well as $\{\theta_2\}$ and $\{\zeta_2, \zeta_3\}$. However, this line of reasoning is flawed. In fact, instead of considering the three possible events $\{\zeta_1, \zeta_2, \zeta_3\}$, one should consider four events. Let event A be "There is life on Sirius" and B the event "There are planets around Sirius". A and B are independent. Based on A and B , there are four events on Ω instead of three: $a = A \wedge B$, $b = A \wedge \neg B$, $c = \neg A \wedge \neg B$, $d = \neg A \wedge B$.

If $P(B) = \alpha$ then we know

$$P(a) = \frac{1}{2}\alpha \quad \text{and} \quad P(b) = \frac{1}{2}(1 - \alpha) \quad \text{and} \quad P(c) = \frac{1}{2}(1 - \alpha) \quad \text{and} \quad P(d) = \frac{1}{2}\alpha$$

Since $\{a, b\} = \theta_1$ and $\{c, d\} = \theta_2$, the solution obtained through the Bayesian method is still consistent and equivalent to the Dempster-Shafer solution. So when working with equivalent formulations, the solutions reached in both the Dempster-Shafer framework and the Bayesian framework are similar. However, the Bayesian framework calls for making assumptions (independence of A and B) and finding out some variables' values ($P(B)$) to reach a solution, when no assumptions or additional variable values besides what is already known are needed to reach a solution in the Dempster-Shafer framework. Reaching a solution in the Bayesian framework when no independence assumption can be made is more difficult.

(Hoffman and Murphy; 1993) compare the use of the Bayesian theory and Dempster-Shafer theory to combine evidence from sensors. They conclude

"Both methods for dealing with uncertainty yield similar results if based on equivalent formulations. [...] [W]e believe that Bayesian theory is best suited to applications where there is no need to represent ignorance, where conditioning is easy to extract through probabilistic representation, and prior odds are available. Dempster-Shafer theory is a more natural representation for situations where uncertainty cannot be assigned probabilistically to a proposition or its complement and when conditioning effects are either impractical to measure separately from the event itself or a simple propositional refinement, and prior odds are not relevant."

In practice, in our case study (the diagnosis process) in particular, ignorance is frequent. And rare are the cases where strong assumptions such as variable independence can be made. Our case study — the diagnostic process — is therefore best represented in the Dempster-Shafer framework rather than the Bayesian framework, since ignorance is a mainstay of the diagnosis process.

Though there have been many studies that show how to successfully model meta-evidence by using Bayesian or Markov networks (de Campos et al.; 2003; Xin and Jin; 2004), we think such models may be unsuitable for our application, because

- the case where new evidence leads to the addition of a new alternative cannot be represented with such networks, because such evidence is not easily represented in a graph
- it would be counter-productive to use two different models (Bayesian/Markov for positive or negative feedback and another model for other types of evidence such as the addition of a new alternative), when we can use one model (based on Dempster-Shafer theory) for all types of evidence.

5.5. Mixing rule versus Dempster combination rule

We use the mixing rule above in our model rather than combination rules such as Dempster’s combination rule, Yager’s combination rule or Zhang’s combination rule, because it allows the combination of evidence coming from sources that may not be equally reliable.

Furthermore, as explained in (Florea et al.; 2009), the classic Dempster combination rule assumes the following:

- the list of alternatives contained in the frame of discernment is an exclusive and exhaustive list of hypotheses,
- all the sources of evidence combined are independent and provide independent evidence, and
- all sources of evidence are homogeneous i.e. equally reliable

All three conditions required for the proper application of the Dempster combination rule do not hold for the medical diagnosis process. The sources’ independence cannot be guaranteed as clinicians(sources) may consult each other while trying to come up with a diagnosis. The sources are not necessarily equally reliable, for instance, in our running toothbrush example, the video-based feedback is more reliable than the EEG interpretation. And finally, the frame of discernment may not necessarily be exhaustive as new alternatives may crop up during the diagnostic process (e.g., the hemochromatosis alternative was considered after the ER-visit by the patient and his pathology-friend in the hemochromatosis case).

6. Using the feedback model: some examples

In this section, we illustrate the usage of our model in practice by applying it to the two examples introduced in Section 1.2 and to Zadeh’s canonical example (introduced in Section 3).

6.1. First example: the toothbrush case

Here we apply our model to the toothbrush example from Section 1.2. The chronology of events in this case can be found in Figure 3. After the start of EEG monitoring, the appearance of a strange EEG pattern was observed. This alone does not carry any evidence that points to a possible diagnosis.

Then several clinicians debate the issue, but they do not arrive at a consensus. They are split between epilepsy and artifact. Besides that there are a few clinicians (2%) that think it is something else altogether, i.e., not epilepsy nor artifact. The new evidence resulting from the debate can be represented with the mass function below.

$$\begin{aligned} \mathcal{L} &= \{\text{epilepsy}, \text{artifact}, \text{other}_1\} \\ w_1 &= 1 \\ m_1(A) &= \begin{cases} 0.98 & \text{if } A = \{\text{epilepsy}, \text{artifact}\} \\ 0.02 & \text{if } A = \{\text{other}_1\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

The next piece of evidence is watching the video which clearly points to a diagnosis of artifact. We may treat this evidence in two ways. It could be seen as new evidence which is much much more reliable than the other two, for example, with a weight of $w_2 = 100$. Or, we may interpret the evidence as including the meta-evidence that earlier diagnoses are completely bogus, because they assumed a normal eyes closed EEG, while in fact the patient was brushing his teeth. Let us work out the former treatment.

$$\begin{aligned} \mathcal{L} &= \{\text{epilepsy}, \text{artifact}, \text{other}_1\} \\ w_2 &= 100 \\ m_2(A) &= \begin{cases} 1 & \text{if } A = \{\text{artifact}\} \\ 0 & \text{if } A = \{\text{epilepsy}\} \\ 0 & \text{if } A = \{\text{other}_1\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

According to the mixing rule, we can now determine a combined diagnosis.

$$\begin{aligned} \mathcal{L} &= \{\text{epilepsy}, \text{artifact}, \text{other}_1\} \\ W_{12} &= 101 \\ m_{12}(A) &= \begin{cases} 0 & \text{if } A = \{\text{epilepsy}\} \\ 0.990099 & \text{if } A = \{\text{artifact}\} \\ 0.009703 & \text{if } A = \{\text{epilepsy}, \text{artifact}\} \\ 0.000198 & \text{if } A = \{\text{other}_1\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

Because of the occurrence of singular alternatives $\{\text{epilepsy}\}$ and $\{\text{artifact}\}$ as well as a combined alternative $\{\text{epilepsy}, \text{artifact}\}$, the situation is not immediately clear. Here, the notions of belief and plausibility help to obtain a lower and

upper bound:

$$\begin{aligned}
Bel(\{\text{artifact}\}) &= \sum_{A \subseteq \{\text{artifact}\}} m_{12}(A) \\
&= m_{12}(\emptyset) + m_{12}(\{\text{artifact}\}) = 0.990099 \\
Pl(\{\text{artifact}\}) &= \sum_{A \cap \{\text{artifact}\} \neq \emptyset} m_{12}(A) \\
&= m_{12}(\{\text{artifact}\}) + m_{12}(\{\text{epilepsy}, \text{artifact}\}) = 0.999802 \\
Bel(\{\text{epilepsy}\}) &= \sum_{A \subseteq \{\text{epilepsy}\}} m_{12}(A) \\
&= m_{12}(\emptyset) + m_{12}(\{\text{epilepsy}\}) = 0 \\
Pl(\{\text{epilepsy}\}) &= \sum_{A \cap \{\text{epilepsy}\} \neq \emptyset} m_{12}(A) \\
&= m_{12}(\{\text{epilepsy}\}) + m_{12}(\{\text{epilepsy}, \text{artifact}\}) = 0.009703
\end{aligned}$$

In other words, the likelihood of an artifact lies somewhere between 0.99099 and 0.999802. There is still some plausibility remaining for an epilepsy diagnosis originating from the debate, but it is very small because of the low relative reliability in comparison with the video evidence.

Note that if we had known from the debate which proportion of the clinicians supported the epilepsy diagnosis and which proportion the artifact diagnosis, m_1 would have distinguished the two cases as singular alternatives with the proportion as confidence (provided that we assume all clinicians participating in the debate carry the same weight). Alternatively, one could also include the opinion of each clinician participating in the debate as a separate piece of evidence. The mixing rule would produce a similar combined result.

6.2. Second example: the hemochromatosis case

Here is how we abbreviate the names of the diseases used in this example: diabetes as *diab*, alcoholism as *alc*, hemochromatosis as *hemo*, hepatitis C as *hepC* and infection as *inf*.

A few days before the patient lands in the ER, a first diagnosis of diabetes type I is made. Our frame of discernment at this point contains labels *diab* and *other₁*.

$$\begin{aligned}
\mathcal{L} &= \{\text{diab}, \text{other}_1\} \\
w_1 &= 1 \\
m_1(A) &= \begin{cases} 1 & \text{if } A = \{\text{diab}\} \\ 0 & \text{if } A = \{\text{other}_1\} \\ 0 & \text{if } A = \emptyset \end{cases}
\end{aligned}$$

The patient lands in the ER and some hypotheses are first considered: hepatitis C or infection. This corresponds to a refinement of the frame of discernment to include both hepatitis C and infection, i.e., $\text{other}_1 \equiv \text{hepC} \sqcup \text{inf} \sqcup \text{other}_2$. We need

to adapt m_1 to the refined frame of discernment.

$$\begin{aligned} \mathcal{L}' &= \{\text{diab}, \text{hepC}, \text{inf}, \text{other}_2\} \\ w'_1 &= 1 \\ m'_1(A) &= \begin{cases} 1 & \text{if } A = \{\text{diab}\} \\ 0 & \text{if } A = \{\text{hepC}, \text{inf}, \text{other}_2\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

The initial consideration of hepatitis C or infection as opposed to diabetes given at the ER can be interpreted as full confidence in hepatitis C or infection. Let us suppose this interpretation is considered more reliable than the initial diabetes evidence, say twice as reliable.

$$\begin{aligned} \mathcal{L}' &= \{\text{diab}, \text{hepC}, \text{inf}, \text{other}_2\} \\ w_2 &= 2 \\ m_2(A) &= \begin{cases} 1 & \text{if } A = \{\text{hepC}, \text{inf}\} \\ 0 & \text{if } A = \{\text{diab}, \text{other}_2\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

Applying the mixing rule gives the following combined diagnosis.

$$\begin{aligned} \mathcal{L}' &= \{\text{diab}, \text{hepC}, \text{inf}, \text{other}_2\} \\ W_{12} &= 3 \\ m_{12}(A) &= \begin{cases} \frac{1}{3} & \text{if } A = \{\text{diab}\} \\ 0 & \text{if } A = \{\text{diab}, \text{other}_2\} \\ \frac{2}{3} & \text{if } A = \{\text{hepC}, \text{inf}\} \\ 0 & \text{if } A = \{\text{hepC}, \text{inf}, \text{other}_2\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

Both the hepatitis C and infection are quickly ruled out at the ER by means of tests and the diagnosis retained is that of diabetes combined with severe alcoholism. Ruling out the initial ER diagnosis can be achieved by updating its weight w_2 to $w'_2 = 0$. The alternative “diabetes combined with severe alcoholism” is a subconcept of diabetes. Therefore, we need to apply operation 3a of Section 5.2.3 to split the label for diabetes: $\text{diab} \equiv \text{diab_alc} \sqcup \text{diab_no_alc}$. We update m'_1 again (we omit m_2 , because its weight is $w_2 = 0$, hence it doesn't count anymore)

$$\begin{aligned} \mathcal{L}'' &= \{\text{diab_alc}, \text{diab_no_alc}, \text{hepC}, \text{inf}, \text{other}_2\} \\ w''_1 &= 1 \\ m''_1(A) &= \begin{cases} 1 & \text{if } A = \{\text{diab_alc}, \text{diab_no_alc}\} \\ 0 & \text{if } A = \{\text{hepC}, \text{inf}, \text{other}_2\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

Because of the additional tests, we may consider the new evidence more reliable to a degree of $w_3 = 4$. The evidence of retaining the diagnosis of diabetes but

combined with alcoholism can be represented as

$$\begin{aligned}\mathcal{L}'' &= \{\text{diab_alc}, \text{diab_no_alc}, \text{hepC}, \text{inf}, \text{other}_2\} \\ w_3 &= 4 \\ m_3(A) &= \begin{cases} 1 & \text{if } A = \{\text{diab_alc}\} \\ 0 & \text{if } A = \{\text{diab_no_alc}, \text{hepC}, \text{inf}, \text{other}_2\} \\ 0 & \text{if } A = \emptyset \end{cases}\end{aligned}$$

However, after research by the patient and a pathologist friend, a different explanation of the symptoms comes to the scene: hemochromatosis. After several tests, this diagnosis is confirmed. This turn of events first calls for yet another expansion of the frame of discernment: $\text{other}_2 \equiv \text{hemo} \sqcup \text{other}_3$. Moreover, the positive testing for hemochromatosis should not be seen as a case of just some more evidence to add to the mix, but rather as evidence that overrules all previous evidence. We therefore set all weights $w_1 = w_2 = w_3 = 0$. The only remaining evidence that counts, is:

$$\begin{aligned}\mathcal{L}''' &= \{\text{diab_alc}, \text{diab_no_alc}, \text{hepC}, \text{inf}, \text{hemo}, \text{other}_3\} \\ w_4 &= 1 \\ m_4(A) &= \begin{cases} 1 & \text{if } A = \{\text{hemo}\} \\ 0 & \text{if } A = \{\text{diab_alc}, \text{diab_no_alc}, \text{hepC}, \text{inf}, \text{other}_3\} \\ 0 & \text{if } A = \emptyset \end{cases}\end{aligned}$$

6.3. Evidence combination model applied to Zadeh's counterexample

Zadeh's example (introduced in (Zadeh; 1984) and explained in Section 3) has become the canonical example to show that the classic Dempster-Shafer evidence combination rule is not suitable for combining highly conflicting pieces of evidence. Haenni, however, contends that the apparent counter-intuitive result of the example is due to poor modelling of the problem. While the criticism leveled by (Haenni; 2005) may be founded, we show how our evidence combination model makes the modelling of Zadeh's example very simple and leads to a logical result.

In Zadeh's example, we have 2 sources of evidence, two clinicians giving conclusions, denoted as clinicians c_1 and c_2 , and 3 alternatives (meningitis abbreviated with *men*, brain tumor abbreviated with *tumor* and concussion abbreviated with *conc*). The diagnosis of clinician c_1 is

$$\begin{aligned}\mathcal{L} &= \{\text{men}, \text{tumor}, \text{conc}, \text{other}\} \\ w_{c_1} &= 1 \\ m_{c_1}(A) &= \begin{cases} 0.99 & \text{if } A = \{\text{men}\} \\ 0.01 & \text{if } A = \{\text{tumor}\} \\ 0 & \text{if } A = \{\text{conc}, \text{other}\} \\ 0 & \text{if } A = \emptyset \end{cases}\end{aligned}$$

We have $m_{c_1}(\{\text{conc}\}) = 0$ as there is no evidence for the concussion alternative

at this point. The conclusions drawn by clinician c_2 point to a different direction.

$$\begin{aligned} \mathcal{L} &= \{\text{men, tumor, conc, other}\} \\ w_{c_2} &= 1 \\ m_{c_2}(A) &= \begin{cases} 0.99 & \text{if } A = \{\text{conc}\} \\ 0.01 & \text{if } A = \{\text{tumor}\} \\ 0 & \text{if } A = \{\text{men, other}\} \\ 0 & \text{if } A = \emptyset \end{cases} \end{aligned}$$

Applying the mixing rule gives us the following:

$$m_{12}(A) = \begin{cases} 0.495 & \text{if } A = \{\text{men}\} \\ 0.01 & \text{if } A = \{\text{tumor}\} \\ 0.495 & \text{if } A = \{\text{conc}\} \\ 0 & \text{if } A = \{\text{men, other}\} \\ 0 & \text{if } A = \{\text{conc, other}\} \\ 0 & \text{if } A = \emptyset \end{cases}$$

The brain tumor alternative is, as expected, extremely unlikely. And since both clinicians (supposed equally reliable) give it the same likelihood, the final basic probability assignment associated with it, $m(\{\text{tumor}\}) = 0.01$ is not wholly unexpected. That both the *concussion* and *meningitis* alternatives are equally likely after both clinicians' conclusions also makes sense, since at this point, there is no way to say that one alternative is more likely than the other. There is no reason to trust one clinician more than the other. Note that, in our modeling of Zadeh's example, the frame of discernment is still $\{\text{men, tumor, conc}\}$ as in (Zadeh; 1984) and not one of the more complex frames of discernment used in (Haenni; 2005).

7. Analytical validation

In this section, we analytically validate the evidence representation and combination model. We formulate several correctness, monotonicity, and convergence properties and prove them. An experimental validation, i.e., a user study that shows that the model improve the decision quality in diagnostic processes, is beyond the scope of this paper.

7.1. Validation of correctness properties

7.1.1. The mixing rule produces a mass function

To prove

The intention of the mixing rule is to combine several mass functions into a combined mass function that represents the combined evidence. Therefore, the result of the mixing rule should be a proper mass function:

$$\begin{aligned} m_{1..n}(\emptyset) &= 0 \\ \sum_{A \in \mathcal{F}} m_{1..n}(A) &= 1 \end{aligned}$$

Proof

Since all m_i are mass functions, $m_i(\emptyset) = 0$ ($i \in [1..n]$). Therefore,

$$m_{1..n}(\emptyset) = \frac{1}{W} \sum_{i=1}^n w_i m_i(\emptyset) = \frac{1}{W} \sum_{i=1}^n w_i 0 = 0$$

The other requirement is that the sum of masses of all alternatives should be 1.

$$\begin{aligned} \sum_{A \in \mathcal{F}} m_{1..n}(A) &= \sum_{A \in \mathcal{F}} \frac{1}{W} \sum_{i=1}^n w_i m_i(A) = \frac{1}{W} \sum_{A \in \mathcal{F}} \sum_{i=1}^n w_i m_i(A) \\ &= \frac{1}{W} \sum_{i=1}^n \sum_{A \in \mathcal{F}} w_i m_i(A) = \frac{1}{W} \sum_{i=1}^n w_i \sum_{A \in \mathcal{F}} m_i(A) \\ &= \frac{1}{W} \sum_{i=1}^n w_i \cdot 1 = \frac{1}{W} W = 1 \end{aligned}$$

*7.1.2. Agreement between evidence of varying reliability***To prove**

If several pieces of evidence agree with each other, their weights should not be relevant. In other words, let all mass functions m_i ($i \in [1..n]$) be equal, i.e., $\forall i \in [1..n] : m_i = m$. Then $m_{1..n}$ should be equal to m irrespective of the individual weights w_i .

Proof

$$\begin{aligned} (\forall A \in \mathcal{F}) \quad m_{1..n}(A) &= \frac{1}{W} \sum_{i=1}^n w_i m_i(A) = \frac{1}{W} \sum_{i=1}^n w_i m(A) \\ &= \frac{1}{W} m(A) \sum_{i=1}^n w_i = \frac{1}{W} m(A) W = m(A) \end{aligned}$$

Therefore, $\forall i \in [1..n] : m_i = m \Rightarrow m_{1..n} = m$

*7.1.3. Weights are relative***To prove**

Weights are intended to be relative, i.e., if two evidences m_1 and m_2 are given, then it should not matter for $m_{1,2}$ whether $w_1 = 1$ and $w_2 = 3$, or that $w_1 = 5$ and $w_2 = 15$. In other words, if all weights w_i are updated with the same factor $f \neq 0$ using operation 2 (Section 5.2.2), the same $m_{1..n}$ is produced by the mixing rule.

Proof

Let $J = [1..n]$, $w'_j = f w_j$ ($j \in J$). First observe that $W' = \sum_{j=1}^n w'_j = \sum_{j=1}^n f w_j = f W$.

According to operation 2

$$(\forall A \in \mathcal{F})$$

$$\begin{aligned}
m_{new}(A) &= \frac{W}{W'} m_{old}(A) + \frac{1}{W'} \sum_{j \in J} (w'_j - w_j) m_j(A) \\
&= \frac{W}{fW} m_{old}(A) + \frac{1}{fW} \sum_{j \in J} (fw_j - w_j) m_j(A) \\
&= \frac{1}{f} m_{old}(A) + \frac{1}{fW} \sum_{j=1}^n w_j (f-1) m_j(A) \\
&= \frac{1}{f} m_{old}(A) + \frac{1}{f} \frac{1}{W} (f-1) \sum_{j=1}^n w_j m_j(A) \\
&= \frac{1}{f} m_{old}(A) + \frac{f-1}{f} m_{old}(A) \\
&= m_{old}(A) \left(\frac{1}{f} + \frac{f-1}{f} \right) \\
&= m_{old}(A) \left(\frac{1+f-1}{f} \right) \\
&= m_{old}(A)
\end{aligned}$$

7.1.4. Refinement of the frame of discernment does not modify existing evidence

To prove

The refinement of the frame of discernment (operations 3a and 3b of Sections 5.2.3 and 5.2.4) is a three-step procedure which updates all existing mass functions. Let us denote the refinement of a mass function with *refine* and the mixing rule with *mix*. Since it is only a refinement of a label, the mass function resulting from the mixing rule should be equal to a direct refinement of the combined evidence. In other words

$$refine(m_{1..n}) = mix(refine(m_1), \dots, refine(m_n))$$

Proof

refine(m_i) produces a m'_i in the following way:

$$(\forall A' \in \mathcal{F}') \quad \begin{cases} m'(A') = m(A) & \text{if } \exists A \in \text{dom}(m) : A' = refine(A) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{where } (\forall A \in \mathcal{F}) \quad refine(A) = \begin{cases} (A \setminus \{a\}) \cup \{a_1, \dots, a_m\} & \text{if } a \in A \\ A & \text{otherwise} \end{cases}$$

Note that $\forall A' \in \mathcal{F}'$, this A' either has a corresponding $A \in \mathcal{F}$ or not. In the former case, the corresponding A is the inverse refinement, i.e., $A = (A' \setminus \{a_1, \dots, a_m\}) \cup \{a\}$. In this case, $\forall A \in \mathcal{F} \forall i \in [1..n] : m_i(A) = m'_i(A')$. In the second case where $\exists a_j \in A'$ but $\{a_1, \dots, a_m\} \not\subseteq A'$, $m'_i(A') = 0$.

Applying the mixing rule

$$\begin{aligned}
&(\forall A' \in \mathcal{F}') \\
m'_{1..n}(A') &= \frac{1}{W} \sum_{i=1}^n w_i refine(m_i)(A')
\end{aligned}$$

$$\begin{aligned}
&= \begin{cases} \frac{1}{W} \sum_{i=1}^n w_i m_i(A) & \text{if } \exists A \in \text{dom}(m_i) : A' = \text{refine}(A) \\ \frac{1}{W} \sum_{i=1}^n w_i \cdot 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} m_{1..n}(A) & \text{if } \exists A \in \text{dom}(m_i) : A' = \text{refine}(A) \\ 0 & \text{otherwise} \end{cases} \\
&= \text{refine}(m_{1..n})(A')
\end{aligned}$$

7.2. Validation of monotonicity properties

7.2.1. Support for an alternative should increase its confidence

To prove

If new evidence (operation 1 from Section 5.2.1) positively supports an alternative A , i.e., $m_u(A) > m_{old}(A)$, then the confidence in A should increase irrespective of the weight w_u , i.e., $m_{new}(A) > m_{old}(A)$.

Proof

Since $W' = W + w_u$ and weights are positive, we know that $0 < \frac{W}{W'} < 1$ and $0 < \frac{w_u}{W'} < 1$.

$$\begin{aligned}
m_{new}(A) &= \frac{W}{W'} m_{old}(A) + \frac{w_u}{W'} m_u(A) \\
&> \frac{W}{W'} m_{old}(A) + \frac{w_u}{W'} m_{old}(A) \\
&= m_{old}(A) \left(\frac{W + w_u}{W'} \right) \\
&= m_{old}(A)
\end{aligned}$$

7.2.2. Increase in weight for evidence supporting an alternative should increase its confidence

To prove

If a weight is increased (operation 2 from Section 5.2.2) for evidence that positively supports an alternative A , i.e., $w'_j > w_j$ for some $j \in [1..n]$ and $m_j(A) > m_{old}(A)$, then the confidence in A should increase, i.e., $m_{new}(A) > m_{old}(A)$. Let $J = \{j\}$.

Proof

$$\begin{aligned}
m_{new}(A) &= \frac{W}{W'} m_{old}(A) + \frac{1}{W'} \sum_{j \in J} (w'_j - w_j) m_j(A) \\
&= \frac{W}{W'} m_{old}(A) + \frac{(w'_j - w_j) m_j(A)}{W'} \\
&> \frac{W}{W'} m_{old}(A) \quad \text{because } w'_j > w_j \wedge m_j(A) > 0 \wedge W' > 0 \\
&> m_{old}(A) \quad \text{because } W' > W
\end{aligned}$$

7.3. Validation of convergence properties

7.3.1. Continuous positive or negative evidence cancels out initial uncertainty

To prove

Suppose we have an initial uncertain evidence, i.e., $0 < m_1(A) < 1$ for some alternative A (hence also for one or more other alternatives in $\text{dom}(m_1)$). If we were to give continuous positive evidence on that alternative, i.e., $\forall i > 1 : m_i(A) = 1$, then the initial uncertainty for that alternative cancels out, i.e., $\lim_{n \rightarrow \infty} m_{1..n}(A) = 1$. Analogously, continuous negative evidence on that alternative will also cancel out the uncertainty but towards 0, i.e., if $\forall i > 1 : m_i(A) = 0$, then $\lim_{n \rightarrow \infty} m_{1..n}(A) = 0$. Let $W_n = \sum_{i=1}^n w_i$.

Proof

According to operation 1 (Section 5.2.1)

$$\begin{aligned}
 (\forall A \in \mathcal{F}) \quad \lim_{n \rightarrow \infty} m_{1..n}(A) &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \sum_{i=1}^n w_i m_i(A) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \left(w_1 m_1(A) + \sum_{i=2}^n w_i m_i(A) \right) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \left(w_1 m_1(A) + \sum_{i=2}^n w_i \right) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{W_n} (w_1 m_1(A) + (W_n - w_1)) \\
 &= \lim_{n \rightarrow \infty} \left(\frac{w_1 m_1(A)}{W_n} + \frac{W_n - w_1}{W_n} \right) \\
 &= 0 + 1 = 1
 \end{aligned}$$

Because in the limit of $n \rightarrow \infty$ the term $w_1 m_1(A)$ remains constant, while $\lim_{n \rightarrow \infty} W_n = \infty$ and $\lim_{n \rightarrow \infty} \frac{W_n - w_1}{W_n} = 1$.

For the continuous negative evidence case, we get

$$\begin{aligned}
 (\forall A \in \mathcal{F}) \quad \lim_{n \rightarrow \infty} m_{1..n}(A) &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \sum_{i=1}^n w_i m_i(A) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \left(w_1 m_1(A) + \sum_{i=2}^n w_i m_i(A) \right) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{W_n} (w_1 m_1(A) + 0) \\
 &= \lim_{n \rightarrow \infty} \frac{w_1 m_1(A)}{W_n} \\
 &= 0
 \end{aligned}$$

Because in the limit of $n \rightarrow \infty$ the term $w_1 m_1(A)$ remains constant, while $\lim_{n \rightarrow \infty} W_n = \infty$.

7.3.2. Random pick

To prove

Suppose we have some frame of discernment \mathcal{F} and we continuously give random evidence with the same weight, then we expect the combined evidence to even out, i.e., evaluate each possible diagnosis as equally likely. We focus on the following form of giving random evidence. We pick a label a randomly and provide new evidence

$$m(A) = \begin{cases} 1 & \text{if } A = \{a\} \\ 0 & \text{otherwise} \end{cases}$$

Proof

Note that we only give evidence on singular alternatives $\{a\}$. Therefore, $m_i(A) = 0$ for non-singular alternatives, hence $m_{1..n}(A) = 0$ as well. Let k be the number of singular alternatives, i.e., $k = |\mathcal{L}|$. Let $J_a = \{i \in [1..n] \mid m_i(\{a\}) = 1\}$. According to operation 1 (Section 5.2.1)

$$\begin{aligned} (\forall a \in \mathcal{L}) \quad \lim_{n \rightarrow \infty} m_{1..n}(\{a\}) &= \lim_{n \rightarrow \infty} \frac{1}{W_n} \sum_{i=1}^n w_i m_i(\{a\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in J_a} m_i(\{a\}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i \in J_a} 1 \\ &= \lim_{n \rightarrow \infty} \frac{|J_a|}{n} \\ &= \frac{1}{k} \end{aligned}$$

because we pick a randomly out of k possibilities, hence $|J_a| = \frac{n}{k}$ when $n \rightarrow \infty$.

8. Storing evidence with lineage and versioning in a probabilistic database

8.1. Probabilistic databases

In general, a DBMS is a robust and safe place for storing, querying and manipulating important data. Medical data, in particular the EEGs in this paper, as well as the pieces of evidence collected during the diagnosis process would benefit from residing in a database. We argue that probabilistic database technology is well suited to store medical data and associated evidence because of its native handling of uncertainty.

There are several representation systems for uncertain data. The ULDB model (Uncertainty-Lineage Databases) is an extension of the classic RDBMS model where data uncertainty and lineage are stored alongside the data itself (Benjeloun et al.; 2006, 2008; Widom; 2009). This model, as well as several others in the database community such as MayBMS (Antova et al.; 2009, 2008), is based

on possible worlds semantics. An important application domain for probabilistic databases is data integration (Magnani and Montesi; 2010), because the explicit treatment of the uncertainty surrounding (integrated) data is an important technique for achieving better data quality (van Keulen; 2012).

While the ULDB-model has been shown to be adequate for storing aleatory uncertainty, its quantification of uncertainty in terms of probabilities doesn't allow it to store data with epistemic uncertainty. Examples of epistemic uncertainty are missing confidence values and imprecisely or coarsely defined confidence values. (Agrawal and Widom; 2010) introduces a generalization of the ULDB-model in which confidence values are taken to be basic probability assignments as introduced by the Dempster-Shafer theory. The generalized model still follows the possible worlds semantics. Since, in practice, medical data, in particular medical diagnosis, is rarely defined in terms of precise probabilities, we believe that the generalized ULDB model introduced in (Agrawal and Widom; 2010) is suitable for storing evidence data in compliance with the model described in this paper.

Being able to incorporate meta-evidence, i.e., evidence that says something about earlier evidence, such as invalidating it or reducing its reliability, is useful and reflects the true evolution of a diagnosis during the diagnosis determination process. The importance of meta-evidence has been confirmed in other domains. Several feedback models have been proposed. (de Campos et al.; 2003; Xin and Jin; 2004) use Bayesian networks to implement relevance feedback. (de Keijzer and van Keulen; 2007) and (van Keulen and de Keijzer; 2009) show how user feedback (positive and negative feedback) can be used to improve the quality of data within their proposed probabilistic XML data integration framework (based on the possible worlds semantics). It is important to observe that these models are based on classic probability theory and are restricted to positive and negative feedback. Feedback that introduces a new alternative, which is a frequent type of feedback as shown in Section 5, is not supported. Moreover, epistemic uncertainty cannot be properly represented. Therefore, we consider that such models, although supporting meta-evidence in the form of feedback, are unsuitable for our purpose of supporting the incremental diagnostic process.

8.2. Motivating example

A feature commonly supported in probabilistic databases is *lineage* or *data provenance*. By *lineage* or *data provenance*, we mean meta-data on which piece of data was derived from which other pieces of data.

Figure 6 illustrates the importance of lineage for a medical diagnosis support system. Suppose a clinician has two patients: patient p_1 and patient p_2 . Patient p_1 is initially diagnosed with d_1 based on evidence e_1 and subsequently his diagnosis was revised to d_2 based on evidence e_2 . At this point, the clinician seeking to diagnose patient p_2 notices similarities between patient p_1 and patient p_2 's cases and reaches a first diagnosis (d'_1) for patient p_2 by comparison with patient p_1 's current diagnosis (d_2). This comparison is in essence a piece of evidence as well (e_3). Patient p_2 's diagnosis is subsequently revised (d'_2 and d'_3) based on evidences e_4 and e_5 . At one point in time, new evidence e_6 leads to a new diagnosis d_3 for patient p_1 . Given that patient p_2 's diagnosis was derived from patient p_1 's diagnosis, it is important to trace back these derivations for determining how

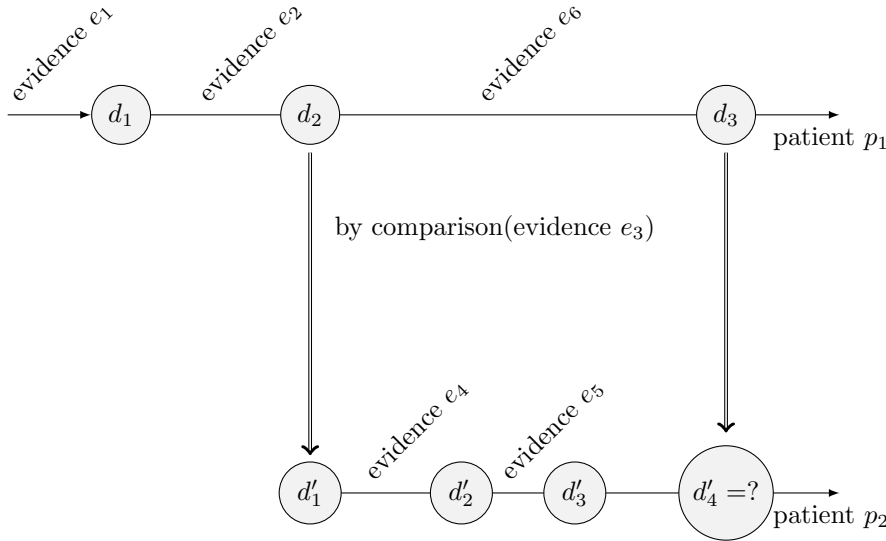


Fig. 6. Motivation

the new evidence leading to d_3 affects patient p_2 's diagnosis. Lineage is needed in the database to be able to do this retracing.

Furthermore, for auditing and quality assurance purposes as well as for rolling back wrong modifications (e.g., faulty evidence due to material problems distorting test results or fraudulent behavior by a clinician), it is important to be able to review all modifications to pieces of evidence. Operations 2, 3a, and 3b all modify the evidence data in the databases. Standard versioning support is sufficient to allow the retracing not only of derivation chains, but also of all modifications that happened during a diagnostic process.

8.3. Data representation

All the data gathered throughout the diagnostic process corresponding to the example described above can be stored with the table structure outlined in Figure 2. The data structure separates lineage (which tuples are derived from which other tuples), evidence versions (for example, updating weights or splitting labels gives rise to new versions of evidence tuples), and a current diagnosis per patient (in essence data derived from the evidence table). Note that we focus on lineage and versioning, hence have not added tables and attributes for storing data from measurements, tests, EEGs and other types of evidence. These can be added in a straightforward manner.

We have chosen the generalized ULDB-model as described in (Agrawal and Widom; 2010) as a database model. While the generalized uncertain database model preserves the concepts of *alternatives*, *x-tuples* and *lineage*, it substitutes the probabilities assigned to each alternative with a basic probability assignment (bpa). This allows us to store a whole mass function in the **bpa** attribute. We refer to (Agrawal and Widom; 2010) for the representation of the other cases

Table	Attribute	Description
diagnosis	Current diagnosis as combined bpa	
	pid	patient identifier
	totweight	total weight
	bpa	combined basic probability assignment
evidence	Current versions of pieces of evidence	
	pid	patient identifier
	evid	evidence identifier
	version	evidence version identifier
	weight	weight (reliability score)
	bpa	basic probability assignment
evidence_version	Versions of pieces of evidence (all but current)	
	pid	patient identifier
	evid	evidence identifier
	version	evidence version identifier
	weight	weight (reliability score)
	bpa	basic probability assignment
lineage	Provenance relationships	
	from	Tuple identifier
	to	Tuple identifier of tuple where from is derived from
	type	Lineage type (e.g., new-version-of, new-diagnosis)
	explanation	description

Table 2. Data representation for diagnosis derivation storage based on lineage and versioning.

(coarse confidence values and imprecise confidence values). In everything that follows, we will not consider ‘maybe’ annotations although they can in theory exist in a generalized uncertain database model. Note that our model is not really specific to a relational database and may be used with few alterations with other types of databases, most notably XML databases (van Keulen and de Keijzer; 2009).

8.4. Data generated for the motivating example

In what follows, we explain how lineage and versioning come into play when new evidence is gathered by showing how the tables are filled for the example outlined in Figure 6.

Step 1: Evidence e_1 for patient 1 (p_1) and initial diagnosis d_1

Before patient p_1 is first diagnosed, some piece of evidence e_1 is gathered and on the basis of e_1 , a first diagnosis for patient p_1 is formed: diagnosis d_1 . The resulting database is depicted in Figure 7. Information about evidence e_1 is added to table **evidence** and information about diagnosis d_1 to table **diagnosis**. Since e_1 is the most recent evidence available about patient p_1 ’s condition, information about e_1 does not need to be copied into table **evidence_version** at this point. Table **lineage** stores the provenance of diagnosis d_1 . It consists of one relationship (arrow l_1) representing the fact that diagnosis d_1 was based on version 1 of evidence e_1 . This is represented in the table **lineage** with **from** and **to** fields containing a tuple identification of the form *table.pid.evid* optionally followed by an evidence version.

Step 2: New evidence e_2 on patient p_1 leading to new diagnosis d_2

New evidence e_2 is gathered on patient p_1 ’s condition and on the basis of e_2 ,

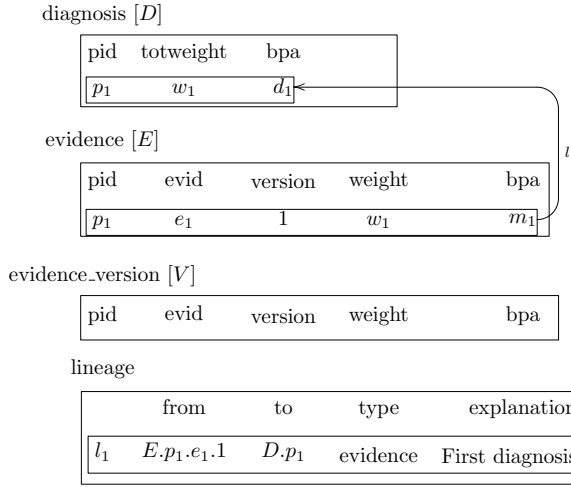


Fig. 7. Table contents after step 1: Evidence e_1 for patient p_1 and initial diagnosis d_1

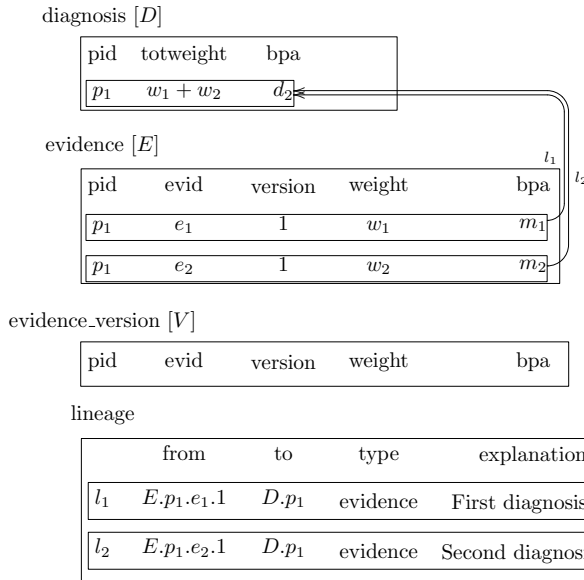


Fig. 8. Table contents after step 2: New evidence e_2 on patient p_1 leading to new diagnosis d_2

diagnosis d_1 is revised to become diagnosis d_2 . The resulting database is depicted in Figure 8. At this point, there are two pieces of evidence and the current diagnosis d_2 is the combination of the bpa's m_1 and m_2 associated with evidences e_1 and e_2 , respectively.

Step 3: Initial diagnosis d'_1 for patient p_2 obtained by comparison with diagnosis d_2 for patient p_1

When patient p_2 presents him/herself for diagnosis, the clinician notices similar-

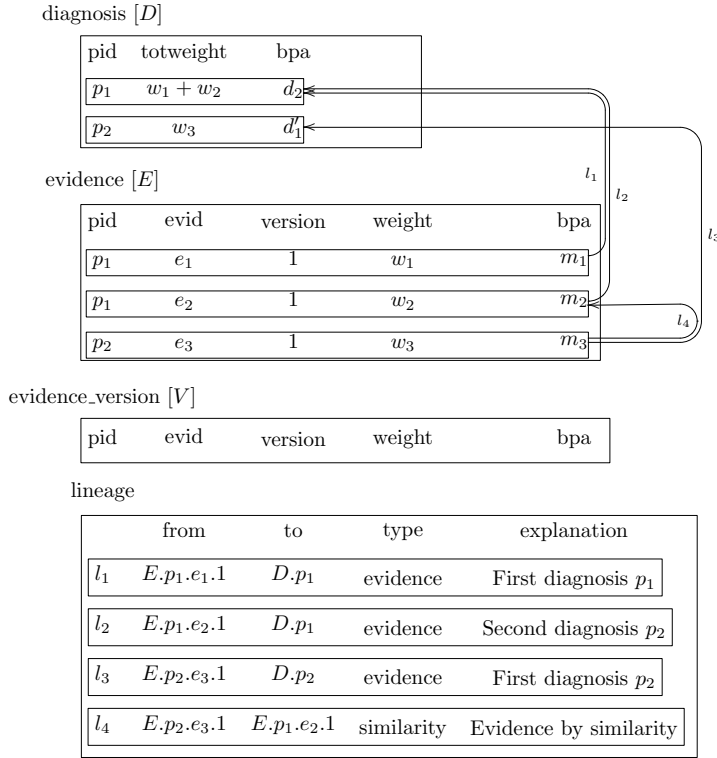


Fig. 9. Table contents after step 3: Initial diagnosis d'_1 for patient p_2 obtained by comparison with diagnosis d_2 for patient p_1 .

ities between patient p_1 and patient p_2 's conditions (evidence e_3) and establishes diagnosis d'_1 for patient p_2 by comparison with diagnosis d_2 of patient 1. The resulting database is depicted in Figure 9. The 'evidence by similarity' gives rise to lineage l_4 of another type **similarity**.

Step 4: New evidence e_4 and e_5 for patient p_2

The diagnosis for patient p_2 is revised successively into diagnosis d'_2 and d'_3 when new pieces of evidence e_4 followed by e_5 are gathered. Suppose for illustration purposes that evidence e_5 shows that the interpretation of evidence e_4 is incorrect and should be discarded analogous to the video showing the toothbrush activity causing the artifact in our running example. The resulting database is depicted in Figure 10. The discarding of evidence e_4 gives rise to lineage l_7 of yet another type **version**: the evidence is cancelled out by updating weight w_4 and set it to $w'_4 = 0$. This creates a new version of the tuple.

Step 5: New evidence e_6 on patient p_1

Finally, new evidence e_6 is obtained for patient p_1 . This not only leads to a revised diagnosis for patient p_1 , but also for patient p_2 , since evidence e_3 depends on the case of patient p_1 by means of similarity. The resulting database is depicted in Figure 11. Lineage l_4 points to the fact that evidence e_3 needs to be revised. We

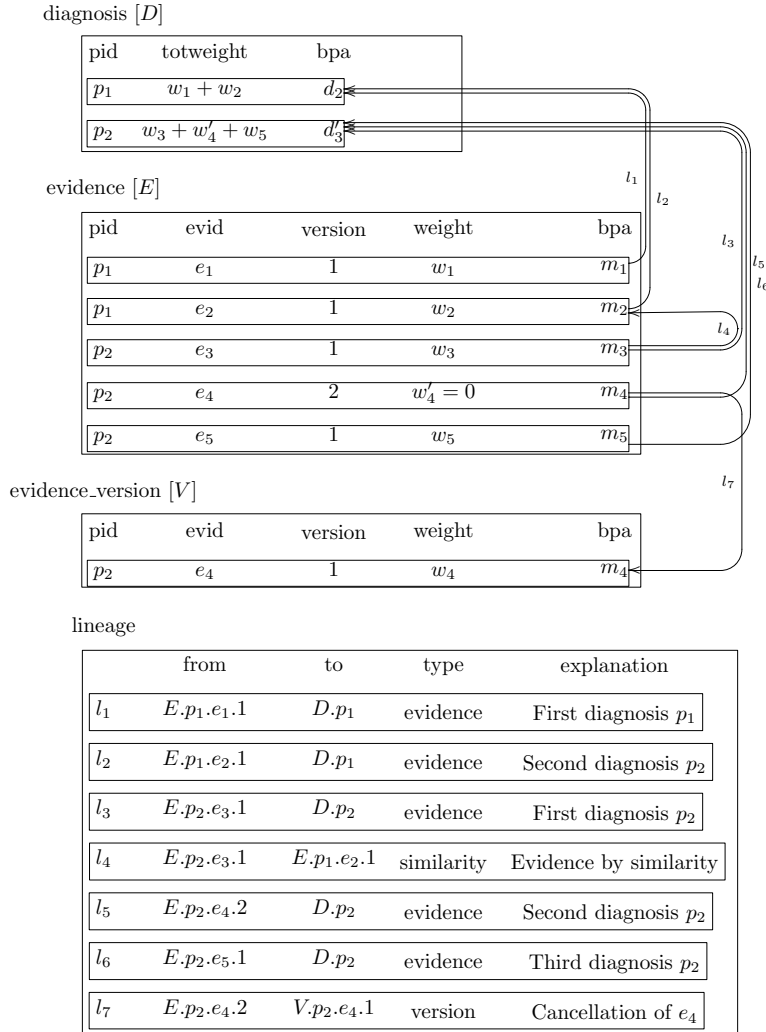


Fig. 10. Table contents after step 4: New evidence e_4 and e_5 for patient p_2 .

create a new version for this evidence and update its bpa to m'_3 accordingly. This revision is probably a manual action, i.e., the clinician receives a warning from the system that the interpretation of e_3 needs to be revised, because it depends on similarity with a patient whose diagnosis has changed.

8.5. Conclusion

In this Section, we have shown how to store and manage evidence and diagnosis in a probabilistic database with support for lineage and versioning. Based on an elaborate example where the diagnosis of one patient depends on the similarity with another patient, we show how versioning and lineage data is gathered and

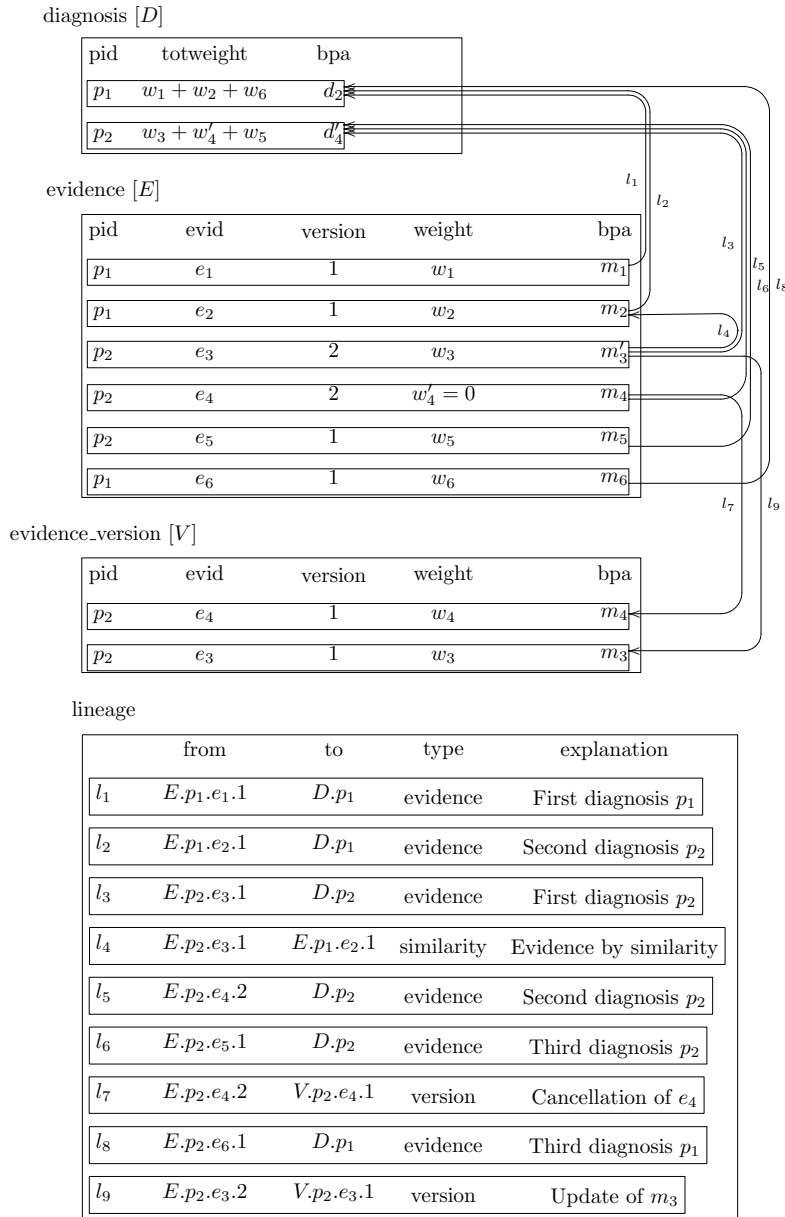


Fig. 11. Table contents after step 5: New evidence e_6 on patient p_1 .

used. Versioning is rather straightforward: whenever a tuple needs updating either because of a new weight or bpa, it is copied to the `evidence_version` table and its version number is increased, thus enabling one to retrace all historic versions of the tuple. Lineage can be stored with an additional table holding data on different types of dependencies between tuples. We distinguished three:

1. **evidence**: lineage from a combined diagnosis to its underlying pieces of evidence.
2. **similarity**: lineage from a piece of evidence to another indicating that the evidence was obtained by similarity with another patient’s condition.
3. **version**: lineage from a piece of evidence to its previous version.

Note that there are many variations possible on the table structure proposed in this section.

9. Conclusion

In this paper, we propose an evidence combination model targeted at medical diagnostic processes that (a) is based on Dempster-Shafer theory, (b) provides a combined diagnosis constructed from all known evidence and opinions known so far at a point in time, (c) allows refinement of considered alternatives as well as addition of new alternatives during the diagnostic process, and (d) supports the inclusion of meta-evidence, i.e., feedback from one clinician on the diagnosis of another or evidence that reduces or nullifies the reliability of an earlier diagnosis. Furthermore, it can handle situations where confidence values are either missing (by far the most frequent case in practice), or coarsely or imprecisely defined. In other words, it tackles both aleatory and epistemic uncertainty.

The model itself is based on evidence being represented by mass functions, a mixing rule based on a normalized weighted average, and 3 atomic operations: adding new evidence, updating weights, and refining the considered alternatives by either splitting a label or adding new alternatives. The model has been validated analytically by proving several correctness, monotonicity, and convergence properties. We also discussed how the model can be implemented with probabilistic database technology. We have illustrated the model with two running examples: 1. the toothbrush case where an initial disagreement between clinicians was settled by a later viewing of a video recording which proved that the abnormal EEG segment was the result of the patient brushing his teeth. 2. the hemochromatosis case where a repeated diagnosis for diabetes which didn’t even really match all symptoms was only rather late turned to a diagnosis for an initially not considered rare condition called hemochromatosis.

In the introduction, we pointed out the fact that the evidence combination model was a first step into the process of building a system to support the medical diagnosis process. The goal of the diagnosis support system is to increase the chances for an early correct diagnosis. An important building block is an automatic component (for instance, a rule-based, similarity search-based or machine learning-based component) that, given the often disregarded patient history in his/her dossier and all the clues at hand, outputs a longer list of likely enough to be considered alternatives — be they rare conditions (“zebras”) or common ones. Rare conditions may enter the scene is multiple clues can be found that correlate with the rare condition and which are unlikely to occur together without the condition. Such an automatic diagnosis can be mixed with the other evidences using

the model in this paper. In this way, it can early on bring history-based evidence and rare conditions worthy of consideration to the attention of the clinician in an unobtrusive way, so that correct diagnoses and treatments are deduced faster in complex and ambiguous situations. Overlooking “zebras” is less likely and at least will occur in shorter periods of time.

Acknowledgements. We would like to thank Michel van Putten from the Technical Medicine department at the University of Twente for providing the EEG data shown in this paper as well as the data and specifics linked to the toothbrush case example used in this paper.

References

- Agrawal, P. and Widom, J. (2010). Generalized Uncertain Databases: First Steps, in A. de Keijzer and M. van Keulen (eds), *MUD*, Vol. WP10-04 of *CTIT Workshop Proceedings Series*, Centre for Telematics and Information Technology (CTIT), University of Twente, The Netherlands, pp. 99–111.
- American Clinical Neurophysiology Society (2006). Guideline 8: Guidelines for recording clinical eeg on digital media, *Journal of clinical neurophysiology* **23**: 122–124.
- Antova, L., Jansen, T., Koch, C. and Olteanu, D. (2008). Fast and simple relational processing of uncertain data, *Proceedings of the 24th International IEEE Conference on Data Engineering (ICDE 2008)*, IEEE, pp. 983–992.
- Antova, L., Koch, C. and Olteanu, D. (2009). 10^{10^6} worlds and beyond: Efficient representation and processing of incomplete information, *The VLDB Journal* **18**(5): 1021–1040.
URL: <http://dx.doi.org/10.1007/s00778-009-0149-y>
- Benjelloun, O., Das Sarma, A., Halevy, A., Theobald, M. and Widom, J. (2008). Databases with uncertainty and lineage, *The VLDB Journal* **17**: 243–264.
 10.1007/s00778-007-0080-z.
- Benjelloun, O., Sarma, A. D., Halevy, A. Y. and Widom, J. (2006). ULDBs: Databases with Uncertainty and Lineage, *VLDB*, pp. 953–964.
- Boodman, S. G. (2011). Medical mystery: Alcoholism didnt cause mans diabetes and cirrhosis, *The Washington Post* .
URL: http://www.washingtonpost.com/national/medical-mystery-alcoholism-didnt-cause-mans-diabetes-and-cirrhosis/2011/05/19/AGd0hdTH_story.html
- Cloostermans, M., de Vos, C., Heida, T., de Keijzer, A. and van Putten, M. (2009). Monitoring the brain in the adult ICU, *Proceedings of the 4th Annual Symposium of the IEEE/EMBS Benelux Chapter*, pp. 128–130.
URL: <http://doc.utwente.nl/70365/>
- Cooke, R. M. and Goossens, L. L. (2008). TU Delft expert judgment data base , *Reliability Engineering & System Safety* **93**(5): 657 – 674. `je:ce:title;Expert Judgement;je:ce:title;`
- Croskerry, P. (2003). The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them, *Academic Medicine* **78**(8): 775–780.
- de Campos, L. M., Fernández-Luna, J. M. and Huete, J. F. (2003). Implementing relevance feedback in the bayesian network retrieval model, *J. Am. Soc. Inf. Sci. Technol.* **54**(4): 302–313.

- de Keijzer, A. and van Keulen, M. (2007). User Feedback in Probabilistic Integration, *Second International Workshop on Flexible Database and Information System Technology (FlexDBIST 2007), Regensburg, Germany*, IEEE Computer Society Press, Los Alamitos, pp. 377–381.
- Dwyer, J. (2012a). Death of a boy prompts new medical efforts nationwide, *The New York Times* .
URL: <http://www.nytimes.com/2012/10/26/nyregion/tale-of-rory-stauntons-death-prompts-new-medical-efforts-nationwide.html>
- Dwyer, J. (2012b). An infection, unnoticed, turns unstoppable, *The New York Times* .
URL: [http://bmjopen.bmj.com/content/4/3/e004007.abstract](http://www.nytimes.com/2012/07/12/nyregion/in-rory-stauntons-fight-for-his-life-signs-that-went-unheeded.html?pagewanted=all&_r=2&_t=2&_u=2&_v=2&_w=2&_x=2&_y=2&_z=2&_aa=2&_ab=2&_ac=2&_ad=2&_ae=2&_af=2&_ag=2&_ah=2&_ai=2&_aj=2&_ak=2&_al=2&_am=2&_an=2&_ao=2&_ap=2&_aq=2&_ar=2&_as=2&_at=2&_au=2&_av=2&_aw=2&_ax=2&_ay=2&_az=2&_ba=2&_bb=2&_bc=2&_bd=2&_be=2&_bf=2&_bg=2&_bh=2&_bi=2&_bj=2&_bk=2&_bl=2&_bm=2&_bn=2&_bo=2&_bp=2&_bq=2&_br=2&_bs=2&_bt=2&_bu=2&_bv=2&_bw=2&_bx=2&_by=2&_bz=2&_ca=2&_cb=2&_cc=2&_cd=2&_ce=2&_cf=2&_cg=2&_ch=2&_ci=2&_cj=2&_ck=2&_cl=2&_cm=2&_cn=2&_co=2&_cp=2&_cq=2&_cr=2&_cs=2&_ct=2&_cu=2&_cv=2&_cw=2&_cx=2&_cy=2&_cz=2&_da=2&_db=2&_dc=2&_dd=2&_de=2&_df=2&_dg=2&_dh=2&_di=2&_dj=2&_dk=2&_dl=2&_dm=2&_dn=2&_do=2&_dp=2&_dq=2&_dr=2&_ds=2&_dt=2&_du=2&_dv=2&_dw=2&_dx=2&_dy=2&_dz=2&_ea=2&_eb=2&_ec=2&_ed=2&_ee=2&_ef=2&_eg=2&_eh=2&_ei=2&_ej=2&_ek=2&_el=2&_em=2&_en=2&_eo=2&_ep=2&_eq=2&_er=2&_es=2&_et=2&_eu=2&_ev=2&_ew=2&_ex=2&_ey=2&_ez=2&_fa=2&_fb=2&_fc=2&_fd=2&_fe=2&_ff=2&_fg=2&_fh=2&_fi=2&_fj=2&_fk=2&_fl=2&_fm=2&_fn=2&_fo=2&_fp=2&_fq=2&_fr=2&_fs=2&_ft=2&_fu=2&_fv=2&_fw=2&_fx=2&_fy=2&_fz=2&_ga=2&_gb=2&_gc=2&_gd=2&_ge=2&_gf=2&_gg=2&_gh=2&_gi=2&_gj=2&_gk=2&_gl=2&_gm=2&_gn=2&_go=2&_gp=2&_gq=2&_gr=2&_gs=2&_gt=2&_gu=2&_gv=2&_gw=2&_gx=2&_gy=2&_gz=2&_ha=2&_hb=2&_hc=2&_hd=2&_he=2&_hf=2&_hg=2&_hh=2&_hi=2&_hj=2&_hk=2&_hl=2&_hm=2&_hn=2&_ho=2&_hp=2&_hq=2&_hr=2&_hs=2&_ht=2&_hu=2&_hv=2&_hw=2&_hx=2&_hy=2&_hz=2&_ia=2&_ib=2&_ic=2&_id=2&_ie=2&_if=2&_ig=2&_ih=2&_ii=2&_ij=2&_ik=2&_il=2&_im=2&_in=2&_io=2&_ip=2&_iq=2&_ir=2&_is=2&_it=2&_iu=2&_iv=2&_iw=2&_ix=2&_iy=2&_iz=2&_ja=2&_jb=2&_jc=2&_jd=2&_je=2&_jf=2&_jg=2&_jh=2&_ji=2&_jj=2&_jk=2&_jl=2&_jm=2&_jn=2&_jo=2&_jp=2&_jq=2&_jr=2&_js=2&_jt=2&_ju=2&_jv=2&_jw=2&_jx=2&_jy=2&_jz=2&_ka=2&_kb=2&_kc=2&_kd=2&_ke=2&_kf=2&_kg=2&_kh=2&_ki=2&_kj=2&_kk=2&_kl=2&_km=2&_kn=2&_ko=2&_kp=2&_kq=2&_kr=2&_ks=2&_kt=2&_ku=2&_kv=2&_kw=2&_kx=2&_ky=2&_kz=2&_la=2&_lb=2&_lc=2&_ld=2&_le=2&_lf=2&_lg=2&_lh=2&_li=2&_lj=2&_lk=2&_ll=2&_lm=2&_ln=2&_lo=2&_lp=2&_lq=2&_lr=2&_ls=2&_lt=2&_lu=2&_lv=2&_lw=2&_lx=2&_ly=2&_lz=2&_ma=2&_mb=2&_mc=2&_md=2&_me=2&_mf=2&_mg=2&_mh=2&_mi=2&_mj=2&_mk=2&_ml=2&_mm=2&_mn=2&_mo=2&_mp=2&_mq=2&_mr=2&_ms=2&_mt=2&_mu=2&_mv=2&_mw=2&_mx=2&_my=2&_mz=2&_na=2&_nb=2&_nc=2&_nd=2&_ne=2&_nf=2&_ng=2&_nh=2&_ni=2&_nj=2&_nk=2&_nl=2&_nm=2&_nn=2&_no=2&_np=2&_nq=2&_nr=2&_ns=2&_nt=2&_nu=2&_nv=2&_nw=2&_nx=2&_ny=2&_nz=2&_oa=2&_ob=2&_oc=2&_od=2&_oe=2&_of=2&_og=2&_oh=2&_oi=2&_oj=2&_ok=2&_ol=2&_om=2&_on=2&_oo=2&_op=2&_oq=2&_or=2&_os=2&_ot=2&_ou=2&_ov=2&_ow=2&_ox=2&_oy=2&_oz=2&_pa=2&_pb=2&_pc=2&_pd=2&_pe=2&_pf=2&_pg=2&_ph=2&_pi=2&_pj=2&_pk=2&_pl=2&_pm=2&_pn=2&_po=2&_pp=2&_pq=2&_pr=2&_ps=2&_pt=2&_pu=2&_pv=2&_pw=2&_px=2&_py=2&_pz=2&_qa=2&_qb=2&_qc=2&_qd=2&_qe=2&_qf=2&_qg=2&_qh=2&_qi=2&_qj=2&_qk=2&_ql=2&_qm=2&_qn=2&_qo=2&_qp=2&_qq=2&_qr=2&_qs=2&_qt=2&_qu=2&_qv=2&_qw=2&_qx=2&_qy=2&_qz=2&_ra=2&_rb=2&_rc=2&_rd=2&_re=2&_rf=2&_rg=2&_rh=2&_ri=2&_rj=2&_rk=2&_rl=2&_rm=2&_rn=2&_ro=2&_rp=2&_rq=2&_rr=2&_rs=2&_rt=2&_ru=2&_rv=2&_rw=2&_rx=2&_ry=2&_rz=2&_sa=2&_sb=2&_sc=2&_sd=2&_se=2&_sf=2&_sg=2&_sh=2&_si=2&_sj=2&_sk=2&_sl=2&_sm=2&_sn=2&_so=2&_sp=2&_sq=2&_sr=2&_ss=2&_st=2&_su=2&_sv=2&_sw=2&_sx=2&_sy=2&_sz=2&_ta=2&_tb=2&_tc=2&_td=2&_te=2&_tf=2&_tg=2&_th=2&_ti=2&_tj=2&_tk=2&_tl=2&_tm=2&_tn=2&_to=2&_tp=2&_tq=2&_tr=2&_ts=2&_tt=2&_tu=2&_tv=2&_tw=2&_tx=2&_ty=2&_tz=2&_ua=2&_ub=2&_uc=2&_ud=2&_ue=2&_uf=2&_ug=2&_uh=2&_ui=2&_uj=2&_uk=2&_ul=2&_um=2&_un=2&_uo=2&_up=2&_uq=2&_ur=2&_us=2&_ut=2&_uu=2&_uv=2&_uw=2&_ux=2&_uy=2&_uz=2&_va=2&_vb=2&_vc=2&_vd=2&_ve=2&_vf=2&_vg=2&_vh=2&_vi=2&_vj=2&_vk=2&_vl=2&_vm=2&_vn=2&_vo=2&_vp=2&_vq=2&_vr=2&_vs=2&_vt=2&_vu=2&_vv=2&_vw=2&_vx=2&_vy=2&_vz=2&_wa=2&_wb=2&_wc=2&_wd=2&_we=2&_wf=2&_wg=2&_wh=2&_wi=2&_wj=2&_wk=2&_wl=2&_wm=2&_wn=2&_wo=2&_wp=2&_wq=2&_wr=2&_ws=2&_wt=2&_wu=2&_wv=2&_ww=2&_wx=2&_wy=2&_wz=2&_xa=2&_xb=2&_xc=2&_xd=2&_xe=2&_xf=2&_xg=2&_xh=2&_xi=2&_xj=2&_xk=2&_xl=2&_xm=2&_xn=2&_xo=2&_xp=2&_xq=2&_xr=2&_xs=2&_xt=2&_xu=2&_xv=2&_xw=2&_xx=2&_xy=2&_xz=2&_ya=2&_yb=2&_yc=2&_yd=2&_ye=2&_yf=2&_yg=2&_yh=2&_yi=2&_yj=2&_yk=2&_yl=2&_ym=2&_yn=2&_yo=2&_yp=2&_yq=2&_yr=2&_ys=2&_yt=2&_yu=2&_yv=2&_yw=2&_yx=2&_yy=2&_yz=2&_za=2&_zb=2&_zc=2&_zd=2&_ze=2&_zf=2&_zg=2&_zh=2&_zi=2&_zj=2&_zk=2&_zl=2&_zm=2&_zn=2&_zo=2&_zp=2&_zq=2&_zr=2&_zs=2&_zt=2&_zu=2&_zv=2&_zw=2&_zx=2&_zy=2&_zz=2&_aa=2&_ab=2&_ac=2&_ad=2&_ae=2&_af=2&_ag=2&_ah=2&_ai=2&_aj=2&_ak=2&_al=2&_am=2&_an=2&_ao=2&_ap=2&_aq=2&_ar=2&_as=2&_at=2&_au=2&_av=2&_aw=2&_ax=2&_ay=2&_az=2&_ba=2&_bb=2&_bc=2&_bd=2&_be=2&_bf=2&_bg=2&_bh=2&_bi=2&_bj=2&_bk=2&_bl=2&_bm=2&_bn=2&_bo=2&_bp=2&_bq=2&_br=2&_bs=2&_bt=2&_bu=2&_bv=2&_bw=2&_bx=2&_by=2&_bz=2&_ca=2&_cb=2&_cc=2&_cd=2&_ce=2&_cf=2&_cg=2&_ch=2&_ci=2&_cj=2&_ck=2&_cl=2&_cm=2&_cn=2&_co=2&_cp=2&_cq=2&_cr=2&_cs=2&_ct=2&_cu=2&_cv=2&_cw=2&_cx=2&_cy=2&_cz=2&_da=2&_db=2&_dc=2&_dd=2&_de=2&_df=2&_dg=2&_dh=2&_di=2&_dj=2&_dk=2&_dl=2&_dm=2&_dn=2&_do=2&_dp=2&_dq=2&_dr=2&_ds=2&_dt=2&_du=2&_dv=2&_dw=2&_dx=2&_dy=2&_dz=2&_ea=2&_eb=2&_ec=2&_ed=2&_ee=2&_ef=2&_eg=2&_eh=2&_ei=2&_ej=2&_ek=2&_el=2&_em=2&_en=2&_eo=2&_ep=2&_eq=2&_er=2&_es=2&_et=2&_eu=2&_ev=2&_ew=2&_ex=2&_ey=2&_ez=2&_fa=2&_fb=2&_fc=2&_fd=2&_fe=2&_ff=2&_fg=2&_fh=2&_fi=2&_fj=2&_fk=2&_fl=2&_fm=2&_fn=2&_fo=2&_fp=2&_fq=2&_fr=2&_fs=2&_ft=2&_fu=2&_fv=2&_fw=2&_fx=2&_fy=2&_fz=2&_ga=2&_gb=2&_gc=2&_gd=2&_ge=2&_gf=2&_gg=2&_gh=2&_gi=2&_gj=2&_gk=2&_gl=2&_gm=2&_gn=2&_go=2&_gp=2&_gq=2&_gr=2&_gs=2&_gt=2&_gu=2&_gv=2&_gw=2&_gx=2&_gy=2&_gz=2&_ha=2&_hb=2&_hc=2&_hd=2&_he=2&_hf=2&_hg=2&_hh=2&_hi=2&_hj=2&_hk=2&_hl=2&_hm=2&_hn=2&_ho=2&_hp=2&_hq=2&_hr=2&_hs=2&_ht=2&_hu=2&_hv=2&_hw=2&_hx=2&_hy=2&_hz=2&_ia=2&_ib=2&_ic=2&_id=2&_ie=2&_if=2&_ig=2&_ih=2&_ii=2&_ij=2&_ik=2&_il=2&_im=2&_in=2&_io=2&_ip=2&_iq=2&_ir=2&_is=2&_it=2&_iu=2&_iv=2&_iw=2&_ix=2&_iy=2&_iz=2&_ja=2&_jb=2&_jc=2&_jd=2&_je=2&_jf=2&_jg=2&_jh=2&_ji=2&_jj=2&_jk=2&_jl=2&_jm=2&_jn=2&_jo=2&_jp=2&_jq=2&_jr=2&_js=2&_jt=2&_ju=2&_jv=2&_jw=2&_jx=2&_jy=2&_jz=2&_ka=2&_kb=2&_kc=2&_kd=2&_ke=2&_kf=2&_kg=2&_kh=2&_ki=2&_kj=2&_kl=2&_km=2&_kn=2&_ko=2&_kp=2&_kq=2&_kr=2&_ks=2&_kt=2&_ku=2&_kv=2&_kw=2&_kx=2&_ky=2&_kz=2&_la=2&_lb=2&_lc=2&_ld=2&_le=2&_lf=2&_lg=2&_lh=2&_li=2&_lj=2&_lk=2&_ll=2&_lm=2&_ln=2&_lo=2&_lp=2&_lq=2&_lr=2&_ls=2&_lt=2&_lu=2&_lv=2&_lw=2&_lx=2&_ly=2&_lz=2&_ma=2&_mb=2&_mc=2&_md=2&_me=2&_mf=2&_mg=2&_mh=2&_mi=2&_mj=2&_mk=2&_ml=2&_mm=2&_mn=2&_mo=2&_mp=2&_mq=2&_mr=2&_ms=2&_mt=2&_mu=2&_mv=2&_mw=2&_mx=2&_my=2&_mz=2&_na=2&_nb=2&_nc=2&_nd=2&_ne=2&_nf=2&_ng=2&_nh=2&_ni=2&_nj=2&_nk=2&_nl=2&_nm=2&_nn=2&_no=2&_np=2&_nq=2&_nr=2&_ns=2&_nt=2&_nu=2&_nv=2&_nw=2&_nx=2&_ny=2&_nz=2&_oa=2&_ob=2&_oc=2&_od=2&_oe=2&_of=2&_og=2&_oh=2&_oi=2&_oj=2&_ok=2&_ol=2&_om=2&_on=2&_oo=2&_op=2&_oq=2&_or=2&_os=2&_ot=2&_ou=2&_ov=2&_ow=2&_ox=2&_oy=2&_oz=2&_pa=2&_pb=2&_pc=2&_pd=2&_pe=2&_pf=2&_pg=2&_ph=2&_pi=2&_pj=2&_pk=2&_pl=2&_pm=2&_pn=2&_po=2&_pp=2&_pq=2&_pr=2&_ps=2&_pt=2&_pu=2&_pv=2&_pw=2&_px=2&_py=2&_pz=2&_qa=2&_qb=2&_qc=2&_qd=2&_qe=2&_qf=2&_qg=2&_qh=2&_qi=2&_qj=2&_qk=2&_ql=2&_qm=2&_qn=2&_qo=2&_qp=2&_qq=2&_qr=2&_qs=2&_qt=2&_qu=2&_qv=2&_qw=2&_qx=2&_qy=2&_qz=2&_ra=2&_rb=2&_rc=2&_rd=2&_re=2&_rf=2&_rg=2&_rh=2&_ri=2&_rj=2&_rk=2&_rl=2&_rm=2&_rn=2&_ro=2&_rp=2&_rq=2&_rr=2&_rs=2&_rt=2&_ru=2&_rv=2&_rw=2&_rx=2&_ry=2&_rz=2&_sa=2&_sb=2&_sc=2&_sd=2&_se=2&_sf=2&_sg=2&_sh=2&_si=2&_sj=2&_sk=2&_sl=2&_sm=2&_sn=2&_so=2&_sp=2&_sq=2&_sr=2&_ss=2&_st=2&_su=2&_sv=2&_sw=2&_sx=2&_sy=2&_sz=2&_ta=2&_tb=2&_tc=2&_td=2&_te=2&_tf=2&_tg=2&_th=2&_ti=2&_tj=2&_tk=2&_tl=2&_tm=2&_tn=2&_to=2&_tp=2&_tq=2&_tr=2&_ts=2&_tt=2&_tu=2&_tv=2&_tw=2&_tx=2&_ty=2&_tz=2&_ua=2&_ub=2&_uc=2&_ud=2&_ue=2&_uf=2&_ug=2&_uh=2&_ui=2&_uj=2&_uk=2&_ul=2&_um=2&_un=2&_uo=2&_up=2&_uq=2&_ur=2&_us=2&_ut=2&_uu=2&_uv=2&_uw=2&_ux=2&_uy=2&_uz=2&_va=2&_vb=2&_vc=2&_vd=2&_ve=2&_vf=2&_vg=2&_vh=2&_vi=2&_vj=2&_vk=2&_vl=2&_vm=2&_vn=2&_vo=2&_vp=2&_vq=2&_vr=2&_vs=2&_vt=2&_vu=2&_vv=2&_vw=2&_vx=2&_vy=2&_vz=2&_wa=2&_wb=2&_wc=2&_wd=2&_we=2&_wf=2&_wg=2&_wh=2&_wi=2&_wj=2&_wk=2&_wl=2&_wm=2&_wn=2&_wo=2&_wp=2&_wq=2&_wr=2&_ws=2&_wt=2&_wu=2&_wv=2&_ww=2&_wx=2&_wy=2&_wz=2&_xa=2&_xb=2&_xc=2&_xd=2&_xe=2&_xf=2&_xg=2&_xh=2&_xi=2&_xj=2&_xk=2&_xl=2&_xm=2&_xn=2&_xo=2&_xp=2&_xq=2&_xr=2&_xs=2&_xt=2&_xu=2&_xv=2&_xw=2&_xx=2&_xy=2&_xz=2&_ya=2&_yb=2&_yc=2&_yd=2&_ye=2&_yf=2&_yg=2&_yh=2&_yi=2&_yj=2&_yk=2&_yl=2&_ym=2&_yn=2&_yo=2&_yp=2&_yq=2&_yr=2&_ys=2&_yt=2&_yu=2&_yv=2&_yw=2&_yx=2&_yy=2&_yz=2&_za=2&_zb=2&_zc=2&_zd=2&_ze=2&_zf=2&_zg=2&_zh=2&_zi=2&_zj=2&_zk=2&_zl=2&_zm=2&_zn=2&_zo=2&_zp=2&_zq=2&_zr=2&_zs=2&_zt=2&_zu=2&_zv=2&_zw=2&_zx=2&_zy=2&_zz=2</p>
<p>Eta S. Berner, M. L. G. (2008). Overconfidence as a cause of diagnostic error in medicine, <i>Am. J. Med</i> 121(5): S2–S23 (Supplement).</p>
<p>Florea, M. C., Josselme, A.-L., Bossé, É. and Grenier, D. (2009). Robust combination rules for evidence theory, <i>Information Fusion</i> 10(2): 183–197.</p>
<p>Graber, M., Gordon, R. and Franklin, N. (2002). Reducing diagnostic errors in medicine: what’s the goal?, <i>Academic Medicine</i> 77(10): 981–992.</p>
<p>Groopman, J. (2007). Medical dispatches: What’s the trouble? How doctors think., <i>The New Yorker</i> .</p>
<p>Gupta, S., Tran, T., Luo, W., Phung, D., Kennedy, R. L., Broad, A., Campbell, D., Kipp, D., Singh, M., Khasraw, M., Matheson, L., Ashley, D. M. and Venkatesh, S. (2014). Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry, <i>BMJ Open</i> 4(3).

URL: <a href=)
- Haenni, R. (2005). Shedding new light on Zadeh’s criticism of Dempster’s rule of combination, *FUSION’05, 8th International Conference on Information Fusion*, Vol. 2, Philadelphia, USA, pp. 879–884.
- Hoffman, J. C. and Murphy, R. R. (1993). Comparison of bayesian and dempster-shafer theory for sensing: A practitioner’s approach, in *SPIE Proc. on Neural and Stochastic Methods in Image and Signal Processing II*, pp. 266–279.
- Kohn, L. T., Corrigan, J. M. and Molla S. Donaldson editors; Committee on Quality of Health Care in America, I. o. M. (2000). *To Err Is Human: Building a Safer Health System*, The National Academies Press.
URL: http://www.nap.edu/openbook.php?record_id=9728
- Kostopoulou, O., Delaney, B. C. and Munro, C. W. (2008). Diagnostic difficulty and error in primary care—a systematic review, *Family Practice* **25**(6): 400–413.
URL: <http://fampra.oxfordjournals.org/content/25/6/400.abstract>
- Leape, L. L. (2000). Institute of medicine medical error figures are not exaggerated, *JAMA* **284**(1): 95–97.
URL: <http://dx.doi.org/10.1001/jama.284.1.95>
- Leape, L. L., Berwick, D. M. and Bates, D. W. (2002). Counting deaths due to medical errors [reply], *JAMA* **288**(19): 2405.
URL: + <http://dx.doi.org/10.1001/jama.288.19.2405-JLT1120-2-3>
- Lodder, S. S. (2014). *Computer assisted interpretation of the human EEG: improving diagnostic efficiency and consistency in clinical reviews*, PhD thesis,

- Enschede.
URL: <http://doc.utwente.nl/89312/>
- Magnani, M. and Montesi, D. (2010). A survey on uncertainty management in data integration, *J. Data and Information Quality* **2**(1): 5:1–5:33.
URL: <http://doi.acm.org/10.1145/1805286.1805291>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity, *Technical report*, Mc Kinsey Global Institute.
URL: <http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology>
- Nardia, D. and Brachman, R. J. (2002). An introduction to description logics, in F. Baader, D. Calvanese, D. McGuinness, D. Nardi and P. Patel-Schneider (eds), *Description Logic Handbook*, Cambridge University Press, pp. 5–44.
- Niedermeyer, E. and Silva, F. (1999). *Electroencephalography: basic principles, clinical applications, and related fields*, Williams & Wilkins.
- O’Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). *Multiple Experts*, John Wiley & Sons, Ltd, pp. 179–192.
- Ouchi, F. (2004). A literature review on the use of expert opinion in probabilistic risk analysis, *Policy Research Working Paper Series 3201*, The World Bank.
URL: <http://ideas.repec.org/p/wbk/wbrups/3201.html>
- Rougier, J., Sparks, S., Hill, L. and Sparks, R. (2013). *Risk and Uncertainty Assessment for Natural Hazards*, Cambridge University Press.
- Salicone, S. (2007). *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence*, Springer Series in Reliability Engineering, Springer US, chapter The Theory of Evidence, pp. 31–71.
- Schiff, G. D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B. L., Elstein, A. S., Hasler, S., Kabongo, M. L., Krosnjar, N., Odwazny, R., Wisniewski, M. F. and McNutt, R. A. (2009). Diagnostic error in medicine: Analysis of 583 physician-reported errors, *Arch Intern Med* **169**(20): 1881–1887.
- Segal, M. (2007). How doctors think, and how software can help avoid cognitive errors in diagnosis, *Acta Paediatrica* **96**(12): 1720–1722.
URL: <http://dx.doi.org/10.1111/j.1651-2227.2007.00480.x>
- Sentz, K. and Ferson, S. (2002). Combination of evidence in Dempster-Shafer theory, *Technical report*.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press, Princeton.
- Singh, H., davis Giardina, T., Meyer, A. N. D., Forjuoh, S. N., Reis, M. D. and Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings, *JAMA Internal Medicine* **173**(6): 418–425.
URL: + <http://dx.doi.org/10.1001/jamainternmed.2013.2777>
- Užga-Rebrovs, O. and Kuļešova, G. (2008). A Comparative Analysis of Alternative Rules of Belief Combination, *Scientific Journal of RTU* **36**: 93–99.
- van Keulen, M. (2012). Managing uncertainty: The road towards better data interoperability, *IT - Information Technology* **54**(3): 138–146.
URL: <http://dx.doi.org/10.1524/itit.2012.0674>
- van Keulen, M. and de Keijzer, A. (2009). Qualitative Effects of Knowledge Rules and User Feedback in Probabilistic Data Integration, *The VLDB Journal* **18**(5): 1191–1217.

- Webber, W., Litt, B., Lesser, R., Fisher, R. and Bankman, I. (1993). Automatic {EEG} spike detection: what should the computer imitate?, *Electroencephalography and Clinical Neurophysiology* **87**(6): 364 – 373.
URL: <http://www.sciencedirect.com/science/article/pii/001346949390149P>
- Widom, J. (2009). *Managing and Mining Uncertain Data*, Vol. 35 of *Advances in Database Systems*, Springer, chapter Trio: A System for Data, Uncertainty, and Lineage, pp. 113–148.
- Xin, J. and Jin, J. S. (2004). Relevance feedback for content-based image retrieval using bayesian network, *Proceedings of the Pan-Sydney area workshop on Visual information processing, VIP '05*, Australian Computer Society, Inc., Darlinghurst, Australia, Australia, pp. 91–94.
- Zadeh, L. A. (1984). Book Review: A Mathematical Theory of Evidence, *AI Magazine* **5**(3): 81–83.

Correspondence and offprint requests to: Maurice van Keulen, Faculty of EEMCS, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. Email: m.vankeulen@utwente.nl