LIBERTAS MATHEMATICA, vol XXX (2010)

Several Remarks on Metrics on Partition Lattices and Their Applications in Data Mining

Dan A. SIMOVICI

Abstract. We develop an axiomatization of a class of metrics on lattices of partitions of finite sets, which leads to a new metric axiomatization of the notion of entropy in an algebraic framework. We point to the application that this type of metrics has in data mining and machine learning.

Keywords: entropy, modular lattice, metric.

1 Introduction

Partitions play a crucial role in data mining. A clustering can be regarded as a partition. Attributes of sets of objects in relational databases induce equivalence relations on these sets, and these, in turn, are described by partitions. Partitions serve in data discretization, and can be construct to design decision trees, a mainstay of classifiers. Frequently, data mining requires an evaluation of the dissimilarity between partitions and this can be best achieved by defining metrics on sets of partitions that are compatible with their natural lattice structure and have other properties required by specific applications. Thus, the field of data mining has focused on the metric space of partitions. In [1] we developed an axiomatization of a generalization of Shannon entropy known as Havrda-Charvat entropy [2].

The main goal of this paper is to axiomatize a metric on the space of partitions of a finite set that is closely related to generalized entropy. Related results motivated by desirability of certain properties of clustering distances were obtained in [3]. Here we are primarily interested in distance axiomatizations that lead to entropy axiomatizations.

Unless stated otherwise, all sets are assumed to be finite. The cardinality of a set S is denoted by $|S|$. A *partition of a set* S is a non-empty collection of non-empty subsets of S, $\pi = \{B_i \mid i \in I\}$ such that $\bigcup \pi = S$ and $B_i \cup B_j = \emptyset$ when $i \neq j$ for $i, j \in I$. The sets B_i are the *blocks* of π. The set of partitions of S is denoted by $\mathsf{PART}(S)$.

A partial order relation on $\mathsf{PART}(S)$ is defined by $\pi \leq \sigma$ for $\pi, \sigma \in \mathsf{PART}(S)$ if every block of B is included in a block of σ. This is easily seen to be equivalent to requiring that each block of σ is a union of blocks of π.

The partially ordered set $(\mathsf{PART}(S), \leq)$ is actually a bounded lattice. The infimum $\pi \wedge \pi'$ of two partitions π and π' is the partition that consists of non-empty intersections of blocks of π and π'. For a description of the supremum $\pi \vee \pi'$ of the partitions π, π' see [4]. The least element of this lattice is the partition $\alpha_S = \{\{s\} \mid s \in S\}$; the largest is the partition $\omega_S = \{S\}$.

The partition σ *covers* the partition π if σ is obtained from π by fusing two blocks of this partition. This is denoted by $\pi \prec \sigma$. We have $\pi \leq \pi'$, if and only if there exists a sequence of partitions $\sigma_0, \sigma_1, \ldots, \sigma_r$ such that $\pi = \sigma_0 \prec \sigma_1 \prec \cdots \prec \sigma_r = \pi'$.

Let C be a subset of the set S and let $\pi = \{B_i \mid i \in I\} \in \mathsf{PART}(S)$ be a partition. The *trace* of π on C is the partition $\pi_C = \{B_i \cap C \mid B_i \cap C \neq \emptyset \text{ and } i \in I\}$.

Let $(L, \wedge, \vee)$ be a lattice. A function $f : L \longrightarrow \mathbb{R}$ is *sub-modular* if $f(x \wedge y) + f(x \vee y) \leq f(x) + f(y)$ for every $x, y \in L$.

Let π, σ be two partitions in $\mathsf{PART}(S)$. We say that σ *covers* π if $\pi \leq \sigma$ and there exists no partition $\theta \in \mathsf{PART}(S)$ such that $\pi < \theta < \sigma$. This is denoted by $\pi \prec \sigma$. One can show that $\pi \prec \sigma$ if and only if the blocks of σ are the same as the blocks of π, with the exception of a single block of σ that is the union of two blocks of π.

2 Axiomatization of Metrics on Partitions

Let $\Phi : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{\geq 0}$ be a function such that $\Phi(1) = 1$ and $\Phi(x) = 0$ implies $x = 0$. The function d_Φ is defined by the following systems of axioms:

(D1) d_Φ is symmetric, that is, $d_\Phi(\pi, \sigma) = d_\Phi(\sigma, \pi)$;

(D2) $d_\Phi(\alpha_S, \sigma) + d_\Phi(\sigma, \omega_S) = d_\Phi(\alpha_S, \omega_S)$;

(D3) $d_\Phi(\pi, \sigma) = d_\Phi(\pi, \pi \wedge \sigma) + d_\Phi(\pi \wedge \sigma, \sigma)$;

(D4) if $\sigma, \theta \in \mathsf{PART}(S)$ such that $\theta = \{D_1, \ldots, D_h\}$ and $\sigma \leq \theta$, then

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

(D5) $d_\Phi(\omega_T, \alpha_T)$ is a characteristic of the set T and there exists a positive constant k such that for every set S such that $T \subseteq S$ we have

$$d_\Phi(\omega_T, \alpha_T) = k\left(1 - \frac{|T|}{\Phi(|S|)\Phi\left(\frac{|T|}{|S|}\right)}\right) \geq 0,$$

for every $T \subseteq S$.

Our goal is to obtain conditions under which the function d_Φ defined in the next theorem is a metric on the set of partitions of a finite set.

Let S and U be two sets such that $T \subseteq S \cap U$, where $T \neq \emptyset$. Since $d_\Phi(\alpha_T, \omega_T)$ is the same regardless whether T is considered as a subset of S or of U we have

$$d_\Phi(\alpha_T, \omega_T) = k\left(1 - \frac{|T|}{\Phi(|S|)\Phi\left(\frac{|T|}{|S|}\right)}\right) = k\left(1 - \frac{|T|}{\Phi(|U|)\Phi\left(\frac{|U|}{|S|}\right)}\right).$$

This implies
$$\frac{1}{\Phi(|S|)\Phi\left(\frac{|T|}{|S|}\right)} = \frac{1}{\Phi(|U|)\Phi\left(\frac{|T|}{|U|}\right)}.$$
If $T = U$ in the above equality we obtain:
$$\Phi\left(\frac{|T|}{|S|}\right) = \Phi(1)\frac{\Phi(|T|)}{\Phi(|S|)} = \frac{\Phi(|T|)}{\Phi(|S|)}. \tag{1}$$

Since $d_\Phi(\alpha_T, \omega_T) \geq 0$ we have
$$\frac{|T|}{\Phi(|S|)\Phi\left(\frac{|T|}{|S|}\right)} \leq 1,$$
for any set S such that $T \subseteq S$. Choosing $S = T$ we obtain $|T| \leq \Phi(T)$ for any finite set T.

Theorem 1. *Let $d_\Phi : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ be a function that satisfies Axioms **(D1)-(D5)**. We have:*
$$\begin{aligned} d_\Phi(\pi, \sigma) &= k\left(\sum_{i=1}^{m}\Phi\left(\frac{|B_i|}{|S|}\right) + \sum_{j=1}^{n}\Phi\left(\frac{|C_j|}{|S|}\right)\right. \\ &\quad \left. -2\sum_{i=1}^{m}\sum_{j=1}^{n}\Phi\left(\frac{|B_i \cap C_j|}{|S|}\right)\right). \end{aligned}$$

Proof. For $\theta = \{D_1, \ldots, D_h\}$ we have
$$\begin{aligned} d_\Phi(\alpha_S, \theta) &= \sum_{i=1}^{h}\Phi\left(\frac{|D_i|}{|S|}\right) d_\Phi(\omega_{D_i}, \alpha_{D_i}) \\ &= k\sum_{i=1}^{h}\Phi\left(\frac{|D_i|}{|S|}\right)\left(1 - \frac{|D_i|}{\Phi(|S|)\Phi\left(\frac{|D_i|}{|S|}\right)}\right) \\ &= k\left(\sum_{i=1}^{h}\Phi\left(\frac{|D_i|}{|S|}\right) - \frac{|S|}{\Phi(|S|)}\right), \end{aligned}$$
by taking $\sigma = \alpha_S$ in Axiom **D4**, and by applying **(D5)**.

By Axiom **(D2)** we have
$$\begin{aligned} d_\Phi(\theta, \omega_S) &= d_\Phi(\alpha_S, \omega_S) - d_\Phi(\alpha_S, \theta) \\ &= k\left(1 - \frac{|S|}{\Phi(|S|)\Phi(1)}\right) - k\left(\sum_{i=1}^{h}\Phi\left(\frac{|D_i|}{|S|}\right) - \frac{|S|}{\Phi(|S|)}\right) \\ &= k\left(1 - \sum_{i=1}^{h}\Phi\left(\frac{|D_i|}{|S|}\right)\right), \end{aligned} \tag{2}$$

because $\Phi(1) = 1$.

Let now $\pi, \sigma \in \mathsf{PART}(S)$, where $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$. Since $\pi \wedge \sigma \leq \pi$, an application of **(D4)** yields

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

Similarly, we can derive

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

so we have (by Axiom **(D3)**):

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

which is the desired equality. □

By taking $\theta = \omega_S$ in Axiom **(D4)** we obtain

$$d_\Phi(\pi, \omega_S) = d_\Phi(\pi, \pi) + d_\Phi(\pi, \omega_S),$$

which implies $d_\Phi(\pi, \pi) = 0$ for every $\pi \in \mathsf{PART}(S)$.

The equality of Theorem 1 can be written as

$$\begin{aligned} d_\Phi(\pi, \sigma) &= k \left(\sum_{i=1}^{m} \Phi\left(\frac{|B_i|}{|S|}\right) \left(1 - \sum_{j=1}^{n} \Phi\left(\frac{|B_i \cap C_j|}{|B_i|}\right)\right)\right. \\ &\quad \left. + \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) \left(1 - \sum_{i=1}^{m} \Phi\left(\frac{|B_i \cap C_j|}{|C_j|}\right)\right)\right) \\ &= \sum_{i=1}^{m} \Phi\left(\frac{|B_i|}{|S|}\right) d_\Phi(\sigma_{B_i}, \omega_{B_i}) + \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) d_\Phi(\pi_{C_j}, \omega_{C_j}). \end{aligned}$$

Let $C_\Phi : \mathsf{PART}(S)^2 \longrightarrow \mathbb{R}_{\geq 0}$ be the function defined by

$$C_\Phi(\pi, \sigma) = \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) d_\Phi(\pi_{C_j}, \omega_{C_j}) \tag{3}$$

for $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$. We have

$$\begin{aligned} C_\Phi(\pi, \sigma) &= k \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right)\left(1 - \sum_{i=1}^{m} \Phi\left(\frac{|B_i \cap C_j|}{|C_j|}\right)\right) \\ &= k\left(\sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) - \sum_{i=1}^{m}\sum_{j=1}^{n} \Phi\left(\frac{|B_i \cap C_j|}{|S|}\right)\right), \end{aligned}$$

which implies

$$d_\Phi(\pi, \sigma) = C_\Phi(\pi, \sigma) + C_\Phi(\sigma, \pi). \tag{4}$$

Observe that

$$C_\Phi(\pi, \sigma) = d_\Phi(\pi \wedge \sigma, \omega_S) - d_\Phi(\sigma, \omega_S), \tag{5}$$

$$d_\Phi(\pi, \sigma) = 2d_\Phi(\pi \wedge \sigma, \omega_S) - d_\Phi(\pi, \omega_S) - d_\Phi(\sigma, \omega_S) \tag{6}$$

Theorem 2. *If Φ is a supra-additive function, then $d_\Phi(\pi, \omega_S)$ is anti-monotonic relative to π.*

Proof. Let η and θ be two partitions of S such that $\eta \leq \theta$, $\eta = \{E_1, \ldots, E_p\}$ and $\theta = \{D_1, \ldots, D_h\}$. For each block D_i there exists a index set K_i such that $D_i = \bigcup\{E_k \in \eta \mid k \in K_i\}$. Since the sets E_k are disjoint, we have $|D_i| = \sum\{|E_k| \mid k \in K_i\}$, which implies $\Phi(|D_i|) \geq \sum\{\Phi(|E_k|) \mid k \in K_i\}$ because Φ is supra-additive. Therefore, $d_\Phi(\theta, \omega_S) \leq d_\Phi(\eta, \omega_S)$. □

Lemma 3. *Let $\Phi : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}$ be a convex, supra-additive function. If C and D are two disjoint subsets of S and $\pi \in \mathsf{PART}(S)$, then we have:*

$$\begin{aligned} &\Phi\left(\frac{|C \cup D|}{|S|}\right) d_\Phi(\pi_{C\cup D}, \omega_{C\cup D}) \\ &\geq \Phi\left(\frac{|C|}{|S|}\right) d_\Phi(\pi_C, \omega_C) + \Phi\left(\frac{|D|}{|S|}\right) d_\Phi(\pi_D, \omega_D). \end{aligned}$$

Proof. Suppose that $\pi = \{B_1, \ldots, B_m\}$. Let

$$x_i = \frac{|B_i \cap C|}{|C|} \text{ and } y_i = \frac{|B_i \cap D|}{|D|},$$

for $1 \leq i \leq m$. Clearly, $\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i = 1$. Note that

$$\frac{|B_i \cap (C \cup D)|}{|C \cup D|} = px_i + qy_i,$$

where $p = \frac{|C|}{|C \cup D|}$ and $q = \frac{|D|}{|C \cup D|}$. Since Φ is convex we have

$$\Phi(px_i + qy_i) \leq p\Phi(x_i) + q\Phi(y_i),$$

so

$$1-\sum_{i=1}^{m}\Phi(px_i+qy_i)\geq 1-\sum_{i=1}^{m}p\Phi(x_i)-q\sum_{i=1}^{m}\Phi(y_i).$$

By Equality (2) we have

$$\begin{aligned}\Phi\left(\frac{|C\cup D|}{|S|}\right)d_\Phi(\pi_{C\cup D},\omega_{C\cup D}) &= \Phi\left(\frac{|C\cup D)}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi\left(\frac{|B_i\cap(C\cup D)|}{|C\cup D|}\right)\right)\\ &= \Phi\left(\frac{|C\cup D|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi(px_i+qy_i)\right),\end{aligned}$$

and

$$\begin{aligned}\Phi\left(\frac{|C|}{|S|}\right)d_\Phi(\pi_C,\omega_C) &= \Phi\left(\frac{|C|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi\left(\frac{|B_i\cap C|)}{|C|}\right)\right),\\ &= \Phi\left(\frac{|C|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi(x_i)\right),\\ \Phi\left(\frac{|D|}{|S|}\right)d_\Phi(\pi_D,\omega_D) &= \Phi\left(\frac{|D|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi\left(\frac{|B_i\cap D|}{|C\cup D|}\right)\right)\\ &= \Phi\left(\frac{|D|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi(y_i)\right).\end{aligned}$$

This allows us to write

$$\begin{aligned}&\Phi\left(\frac{|C\cup D|}{|S|}\right)d_\Phi(\pi_{C\cup D},\omega_{C\cup D})\\ &= \Phi\left(\frac{|C\cup D|}{|S|}\right)\left(1-\sum_{i=1}^{m}\Phi(px_i+qy_i)\right)\\ &\geq \Phi\left(\frac{|C\cup D|}{|S|}\right)\left(1-p\sum_{i=1}^{m}\Phi(x_i)-q\sum_{i=1}^{m}\Phi(y_i)\right)\\ &\geq \Phi\left(\frac{|C|}{|S|}\right)+\Phi\left(\frac{|D|}{|S|}\right)-p\Phi\left(\frac{|C\cup D|}{|S|}\right)\sum_{i=1}^{m}\Phi(x_i)-q\Phi\left(\frac{|C\cup D|}{|S|}\right)\sum_{i=1}^{m}\Phi(y_i)\\ &\qquad \text{(due to the supra-additivity of }\Phi)\\ &\geq \Phi\left(\frac{|C|}{|S|}\right)+\Phi\left(\frac{|D|}{|S|}\right)-\Phi\left(\frac{|C|}{|S|}\right)\sum_{i=1}^{m}\Phi(x_i)-\Phi\left(\frac{|D|}{|S|}\right)\sum_{i=1}^{m}\Phi(y_i)\\ &= \Phi\left(\frac{|C|}{|S|}\right)d_\Phi(\pi_C,\omega_C)+\Phi\left(\frac{|D|}{|S|}\right)d_\Phi(\pi_D,\omega_D),\end{aligned}$$

which concludes the argument. □

Theorem 4. *If Φ is a supra-additive function, then C_Φ is anti-monotonic in its first argument and is monotonic in its second argument.*

Proof. The first part of the theorem follows immediately from Theorem 2. For the second part it suffices to show that if $\sigma \prec \tau$, then $C_\Phi(\pi,\sigma) \leq C_\Phi(\pi,\tau)$ for $\pi,\sigma,\tau \in \mathsf{PART}(S)$. Suppose that $\pi = \{B_1,\ldots,B_m\}$, $\sigma = \{C_1,\ldots,C_{n-2},C_{n-1},C_n\}$ and that $\tau = \{C_1,\ldots,C_{n-2}, C_{n-1}\cup C_n\}$.

We have

$$\begin{aligned}
& C_\Phi(\pi,\tau) \\
= & \sum_{j=1}^{n-2} \frac{\Phi(|C_j|)}{\Phi(|S|)} d_\Phi(\pi_{C_j},\omega_S) + \frac{\Phi(|C_{n-1}\cup C_n|}{\Phi(|S|)} d_\Phi(\pi_{C_{n-1}\cup C_n},\omega_S) \\
\geq & \sum_{j=1}^{n-2} \frac{\Phi(|C_j|)}{\Phi(|S|)} d_\Phi(\pi_{C_j},\omega_S) + \frac{\Phi(|C_{n-1})}{\Phi(|S|)} d_\Phi(\pi_{C_{n-1}},\omega_S) + \frac{\Phi(|C_n)}{\Phi(|S|)} d_\Phi(\pi_{C_n},\omega_S) \\
= & C_\Phi(\pi,\sigma).
\end{aligned}$$

□

Lemma 5. *We have*

$$C_\Phi(\pi,\sigma\wedge\tau) + C_\Phi(\sigma,\tau) = C_\Phi(\pi\wedge\sigma,\tau).$$

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

Proof. Let $\pi = \{B_1,\ldots,B_m\}$, $\sigma = \{C_1,\ldots,C_m\}$, $\tau = \{D_1,\ldots,D_h\}$. By Equality **??** we have

$$C_\Phi(\pi,\sigma\wedge\tau) = k\left(\sum_{j=1}^{n}\sum_{\ell=1}^{h}\Phi\left(\frac{|C_j\cap D_\ell|}{|S|}\right) - \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{\ell=1}^{h}\Phi\left(\frac{|B_i\cap C_j\cap D_\ell|}{|S|}\right)\right)$$

and

$$C_\Phi(\sigma,\tau) = k\left(\sum_{\ell=1}^{h}\Phi\left(\frac{|D_\ell|}{|S|}\right) - \sum_{j=1}^{h}\sum_{\ell=1}^{h}\Phi\left(\frac{|C_j\cap D_\ell|}{|S|}\right)\right),$$

so

$$C_\Phi(\pi,\sigma\wedge\tau) + C_\Phi(\sigma,\tau) = C_\Phi(\pi\wedge\sigma,\tau).$$

□

Theorem 6. *If Φ is a supra-additive function, then for every partitions π,σ,τ in $\mathsf{PART}(S)$ we have*

$$C_\Phi(\pi,\sigma) + C_\Phi(\sigma,\tau) \geq C_\Phi(\pi,\tau).$$

Proof. The monotonicity of C_Φ in its second argument and the anti-monotonicity of the same in the first argument imply

$$\begin{aligned} C_\Phi(\pi,\sigma) + C_\Phi(\sigma,\tau) &\geq C_\Phi(\pi, \sigma\wedge\tau) + C_\Phi(\sigma,\tau) \\ &= C_\Phi(\pi\wedge\sigma, \tau) \\ &\quad \text{(by Lemma 5)} \\ &\geq C_\Phi(\pi,\tau), \end{aligned}$$

which is the desired inequality. □

Corollary 7. *If Φ is a supra-additive function, then d_Φ satisfies the triangular inequality, that is,*

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

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

Proof. This statement follows immediately from Theorem 6 and Equality (4). □

Theorem 8. *Let π,σ be two partitions of the set S. We have $C_\Phi(\pi,\sigma) = 0$ if and only if $\sigma \leq \pi$.*

Proof. Suppose that $\sigma = \{C_1,\ldots,C_n\}$. If $\sigma \leq \pi$, then $\pi_{C_j} = \omega_{C_j}$ and, therefore, by Equality 3, we have $C_\Phi(\pi,\sigma) = 0$.

Conversely, suppose that

$$C_\Phi(\pi,\sigma) = \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) d_\Phi(\pi_{C_j}, \omega_{C_j}) = 0.$$

This implies $d_\Phi(\pi_{C_j}, \omega_{C_j}) = 0$, so $\pi_{C_j} = \omega_{C_j}$ for $1 \leq j \leq n$, which means that every block C_j of σ is included in a block of π. Thus, $\sigma \leq \pi$. □

Corollary 9. *If Φ is convex and supra-additive, then d_Φ is a metric on $\mathsf{PART}(S)$.*

Proof. The function d_Φ is symmetric and, by Corollary 7 it has the triangular property. We saw that $d_\Phi(\pi,\pi) = 0$ for every $\pi \in \mathsf{PART}(S)$. Suppose now that $d_\Phi(\pi,\sigma) = 0$. Equality 4 implies $C_\Phi(\pi,\sigma) = C_\Phi(\sigma,\pi)$, so $\sigma = \pi$ by Theorem 8. □

The next result shows that the diameter of the metric space $(\mathsf{PART}(U), d_\Phi)$, where $U \subseteq S$ increases monotonically with the cardinality of U.

Theorem 10. *Let U,V be two subsets of a set S such that $|U| \leq |V|$. If the function $\varphi : \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{\geq 0}$ given by $\varphi(x) = \frac{\Phi(x)}{x}$ is monotonic on $[0,1]$, then $d_\Phi(\alpha_U,\omega_U) \leq d_\Phi(\alpha_V,\omega_V)$.*

Proof. Since $|U| \leq |V|$, we have $\varphi(|U|) \leq \varphi(|V|)$, so

$$\frac{\Phi(|U|)}{|U|} \leq \frac{\Phi(|V|)}{|V|},$$

which, by Axiom **(D5)** implies $d_\Phi(\alpha_U,\omega_U) \leq d_\Phi(\alpha_V,\omega_V)$. □

Example 11. Let $\Phi : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}$ be defined by $\Phi(x) = x^\beta$, where $\beta > 1$. Φ is clearly monotonic, multiplicative and supra-additive, as it can be easily be verified. In addition, note that $\varphi(x) = \frac{\Phi(x)}{x} = x^{\beta-1}$ is also monotonic.

As a result, the metric d_Φ is given by

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

For $\beta = 2$ we obtain a metric that is identical to the Mirkin metric [5] (up to a multiplicative constant).

Choose now the constant k as $k = \frac{1}{1-2^{1-\beta}}$ for $\Phi(x) = x^\beta$, where $\beta > 1$. We have

$$\begin{aligned}
&\lim_{\beta\to 1} d_\Phi(\pi,\sigma)\\
= &\lim_{\beta\to 1}\frac{\sum_{i=1}^{m}\left(\frac{|B_i|}{|S|}\right)^\beta + \sum_{j=1}^{n}\left(\frac{|C_j|}{|S|}\right)^\beta - 2\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{|B_i\cap C_j|}{|S|}\right)^\beta}{1-2^{1-\beta}}\\
= &\lim_{\beta\to 1}\frac{\sum_{i=1}^{m}\left(\frac{|B_i|}{|S|}\right)^\beta \ln\frac{|B_i|}{|S|} + \sum_{j=1}^{n}\left(\frac{|C_j|}{|S|}\right)^\beta \ln\frac{|C_j|}{|S|} - 2\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{|B_i\cap C_j|}{|S|}\right)^\beta \ln\frac{|B_i\cap C_j|}{|S|}}{2^{1-\beta}\ln 2}\\
&\text{(by applying l'Hôpital rule)}\\
= &\frac{\sum_{i=1}^{m}\frac{|B_i|}{|S|}\ln\frac{|B_i|}{|S|} + \sum_{j=1}^{n}\frac{|C_j|}{|S|}\ln\frac{|C_j|}{|S|} - 2\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|B_i\cap C_j|}{|S|}\ln\frac{|B_i\cap C_j|}{|S|}}{\ln 2}\\
= &\sum_{i=1}^{m}\frac{|B_i|}{|S|}\log_2\frac{|B_i|}{|S|} + \sum_{j=1}^{n}\frac{|C_j|}{|S|}\log_2\frac{|C_j|}{|S|} - 2\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{|B_i\cap C_j|}{|S|}\log_2\frac{|B_i\cap C_j|}{|S|}.
\end{aligned}$$

This metric d_Φ, where $\Phi(x) = x^\beta$ for $x \in \mathbb{R}_{\geq 0}$ and $\beta > 1$ was used by R. López de Màntaras in [6] for decision tree induction.

3 Partition Metrics and Lattice Valuations

Metrics on lattices are closely related to lower valuations of the upper semi-modular lattices of partitions of finite sets. This connection was established in [7] and studied in [8, 9, 10].

A *lower valuation* on a lattice $(L, \vee, \wedge)$ is a mapping $v : L \longrightarrow \mathbb{R}$ 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 12. *Let $\pi, \sigma \in \mathsf{PART}(S)$ be two partitions. We have:*

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

Proof. The equalities of the theorem can be immediately verified by using the definition of d_Φ. ☐

Corollary 13. *Let $\theta, \tau \in \mathsf{PART}(S)$. If $\theta \leq \tau$ and we have either $d_\Phi(\theta, \omega_S) = d_\Phi(\tau, \omega_S)$ or $d_\Phi(\alpha_S, \theta) = d_\Phi(\alpha_S, \tau)$, then $\theta = \tau$.*

Proof. Note that if $\theta \leq \tau$, then Theorem 12 implies

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

and

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

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

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

It is known [7] 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 ([7]) it is clear that positive valuations do not exist on partition lattices. However, lower and upper valuations do exist, as we show next.

Theorem 14. *Let S be a finite set. Define the mappings $\mathcal{K}_\Phi : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ and let $\mathcal{H}_\Phi : \mathsf{PART}(S) \longrightarrow \mathbb{R}$ be by $\mathcal{K}_\Phi(\pi) = d_\Phi(\alpha_S, \pi)$ and $\mathcal{H}_\Phi(\pi) = d_\Phi(\pi, \omega_S)$, respectively, for $\pi \in \mathsf{PART}(S)$. Then, $\mathcal{K}_\Phi$ is a lower valuation and $\mathcal{H}_\Phi$ is an upper valuation on the lattice $(\mathsf{PART}(S), \vee, \wedge)$.*

Proof. Theorem 12 allows us to write:

$$\begin{aligned} d_\Phi(\pi, \sigma) &= \mathcal{K}_\Phi(\pi) + \mathcal{K}_\Phi(\sigma) - 2\mathcal{K}_\Phi(\pi \wedge \sigma) \\ &= 2\mathcal{H}_\Phi(\pi \wedge \sigma) - \mathcal{H}_\Phi(\pi) - \mathcal{H}_\Phi(\sigma), \end{aligned}$$

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

If we rewrite the triangular inequality $d_\Phi(\pi, \tau) + d_\Phi(\tau, \sigma) \geq d_\Phi(\pi, \sigma)$ using the valuations $\mathcal{K}_\Phi$ and $\mathcal{H}_\Phi$ we obtain:

$$\begin{aligned} \mathcal{K}_\Phi(\tau) + \mathcal{K}_\Phi(\pi \wedge \sigma) &\geq \mathcal{K}_\Phi(\pi \wedge \tau) + \mathcal{K}_\Phi(\tau \wedge \sigma), \\ \mathcal{H}_\Phi(\pi \wedge \tau) + \mathcal{H}_\Phi(\tau \wedge \sigma) &\geq \mathcal{H}_\Phi(\tau) + \mathcal{H}_\Phi(\pi \wedge \sigma), \end{aligned}$$

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

$$\begin{aligned} \mathcal{K}_\Phi(\pi) + \mathcal{K}_\Phi(\sigma) &\leq \mathcal{K}_\Phi(\pi \vee \sigma) + \mathcal{K}_\Phi(\pi \wedge \sigma) \\ \mathcal{H}_\Phi(\pi) + \mathcal{H}_\Phi(\sigma) &\geq \mathcal{H}_\Phi(\pi \vee \sigma) + \mathcal{H}_\Phi(\pi \wedge \sigma), \end{aligned}$$

for every $\pi, \sigma \in \mathsf{PART}(S)$, which shows that $\mathcal{K}_\Phi$ is a lower valuation and $\mathcal{H}_\Phi$ is an upper valuation on the lattice $(\mathsf{PART}(S), \vee, \wedge)$. ☐

4 Entropy and Conditional Entropy

Let θ be a partition of a set S. For reasons that will become apparent in this section we will refer to $d_\Phi(\theta, \omega_S)$ as the Φ-*entropy* of the partition θ and will denote $d_\Phi(\theta, \omega_S)$ by $\mathcal{H}_\Phi(\theta)$.

We observed in Theorem 2 that, under certain conditions, the function $\mathcal{H}_\Phi(\theta)$ is anti-monotonic relative to θ. In other words, $\mathcal{H}_\Phi(\alpha_S)$ has the largest value for any partition π of the set S.

Theorem 15. *Let r be a divisor of $|S|$ and let $\rho \in \mathsf{PART}(S)$ be a partition that contains r blocks of equal size $|S|/r$. If $\theta = \{A_1, \ldots, A_r\}$ is a partition of S that consists of r blocks, then $\mathcal{H}_\Phi(\kappa) \geq \mathcal{H}_\Phi(\theta)$.*

Proof. Since Φ is a convex function, by Jensen's inequality we have

$$\Phi\left(\sum_{i=1}^{r} \frac{1}{r}|D_i|\right) \leq \sum_{i=1}^{r} \frac{1}{r}\Phi(|D_i|),$$

which amounts to

$$r\Phi\left(\frac{|S|}{r}\right) \leq \sum_{i=1}^{r} \Phi(|D_i|).$$

In turn, this yields

$$r\Phi\left(\frac{1}{r}\right) \leq \sum_{i=1}^{r} \Phi\left(\frac{|D_i|}{|S|}\right).$$

Therefore,

$$\begin{aligned}
\mathcal{H}_\Phi(\rho) &= k\left(1 - r\Phi\left(\frac{1}{r}\right)\right) \\
&\geq k\left(1 - \sum_{i=1}^{r} \Phi\left(\frac{|D_i|}{|S|}\right)\right) = \mathcal{H}_\Phi(\theta).
\end{aligned}$$

□

Thus, $\mathcal{H}_\Phi$ retains one of the most important property of the entropy: the capability of measuring concentration of objects in the blocks of a partition. The more balanced a partition is, the higher the value of the entropy is. Thus, partitions with low entropy suggest that certain blocks contain substantial groups of objects, which are identified with clusters.

By Equality (3) we can write

$$C_\Phi(\pi, \sigma) = \sum_{j=1}^{n} \Phi\left(\frac{|C_j|}{|S|}\right) \mathcal{H}_\Phi(\pi_{C_j}),$$

which justifies using the notation $\mathcal{H}_\Phi(\pi|\sigma)$ for $C_\Phi(\pi, \sigma)$ as the *entropy of π conditioned on σ*.

Theorem 16. *Let π, σ be two partitions of the finite set S. Then, we have:*

$$\mathcal{H}_\Phi(\pi \vee \sigma) + \mathcal{H}_\Phi(\pi \wedge \sigma) \leq \mathcal{H}_\Phi(\pi) + \mathcal{H}_\Phi(\sigma).$$

Proof. By Theorem 6 we have $\mathcal{H}_\Phi(\pi|\sigma) \leq \mathcal{H}_\Phi(\pi|\tau) + \mathcal{H}_\Phi(\tau|\sigma)$. Then, by applying Equality 5, we obtain

$$\mathcal{H}_\Phi(\pi \wedge \sigma) - \mathcal{H}_\Phi(\sigma) \leq \mathcal{H}_\Phi(\pi \wedge \tau) - \mathcal{H}_\Phi(\tau) + \mathcal{H}_\Phi(\tau \wedge \sigma) - \mathcal{H}_\Phi(\sigma),$$

hence

$$\mathcal{H}_\Phi(\tau) + \mathcal{H}_\Phi(\pi \wedge \sigma) \leq \mathcal{H}_\Phi(\pi \wedge \tau) + \mathcal{H}_\Phi(\tau \wedge \sigma).$$

Choosing $\tau = \pi \vee \sigma$ implies immediately the inequality of the theorem. □

Theorem 16 shows that the entropy $\mathcal{H}_\Phi$ is submodular. This result generalizes the modularity of the Gini index proven in [11] and gives an elementary proof of a result shown in [12] concerning Shannon's entropy.

5 Conclusions

We introduced an axiomatization of a class of metrics on lattices of partitions of finite sets. These metrics are used for a variety of data mining tasks ranging from clustering [3, 13, 14] to classification [15, 16], feature extraction [17], discretization [18], and the study of genetic codes [19].

A secondary result of this axiomatization is the possibility to axiomatize the notion of entropy, or the notion of conditional entropy in an algebraic framework. This approach takes advantage of the natural lattice structure of the collection of partitions of a finite set, a structure that is not natural on finite probability fields.

References

[1] D. A. Simovici and S. Jaroszewicz, "An axiomatization of partition entropy," *IEEE Transactions on Information Theory*, vol. 48, pp. 2138–2142, 2002.

[2] J. H. Havrda and F. Charvat, "Quantification methods of classification processes: Concepts of structural α-entropy," *Kybernetica*, vol. 3, pp. 30–35, 1967.

[3] M. Meilă, "Comparing clusterings: an axiomatic view," in *ICML*, pp. 577–584, 2005.

[4] D. A. Simovici and C. Djeraba, *Mathematical Tools for Data Mining.* London: Springer-Verlag, 2008.

[5] B. Mirkin, *Mathematical Classification and Clustering.* Boston: Kluwer, 1996.

[6] R. L. de Màntaras, "A distance-based attribute selection measure for decision tree induction," *Machine Learning*, vol. 6, pp. 81–92, 1991.

[7] G. Birkhoff, *Lattice Theory.* Providence: American Mathematical Society, 1973.

[8] J. Barthélemy and B. Leclerc, "The median procedure for partitions," in *Partitioning Data Sets*, (Providence), pp. 3–34, American Mathematical Society, 1995.

[9] J. Barthélemy, "Remarques sur les propriétés metriques des ensembles ordonnés," *Math. Sci. hum.*, vol. 61, pp. 39–60, 1978.

[10] B. Monjardet, "Metrics on parially ordered sets – a survey," *Discrete Mathematics*, vol. 35, pp. 173–184, 1981.

[11] B. Sayrafi, *A Measure-theoretic Framework for Constraints and Bounds on Measurements of Data*. PhD thesis, Indiana University, 2005.

[12] E. H. Lieb and M. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *Journal of Mathematical Physics*, vol. 14, pp. 1938–1941, 1973.

[13] D. A. Simovici, N. Singla, and M. Kuperberg, "Metric incremental clustering of nominal data," in *Proceedings of ICDM 2004*, (Brigton, UK), pp. 523–527, 2004.

[14] "Metric methods in data mining," in *Data Mining Patterns - New Methods and Applications* (F. M. P. Poncelet. M. Teisseire, ed.), pp. 1–31, Hershey, PA: Information Science Reference, 2007.

[15] D. Simovici and S. Jaroszewicz, "Generalized conditional entropy and decision trees," in *Extraction et Gestion des connaissances - EGC 2003*, (Paris), pp. 363–380, Lavoisier, 2003.

[16] D. A. Simovici and S. Jaroszewicz, "A new metric splitting criterion for decision trees," *Journal of Parallel, Emerging and Distributed Computing*, vol. 21, pp. 239–256, 2006.

[17] R. Butterworth, D. A. Simovici, and G. Piatetsky-Shapiro, "On feature extraction through clustering," in *Proceedings of ICDM 2005*, (Houston, Texas), pp. 581–584, 2005.

[18] R. Butterworth, D. A. Simovici, D. S. Santos, and L. Ohno-Machado, "A greedy algorithm for supervised discretization," *Journal of Biomedical Informatics*, vol. 37, pp. 287–292, 2004.

[19] D. Weisman and D. A. Simovici, "Several remarks on the metric space of genetic codes." to appear.