Tsallis Mutual Information for Document Classification

Preetham Reddy Armoor

Department of Mathematics
Hochschule,Mittweida

24 January, 2018

Outline

Entropy

Entropy: Measure the amount of uncertainty about the random variable
Given a random variable X with distribution $(x_1, x_2, .., x_n)$

Definition

Shannon entropy $H(X)$ of a random variable X is defined as

$$H(X) = -\sum_{x \in X} P(x) \log P(x)$$

Tsallis Entropy

Generalization of Shannon Entropy.

Definition

Tsallis-entropy is defined by

$$H_\alpha^T(X) = \frac{1 - \sum_{x\in X} P_x^\alpha}{\alpha - 1}$$

where the parameter α is called Entropic Index ,$\alpha > 0$ and $\neq 1$
where $\sum_{x\in X} P(x) = 1$

$H_\alpha^T(X)$ is concave function of p for $\alpha > 0$
$H_\alpha^T(X) = H(X)$ when $\alpha \to 1$

Given a random variable P with bernoulli distribution $(p, 1-p)$

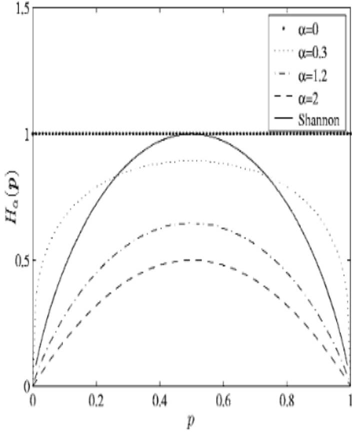

Figure: Tsallis Entropy of bernoulli distribution $(p, 1-p)$ for different α values

Tsallis Entropy

non-extensive (pseudoadditvity) property

Given two random variables X and Y with distribution $(x_1, x_2, .., x_n)$ and $(y_1, y_2, .., y_n)$
Shannon Joint Entropy:
if the variables are independent

$$H(X, Y) = H(X) + H(Y)$$

if the variables are correlated

$$H(X, Y) = H(X) + H(Y|X)$$

Tsallis Entropy

non-extensive (pseudoadditvity) property

Given two random variables X and Y with distribution $(x_1, x_2, .., x_n)$ and $(y_1, y_2, .., y_n)$
Shannon Joint Entropy:
if the variables are independent

$$H(X, Y) = H(X) + H(Y)$$

if the variables are correlated

$$H(X, Y) = H(X) + H(Y|X)$$

Tsallis Joint Entropy:

$$H_\alpha^T(X, Y) = H_\alpha^T(X) + H_\alpha^T(Y) + (1 - \alpha) H_\alpha^T(X) H_\alpha^T(Y)$$

Mutual Information

Mutual Information measures the information that X shares with Y

$$MI(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$MI(X;Y) = H(X) + H(Y) - H(X,Y)$$

Tsallis Mutual Entropy is generalization of Mutual Information
Tsallis Mutual Entropy is defined for $\alpha > 1$

$$\begin{aligned} ME_\alpha^T(X;Y) &= H_\alpha^T(X) - H_\alpha^T(X|Y) = H_\alpha^T(Y) - H_\alpha^T(Y|X) \\ &= H_\alpha^T(X) + H_\alpha^T(Y) - H_\alpha^T(X,Y) \end{aligned}$$

Tsallis Mutual Entropy is generalization of Mutual Information
Tsallis Mutual Entropy is defined for $\alpha > 1$

$$ME_\alpha^T(X;Y) = H_\alpha^T(X) - H_\alpha^T(X|Y) = H_\alpha^T(Y) - H_\alpha^T(Y|X)$$
$$= H_\alpha^T(X) + H_\alpha^T(Y) - H_\alpha^T(X,Y)$$

Upper bound of Mutual Entropy

$$ME_\alpha^T(X;Y) \leq H_\alpha^T(X,Y)$$

Measure is always positive and symmetric

Normalization of Mutual Entropy

$$NME_{\alpha}^{T}(X;Y) = \frac{ME_{\alpha}^{T}(X;Y)}{H_{\alpha}^{T}(X,Y)}$$

Normalization of Mutual Entropy:
takes values 0 if and only if X and Y are independent and $\alpha = 1$
takes values 1 if and only if X = Y

kullback-leibler distance or informational divergence

$$KL(p, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

the conventions that $0 \log \frac{0}{0} = 0$ and $a \log \frac{a}{0} = \infty$ if $a > 0$ are adopted
the kullback-leibler distance satisfies the information inequality
$KL(p, q) \geq 0$ with equality if and only if $p = q$

kullback-leibler distance or informational divergence

$$KL(p, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

the conventions that $0 \log \frac{0}{0} = 0$ and $a \log \frac{a}{0} = \infty$ if $a > 0$ are adopted
the kullback-leibler distance satisfies the information inequality
$KL(p, q) \geq 0$ with equality if and only if $p = q$
mutual information can be obtained from the kullback-leibler distance as follows

$$MI(X; Y) = KL(p(x, y), p(x)p(y))$$

Tsallis generalization of kullback-leibler distance:

$$KL_{\alpha}^{T}(p, q) = \frac{1}{1-\alpha}(1 - \sum_{x \in X} \frac{p(x)^{\alpha}}{q(x)^{\alpha-1}})$$

Tsallis generalization of kullback-leibler distance:

$$KL_{\alpha}^{T}(p,q) = \frac{1}{1-\alpha}(1 - \sum_{x \in X} \frac{p(x)^{\alpha}}{q(x)^{\alpha-1}})$$

$$MI_{\alpha}^{T}(X;Y) = KL_{\alpha}^{T}(p(x,y), p(x)p(y))$$

$$MI_{\alpha}^{T}(X;Y) = \frac{1}{1-\alpha}(1 - \sum_{x \in X} \sum_{y \in Y} \frac{p(x,y)^{\alpha}}{p(x)^{\alpha-1}p(y)^{\alpha-1}})$$

Normalization of Mutual Information:

$$NMI_{\alpha}^{T}(X;Y) = \frac{MI_{\alpha}^{T}(X;Y)}{H_{\alpha}^{T}(X,Y)}$$

$NMI_{\alpha}^{T}(X;Y)$ is a normalize measure for $\alpha \to 1$
Not true for other α values as $NMI_{\alpha}^{T} > 1$
measure is always positive and symmetric

from Jensen's inequality ,we obtain the jensen-shannon inequality

$$JS(\pi_1, ...\pi_n; p_1, ..., p_n) = H(\sum_{i=1}^{n} \pi_i p_i) - \sum_{i=1}^{n} \pi_i H(p_i) \geq 0$$

where $JS(\pi_1, ...\pi_n; p_1, ..., p_n)$ is the Jensen-shannon Divergence of probability distributions $p_1, ..., p_n$ with prior probability or weights $\pi_1, ...\pi_n$ fulfilling $\sum_{i=1}^{n} \pi_i = 1$

Mutual Information can also be expressed as Jensen-shannon Divergence

$$I(X; Y) = JS(p(x_1), ..., p(x_n); p(Y|x_1), ..., p(Y|x_n))$$

Jensen-Shannon divergence can be extended to define the Jensen-Tsallis divergence

$$JS_\alpha^T(\pi_1, \ldots\ldots\pi_n; p_1, \ldots\ldots, p_n) = H_\alpha^T(\sum_{i=1}^{n} \pi_i p_i) - \sum_{i=1}^{n} \pi_i H_\alpha^T(p_i)$$

$$JTI_\alpha^T(X \to Y) = JT_\alpha(p(x); p(Y|x))$$

$$JTI_\alpha^T(X \to Y) = H_\alpha^T(\sum_{x \in X} p(x)p(Y|x)) - \sum_{x \in X} p(x) H_\alpha^T(Y|x)$$

$$JTI_\alpha^T(X \to Y) = H_\alpha^T(Y) - \sum_{x \in X} p(x) H_\alpha^T(Y|x)$$

for reverse channel $Y \to X$,we have

$$JTI_{\alpha}^{T}(Y \to X) = H_{\alpha}^{T}(X) - \sum_{y \in Y} p(y) H_{\alpha}^{T}(X|y)$$

measure is positive and non-symmetric
Normalized of JTI_{α}^{T} can be defined as

$$NJTI_{\alpha}^{T}(X \to Y) = \frac{JTI_{\alpha}^{T}(X \to Y)}{H_{\alpha}^{T}(X, Y)}$$

measure will also take values in the interval$[0, 1]$

Metric

For N pixels in the overlap domain $\Omega_{A,B}$ of images A and B

- sum of squared difference

$$SSD = \frac{1}{N} \sum_{i \in \Omega_{A,B}} |A(i) - B(i)|^2$$

  A(i) and B(i) represents the intensity at a pixel i of the images A and B

- correlation coefficient

$$cc = \frac{\sum_{i \in \Omega_{A,B}} (A(i) - A^{'})(B(i) - B^{')}}{[\sum_{i \in \Omega_{A,B}} (A(i) - A^{'})^2 \sum_{i \in \Omega_{A,B}} (B(i) - B^{'})^2]^{\frac{1}{2}}}$$

Image Classification

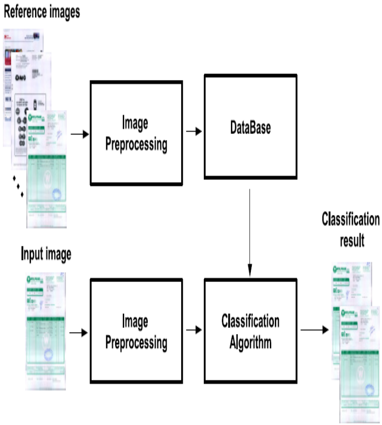

Figure: Document classification

Image Classification
Results

Objective: to calculate the degree of similarity between each input invoice and all reference invoice (a Similarity list is created)

Performance Measures

percentage of classification success: number of correctly classified input invoices over the total number of input

classification error: position of the corresponding reference invoice in the similarity list

Image Classification
Results

α values	MI^T		NMI^T		ME^T		NME^T		JTI^T		$NJTI^T$	
0.2	96.84	(1.67)	92.63	(1.29)					71.58	(6.26)	69.47	(9.00)
0.4	98.95	(2.00)	100.0	(0.00)					80.00	(2.58)	81.05	(3.39)
0.6	98.95	(1.00)	100.0	(0.00)					90.53	(1.67)	89.47	(1.30)
0.8	98.95	(1.00)	100.0	(0.00)					94.74	(1.40)	94.74	(1.00)
1.0	98.95	(2.00)	100.0	(0.00)	98.95	(2.00)	100.0	(0.00)	98.95	(2.00)	100.0	(0.00)
1.2	98.95	(2.00)	100.0	(0.00)	87.37	(2.58)	100.0	(0.00)	97.89	(1.00)	100.0	(0.00)
1.4	97.89	(1.50)	97.89	(1.00)	78.95	(6.40)	100.0	(0.00)	97.89	(1.50)	100.0	(0.00)
1.6	94.74	(1.40)	94.74	(1.00)	72.63	(8.27)	97.89	(1.50)	96.84	(1.33)	97.89	(1.00)
1.8	89.47	(2.10)	90.53	(1.56)	67.37	(9.65)	93.68	(2.50)	96.84	(1.33)	97.89	(1.00)
2.0	87.37	(2.33)	86.32	(2.15)	63.16	(10.66)	91.58	(4.86)	96.84	(1.33)	97.89	(1.00)
2.2	75.79	(2.13)	77.89	(2.10)	54.74	(10.23)	88.42	(6.64)	96.84	(1.33)	97.89	(1.00)
2.4	67.37	(2.61)	70.53	(2.50)	52.63	(10.78)	86.32	(8.31)	96.84	(1.33)	97.89	(1.00)
2.6	65.26	(3.06)	66.32	(2.97)	46.32	(10.55)	85.26	(9.50)	96.84	(1.33)	97.89	(1.00)
2.8	64.21	(3.50)	64.21	(3.38)	42.11	(10.60)	81.05	(8.67)	96.84	(1.33)	97.89	(1.00)
3.0	63.16	(3.80)	64.21	(3.79)	38.95	(10.93)	77.89	(8.67)	97.89	(1.50)	100.0	(0.00)

Figure: Table for the percentage of classification success and mean of classification error of the misclassified input invoices for different α values

Image Registration

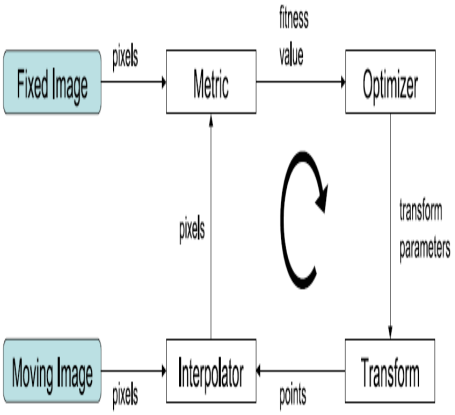

Figure: Main components of the registration process

Two different measures

Robustness

Accuracy

Robustness: evaluated in terms of the partial image overlap

AFA(area of function attraction) evaluates the range of convergence of a registration measure to its global maximum, counting the number of pixels from which the global maximum is reached by applying a maximum gradient method

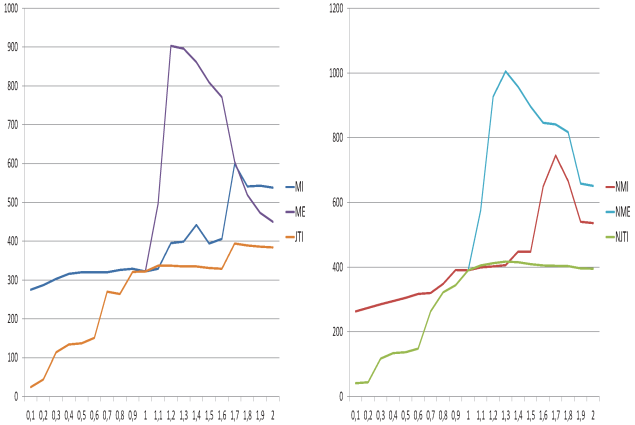

Figure: AFA parameter values with respect to the α value for the MI_T,ME_T,JTI_T measures(left) and the corresponding normalized measures(right)

Accuracy

powell's method optimizer,a rigid transform

for each image with original resolution 14 points manually identified and converted to the scaled space of a height 800 pixels

mean error: Average euclidean distance between these moved points and the corresponding points in the template

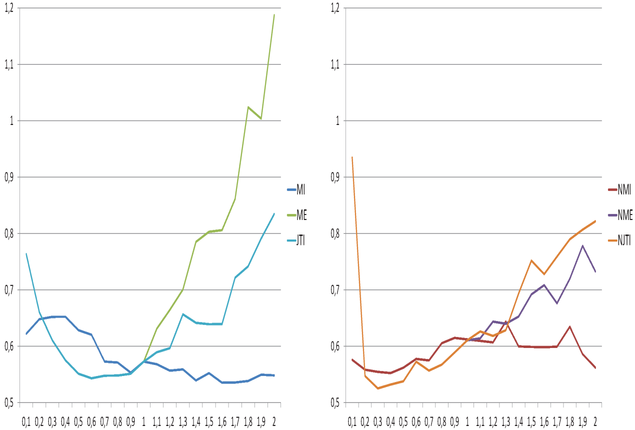

Figure: Mean error at the final registration position for different measure and α value for the MI_T,ME_T,JTI_T measures(left) and the corresponding normalized measures(right)

Conclusion

- for document classification the best results are obtained with normalized measure using α values between 0.4 and 1.2 for NMI_T and between 1 and 1.4 for NME_T and $NJTI_T$
- for document registration, the most robust results have been obtained by NME_T with $\alpha = 1.3$ and the most accurate ones have been achieved by $NJTI_T$ with$\alpha = 0.3$

Reference I

Marius Villa,Anton Bardera,Miquel Feixas and Mateu Sbert
Tsallis Mutual Information for Document Classification
Institut d'informatica i Applicacions,Universitat de Girona