Computing the Decomposable Entropy of Graphical Belief Function Models

Radim Jiroušek[1,2], Václav Kratochvíl[1,2], and Prakash P. Shenoy[3]

[1]Faculty of Management, University of Economics, Jindřichův Hradec, Czechia
[2]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
[2] *radim@utia.cas.cz, velorex@utia.cas.cz*
[3]School of Business, University of Kansas, Lawrence, KS, USA
[3] *pshenoy@ku.edu*

April 29, 2022

Abstract

In 2018, Jiroušek and Shenoy proposed a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy*. Here, we provide an algorithm for computing the decomposable entropy of directed graphical D-S belief function models. For undirected graphical belief function models, assuming that each belief function in the model is non-informative to the others, no algorithm is necessary. We compute the entropy of each belief function and add them together to get the decomposable entropy of the model. Finally, the decomposable entropy generalizes Shannon's entropy not only for the probability of a single random variable but also for multinomial distributions expressed as directed acyclic graphical models called Bayesian networks.

1 Introduction

Jiroušek and Shenoy (2018a) propose a definition of entropy for Dempster-Shafer (D-S) belief functions called *decomposable entropy*. Some basic properties of the decomposable entropy are described in (Jiroušek and Shenoy, 2020). One of the main properties of this entropy is as follows. Suppose we have a joint basic probability assignment (BPA) $m_{X,Y}$ for $\{X, Y\}$ that decomposes as follows: $m_{X,Y} = m_X \oplus m_{Y|X}$, where m_X is the marginal of $m_{X,Y}$ for X, $m_{Y|X}$ is a conditional BPA for Y given X, and $\oplus$ is Dempster's combination rule. Then, the joint decomposable entropy of $m_{X,Y}$, denoted by $H(m_{X,Y})$, is equal to $H(m_X) + H(m_{Y|X})$, where $H(m_{Y|X})$ denotes the conditional decomposable

entropy of $m_{Y|X}$. This decomposable property is analogous to the decomposable property of Shannon's entropy for joint probability mass functions that is the basis of its definition (Shannon, 1948). There are numerous definitions of entropy for the D-S theory (see (Jiroušek and Shenoy, 2018b) for a review), but none of these satisfy the decomposable property and, therefore, the computation of these entropies for large graphical models may be intractable.

Graphical belief function models can be either directed or undirected. This article provides an algorithm for computing the decomposable entropy of directed graphical belief function models and illustrates it using an example called the *captain's decision problem* (Almond, 1995). This problem has eight variables, and the joint state space of the eight variables has 2,304 states.

Two distinct belief functions are said to be *non-informative* if the marginals of these belief functions for the intersection of their domains are vacuous. A set of distinct belief functions is said to be non-informative if every pair of belief functions from the set is non-informative. No algorithm is necessary for undirected graphical belief function models with non-informative belief functions. We compute the entropy of each belief function in the model and add them together to get the entropy of the model. This is illustrated by using the *communication network* example (Haenni and Lehmann, 2002). This problem has forty-six binary variables with a joint state space of 2^{46} states, and seventy non-informative belief functions.

Finally, the decomposable entropy generalizes Shannon's entropy for the probability of large multinomial distributions expressed as directed acyclic graph models called Bayesian networks. We illustrate this using the *chest clinic* Bayesian network example (Lauritzen and Spiegelhalter, 1988). First, we convert all probability potentials in the example to belief functions. In particular, we use Smets' conditional embedding to convert the conditional probability tables (CPTs) to conditional belief functions. These conditional belief functions are not Bayesian. Next, we compute the decomposable entropy of the directed graphical belief function model and show that it is the same as Shannon's entropy of this probability model. This example has eight binary variables with a joint state space of $2^8 = 256$ states.

An outline of the remainder of the article is as follows. Section 2 sketches the basic definitions in the D-S theory and also reviews conditional belief functions. Section 3 reviews the basic definitions and properties of decomposable entropy. This section also contains a new property of decomposable entropy for two non-informative belief functions. Section 4 describes an algorithm for computing the decomposable entropy of large directed graphical belief function models. Section 5 describes three graphical belief function models. Section 6 describes some implementation details and tools used to implement the algorithm. Finally, Section 7 provides a summary and states some unresolved issues.

2 Dempster-Shafer's Belief Function Theory

In this section, we sketch the basics of Dempster-Shafer's theory of belief functions (Dempster, 1968; Shafer, 1976).

2.1 Representations

There are several representations in the D-S theory of belief functions. Here we focus on basic probability assignments and commonality functions.

Basic Probability Assignment Suppose X is a random variable with a finite state space Ω_X. Let 2^{Ω_X} denote the set of all subsets of Ω_X. A basic probability assignment (BPA) m for X is a function $m : 2^{\Omega_X} \to [0, 1]$ such that:

$$m(\emptyset) = 0, \text{and} \tag{1}$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1. \tag{2}$$

$m(\mathsf{a})$ represents the probability mass that is assigned exactly to subset a. Thus, no mass is assigned to the empty subset (Eq. (1)) and the total probability assigned to all non-empty subsets is 1 (Eq. (2)).

The non-empty subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. A BPA m with only one focal element a (with mass 1) is called *determinsitic*. A deterministic BPA with focal element Ω_X is called *vacuous*. We say m is *consonant* if the focal elements of m are nested, i.e., if they can be ordered such that $\mathsf{a}_1 \subset \mathsf{a}_2 \subset ... \subset \mathsf{a}_m$, where $\{\mathsf{a}_1, ..., \mathsf{a}_m\}$ denotes the set of all focal elements of m. Deterministic BPAs are trivially consonant. We say m is *quasi-consonant* if the intersection of all focal elements of m is non-empty. A consonant BPA is also quasi-consonant, but not vice-versa. We say m is *Bayesian* if its focal elements are singleton subsets.

Commonality Function The information in a BPA m for X can also be represented by a corresponding commonality function (CF) Q_m for X that is defined as follows:

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} : \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}), \text{for all } \mathsf{a} \in 2^{\Omega_X}. \tag{3}$$

$Q_m(\mathsf{a})$ represents the probability mass that could possibly move to subset a.

From Eq. (3), it follows that $0 \leq Q_m \leq 1$. From Eqs. (1)–(3), it follows that $Q_m(\emptyset) = 1$. If m is a vacuous BPA for X, then $Q_m(\mathsf{a}) = 1$ for all $\mathsf{a} \in 2^{\Omega_X}$. CFs are non-increasing in the sense that if $\mathsf{a} \subseteq \mathsf{b}$, then $Q_m(\mathsf{a}) \geq Q_m(\mathsf{b})$. The CF Q_m has exactly the same information as in the corresponding BPA m.

2.2 Marginalization and Combination

In the D-S theory, we reduce the domain of a joint belief function using the marginalization operation, and we combine distinct (or independent) belief functions using Dempster's combination rule (Dempster, 1968).

Marginalization Marginalization in D-S theory is the summation of values of BPAs.

Projection of states means dropping extra coordinates; for example, if (x, y) is a state of (X, Y), then the projection of (x, y) to X, denoted by $(x, y)^{\downarrow X}$, is simply x, which is a state of X.

Projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b} \in 2^{\Omega_{(X,Y)}}$. Then $\mathsf{b}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

Suppose m is a BPA for (X, Y). Then, the marginal of m for X, denoted by $m^{\downarrow X}$, is a BPA for X such that for each $\mathsf{a} \in 2^{\Omega_X}$,

$$m^{\downarrow X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{(X,Y)}} : \mathsf{b}^{\downarrow X} = \mathsf{a}} m(\mathsf{b}). \tag{4}$$

It follows from Eq. (4), that if $m(\mathsf{b}) > 0$, then $m^{\downarrow X}(\mathsf{b}^{\downarrow X}) > 0$, for all $\mathsf{b} \in 2^{\Omega_{(X,Y)}}$.

Dempster's Combination Rule We will define Dempster's combination rule in terms of CFs. Suppose $\mathcal{X}_1$ and $\mathcal{X}_2$ are arbitrary (finite) sets of variables, and Q_1 and Q_2 are distinct CFs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then $Q_1 \oplus Q_2$ is a CF for $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$ given by:

$$(Q_1 \oplus Q_2)(\mathsf{a}) = \begin{cases} 1 & \text{if } \mathsf{a} = \emptyset, \\ K^{-1} Q_1(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_2(\mathsf{a}^{\downarrow \mathcal{X}_2}) & \text{otherwise,} \end{cases} \tag{5}$$

for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, where K is a normalization constant given by:

$$K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_{\mathcal{X}_1 \cup \mathcal{X}_2}}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow \mathcal{X}_1})\, Q_2(\mathsf{a}^{\downarrow \mathcal{X}_2}). \tag{6}$$

$(1 - K)$, where K is the normalization constant in Eq. (6), can be interpreted as a measure of conflict in the two CFs. The definition of Dempster's rule assumes that the normalization constant K is non-zero. If $K = 0$, i.e., $1 - K = 1$, then the two CFs Q_1 and Q_2 are said to be in *total conflict* and cannot be combined. If $K = 1$, i.e., $1 - K = 0$, we say Q_1 and Q_2 are *non-conflicting*.

Non-informative Belief Functions Suppose m_1 and m_2 are two distinct BPAs for $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. We say m_1 and m_2 are *non-informative* to each other if $m_1^{\downarrow \mathcal{X}_1 \cap \mathcal{X}_2}$ and $m_2^{\downarrow \mathcal{X}_1 \cap \mathcal{X}_2}$ are vacuous BPAs for $\mathcal{X}_1 \cap \mathcal{X}_2$. Notice that if m_1 and m_2 are non-informative to each other, then $(m_1 \oplus m_2)^{\downarrow \mathcal{X}_1} = m_1$ and $(m_1 \oplus m_2)^{\downarrow \mathcal{X}_2} = m_2$. This follows from the definition of non-informative belief functions and the local computation property (Shenoy and Shafer, 1990).

Intuitively, Q_1 doesn't tell us anything about Q_2 and vice-versa. If $\mathcal{X}_1$ and $\mathcal{X}_2$ are disjoint, then they are trivially non-informative to each other. The definition of non-informative belief functions can be generalized to sets of belief functions. A set of belief functions is non-informative if every pair of belief functions from the set is non-informative to each other. Of course, it is sufficient to check only those pairs with a non-empty intersection of their domains.

2.3 Conditional Belief Functions

Conditional belief functions were initially studied by Smets (1978) who introduced the notion of conditional embedding. They have been further explored in (Shafer, 1982; Almond, 1995; Xu and Smets, 1996). Here we review the basics.

Consider a BPA m for X and $\mathsf{a} \in 2^{\Omega_X}$. Suppose that there is a BPA for Y expressing our belief about Y if we know that $X \in \mathsf{a}$, and denote it by $m_{Y|\mathsf{a}}$. Notice that $m_{Y|\mathsf{a}} : 2^{\Omega_Y} \to [0,1]$ is a BPA for Y. We can embed this BPA for Y into a conditional BPA for (X,Y), which is denoted by $m_{\mathsf{a},Y}$, so that the following two conditions hold:

1. $m_{\mathsf{a},Y}$ tells us nothing about X, i.e., $m_{\mathsf{a},Y}^{\downarrow X}(\Omega_X) = 1$.
2. If we combine $m_{\mathsf{a},Y}$ with the deterministic BPA $m_{X\in\mathsf{a}}$ for X such that $m_{X\in\mathsf{a}}(\mathsf{a}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain $m_{Y|\mathsf{a}}$, i.e., $(m_{\mathsf{a},Y} \oplus m_{X\in\mathsf{a}})^{\downarrow Y} = m_{Y|\mathsf{a}}$.

One way to obtain such an embedding is suggested by Smets (1978) (see also, (Shafer, 1982), Xu and Smets (1996), and Almond (1995)), called *conditional embedding*. It consists of taking each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of $m_{Y|\mathsf{a}}$, and converting it to the corresponding focal element

$$(\mathsf{a} \times \mathsf{b}) \cup ((\Omega_X \setminus \mathsf{a}) \times \Omega_Y) \in 2^{\Omega_{X,Y}} \tag{7}$$

of $m_{\mathsf{a},Y}$ with the same mass. It is easy to confirm that this method of embedding satisfies both conditions mentioned above.

When does a belief function qualify as a conditional? For example, suppose we have a BPA m for $\{Y\} \cup \mathcal{X}$ where $\{Y\} \cap \mathcal{X} = \emptyset$. Under what conditions does m constitute a conditional for Y given $\mathcal{X}$? Analogous to conditional probability tables in Bayesian networks, the answer is straightforward. Any BPA m for $\{Y\} \cup \mathcal{X}$ such that $m^{\downarrow\mathcal{X}}$ is the vacuous BPA for $\mathcal{X}$ constitutes a conditional for Y given $\mathcal{X}$. Sometimes, we will let $m_{Y|\mathcal{X}}$ denote such conditionals.

3 Decomposable Entropy of D-S Belief Functions

This section reviews the definitions of decomposable entropy and conditional decomposable entropy of belief functions in the D-S theory (Jiroušek and Shenoy, 2018a) and describes its properties (Jiroušek and Shenoy, 2020). We also describe a new property of decomposable entropy motivated by the need to compute the decomposable entropy of an undirected graphical belief function model.

3.1 Decomposable Entropy

Definition 1 (Entropy of a CF Q) *Suppose Q is a CF for $\mathcal{X}$ with state-space $\Omega_{\mathcal{X}}$. Then, the decomposable entropy of Q, denoted by $H(Q)$, is defined as*

$$H(Q) = \sum_{\mathsf{a}\in 2^{\Omega_{\mathcal{X}}}} (-1)^{|\mathsf{a}|} Q(\mathsf{a}) \log(Q(\mathsf{a})). \tag{8}$$

The definition of entropy of Q in Eq. (8) is well-defined as it follows from the definition of a CF in Eq. (3) that for all $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$, $Q(\mathsf{a}) \geq 0$. If $Q(\mathsf{a}) = 0$, we will follow the convention that $Q(\mathsf{a}) \log(Q(\mathsf{a})) = 0$ as $\lim_{\theta \to 0^+} \theta \log(\theta) = 0$. Thus, in computing the entropy $H(Q)$ as defined in Def. 1, it is sufficient that the summation in the right-hand side of Eq. (8) is restricted to $\mathsf{a} \in 2^{\Omega_{\mathcal{X}}}$ such that $Q(\mathsf{a}) > 0$.

3.2 Conditional Decomposable Entropy

Definition 2 (Conditional entropy of $Q_{Y|X}$) *Suppose Q_X is a CF for X, and suppose $Q_{Y|X}$ is a conditional CF for (X, Y). Then, the conditional decomposable entropy of $Q_{Y|X}$, denoted by $H(Q_{Y|X})$, is defined as follows:*

$$H(Q_{Y|X}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|} \, Q_X(\mathsf{a}^{\downarrow X}) \, Q_{Y|X}(\mathsf{a}) \, \log(Q_{Y|X}(\mathsf{a})). \tag{9}$$

Notice that as $Q_X(\mathsf{a}^{\downarrow X}) \, Q_{Y|X}(\mathsf{a}) = Q_{X,Y}(\mathsf{a})$ for all $\mathsf{a} \in 2^{\Omega_{X,Y}}$, we can rewrite Eq. (9) as follows:

$$H(Q_{Y|X}) = \sum_{\mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|} \, Q_{X,Y}(\mathsf{a}) \, \log(Q_{Y|X}(\mathsf{a})) \tag{10}$$

3.3 Properties of Decomposable Entropy

A list of relevant properties of the decomposable entropy is as follows. For formal proofs, see (Jiroušek and Shenoy, 2020).

Property 1 (Compound distributions) *Suppose Q_X is a CF for X, and suppose $Q_{Y|X}$ is a conditional CF for (X, Y). Let $Q_{X,Y} = Q_X \oplus Q_{Y|X}$. Then,*

$$H(Q_{X,Y}) = H(Q_X) + H(Q_{Y|X}). \tag{11}$$

Property 2 (Quasi-consonant BPAs have 0 decomposable entropy) *Suppose m is a quasi-consonant BPA. Then $H(m) = 0$. As vacuous, deterministic, and consonant BPAs are also quasi-consonant, their decomposable entropies are also 0.*

Suppose P_X is a probability mass function (PMF) for X such that $P_X(x) > 0$ for all $x \in \Omega_X$, and $P_{Y|X}$ is a conditional probability table (CPT) for Y given X, i.e., $P_{Y|X}(x, y) = P_{Y|x}(y)$, where $P_{Y|x}$ is the conditional PMF for Y given $X = x$ for all $(x, y) \in \Omega_{X,Y}$. Let $P_{X,Y} = P_X \otimes P_{Y|X}$ ($\otimes$ denotes probabilistic combination, which is pointwise multiplication followed by normalization). Let m_X denote the Bayesian BPA corresponding to P_X, let $m_{Y|x}$ denote the Bayesian conditional BPA for Y corresponding to the conditional PMF $P_{Y|x}$ for Y given $X = x$. Let $m_{x,Y}$ denote the conditional BPA for (X, Y) obtained by conditional embedding of $m_{Y|x}$. Let $m_{Y|X}$ denote $\bigoplus_{x \in \Omega_X} m_{x,Y}$. Let $m_{X,Y}$ denote $m_X \oplus m_{Y|X}$. Notice that $m_{x,Y}$ and $m_{Y|X}$ are not Bayesian BPAs.

Property 3 (Strong probability consistency) *Consider the situation described in the preceding paragraph. Let $H_s(P_{X,Y})$ and $H_s(P_X)$ denote Shannon's entropy of PMFs $P_{X,Y}$ and P_X, respectively, and let $H_s(P_{Y|X})$ denote Shannon's conditional entropy of the CPT $P_{Y|X}$. Then, $m_{X,Y}$ is a Bayesian BPA for (X,Y) corresponding to PMF $P_{X,Y}$ such that:*

$$H(m_{X,Y}) = H_s(P_{X,Y}), \tag{12}$$

$$H(m_X) = H_s(P_X), \tag{13}$$

$$H(m_{Y|X}) = H_s(P_{Y|X}). \tag{14}$$

The following theorem generalizes Property 1. It is a new property not discussed in (Jiroušek and Shenoy, 2020). It is motivated by the need to compute the entropy of an undirected belief function graphical model.

Theorem 1 (Entropy of non-informative belief functions) *Suppose Q_1 and Q_2 are distinct CFs for $\mathcal{X}_1$, and $\mathcal{X}_2$, respectively, such that they are non-informative for each other. Then,*

$$H(Q_1 \oplus Q_2) = H(Q_1) + H(Q_2) \tag{15}$$

A proof of this property can be found in a longer version of this paper (Jiroušek et al., 2022).

4 An Algorithm

This section describes an algorithm for computing the decomposable entropy of a directed graphical belief function.

Suppose we have a directed acyclic graph G consisting of a set of variables $\{X_1, \ldots, X_n\}$ as nodes, and a set of directed edges. Let $Pa_G(X_k)$ denote the parents of X_k in graph G. Associated with each node X_k is a conditional BPA m_k for $X_k \cup Pa_G(X_k)$ that is a conditional for X_k given $Pa_G(X_k)$. If $Pa_G(X_k) = \emptyset$, then the conditional for X_k is the prior belief function for X_k. If $Pa_G(X_k) \neq \emptyset$, then we will assume that m_k is a conditional BPA for $X_k \cup Pa_G(X_k)$, i.e., $m_k^{\downarrow X_k}$ is a vacuous BPA for X_k.

Notice that if we have evidence for a variable that is different from priors or conditionals in a directed graphical belief function model, we need to disregard such evidence. For example, suppose we have a directed acyclic graph $X \to Y$ with a BPA m_1 for X, a conditional BPA m_2 for $\{X,Y\}$ that constitutes a conditional for $Y|X$ so that $m_2^{\downarrow X}$ is the vacuous BPA for X, and a BPA m_3 for Y that represents some evidence for Y. It follows from the compound distributions property that $H(m_1 \oplus m_2) = H(m_1) + H(m_2)$. But, in general, $H(m_1 \oplus m_2 \oplus m_3) \neq H(m_1) + H(m_2) + H(m_3)$. For this reason, we need to disregard evidence in computing the decomposable entropy of a directed graphical belief function model.

Algorithm First, we start with a sequence $(X_1, \ldots X_n)$ such that if there is a directed arc $X_i \to X_j$ in G, then X_i precedes X_j in the sequence. As G is acyclic, such a sequence always exists, but it may not be unique.
Do $k = 1, \ldots, n$:

- If $Pa_G(X_k) = \emptyset$, then $H(m_k)$ is computed using Definition 1.
- If $Pa_G(X_k) \neq \emptyset$, then first we find the marginal $(\bigoplus_{i=1}^{k-1} m_i)^{\downarrow Pa_G(X_k)}$ using local computation (Shenoy and Shafer, 1990). Next, we find the conditional decomposable entropy of m_k, $H(m_k)$, using Definition 2.

End Do;

The decomposable entropy of the joint belief function $H(\bigoplus_{k=1}^{n} m_k) = \sum_{k=1}^{n} H(m_k)$. This follows from the compound distributions property of decomposable entropy.

5 Three Examples

This section computes the decomposable entropy of three graphical belief function models.

Captain's Problem The captain's problem is from Almond (1995). A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination. The arrival delay is the sum of departure delay and sailing delay. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). Sailing delay may result from bad weather (at most one day) and whether repairs are needed at sea (at most one day). If maintenance is done before sailing, chances of repairs at sea are less likely. The weather forecast says a slight chance of bad weather (0.2) and a good chance of good weather (0.6). The forecast is 80% reliable. The captain knows the loading delay and whether maintenance is done before departure. Fig. 1 shows the directed acyclic graph associated with this problem. Table 1 shows the variables, their state spaces, and the associated conditionals. What is the decomposable entropy of this belief function model?

As ϕ_2 is an evidence for F, we ignore this belief function. First, notice that ϕ_1 and σ are consonant, and μ, δ, and α are deterministic. So the decomposable entropies of these BPAs are zeroes. The decomposable entropies of the remaining BPAs are as follows. $H(\lambda) \approx 0.3958$, $H(\rho_1 \oplus \rho_2) \approx 0.0729$, Thus, the decomposable entropy of the captain's problem (ignoring the evidence ϕ_2) is $0.3958 + 0.0729 = 0.4687$.

Communication Network This example is from Haenni and Lehmann (2002). Fig. 2 shows an undirected graph associated with this example. We have a grid of $44 = 8 + 9 + 10 + 9 + 8$ communication nodes arranged in 19 columns and 5 rows. There are 68 links, and each link has 90% reliability. Nodes A and B are connected to the grid with links having 80% reliability. What is the decomposable entropy of this graphical model?

Consider the variables in the grid with 19 columns and 5 rows. Let X_{13} denote the variable in column 1, row 3 and let X_{22} denote the variable in column 2 and row 2. Let

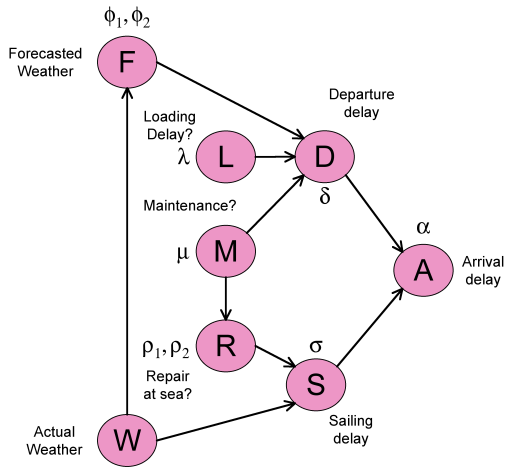


Figure 1: The directed acyclic graph for the captains's problem. The Greek alphabets adjacent to a variable denote the prior or conditional or evidence associated with the variable.

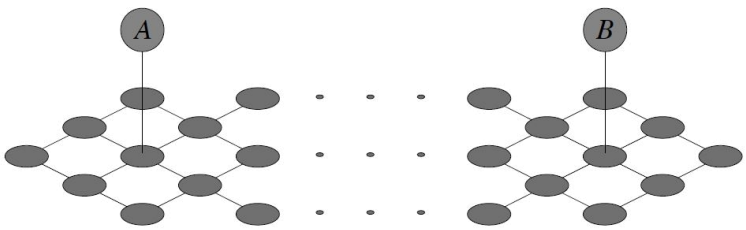


Figure 2: The undirected graph for the communication network example.

$\Omega_{13} = \{t_{13}, f_{13}\}$, and and let $\Omega_{22} = \{t_{22}, f_{22}\}$. The BPA m_{13-22} associated with the edge between X_{13} and X_{22} is as follows:

$$m_{13-22}(\{(t_{13}, t_{22}), (f_{13}, f_{22})\}) = 0.9, m_{13-22}(\Omega_{13} \times \Omega_{22}) = 0.1.$$

The BPAs associated with the remaining 67 links are similar. The edges between A and X_{33} and between B and X_{38} are also similar, except that the reliability is 0.8 instead of 0.9. As these BPAs are consonant, the decomposable entropy of all 70 BPAs are zeroes. Also, notice that the BPAs m_{13-22} and m_{12-24} associated with the corresponding edges satisfy the conditions in Theorem 1. As all the BPAs in this example have the same structure, it follows that the set of all BPAs is non-informative. Thus, the decomposable entropy of the communication network model is 0.

Table 1: The variables, their state spaces, and associated conditionals in the captain's problem.

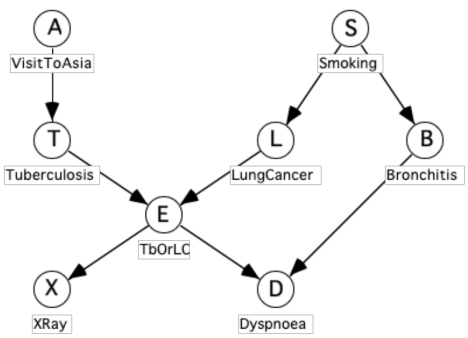
Variable	Name	State Space, Ω	Associated Conditional
W	Actual weather	$\{g_w, b_w\}$	vacuous for W
F	Forecasted weather	$\{g_f, b_f\}$	ϕ_1 for $F\|W$ (consonant)
L	Loading delay?	$\{t_l, f_l\}$	λ for L
M	Maintenance done?	$\{t_m, f_m\}$	μ for M (deterministic)
R	Repair at sea needed?	$\{t_r, f_r\}$	$\rho_1 \oplus \rho_2$ for $R\|M$
D	Departure delay (in days)	$\{0, 1, 2, 3\}$	δ for $D\|\{F, L, M\}$ (deterministic)
S	Sailing delay (in days)	$\{0, 1, 2, 3\}$	σ for $S\|\{W, R\}$ (consonant)
A	Arrival delay (in days)	$\{0, 1, 2, 3, 4, 5, 6\}$	α for $A\|\{D, S\}$ (deterministic)

Chest Clinic This example is from Lauritzen and Spiegelhalter (1988). Fig. 3 shows a Bayesian network that is represented as a directed graphical belief function model. There are eight binary variables, and not all probabilities in the joint probability distribution are positive. Fig. 3 also shows the conditional probability tables (CPTs). These are represented as BPAs using conditional embedding, and most of these BPAs are not Bayesian. The decomposable entropies of the conditionals are as follows (computed using the algorithm in Section 4):

$$H(P(A)) \approx 0.0808, H(P(T|A)) \approx 0.0828, H(P(S)) = 1, H(P(L|S)) \approx 0.2749,$$
$$H(P(B|S)) \approx 0.9261, H(P(E|L,T)) = 0, H(P(X|E)) \approx 0.2770, H(P(B|DE)) \approx 0.6471.$$

Thus, the decomposable entropy of the directed graphical belief function model is approximately 3.2887, which is the same as Shannon's entropy of the corresponding Bayesian network.

Figure 3: The directed acyclic graph and the CPTs for the chest clinic example.



P(A):	p(a) = .01	P(E|L,T):	p(e | l, t) = 1 p(e | l, ~t) = 1 p(e | ~l, t) = 1 p(e | ~l, ~t)=0
P(T|A):	p(t | a) = .05 p(t | ~a) = .01		
P(S):	p(s) = .50	P(X|E):	p(x | e) = .98 p(x | ~e) = .05
P(L|S):	p(l | s) = .10 p(l | ~s) = .01	P(D|E,B):	p(d | e, b) = .90 p(d | e, ~b) = .70 p(d |~e, b) = .80 p(d |~e, ~b)=.10
P(B|S):	p(b | s) = .60 p(b | ~s) = .30		

6 Notes on Implementation

We performed all experiments in R. We have created an R package to work with belief functions, which we plan to complete and publish for use by other users. The package is based on relational databases as implemented in the R package *data.table* (Dowle and Srinivasan, 2021). Each belief function is an object with three different tables. The first table, called the *coding table*, consists of random variables and their states. The columns correspond to the random variables in the domain $\mathcal{X}$ of the belief function, the rows to the elements of their joint state space $\times_{X \in \mathcal{X}} \Omega_X$. Each row is labeled with a unique identifier. The second table, called *focal element table*, stores each focal element as a set of states using identifiers from the *coding table*. The third table, called *mass table*, assigns a probability mass to each focal element.

Regarding computing the marginal of the joint in the algorithm described in Section 4, an implementation using local computation is available in the *Belief Function Machine* environment in Matlab (Giang and Shenoy, 2003). We implemented this algorithm in R.

7 Summary & Conclusions

The primary goal of this article is to describe an algorithm for computing the decomposable entropy of directed graphical belief function models. The decomposable entropy has a property that if we construct a joint BPA for two variables (X, Y) by Dempster's combination of a BPA for m_X for X and a conditional BPA $m_{Y|X}$ for Y given X, then the decomposable entropy of $m_{X,Y} = m_X \oplus m_{Y|X}$ is equal to the decomposable entropy of m_X plus the decomposable conditional entropy of $m_{Y|X}$.

The decomposable entropy is defined using commonality functions. If a graphical model has a clique whose state space is large, then computing the decomposable entropy of the clique may be intractable. For example, in the captain's problem, the conditional for arrival delay has three variables with a joint space of $4 \times 4 \times 7 = 112$ states. Fortunately this conditional is deterministic, and the decomposable entropy of deterministic BPAs is 0. If this conditional wasn't deterministic or consonant or quasi-consonant, and the joint commonality function for these three variables had non-zero values for each of the 2^{112} subsets, then the computation of the exact decomposable entropy of the conditional would be intractable. In such cases, we may have to resort to some approximate methods. This is yet to be done.

References

R. G. Almond. *Graphical Belief Modeling.* Chapman & Hall, London, UK, 1995.

A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30(2):205–247, 1968.

M. Dowle and A. Srinivasan. *data.table: Extension of 'data.frame'*, 2021. URL `https://CRAN.R-project.org/package=data.table`. R package version 1.14.2.

P. Giang and S. Shenoy. The belief function machine: An environment for reasoning with belief functions in Matlab. Working paper, University of Kansas School of Business, Lawrence, KS 66045, 2003. URL `https://pshenoy.ku.edu/Papers/BFM072503.zip`.

R. Haenni and N. Lehmann. Resource bounded and anytime approximation of belief function computation. *International Journal of Approximate Reasoning*, 31(1–2):103–154, 2002.

R. Jiroušek and P. P. Shenoy. A decomposable entropy for belief functions in the Dempster-Shafer theory. In S. Destercke, T. Denoeux, F. Cuzzolin, and A. Martin, editors, *Belief Functions: Theory and Applications*, volume 11069 of *Lecture Notes in Artificial Intelligence*, pages 146–154. Springer Nature, Switzerland, 2018a. doi: 10.1007/978-3-319-99383-6_19.

R. Jiroušek and P. P. Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92(1):49–65, 2018b.

R. Jiroušek and P. P. Shenoy. On properties of a new decomposable entropy of Dempster-Shafer belief functions. *International Journal of Approximate Reasoning*, 119(4):260–279, 2020.

R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Computing the decomposable entropy of graphical belief function models. Working Paper 340, University of Kansas School of Business, Lawrence, KS 66045, USA, 2022. URL `https://pshenoy.ku.edu/Papers/WP340.pdf`.

S. L. Lauritzen and D. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, series B*, 50(2):157–224, 1988.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, 44(3):322–352, 1982.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, Machine Intelligence and Pattern Recognition Series, vol. 9, pages 169–198, Amsterdam, 1990. North-Holland.

P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical*. PhD thesis, Free University of Brussels, 1978.

H. Xu and P. Smets. Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, 14(2–3):155–185, 1996.