BELIEF THEORETIC METHODS FOR SOFT AND HARD DATA FUSION

T. L. Wickramarathne†, K. Premaratne†, M. N. Murthi†, M. Scheutz‡, S. Kübler*, and M. Pravia†

†Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33146 USA
t.wickramarathne@miami.edu
{kamal.mmurthi}@miami.edu

‡Computer Science, †Informatics,
Indiana University
Bloomington, IN 47406 USA
{mscheutz,skuebler}@indiana.edu

*BAE Systems
6 New England Executive Park
Burlington, Massachusetts, USA
marco.pravia@baesystems.com

ABSTRACT

In many contexts, one is confronted with the problem of extracting information from large amounts of different types of soft data (e.g., text) and hard data (e.g., physics-based sensing systems). In handling hard data, signal and data processing offers a wealth of methods related to modeling, estimation, tracking, and inference tasks. However, soft data present several challenges that necessitate the development of new data processing methods. For example, with suitable statistical natural language processing (NLP) methods, text can be converted into logic statements that are associated with various forms of associated uncertainty related to the credibility of the statement, the reliability of the text source, and so forth. In combining or fusing soft data with either soft or hard data, one must deploy methods that can suitably preserve and update the uncertainty associated with the data, thereby providing uncertainty bounds related to any inferences regarding semantics.

Since standard Bayesian probabilistic approaches have problems with suitably handling uncertain logic statements, there is an emerging need for new methods for processing heterogeneous data. In this paper, we describe a framework for fusing soft and hard data based on the Dempster-Shafer (DS) belief theoretic approach which is well-suited to the task of capturing the types of models and uncertain rules that are more typical of soft data. Since the effectiveness of traditional DS methods has been hampered by high computational requirements, we base the processing framework on our new conditional approach to DS theoretic evidence updating and fusion. We address the issue of laying the foundation for a theoretically justifiable, and computationally efficient framework for fusing soft and hard data taking into account the inherent data uncertainty such as reliability and credibility. Moreover, we present an illustrative example that highlights the potential for the DS conditional approach for fusing heterogeneous data.

Index Terms— Dempster-Shafer belief theory, evidence updating, evidence fusion, soft information, soft information.

1. INTRODUCTION

Motivation: In many contexts, one is confronted with large amounts of different types of soft data (e.g., text from interview transcriptions, written expert opinions, case histories, internet blogs) and hard data (from physics-based sensing systems). For example, in a medical application context, a physician can take into account both soft evidence in the form of text transcriptions of patient statements and text from expert opinions in journal articles, and hard evidence generated from various sensor-based data such as blood-pressure readings to render a judgment about a course of treatment. In a consumer product scenario, a company may want to make sense of data from a large number of customer opinions or complaints about a product, and data from a large number of test measurements associated with the product. In a defense scenario, an intelligence analyst may be confronted with soft data such as COMINT (e.g., communication chatter, telephone records), HUMINT (e.g., informant and interrogation statements, domain expert statements), and OSINT (e.g., newspapers, internet blogs, TV) as well as hard data (e.g., radar, images).

In most of these applications, the amount of soft and hard data is enormous, and the data often are imperfect, thus overwhelming the analyst who must make sense of the data. Consequently, there is interest in developing automated methods for fusing and analyzing soft and hard data to extract meaning. The question of how the more ‘qualitative’ information in soft evidence can be captured and fused with the more ‘quantitative’ information in hard evidence for increased automation of the decision-making process is attracting considerable attention from the data fusion community [1,2].

Challenges: Signal/data processing offers numerous methods for modeling and processing hard data. However, the nature of soft data presents several challenges for automated processing. In particular, often soft data is in the form of text. Consequently, statistical natural language processing (NLP) methods must be utilized to parse the text and subsequently convert the text into a form such as logic statements suitable for fusion. In many contexts, there is inherent uncertainty associated with the text statement, and this uncertainty must be modeled. For example, NLP can result in semantic uncertainty about a statement. In addition to the semantic uncertainty, the source (e.g., informant, blogger) of the text may not be reliable; in addition, the text statement itself may not be credible. Consequently, even if the semantic uncertainty is ignored and the NLP analysis is perfect, the resulting logic statements (e.g., in the form of propositional logic, first-order logic) will have to incorporate an uncertainty that captures the underlying imperfections and this uncertainty must be accounted for as data fusion progresses.

Moreover, one can apply rule mining methods, reasoning on the information in the text, and textual entailment to extract association rules that consist of uncertain implications. Therefore, an automated method for fusing data must be able to successfully grapple with uncertain logic statements and uncertain implications rules, allowing for the calculation of inferences that illustrate the effects of the uncertainties in the original data.

The Bayesian probabilistic framework, often the starting point for many statistical data processing algorithms, has difficulty in capturing the ‘non-numerical’ models that are more typical of soft ev-
idence [3]. Probabilistic models’ inability to preserve the material implications of propositional logic statements that are represented by uncertain rules [4, 5] constitutes a serious drawback that limits their utility in capturing soft evidence. On the other hand, models that are based on the Dempster-Shafer (DS) belief theoretic framework [6] can capture such uncertain rules while preserving the material implications of propositional logic statements that such rules represent, viz., reflexivity, transitivity, and contra-positivity [5].

To apply DS theoretic fusion to soft data, one must confront several additional challenges. First, DS methods must account for sources that may not have the same scope, or frame of discernment (FoD) in DS theoretic jargon. This is due to the fact that soft evidence is likely to be generated from a variety of sources having dissimilar FoDs. For example, the information contained in a public database of vehicles belonging to town residents would have a much larger, but not completely disjoint, scope than the vehicles that had been recorded at a checkpoint. Conventional DS methods based on Dempster’s combination rule (DCR) are not suitable for such problems. In addition, soft and hard data fusion can involve evidence that is contradictory (e.g., two witness statements that differ). The DCR tends to produce counter-intuitive results when it encounters contradictory evidence.

Furthermore, the computational complexity of DS theoretic methods exponentially increases with increasing cardinality of the FoD (which equals the total set of mutually exclusive and exhaustive events that constitutes an evidence source’s ‘scope of expertise’). As a result, many DS-based applications, even the most common and fundamental task of conditioning can quickly become computationally prohibitive, especially in the presence of FoDs with high cardinality.

Contributions: In this paper, we present the foundation of an emerging DS theoretic framework for fusing soft and hard data. Instead of using conventional DS theoretic methods, we base our approach on the conditional approach that possesses a theoretically justified and computationally efficient method for modeling and processing soft data. In particular, the conditional approach can model and account for dissimilar scopes and contradictory evidence in fusion. In addition, the recently identified Conditional Core Theorem (see Theorem 1) provides a valuable basis for carrying out evidence fusion in a significantly more efficient manner. We present an illustrative example which captures the essence and potential of the DS conditional approach to evidence updating and fusion. By providing inferences with associated bounds that account for the inherent uncertainty in the soft data, this DS conditional approach has the potential for emerging soft and hard data fusion applications. Some preliminary work appears in [7, 8].

2. PRELIMINARIES

Basic Notions: In DS theory, the total set of mutually exclusive and exhaustive propositions of interest (i.e., the ‘scope of expertise’) is referred to as the frame of discernment (FoD) \( \Theta = \{ \theta_1, \ldots, \theta_n \} \) [6]. A singleton proposition \( \theta_i \) represents the lowest level of discernible information. Elements in the power set of the FoD, \( 2^{\Theta} \), form all the propositions of interest. We use \( A \setminus B \) to denote all singletons in \( A \) that are not in \( B \); \( A \) denotes \( \Theta \setminus A \).

Definition 1. Consider the FoD \( \Theta \) and \( A \subseteq \Theta \).

(i) The mapping \( m_{\Theta} (\cdot) : 2^{\Theta} \rightarrow [0, 1] \) is a basic belief assignment (BBA) or mass assignment if \( m_{\Theta} (\emptyset) = 0 \) and \( \sum_{A \subseteq \Theta} m_{\Theta} (A) = 1 \). The BBA is said to be vacuous if the only proposition receiving a non-zero mass is \( \Theta \).

(ii) The belief of \( A \) is \( B_{\Theta} (A) = \sum_{B \subseteq A} m_{\Theta} (B) \).

(iii) The plausibility of \( A \) is \( P_{\Theta} (A) = 1 - B_{\Theta} (A^{\complement}) \).

A proposition that possesses non-zero mass is referred to as a focal element; the set of focal elements is the core \( S_{\Theta} \). The triple \( \{ \Theta, S_{\Theta}, m_{\Theta} (\cdot) \} \) is the corresponding body of evidence (BoE). DS theory models the notion of ignorance by allowing composite propositions (i.e., a non-singleton propositions) to be focal elements. While \( m_{\Theta} (A) \) measures the support assigned to proposition \( A \) only, the belief represents the total support that can move into \( A \) without any ambiguity; \( P_{\Theta} (A) \) represents the extent to which one finds \( A \) plausible. When focal elements are constituted of singletons only, the BBA, belief and plausibility all reduce to a probability assignment.

Definition 2 (Dempster’s Combination Rule (DCR)). The DCR-fused BoE \( E = \{ E_1, E_2 \} \) generated from the BoEs \( E_i = \{ \Theta_i, S_{\Theta_i}, m_{\Theta_i} (\cdot) \} \), \( i = 1, 2 \), when \( \Theta_1 \cap \Theta_2 = \emptyset \), is

\[
\begin{align*}
m_{\Theta} (A) & = \sum_{C \cap D = A} m_{\Theta_1} (C) m_{\Theta_2} (D) / (1 - K), \forall A \subseteq \Theta, \\
\text{wherever } K & = \sum_{C \cap D = A} m_{\Theta_1} (C) m_{\Theta_2} (D) \neq 1.
\end{align*}
\]

Note that \( K \in [0, 1] \) is an indication of the conflict between the evidence provided by the BoEs. Hence, \( K \) is referred to as the conflict between the BoEs being fused. The DCR’s difficulties in fusing conflicting BoEs are well documented. Another drawback of the DCR is that it requires the FoDs being fused to be identical.

To fuse evidence generated from non-identical FoDs \( \Theta_1 \) and \( \Theta_2 \) (so that \( \Theta_1 \neq \Theta_2 \) and \( \Theta_1 \cap \Theta_2 \neq \emptyset \)), one can simply ignore the differences in the FoDs by having each source allocate zero mass to propositions that are not within its own FoD and continue applying DCR. In essence, this approach assumes that each source can discern \( \Theta_1 \cup \Theta_2 \) and ignores the fact that some propositions are not within its scope of expertise. The counter-intuitive conclusions this approach may generate are well documented [9]. In the deconditioning approaches, each source would artificially introduce ambiguities or implement ‘ballooning’ extensions into its evidence so that its own FoD is ‘expanded’ to \( \Theta_1 \cup \Theta_2 \).

Conditional Update Equation (CUE): The conditional approach to fusing evidence ‘conditions’ or ‘updates’ the already available evidence with respect to what both FoDs can discern [7, 10]. Once the conditioning operation is performed, each source invokes a strategy to incorporate its originally cast evidence that does not belong to \( \Theta_1 \cap \Theta_2 \). This approach enables a source to update its own knowledge base and exchange information with other sources for the express purpose of refining its own knowledge without having to continually ‘expand’ its FoD.

Definition 3. For the BoEs \( E_i \), the CUE that updates \( E_1 \) with the evidence in \( E_2 \) is \( E_1 [k + 1] = E_1 [k] \cup E_2 [k] \), \( \forall k \geq 0 \), where

\[
\begin{align*}
B_{\Theta_2} (B)[k + 1] & = \alpha [k] B_{\Theta_1} (B)[k] + \frac{\beta (A)[k]}{2} \left\{ \begin{array}{l}
B_{\Theta_2} (B|A)[k] \\
B_{\Theta_2} (B|A)[k] - B_{\Theta_2} (\Theta_2 |A)[k]
\end{array} \right\}, \\
\text{where } \Theta_2 & = \Theta_2 \setminus \Theta_1, D_B = B \cup \Theta_2, \text{ and } B_{\Theta_2} (A)[k] > 0. \text{ The CUE parameters } \{ \alpha [k], \beta (A)[k] \} \text{ are non-negative and satisfy }
\end{align*}
\]

\[
\alpha [k] + \frac{\beta (A)[k]}{2} \left\{ B_{\Theta_2} (\Theta_1 |A)[k] + P_{\Theta_2} (\Theta_1 |A)[k] \right\} = 1,
\]

\[
\alpha [k] + \frac{\beta (A)[k]}{2} \left\{ B_{\Theta_2} (\Theta_1 |A)[k] + P_{\Theta_2} (\Theta_1 |A)[k] \right\} = 1.
\]
with $\beta(A)[k] = 0, \forall A \notin \mathcal{F}_{\Theta_2}$.

Fig. 1. Updating the BoE $E_1 = \{\Theta_1, \mathcal{F}_{\Theta_1}, m_{\Theta_1}(\cdot)\}$ with the evidence of BoE $E_2 = \{\Theta_2, \mathcal{F}_{\Theta_2}, m_{\Theta_2}(\cdot)\}$ when $\Theta_1 \neq \Theta_2$ and $\Theta_1 \cap \Theta_2 \neq \emptyset$. The terms that contribute towards updating of $m_{\Theta_1}(B)$ are also shown.

The CUE can be used to update a given BoE from another BoE which does not necessarily span the same FoD. However, the CUE can still compute the contribution from the propositions that may have an impact on the updated BoE. Fig. 1 shows the masses that contribute towards updating $m_{\Theta_1}(B)[k]$. Notice that a proposition that does not occur in $\Theta_1$ can still have an impact on it depending on the characteristics of the two FoDs and the conditioning event.

The conditional operations in the above definitions are implemented using the Fagin-Halpern (FH) DS conditional operators.

Definition 4. [11] For $E = \{\Theta, \mathcal{F}_{\Theta}, m_{\Theta}(\cdot)\}$, $A, B \subseteq \Theta$ with $Bl_{\Theta}(A) > 0$, the conditional belief of $B$ given $A$ is

$$Bl_{\Theta}(B|A) = \frac{Bl_{\Theta}(A \cap B)}{Bl_{\Theta}(A)} = Bl_{\Theta}(A \cap B) + Pl_{\Theta}(A \setminus B).$$

To efficiently compute the FH conditionals, and hence the CUE, one can invoke the following theorem [8]:

Theorem 1 (Conditional Core Theorem (CCT)). Consider the BoE $E = \{\Theta, \mathcal{F}_{\Theta}, m_{\Theta}(\cdot)\}$. Then, $m(B|A) > 0$, $Bl_{\Theta}(A) > 0$, iff $B$ can be expressed as $B = X \cup Y$, for some $X \in in(A)$, $Y \in out(A) \cup \{\emptyset\}$, where $in(A) = \{B \subseteq A | B \in \mathcal{F}_{\Theta_0}\}$ and $out(A) = \{B \subseteq A | B \cup C \in \mathcal{F}_{\Theta_0}, \emptyset \neq B, \emptyset \not C \subseteq \mathcal{F}_{\Theta_0}\}$.

The CCT identifies the propositions receiving a positive mass after conditioning without any numerical computations. This enables one to avoid computing all the $2^{|A|}$ propositions that otherwise would have to be computed to evaluate the conditional masses. In real application settings, the CCT may yield computational savings of 80% or more (see [8] for further details). The CCT thus lays the foundation to efficiently implement the above mentioned evidence updating strategies that are based on the conditionals.

3. AN ILLUSTRATIVE EXAMPLE

Here, we illustrate how the above mentioned techniques can be used for the task of incorporating soft evidence into the fusion task. The issues of NLP parsing of text, logical form extraction, and conversion to DS theoretic forms are not within the scope of this paper; they are to be addressed in forthcoming papers. We use this example to illustrate how the CUE can be used for the pertinent fusion tasks.

Scenario: A suspicious activity in the proximity of a military base was reported. Reports from the various hard sensors (e.g., metal, magnetic, IR) deployed around the perimeter of the base conforms vehicle activity; and analysis of night vision cameras confines the set of possible vehicles to one in $\Theta_V = \{\text{Jeep, Truck, Car}\} \equiv \{\text{Jp, Tk, Cr}\}$. The task of the base commander is to determine the most probable suspect and the vehicle driven by the suspect.

Setup: The base commander maintains a ‘blacklisted’ group of personnel in $\Theta_x = \{\text{Andy, Bob, Ken, Larry}\} \equiv \{A, B, K, L\}$, from which he usually picks the initial suspects. The commander gets soft evidence from two human witnesses $W_{S1}$ and $W_{S2}$ and also from a public database $DB_3$ containing demographic information on $\Theta_S \times \Theta_II$, where $\Theta_S = \{\text{Tall, Med, Short}\}$ and $\Theta_II = \{\text{Andy, Bob, Chuck, . . . , Jude}\} \equiv \{A, B, C, . . . , J\}$.

$W_{S1}$: “A tall man was driving a truck or jeep” [0.7]
$W_{S2}$: “Andy drives a truck” [0.9]
$DB_3$: (Height = Tall $\implies$ Person = Bob) [0.5, 0.8]

The values in square brackets indicate the confidence each source places on its own evidence. Table 1 shows the reliability and DS theoretic evidence model corresponding to each evidence source. Note that $|\Theta_{II}| = 10$, $\Theta_x \not \subseteq \Theta_{II}$ and $\Theta_S \cap \Theta_x = \{A, B\}$.

The CUE’s ability to handle non-exhaustive frames without having to expand the FoDs using computationally expensive ballooning extensions becomes very handy in this situation.

<table>
<thead>
<tr>
<th>Source</th>
<th>Source Reliability</th>
<th>DS Model</th>
<th>FoD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_{S1}$</td>
<td>$r_1$</td>
<td>$Bo_{E1}$</td>
<td>$\Theta_S \times \Theta_V \times \Theta_x$</td>
</tr>
<tr>
<td>$W_{S2}$</td>
<td>$r_2$</td>
<td>$Bo_{E2}$</td>
<td>$\Theta_S \times \Theta_V \times \Theta_x$</td>
</tr>
<tr>
<td>$DB_3$</td>
<td>$r_3$</td>
<td>$Bo_{E3}$</td>
<td>$\Theta_S \times \Theta_V \times \Theta_II$</td>
</tr>
</tbody>
</table>

Table 1. Evidence models of $W_{S1}$, $W_{S2}$, and $DB_3$.

Modeling: Let us use the following models for the soft evidence provided by witnesses and the database:

$Bo_{E1}: m_1(Tl \times (Jp,Tk) \times \Theta_x) = 0.7r_1$
$m_1(\Theta_S \times \Theta_V \times \Theta_x) = 1 - 0.7r_1$

$Bo_{E2}: m_2(\Theta_S \times Tk \times A) = 0.9r_2$
$m_2(\Theta_S \times \Theta_V \times \Theta_x) = 1 - 0.9r_2$

$Bo_{E3}: m_3(Tl \times \Theta_V \times B) = r_3c_1$
$m_3((Md,St) \times \Theta_V \times \Theta_x) = r_3(1-c_2)$
$m_3(\Theta_S \times \Theta_V \times \Theta_II) = 1 - r_3(1 + c_1 - c_2)$

Remarks:

(i) We use simple, intuitive models to capture the soft evidence from witnesses $W_{S1}$ and $W_{S2}$. The reliability associated with an evidence source is incorporated by simply discounting the initial mass assignments. For instance, if the reliability $r_1$ of $W_{S1}$ is very low, we may want to give a lower weight to the proposition $Tl \times (Jp,Tk) \times \Theta_x$.

(ii) Textual information or expert opinions are often modeled as logical implication rules. The ability to combine such evidence into hard evidence is of significant importance, especially in military, medical, and other sensitive domains.

Evidence Updating with CUE: We use the following evidence updating strategy (see Fig. 2) to fuse hard and soft data evidence. Let $Bo_{E^{(0)}}(\Theta_S \times \Theta_V \times \Theta_x, \mathcal{F}_{\Theta}[k], m_{\Theta}[k])$ be the evidence BoE at the $k$-th update. We initialize $Bo_{E^{(0)}}$ with a vacuous BoE representing the hard evidence, “The observed vehicle is in $\Theta_V$”. At the $k$-th update cycle, we compute the update $Bo_{E^{(k)}} = Bo_{E^{(k-1)}} \triangleq Bo_{E_k}, k = 1, \ldots, n$. The idea here is to refine (i.e., update) the evidence obtained from hard sensors with the soft evidence from the witnesses and the public database in order to narrow down the possible suspects.
In this paper, we have illustrated a framework based on the DS conditional approach for fusing soft and hard data, a task that is gaining increased attention. The DS theoretic conditional approach appears to be well suited to account for the various forms of uncertainty associated with soft data, including accommodating non-identical FoDs associated with different evidence sources. Moreover, the DS theoretic basis of the conditional approach provides the analyst valuable information regarding the uncertainty associated with conclusions that are being drawn. One can also utilize the pignistic probability transformation [12] to convert DS data to probabilities, thus providing a bridge to more traditional data processing methods. Our current work centers on methods for inferring reliability and meaning from multiple soft and hard sources.

4. CONCLUDING REMARKS

In this paper, we have illustrated a framework based on the DS conditional approach for fusing soft and hard data, a task that is gaining increased attention. The DS theoretic conditional approach appears to be well suited to account for the various forms of uncertainty associated with soft data, including accommodating non-identical FoDs associated with different evidence sources. Moreover, the DS theoretic basis of the conditional approach provides the analyst valuable information regarding the uncertainty associated with conclusions that are being drawn. One can also utilize the pignistic probability transformation [12] to convert DS data to probabilities, thus providing a bridge to more traditional data processing methods. Our current work centers on methods for inferring reliability and meaning from multiple soft and hard sources.

Fusion Results: We obtain the following BoE after updating the initial BoE with all three pieces of evidence:

\[ m(TI \times (Jp, Tk)) = \alpha_{3m} c_2 (1 - \alpha_{1})(0.7\tau_1)(2 - 0.7\tau_1) \]

\[ m(\Theta_S \times Tk \times A) = \alpha_{3m} (1 - \alpha_{2})(0.9r_2)(2 - 0.9r_2) \]

\[ m(\Theta_S \times \Theta_V \times \Theta_r) = \alpha_{3m} c_2 (1 + \alpha_{1})(1 - 0.7\tau_1) \]

\[ + \alpha_{3m} (1 - \alpha_{2})(1 - 0.9r_2) \]

\[ m(TI \times \Theta_V \times B) = r_3 c_1 (1 - \alpha_{3})(2 - r_3(1 + c_1 - c_2)) \]

\[ m((Md, St) \times \Theta_V \times A) = r_3 c_1 (1 - \alpha_{3})(c_2 - 1)(1 + c_1 - c_2) \]

\[ + 2r_3(1 - \alpha_{3})(1 - c_2) \]

\[ m(\Theta_S \times \Theta_V \times (A, B)) = (1 - \alpha_{3})[1 - r_3 (1 + c_1 - c_2)]^2 \].

Note the dependency of these final masses on the source reliabilities and 'inertia' (i.e., the \( \alpha_i \) values).

Analysis of Fusion Results: Table 2 contains the final fusion results for different parameters values. We have taken \( DB_3 \) to be very reliable with \( r_3 = 0.9 \); \( r_1 \) and \( r_2 \) denote the reliabilities of \( WS_1 \) and \( WS_2 \), respectively. \( Pr(\cdot) \) column depicts the pignistic probability [12]; the \([Bl(\cdot), Pl(\cdot)]\) values depict the corresponding belief and plausibility values which can be interpreted as indicating the uncertainty associated with the underlying probability.

The main observation one should make here is that, while one may reach the same conclusion under different circumstances, the uncertainty associated with the decision may vary significantly. For instance, when rule confidence is low, while both scenarios \( r_1 = 0.1, r_2 = 0.9 \) and \( r_1 = 0.9, r_2 = 0.1 \) favor Andy driving the truck, the uncertainty associated with the latter is much smaller because of the higher reliability of \( WS_1 \). When rule confidence is high, a decision favoring Bob driving has to be made with care because the associated uncertainty is very high. This is one main advantage of DS theory: one can make a decision with a better awareness of the associated uncertainties.

Table 2. Final fusion results.

5. REFERENCES