Neurocomputing 249 (2017) 86–94

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Exponential weighted entropy and exponential weighted mutual information

Shiwei Yu, Ting-Zhu Huang*

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, PR China

ARTICLE INFO

Article history:
Received 24 June 2015
Revised 26 February 2017
Accepted 30 March 2017
Available online 6 April 2017

Communicated by Shen Jianbing Shen

Keywords:
Entropy
Mutual information
Concavity
Image registration

ABSTRACT

In this paper, the exponential weighted entropy (EWE) and exponential weighted mutual information (EWMI) are proposed as the more generalized forms of Shannon entropy and mutual information (MI), respectively. They are position-related and causal systems that redefine the foundations of information-theoretic metrics. As the special forms of the weighted entropy and the weighted mutual information, EWE and EWMI have been proved that they preserve nonnegativity and concavity properties similar to Shannon frameworks. They can be adopted as the information measures in spatial interaction modeling. Paralleling with the normalized mutual information (NMI), the normalized exponential weighted mutual information (NEWMI) is also investigated. Image registration experiments demonstrate that EWMI and NEWMI algorithms can achieve higher aligned accuracy than MI and NMI algorithms.

1. Introduction

Since Shannon's original work in [1], entropy has been viewed as a kernel information measure of the uncertainty associated with a random variable. It has gained wide interest in signal processing, coding, data compression, and the other various fields. Mutual information (MI) is a quantity that measures the mutual dependence of the two variables. It is often recognized as an effective similarity measure in signal processing.

In this paper, we consider (X, Y) as the discrete random variable (r.v.) over a state space $\Omega\times\Omega$ with the joint probability distribution function (pdf) $p(i, j)$, marginal pdfs $p_X(i)$ and $p_Y(j)$. We also consider the conditional pdf $p_{X|Y}(i|j)$ of X given Y defined over Ω. Note that $p(i)$, $p(j)$, and $p(i|j)$ are also used to mean $p_X(i)$, $p_Y(j)$, and $p_{X|Y}(i|j)$, respectively.

Shannon entropy [1] of r.v. X was defined by

$$H(X) = -\sum_{i=1}^{\infty} p(i)\log p(i) \tag{1}$$

Based on Shannon entropy, MI and the normalized mutual information (NMI) were proposed as the similar measure in [1,2].

Shannon frameworks have the drawbacks of being position-free and memoryless [3]. They characterize the objective information of events occurrence. In other terms, they only consider the pdf of an event, regardless the distribution of r.v.s. As a consequence, a random variable X possesses the same Shannon entropy as $X + a$, for any $a \in R$. In some applications, besides the objective information, the subjective information or utility about a goal of events occurrence should also be taken into account. This leads to the propositions of the weighted entropy [3] and the weighted MI [4].

Under the axiomatic framework of entropy [5], the weighted entropy [3] of r.v. $X \sim p(i)$ was defined as

$$H^w(X) = -\sum_{i=1}^{\infty} i p(i)\log p(i) \tag{2}$$

There are different versions of the weighted MI and the normalized weighted MI. A typical version of the weighted MI of r.v.s $(X, Y) \sim p(i, j)$ was given by [4]

$$I^w(X,Y) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} w(i, j)p(i, j)\log\frac{p(i, j)}{p(i)p(j)} \tag{3}$$

Since the additivity hypothesis in thermodynamics, Shannon entropy neglects the correlations between the subsystems, whereas non-extensive processes are common at many physical levels in statistical mechanics and atomic physics [6]. There are two ways to overcome this intrinsic drawbacks. The one way is to extend the additivity to nonadditivity, such as Rényi entropy [7] and Tsallis entropy[6,8,9]. The other way is taking some prior statistical information into account [10]. In (2), the weight function i is too simple. In (3), a weight $w(i, j)$ is placed on the probability of each variable value co-occurrence $p(i, j)$, which leads it to be difficult to study their mathematical properties. Till now, there still lack of theoretic

* Corresponding author.
E-mail addresses: shiweiyu365@126.com (S. Yu), tingzhuhuang@126.com, tzhuang@uestc.edu.cn (T.-Z. Huang).

http://dx.doi.org/10.1016/j.neucom.2017.03.075
0925-2312/

research on the weighted entropy and the weighted MI. Since the exponential weighted method has been viewed as an efficient tool in engineering [11], in this paper, the exponential weighted entropy is proposed as the generalized form of the (weighted) entropy [1,3] and exponential weighted mutual information is proposed as the special form of the weighted mutual information introduced in [4]. They are the extensions of Shannon frameworks and generalize the corresponding concepts in [12] that defined in a generalized Euclidean metric space based on fractional calculus.

The rest of this paper is organized as follows. EWE, EWMI, and NEWMI are proposed in Section 2. The concavity properties are studied in Section 3. Section 4 provides applications in image registration. We conclude with a summary of content in Section 5.

2. Exponential weighted entropy and exponential weighted mutual information

Let $f(i)$ be bounded a function and r.v. $X \sim p(i), i = 1, 2, \dots,$ we use $\mathbb{W}_X^\alpha$ as the exponential weighted summation operator with order α on $p(i)$ to mean that

$$\mathbb{W}_X^\alpha(p(i)) = \sum_{i=1}^{\infty} f^{1-\alpha}(i)p(i) \quad (\alpha \in R) \tag{4}$$

It is worth noting that the exponential weighted summation operator $\mathbb{W}_X^\alpha$ in Eq. (4) works on the probability distribution function of r.v. $X \sim p(i), i = 1, 2, \dots,$ and not on a single $p(i)$.

We call $f(i)$ the weight function and $f^{1-\alpha}(i)$ the exponential weight function. To keep consistent with the formula in [12], we use $f^{1-\alpha}$ rather than f^α in this paper. For simply, $\mathbb{W}_X^\alpha$ is often written as $\mathbb{W}^\alpha$ if it is clear that the weight summation iterates through i with respect to r.v. X in context.

Definition 1. The exponential weighted entropy (EWE) of r.v. $X \sim p(i)$ with order α is defined by

$$H_w^\alpha(X) = -\mathbb{W}^\alpha p(i)\log p(i) \quad (\alpha \in R) \tag{5}$$

Generally, a positive real function can be adopted as the weight function in (5). We can normalize $f(i)$, and thus, $f(i)$ is always supposed $0 < f(i) \le 1$ throughout this paper. We also write $H_w^\alpha(p)$ or $H_w^\alpha(p(i))$ for Eq. 5. The log is to the base e.

Considering the property of pdf in probability space, the nonnegativity of EWE is easy to obtain, e.g., $H_w^\alpha(X) \ge 0$.

Definition 2. The joint exponential weighted entropy of r.v.s $(X, Y) \sim p(i, j)$ with order α is defined as

$$H_w^\alpha(X, Y) = -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(i, j) \tag{6}$$

Definition 3. The conditional exponential weighted entropy of r.v.s $(X, Y) \sim p(i, j)$ with order α is defined as

$$H_w^\alpha(Y|X) = -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(j|i) \tag{7}$$

Theorem 1. $H_w^\alpha(X, Y) = H_w^\alpha(X) + H_w^\alpha(Y|X)$.

Proof. We obtain

$$\begin{aligned} H_w^\alpha(Y|X) &= -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log \frac{p(i, j)}{p(i)} \\ &= -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(i, j) + \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(i) \\ &= H_w^\alpha(X, Y) - H_w^\alpha(X) \end{aligned}$$

□

Corollary 1. *$H_w^\alpha(X, Y) \ge H_w^\alpha(X)$ and $H_w^\alpha(X, Y) \ge H_w^\alpha(Y|X)$.*

Definition 4. The exponential weighted mutual information (EWMI) of $(X, Y) \sim p(i, j)$ with order α is defined by

$$I_w^\alpha(X, Y) = \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log \frac{p(i, j)}{p(i)p(j)} \tag{8}$$

Theorem 2. *$I_w^\alpha(X, Y) \ge 0$, with equality if and only if X and Y are independent.*

Proof. Using the Log Sum Inequality [13], we obtain

$$\begin{aligned} I_w^\alpha(X, Y) &= \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log \frac{p(i, j)}{p(i)p(j)} \\ &\ge \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j) \cdot \log \frac{\sum_j p(i, j)}{\sum_j p(i)p(j)} \\ &= \mathbb{W}^\alpha p(i)\log \frac{p(i)}{p(i)} \\ &= 0 \end{aligned} \tag{9}$$

The equality in (9) is true if and only if $p(i, j) = p(i)p(j)$ for all $i = 1, 2, \dots,$ which means X and Y are independent. □

Theorem 3. *For two r.v.s X and Y, we have*

$$I_w^\alpha(X, Y) = K_w^\alpha(Y) - H_w^\alpha(Y|X) \tag{10}$$

where $K_w^\alpha(Y) = -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(j)$.

Proof. We obtain

$$\begin{aligned} I_w^\alpha(X, Y) &= \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log \frac{p(i, j)}{p(i)p(j)} \\ &= \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log \frac{p(j|i)}{p(j)} \\ &= \mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\big(\log p(j|i) - \log p(j)\big) \\ &= K_w^\alpha(Y) - H_w^\alpha(Y|X) \end{aligned}$$

Especially, $I_w^\alpha(X, X) = K_w^\alpha(X) - H_w^\alpha(X|X) = H_w^\alpha(X)$. □

Corollary 2. $I_w^\alpha(X, Y) = H_w^\alpha(X) + K_w^\alpha(Y) - H_w^\alpha(X, Y)$.

NMI often acts as a robust and accurate similarity measure in signal processing [2]. Similarly, the normalized exponential weighted MI can be similarly defined.

Definition 5. The normalized exponential weighted mutual information (NEWMI) of r.v.s (X, Y) with order α is defined as

$$NI_w^\alpha(X, Y) = \frac{H_w^\alpha(X) + K_w^\alpha(Y)}{H_w^\alpha(X, Y)} \tag{11}$$

Theorem 4. $1 \le NI_w^\alpha(X, Y) \le 2$.

Proof. It is easy to verify that $K_w^\alpha(Y) \ge 0$. We obtain

$$K_w^\alpha(Y) \le -\mathbb{W}_X^\alpha \sum_{j=1}^{\infty} p(i, j)\log p(i, j) = H_w^\alpha(X, Y)$$

Since

$$\begin{aligned} H_w^\alpha(X, Y) &= H_w^\alpha(X) + H_w^\alpha(Y|X) \ge H_w^\alpha(X) \\ I_w^\alpha(X, Y) &= H_w^\alpha(X) + K_w^\alpha(Y) - H_w^\alpha(X, Y) \ge 0 \end{aligned}$$

we obtain $H_w^\alpha(X) \le H_w^\alpha(X, Y) \le H_w^\alpha(X) + K_w^\alpha(Y)$.

Similar as $1 \le NI(X, Y) \le 2$ [13], we have $1 \le NI_w^\alpha(X, Y) \le 2$. □

Corollary 3. $K_w^\alpha(Y) \ge H_w^\alpha(Y|X)$.