

On the Axiomatization of Generalized Entropic Metrics

Dan A. Simovici
Univ. of Massachusetts Boston,
Dept. of Comp. Science,
100 Morrissey Blvd.
Boston, Massachusetts
02125 USA
dsim@cs.umb.edu

Abstract

Starting from an axiomatization of a generalization of Shannon entropy we introduce a set of axioms for a parametric family of distances over sets of partitions of finite sets. This family includes some well-known metrics used in data mining and in the study of finite functions.

1 Introduction

This paper is examining an axiomatization of a parameterized family of metrics on sets of partitions of finite sets that generalizes the entropic metric introduced by R. López de Mántaras [5], as well as the Mirkin metric introduced in [8]. This unifies the separate axiom systems for these metrics introduced in [7] and illuminates the relationship of the axiomatization of these metrics with our previous axiomatization of generalized entropy [12, 11].

Metrics on sets of partitions of finite sets are useful because they allow us to study properties of finite functions related to their kernel partitions. In a different direction, these metrics are interesting for data mining because the attributes of a table induce partitions on the sets of tuples of the table. Thus, metrics on partitions allow us to determine interesting relationships between attributes and to use these relationships for classification, data summarization and other applications. Also, exclusive clusterings can be regarded as partitions of the set of clustered objects and partition metrics can be used for evaluating clusterings, a point of view presented in [7].

A natural link exists between random variables and partitions of sets, as we show next. This link allows the transfer of certain probabilistic and information-theoretical notions to partitions of sets, where we can take advantage of the partial order between partitions.

A *partition* of a set S is a non-empty collection π of non-empty subsets of S , $\pi = \{B_i \mid i \in I\}$ such that for every $i, j \in I$, $i \neq j$ implies $B_i \cap B_j = \emptyset$ and $\bigcup_{i \in I} B_i = S$. We refer to the sets B_i as the *blocks* of π .

Let $\text{PART}(S)$ be the set of partitions of a set S . The class of all partitions of finite sets is denoted by PART . The one-block partition of S is denoted by ω_S . The partition $\{\{s\} \mid s \in S\}$ is denoted by α_S .

If $\pi, \pi' \in \text{PART}(S)$, then $\pi \leq \pi'$ if every block of π is included in a block of π' . Clearly, for every $\pi \in \text{PART}(S)$ we have $\alpha_S \leq \pi \leq \omega_S$.

π' covers π if $\pi \leq \pi'$ and there is no partition $\theta \in \text{PART}(S)$ such that $\pi < \theta < \pi'$. This fact is denoted by $\pi < \pi'$. It is known [6] that $\pi < \pi'$ if and only if π' is obtained from π by fusing two blocks of this partition into a new block.

For every two partitions π, σ both $\inf\{\pi, \sigma\}$ and $\sup\{\pi, \sigma\}$ in the partial ordered set $(\text{PART}(S), \leq)$ exist. Namely, if $\pi = \{B_i \mid i \in I\}$ and $\sigma = \{C_j \mid j \in J\}$, then $\inf\{\pi, \sigma\}$ is the partition:

$$\pi \wedge \sigma = \{B_i \cap C_j \mid B_i \cap C_j \neq \emptyset, i, j \in J\}.$$

The supremum $\pi \vee \sigma = \sup\{\pi, \sigma\}$ can be described using a bipartite graph \mathcal{G} having $\{B_i \mid i \in I\} \cup \{C_j \mid j \in J\}$ as set of vertices. An edge (B_i, C_j) exists only if $B_i \cap C_j \neq \emptyset$. If \mathcal{C} is a connected component of \mathcal{G} note that $\bigcup\{B \in \pi \mid B \in \mathcal{C}\} = \bigcup\{C \in \sigma \mid C \in \mathcal{C}\}$; we denote this set by $D_{\mathcal{C}}$. The family of sets $\{D_{\mathcal{C}} \mid \mathcal{C} \text{ is a connected component of } \mathcal{G}\}$ is a partition of the set S . It is easy to verify that this is exactly $\pi \vee \sigma$.

It is not difficult to show that $(\text{PART}(S), \leq)$ is an upper semimodular lattice; in other words if π, σ are two distinct partitions such each covers $\pi \wedge \sigma$, then $\pi \vee \sigma$ covers both π and σ .

If S, T are two disjoint and nonempty sets, $\pi \in \text{PART}(S)$, $\sigma \in \text{PART}(T)$, where $\pi = \{A_1, \dots, A_m\}$,

$\sigma = \{B_1, \dots, B_n\}$, then the partition $\pi + \sigma$ is the partition of $S \cup T$ given by $\pi + \sigma = \{A_1, \dots, A_m, B_1, \dots, B_n\}$.

Whenever the “+” operation is defined, then it is easily seen to be associative. In other words, if S, U, V are pairwise disjoint and nonempty sets, and $\pi \in \text{PART}(S)$, $\sigma \in \text{PART}(U)$, $\tau \in \text{PART}(V)$, then $\pi + (\sigma + \tau) = (\pi + \sigma) + \tau$. Observe that if S, U are disjoint, then $\alpha_S + \alpha_U = \alpha_{S \cup U}$. Also, $\omega_S + \omega_U$ is the partition $\{S, U\}$ of the set $S \cup U$.

If $\pi = \{B_1, \dots, B_m\}$, $\sigma = \{C_1, \dots, C_n\}$ are partitions of two arbitrary sets S, U , respectively, then we denote the partition $\{B_i \times C_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ of $S \times U$ by $\pi \times \sigma$. Note that $\alpha_S \times \alpha_U = \alpha_{S \times U}$ and $\omega_S \times \omega_U = \omega_{S \times U}$.

Let $\pi \in \text{PART}(S)$ and let $C \subseteq S$. Denote by π_C the “trace” of π on C given by $\pi_C = \{B \cap C \mid B \in \pi \text{ such that } B \cap C \neq \emptyset\}$. Clearly, $\pi_C \in \text{PART}(C)$; also, if C is a block of π , then $\pi_C = \omega_C$.

A subset T of S is *pure* relative to a partition $\pi \in \text{PART}(S)$ if $\pi_T = \omega_T$. In other words, T is pure relative to a partition π if T is included in some block of π .

2 An Axiomatization of Generalized Entropy

In [4] the notion of β -entropy of a probability distribution $\mathbf{p} = (p_1, \dots, p_n)$ was defined as:

$$H_\beta(\mathbf{p}) = \frac{1}{2^{1-\beta} - 1} \left(\sum_{i=1}^m p_i^\beta - 1 \right),$$

where $p_1 + \dots + p_n = 1$ and $p_i \geq 0$ for $1 \leq i \leq n$. In the same reference it was observed that Shannon’s entropy $H(\mathbf{p})$ can be obtained as $\lim_{\beta \rightarrow 1} H_\beta(\mathbf{p})$.

In [12] we offered a new interpretation of the notion of entropy for finite distributions as entropies of partitions of finite sets. Our approach takes advantage of the properties of the partial order of the lattice of partitions of a finite set and makes use of operations defined on partitions.

We defined the H_β entropy for $\beta \in \mathbb{R}$, $\beta > 0$ as a function $H_\beta : \text{PART}(S) \rightarrow \mathbb{R}_{\geq 0}$ that satisfies the following axioms:

- (P1) If $\pi, \pi' \in \text{PART}(S)$ are such that $\pi \leq \pi'$, then $H_\beta(\pi') \leq H_\beta(\pi)$.
- (P2) If S, T are two finite sets such that $|S| \leq |T|$, then $H_\beta(\alpha_S) \leq H_\beta(\alpha_T)$.
- (P3) For every disjoint sets S, T and partitions $\pi \in \text{PART}(S)$, and $\sigma \in \text{PART}(T)$ we have:

$$\begin{aligned} H_\beta(\pi + \sigma) &= \left(\frac{|S|}{|S| + |T|} \right)^\beta H_\beta(\pi) + \\ &\quad \left(\frac{|T|}{|S| + |T|} \right)^\beta H_\beta(\sigma) + H_\beta(\{S, T\}). \end{aligned}$$

(P4) If $\pi \in \text{PART}(S)$ and $\sigma \in \text{PART}(T)$, then

$$H_\beta(\pi \times \sigma) = \Phi(H_\beta(\pi), H_\beta(\sigma)),$$

where $\Phi : \mathbb{R}_{\geq 0}^2 \rightarrow \mathbb{R}_{\geq 0}$ be a continuous function such that $\Phi(x, y) = \Phi(y, x)$, $\Phi(x, 0) = x$ for $x, y \in \mathbb{R}_{\geq 0}$.

We have shown in [12] that if $\pi = \{B_1, \dots, B_n\} \in \text{PART}(S)$, then

$$H_\beta(\pi) = \frac{1}{2^{1-\beta} - 1} \left(\sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^\beta - 1 \right).$$

In the special case, when $\beta \rightarrow 1$ we have:

$$H_\beta(\pi) = - \sum_{i=1}^m \frac{|B_i|}{|S|} \cdot \log_2 \frac{|B_i|}{|S|}.$$

The axiomatization also implies a specific form of the function Φ . Namely, if $\beta \neq 1$ it follows that $\Phi(x, y) = x + y + (2^{1-\beta} - 1)xy$. In the case of Shannon entropy, obtained using $\beta = 1$, we have $\Phi(x, y) = x + y$ for $x, y \in \mathbb{R}_{\geq 0}$.

Note that if $|S| = 1$, then $\text{PART}(S)$ consists of a unique partition ($\omega_S = \alpha_S$) and $H_\beta(\omega_S) = 0$. Moreover, as we have shown in [12], for an arbitrary finite set S we have $H_\beta(\pi) = 0$ if and only if $\pi = \omega_S$.

These facts suggest that for a subset T of S the number $H_\beta(\pi_T)$ can be used as a measure of the “purity” of the set T with respect to the partition π . If T is π -pure, then $\pi_T = \omega_T$ and, therefore, $H_\beta(\pi_T) = 0$. Thus, the smaller $H_\beta(\pi_T)$, the more pure the set T is.

The largest value of $H_\beta(\pi)$ when $\pi \in \text{PART}(S)$ is achieved when $\pi = \alpha_S$; in this case we have:

$$H_\beta(\alpha_S) = \frac{1}{2^{1-\beta} - 1} \left(\frac{1}{|S|^{\beta-1}} - 1 \right).$$

Axiom (P3) can be extended as follows.

Theorem 2.1 *Let S_1, \dots, S_n be n pairwise disjoint finite sets, $S = \bigcup_{i=1}^n S_i$ and let π_1, \dots, π_n be partitions of S_1, \dots, S_n , respectively.*

We have:

$$H_\beta(\pi_1 + \dots + \pi_n) = \sum_{i=1}^n \left(\frac{|S_i|}{|S|} \right)^\beta H_\beta(\pi_i) + H_\beta(\theta),$$

where θ is the partition $\{S_1, \dots, S_n\}$ of S .

The β -entropy defines naturally a conditional entropy of partitions. We note that the definition introduced here is an improvement over our previous definition given in [12]. Starting from conditional entropies we will be able to define a family of metrics on the set of partitions of a finite set.

Definition 2.2 Let $\pi, \sigma \in \text{PART}(S)$ and let $\sigma = \{C_1, \dots, C_n\}$. The β -conditional entropy of the partitions

$\pi, \sigma \in \text{PART}(S)$ is the function $H : \text{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ defined by:

$$H_\beta(\pi|\sigma) = \sum_{j=1}^n \left(\frac{|C_j|}{|S|} \right)^\beta H_\beta(\pi_{C_j})$$

□

The conditional entropy can be written explicitly as:

$$\begin{aligned} H_\beta(\pi|\sigma) &= \sum_{j=1}^m \left(\frac{|C_j|}{|S|} \right)^\beta \sum_{i=1}^n \frac{1}{2^{1-\beta} - 1} \left[\left(\frac{|B_i \cap C_j|}{|C_j|} \right)^\beta - 1 \right] \\ &= \frac{1}{2^{1-\beta} - 1} \sum_{i=1}^m \sum_{j=1}^n \left(\left(\frac{|B_i \cap C_j|}{|S|} \right)^\beta - \left(\frac{|C_j|}{|S|} \right)^\beta \right), \end{aligned}$$

where $\pi = \{B_1, \dots, B_m\}$.

We have shown in [14] that the conditional β -entropy enjoys the property specified next.

Theorem 2.3 *Let π, σ, σ' be three partitions of a finite set S such that $\sigma \leq \sigma'$. We have*

- (i) $H_\beta(\pi|\sigma) = 0$ if and only if $\sigma \leq \pi$;
- (ii) $H_\beta(\pi \wedge \sigma) = H_\beta(\pi|\sigma) + H_\beta(\sigma) = H_\beta(\sigma|\pi) + H_\beta(\pi)$;
- (iii) $H_\beta(\sigma|\pi) \geq H_\beta(\sigma'|\pi)$ and $H_\beta(\pi|\sigma) \leq H_\beta(\pi|\sigma')$.

Corollary 2.4 *Since $H_\beta(\pi) = H_\beta(\pi|\omega_S)$ it follows that if $\pi, \sigma \in \text{PART}(S)$, then $H_\beta(\pi) \geq H_\beta(\pi|\sigma)$.*

The behavior of β -conditional entropies with respect to the ‘‘addition’’ of partitions is discussed in the next statement.

Theorem 2.5 *Let S be a finite set, π, θ be two partitions of S , where $\theta = \{D_1, \dots, D_h\}$. If $\sigma_i \in \text{PART}(D_i)$ for $1 \leq i \leq h$, then*

$$H_\beta(\pi|\sigma_1 + \dots + \sigma_h) = \sum_{i=1}^h \left(\frac{|D_i|}{|S|} \right)^\beta H_\beta(\pi_{D_i}|\sigma_i).$$

If $\tau = \{F_1, \dots, F_k\}$, $\sigma = \{C_1, \dots, C_n\}$ be two partitions of S , and let $\pi_i \in \text{PART}(F_i)$ for $1 \leq i \leq k$. Then,

$$H_\beta(\pi_1 + \dots + \pi_k|\sigma) = \sum_{i=1}^k \left(\frac{|F_i|}{|S|} \right)^\beta H_\beta(\pi_i|\sigma_{F_i}) + H_\beta(\tau|\sigma).$$

3 Generalized Entropic Metrics and Their Axiomatization

In [5] L. de Mántaras proved that Shannon’s entropy generates a metric $d : \text{PART}(S)^2 \longrightarrow \mathbb{R}^2$ given by $d(\pi, \sigma) = H(\pi|\sigma) + H(\sigma|\pi)$, for $\pi, \sigma \in \text{PART}(S)$. His result can be extended to a class of metrics that can be defined by β -entropies, thereby improving our earlier results [13]. To this end we need the following statement:

Theorem 3.1 *Let π, σ, τ be three partitions of the finite set S . We have:*

$$H_\beta(\pi|\sigma \wedge \tau) + H_\beta(\sigma|\tau) = H_\beta(\pi \wedge \sigma|\tau).$$

Proof. Suppose that $\sigma = \{C_1, \dots, C_n\}$ and $\tau = \{D_1, \dots, D_p\}$. We observed already that

$$\sigma \wedge \tau = \sigma_{D_1} + \dots + \sigma_{D_p} = \tau_{C_1} + \dots + \tau_{C_n}.$$

Consequently, by Theorem 2.5, we have

$$\begin{aligned} H_\beta(\pi|\sigma \wedge \tau) &= H_\beta(\pi|\sigma_{D_1} + \dots + \sigma_{D_p}) \\ &= \sum_{l=1}^p \left(\frac{|D_l|}{|S|} \right)^\beta H_\beta(\pi_{D_l}|\sigma_{D_l}). \end{aligned}$$

Also, we have

$$H_\beta(\sigma|\tau) = \sum_{l=1}^p \left(\frac{|D_l|}{|S|} \right)^\beta H_\beta(\sigma_{D_l}),$$

which implies

$$H_\beta(\pi|\sigma \wedge \tau) + H_\beta(\sigma|\tau) = H_\beta(\pi \wedge \sigma|\tau).$$

■

Corollary 3.2 *Let π, σ, τ be three partitions of the finite set S . Then, we have:*

$$H_\beta(\pi|\sigma) + H_\beta(\sigma|\tau) \geq H_\beta(\pi|\tau).$$

Proof. By Theorem 3.1, the monotonicity of β -conditional entropy in its second argument and the dual monotonicity of the same in its first argument we can write:

$$\begin{aligned} H_\beta(\pi|\sigma) + H_\beta(\sigma|\tau) &\geq H_\beta(\pi|\sigma \wedge \tau) + H_\beta(\sigma|\tau) \\ &= H_\beta(\pi \wedge \sigma|\tau) \\ &\geq H_\beta(\pi|\tau), \end{aligned}$$

which is the desired inequality. ■

We can show now a central result:

Theorem 3.3 *The mapping $d_\beta : \text{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ defined by: $d_\beta(\pi, \sigma) = H_\beta(\pi|\sigma) + H_\beta(\sigma|\pi)$ for $\pi, \sigma \in \text{PART}(S)$ is a metric on $\text{PART}(S)$.*

Proof. A double application of Corollary 3.2 yields:

$$\begin{aligned} H_\beta(\pi|\sigma) + H_\beta(\sigma|\tau) &\geq H_\beta(\pi|\tau), \\ H_\beta(\sigma|\pi) + H_\beta(\tau|\sigma) &\geq H_\beta(\tau|\pi). \end{aligned}$$

Adding these inequality gives

$$d_\beta(\pi, \sigma) + d_\beta(\sigma, \tau) \geq d_\beta(\pi, \tau),$$

which is the triangular inequality for d_β .

The symmetry of d_β is obvious and it is clear that $d_\beta(\pi, \pi) = 0$ for every $\pi \in \text{PART}(S)$.

Suppose now that $d_\beta(\pi, \sigma) = 0$. Since the values of β -conditional entropies are non-negative this implies $H_\beta(\pi|\sigma) = H_\beta(\sigma|\pi) = 0$. By Theorem 2.3 we have both $\sigma \leq \pi$ and $\pi \leq \sigma$, respectively, so $\pi = \sigma$. Thus, d_β is a metric on $\text{PART}(S)$. ■

It is clear that $d_\beta(\pi, \omega_S) = H_\beta(\pi)$ and $d_\beta(\pi, \alpha_S) = H_\beta(\alpha_S|\pi)$.

Another useful form of d_β can be obtained starting from the equalities $H_\beta(\pi|\sigma) = H_\beta(\pi \wedge \sigma) - H_\beta(\sigma)$ and $H_\beta(\sigma|\pi) = H_\beta(\pi \wedge \sigma) - H_\beta(\pi)$. Thus, we have:

$$d_\beta(\pi, \sigma) = 2H_\beta(\pi \wedge \sigma) - H_\beta(\pi) - H_\beta(\sigma), \quad (1)$$

for $\pi, \sigma \in \text{PART}(S)$.

The behavior of the distance d_β with respect to partition addition is discussed in the next statement.

Theorem 3.4 *Let S be a finite set, π, θ be two partitions of S , where $\theta = \{D_1, \dots, D_h\}$. If $\sigma_i \in \text{PART}(D_i)$ for $1 \leq i \leq h$, then*

$$d_\beta(\pi, \sigma_1 + \dots + \sigma_h) = \sum_{i=1}^h \left(\frac{|D_i|}{|S|} \right)^\beta d_\beta(\pi_{D_i}, \sigma_i) + H_\beta(\theta|\pi).$$

Proof. This statement follows directly from Theorem 2.5. ■

The next statement is a generalization of the axiom system proposed in [7] for the Shannon entropic metric and for the Mirkin metric.

Theorem 3.5 *The following properties hold in the metric space $(\text{PART}(S), d_\beta)$:*

- (i) if $\sigma \leq \pi$, then $d_\beta(\pi, \sigma) = H_\beta(\sigma) - H_\beta(\pi)$;
 - (ii) $d_\beta(\alpha_S, \sigma) + d_\beta(\sigma, \omega_S) = d_\beta(\alpha_S, \omega_S)$;
 - (iii) $d_\beta(\pi, \pi \wedge \sigma) + d_\beta(\pi \wedge \sigma, \sigma) = d_\beta(\pi, \sigma)$,
- for every partitions $\pi, \sigma \in \text{PART}(S)$.

Furthermore, we have $d(\omega_T, \alpha_T) = \frac{1-|T|^{1-\beta}}{1-2^{1-\beta}}$, for every subset T of S .

Proof. The first three statements of the theorem follow immediately from Equality 1; the last part is an application of the definition of d_β . ■

A generalization of a result obtained in [7] is contained in the next statement, which gives an axiomatization of the metric d_β .

Theorem 3.6 *Let $d : \text{PART}(S)^2 \rightarrow \mathbb{R}_{\geq 0}$ be a function that satisfies the following conditions:*

- (D1) d is symmetric, that is, $d(\pi, \sigma) = d(\sigma, \pi)$;
- (D2) $d(\alpha_S, \sigma) + d(\sigma, \omega_S) = d(\alpha_S, \omega_S)$;
- (D3) $d(\pi, \sigma) = d(\pi, \pi \wedge \sigma) + d(\pi \wedge \sigma, \sigma)$;
- (D4) if $\sigma, \theta \in \text{PART}(S)$ such that $\theta = \{D_1, \dots, D_h\}$ and $\sigma \leq \theta$ then we have:

$$d(\theta, \sigma) = \sum_{i=1}^h \left(\frac{|D_i|}{|S|} \right)^\beta d(\omega_{D_i}, \sigma_{D_i});$$

(D5) $d(\omega_T, \alpha_T) = \frac{1-|T|^{1-\beta}}{1-2^{1-\beta}}$, for every $T \subseteq S$.
Then, $d = d_\beta$.

Proof. Choosing $\sigma = \alpha_S$ in axiom (D4) and using (D5) we can write:

$$\begin{aligned} d(\alpha_S, \theta) &= \sum_{i=1}^h \left(\frac{|D_i|}{|S|} \right)^\beta d(\omega_{D_i}, \alpha_{D_i}) \\ &= \sum_{i=1}^h \left(\frac{|D_i|}{|S|} \right)^\beta \frac{1 - |D_i|^{1-\beta}}{1 - 2^{1-\beta}} \\ &= \frac{\sum_{i=1}^h |D_i|^\beta - |S|}{(1 - 2^{1-\beta})|S|^\beta}. \end{aligned}$$

From Axioms (D2) and (D5) it follows that

$$\begin{aligned} d(\theta, \omega_S) &= d(\alpha_S, \omega_S) - d(\alpha_S, \theta) \\ &= \frac{1 - |S|^{1-\beta}}{1 - 2^{1-\beta}} - \frac{\sum_{i=1}^h |D_i|^\beta - |S|}{(1 - 2^{1-\beta})|S|^\beta} \\ &= \frac{|S|^\beta - \sum_{i=1}^h |D_i|^\beta}{(1 - 2^{1-\beta})|S|^\beta}. \end{aligned}$$

Let now $\pi, \sigma \in \text{PART}(S)$, where $\pi = \{B_1, \dots, B_m\}$ and $\sigma = \{C_1, \dots, C_n\}$. Since $\pi \wedge \sigma \leq \pi$ and $\sigma_{B_i} = \{C_1 \cap B_i, \dots, C_n \cap B_i\}$, an application of Axiom (D4) yields

$$\begin{aligned} d(\pi, \pi \wedge \sigma) &= \sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^\beta d(\omega_{B_i}, (\pi \wedge \sigma)_{B_i}) \\ &= \sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^\beta d(\omega_{B_i}, \sigma_{B_i}) \\ &= \sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^\beta \frac{|B_i|^\beta - \sum_{j=1}^n |B_i \cap C_j|^\beta}{(1 - 2^{1-\beta})|B_i|^\beta} \\ &= \frac{1}{(1 - 2^{1-\beta})|S|^\beta} \left(\sum_{i=1}^m |B_i|^\beta - \sum_{j=1}^n \sum_{i=1}^m |B_i \cap C_j|^\beta \right), \end{aligned}$$

because $(\pi \wedge \sigma)_{B_i} = \sigma_{B_i}$. ■

By Axiom (D1) we obtain the similar equality:

$$d(\pi \wedge \sigma, \sigma) = \frac{1}{(1 - 2^{1-\beta})|S|^\beta} \left(\sum_{i=1}^m |B_i|^\beta - \sum_{j=1}^n \sum_{i=1}^m |B_i \cap C_j|^\beta \right),$$

which, by Axiom (D3), implies:

$$\begin{aligned} d(\pi, \sigma) &= \frac{1}{(1 - 2^{1-\beta})|S|^\beta} \left(\sum_{i=1}^m |B_i|^\beta + \sum_{j=1}^n |C_j|^\beta \right. \\ &\quad \left. - 2 \sum_{j=1}^n \sum_{i=1}^m |B_i \cap C_j|^\beta \right), \end{aligned}$$

that is $d(\pi, \sigma) = d_\beta(\pi, \sigma)$. \blacksquare

In fact, the Mirkin metric [8] (up to a multiplicative constant) is obtained for $\beta = 2$:

$$d_2(\pi, \sigma) = \frac{2}{|S|^2} \left(\sum_{i=1}^m |B_i|^2 + \sum_{j=1}^n |C_j|^2 - 2 \sum_{i=1}^m \sum_{j=1}^n |B_i \cap C_j|^2 \right).$$

The corresponding generalized entropy $H_2(\pi)$ is double the Gini index of the partition $\pi = \{B_1, \dots, B_m\}$:

$$H_2(\pi) = 2 \left(\sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^2 - 1 \right)$$

It is worth noting that one could axiomatize the entropy starting from the notion metric between partitions. Indeed, if the β -entropy of a partition $\pi \in \text{PART}(S)$ is defined as:

$$H_\beta(\pi) = d_\beta(\pi, \omega_S),$$

then we would retrieve the β -entropy:

$$H_\beta(\pi) = \frac{1}{2^{1-\beta} - 1} \left(\sum_{i=1}^m \left(\frac{|B_i|}{|S|} \right)^\beta - 1 \right).$$

4 Partition Valuations and β -Entropy

Metrics generated by β -conditional entropies are closely related to lower valuations of the upper semi-modular lattices of partitions of finite sets. This connection was established in [3] and studied in [2, 1, 9].

A *lower valuation* on a lattice (L, \vee, \wedge) is a mapping $v : L \rightarrow \mathbb{R}$ such that $v(\pi \vee \sigma) + v(\pi \wedge \sigma) \geq v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$. If the reverse inequality is satisfied, that is, if $v(\pi \vee \sigma) + v(\pi \wedge \sigma) \leq v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$, then v is referred to as an *upper valuation*.

If $v \in L$ is both a lower and upper valuation, that is, if $v(\pi \vee \sigma) + v(\pi \wedge \sigma) = v(\pi) + v(\sigma)$ for every $\pi, \sigma \in L$, then v is a valuation on L .

We have the following result:

Theorem 4.1 *Let $\pi, \sigma \in \text{PART}(S)$ be two partitions. We have:*

$$\begin{aligned} d_\beta(\pi, \sigma) &= 2 \cdot d_\beta(\pi \wedge \sigma, \omega_S) - d_\beta(\pi, \omega_S) - d_\beta(\sigma, \omega_S) \\ &= d_\beta(\alpha_S, \pi) + d_\beta(\alpha_S, \sigma) - 2 \cdot d_\beta(\alpha_S, \pi \wedge \sigma). \end{aligned}$$

Proof. The equalities of the theorem can be immediately verified by using the definition of d_β . \blacksquare

Corollary 4.2 *Let θ, τ be two partitions from $\text{PART}(S)$. If $\theta \leq \tau$ and we have either $d_\beta(\theta, \omega_S) = d_\beta(\tau, \omega_S)$ or $d_\beta(\alpha_S, \theta) = d_\beta(\alpha_S, \tau)$, then $\theta = \tau$.*

Proof. Observe that if $\theta \leq \tau$, then Theorem 4.1 implies

$$d_\beta(\theta, \tau) + d_\beta(\tau, \omega_S) = d_\beta(\theta, \omega_S),$$

and

$$d_\beta(\theta, \tau) = d_\beta(\alpha_S, \tau) - d_\beta(\alpha_S, \theta).$$

Suppose that $d_\beta(\theta, \omega_S) = d_\beta(\tau, \omega_S)$. Since $d_\beta(\tau, \omega_S) = d_\beta(\theta, \omega_S)$ it follows that $d_\beta(\theta, \tau) = 0$, so $\theta = \tau$.

If $d_\beta(\alpha_S, \theta) = d_\beta(\alpha_S, \tau)$ the same conclusion can be reached immediately. \blacksquare

It is known [3] that if there exists a positive valuation v on L , then L must be a modular lattice. Since the partition lattice of a set is an upper-semimodular lattice that is not modular ([3]) it is clear that positive valuations do not exist on partition lattices. However, lower and upper valuations do exist, as shown next:

Theorem 4.3 *Let S be a finite set. Define the mappings $v_\beta : \text{PART}(S) \rightarrow \mathbb{R}$ and let $w_\beta : \text{PART}(S) \rightarrow \mathbb{R}$ be by $v_\beta(\pi) = d_\beta(\alpha_S, \pi)$ and $w_\beta(\pi) = d_\beta(\pi, \omega_S)$, respectively, for $\pi \in \text{PART}(S)$. Then, v_β is a lower valuation and w_β is an upper valuation on the lattice $(\text{PART}(S), \vee, \wedge)$.*

Proof. Theorem 4.1 allows us to write:

$$\begin{aligned} d_\beta(\pi, \sigma) &= v_\beta(\pi) + v_\beta(\sigma) - 2v_\beta(\pi \wedge \sigma) \\ &= 2w_\beta(\pi \wedge \sigma) - w_\beta(\pi) - w_\beta(\sigma), \end{aligned}$$

for every $\pi, \sigma \in \text{PART}(S)$.

If we rewrite the triangular inequality $d_\beta(\pi, \tau) + d_\beta(\tau, \sigma) \geq d_\beta(\pi, \sigma)$ using the valuations v_β and w_β we obtain:

$$\begin{aligned} v_\beta(\tau) + v_\beta(\pi \wedge \sigma) &\geq v_\beta(\pi \wedge \tau) + v_\beta(\tau \wedge \sigma), \\ w_\beta(\pi \wedge \tau) + w_\beta(\tau \wedge \sigma) &\geq w_\beta(\tau) + w_\beta(\pi \wedge \sigma), \end{aligned}$$

for every $\pi, \tau, \sigma \in \text{PART}(S)$. If we choose $\tau = \pi \vee \sigma$ the last inequalities yield:

$$\begin{aligned} v_\beta(\pi) + v_\beta(\sigma) &\leq v_\beta(\pi \vee \sigma) + v_\beta(\pi \wedge \sigma) \\ w_\beta(\pi) + w_\beta(\sigma) &\geq w_\beta(\pi \vee \sigma) + w_\beta(\pi \wedge \sigma), \end{aligned}$$

for every $\pi, \sigma \in \text{PART}(S)$, which shows that v_β is a lower valuation and w_β is an upper valuation on the lattice $(\text{PART}(S), \vee, \wedge)$. \blacksquare

5 Conclusion and Future Work

We introduced a general axiomatization of a family of metrics on the set of partitions of a finite set that is related to generalized entropies. These metrics are used for a variety of data mining tasks ranging from clustering [7, 15] to classification [13, 14] and discretization [10].

The value of the parameter β that gives optimal results depends on the statistical properties of the data set that is analyzed. Developing algorithms that learn the values of β for a specific data set and mining task remains an open problem.

References

- [1] J. Barthélemy. Remarques sur les propriétés métriques des ensembles ordonnés. *Math. Sci. hum.*, 61:39–60, 1978.
- [2] J. Barthélemy and B. Leclerc. The median procedure for partitions. In *Partitioning Data Sets*, pages 3–34. American Mathematical Society, Providence, 1995.
- [3] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, 1973.
- [4] Z. Daróczy. Generalized information functions. *Information and Control*, 16:36–51, 1970.
- [5] R. L. de Mántaras. A distance-based attribute selection measure for decision tree induction. *Machine Learning*, 6:81–92, 1991.
- [6] I. C. Lerman. *Classification et analyse ordinaire des données*. Dunod, Paris, 1981.
- [7] M. Meilă. Comparing clusterings: an axiomatic view. In *Proc. of ICML*, pages 577–584, Bonn, Germany, 2005.
- [8] B. Mirkin. *Mathematical Classification and Clustering*. Kluwer, Boston, 1996.
- [9] B. Monjardet. Metrics on partially ordered sets – a survey. *Discrete Mathematics*, 35:173–184, 1981.
- [10] D. S. S. R. Butterworth, D. A. Simovici and L. Ohno-Machado. A greedy algorithm for supervised discretization. *Journal of Biomedical Informatics*, 37(4):285–292, 2004.
- [11] I. Rosenberg and D. A. Simovici. An abstract axiomatization of the notion of entropy. In *Proceedings of ISMVL*, pages 8–13, Calgary, Canada, 2005.
- [12] D. A. Simovici and S. Jaroszewicz. An axiomatization of partition entropy. *IEEE Transactions on Information Theory*, 48:2138–2142, 2002.
- [13] D. A. Simovici and S. Jaroszewicz. Generalized entropy and decision trees. In *EGC 2003 - Journées francophones d'Extraction et de Gestion de Connaissances*, pages 369–380, Lyon, France, 2003.
- [14] D. A. Simovici and S. Jaroszewicz. A new metric splitting criterion for decision trees. *Journal of Parallel, Emerging and Distributed Computing*, 21:239–256, 2006.
- [15] D. A. Simovici, N. Singla, and M. Kuperberg. Metric incremental clustering of nominal data. In *Proceedings of ICDM 2004*, pages 523–527, Brighton, UK, 2004.