

Article

A Novel Distance Metric: Generalized Relative Entropy

Shuai Liu ^{1,2,*}, Mengye Lu ^{1,2}, Gaocheng Liu ^{1,2} and Zheng Pan ^{1,2}

¹ College of Computer Science, Inner Mongolia University, Hohhot 010010, China;

lmy@mail.imu.edu.cn (M.L.); lgcxy@mail.imu.edu.cn (G.L.); cs_pz@imu.edu.cn (Z.P.)

² Inner Mongolia Key Laboratory of Data Mining and Knowledge Engineering, Hohhot 010010, China

* Correspondence: cs_liushuai@imu.edu.cn; Tel.: +86-0471-4993132

Received: 26 May 2017; Accepted: 7 June 2017; Published: 13 June 2017

Abstract: Information entropy and its extension, which are important generalizations of entropy, are currently applied to many research domains. In this paper, a novel generalized relative entropy is constructed to avoid some defects of traditional relative entropy. We present the structure of generalized relative entropy after the discussion of defects in relative entropy. Moreover, some properties of the provided generalized relative entropy are presented and proved. The provided generalized relative entropy is proved to have a finite range and is a finite distance metric. Finally, we predict nucleosome positioning of fly and yeast based on generalized relative entropy and relative entropy respectively. The experimental results show that the properties of generalized relative entropy are better than relative entropy.

Keywords: relative entropy; generalized relative entropy; upper bound; distance metric; adjusted distance

1. Background

The concept of entropy was proposed by T. Clausius as one of the parameters to reflect the degree of chaos for the object. Later, research found that information was such an abstract concept that was hard to make it clear to obtain its amount. Indeed, it was not until the information entropy was proposed by Shannon that we had a standard measure for the amount of information. Then, some related concepts based on information entropy have been proposed subsequently, such as cross entropy, relative entropy and mutual information, which offered an effective method to solve the complex problems of information processing. Therefore, the study of a novel metric based on information entropy was significant in the research domain of information science.

Information entropy was first proposed by Shannon. Assuming an information source I is composed by n different signals I_i , $H(I)$, the information entropy of I was shown in Equation (1), where $p_i = \frac{\text{amount of } I_i}{\text{signal's amount of } I}$ denotes frequency of I_i , $E()$ means mathematical expectation, $k > 1$ denotes the base of logarithm. When $k = 2$, the unit of $H(I)$ is bit.

$$H(I) = E(-\log_k p_i) = -\sum_{i=1}^n p_i \cdot \log_k p_i \quad (1)$$

Information entropy was a metric of the chaos degree for an information source. The bigger the information entropy was, the more chaotic the information source, and vice versa. Afterwards cross entropy was proposed based on information entropy, the definition was shown in Equation (2) where P

denotes “real” distribution of information source, and Q denotes “unreal” distribution of information source p_i denotes frequency of components of P and q_i denotes frequency of components of Q.

$$H(P, Q) = \sum_{i=1}^n p_i \cdot \log_k \frac{1}{q_i} \quad (2)$$

Cross entropy also can act as the reaction of the similarity degree of component’s distributions of the two information sources. $H(P, Q) = H(P)$ if and only if all components’ distributions were identical. A homologous metric was relative entropy, and this was also known as Kullback–Leibler divergence. Its definition was shown in Equations (3) and (4), where Equation (3) was the definition of relative entropy for the discrete random variables and Equation (4) was the definition of the continuous random variables.

$$D(P||Q) = \sum_{i=1}^n p_i \cdot \log_k \frac{p_i}{q_i} \quad (3)$$

$$D(P||Q) = \int p_x \cdot \log_k \frac{p_x}{q_x} dx \quad (4)$$

Relative entropy reflected the differences of the two information sources with different distributions, the bigger the relative entropy was, the bigger the differences in the information sources were, and vice versa. Subsequently, mutual information, another entropy-based metric was also proposed. This included two random variables X and Y. Mutual information was defined as relative entropy of $p(x)$ and $p(y)$ in $I(X; Y) = \int_{x \in X, y \in Y} p(x, y) \cdot \log_k \frac{p(x, y)}{p(x) \cdot p(y)}$. The value of mutual information was non-negative. If and only if X and Y were independent variables with each other, was the value of mutual information equal to zero.

Recently, there have been many extensions and applications of information entropy based on metric. However, all of these information entropy-based methods have some defects. The two most important defects are: (1) it is not a distance metric; (2) it does not have an upper bound.

So in this paper, Section 2 introduces related works with information entropy-based methods; Section 3 provides a novel generalized relative entropy and prove some properties of the provided entropy; Section 4 predicts nucleosome positioning based on generalized relative entropy and relative entropy respectively; Section 5 summarizes the whole content.

2. Related Work

For years, many scholars have studied the applications of various entropy. Earlier, Białynicki-Birula et al. deduced a new uncertain relationship in quantum mechanics based on information entropy [1]. Uhlmann et al. applied relative entropy in digital integration, and proved some properties of the interpolation theory [2]. Shore et al. deduced the principle of maximum entropy and minimum cross entropy [3]. Fraser et al. analyzed the coordinate of singular factors [4]. Pincus et al. analyzed the complex degree of the system by the entropy [5]. Afterwards, Hyvarinen et al. analyzed independent component and projection pursuit based on entropy [6].

In 2000, Petersen et al. analyzed the optimization problem for the system with constraint of the relative entropy [7]. Kwak et al. classified a sample based on mutual information between the input information and the variable category [8]. Later, Pluim et al. analyzed the image matching in medicine based on mutual information [9]. Arif et al. used the entropy to analyze the changes of the center of gravity between the old and young in order to find a method of improving the walking stability for the old [10]. Phillips et al. analyzed the distribution of species by maximum entropy model [11]. Krishnaveni et al. analyzed the electroencephalogram of humans by mutual information [12]. Afterwards, Wolf et al. researched area laws in quantum systems by using mutual information and correlations [13]. Baldwin et al. utilized a maximum entropy model to find some regularity about the selection of habitat of wild animal [14]. Verdu et al. combined the matching with relative entropy and analyzed the relationship between both of them [15].

In 2011, Batina et al. reviewed mutual information [16]. Audenaert studied the asymmetry of relative entropy [17]. Gong et al. made the best of the scale-invariant feature transform (SIFT) and mutual information to propose a method that can match the object precisely [18]. A novel coarse-to-fine scheme for automatic image registration is proposed Giagkiozis et al. proposed a new method that can take advantage of the knowledge of cross entropy to solve the problem of multi-object programming [19]. Tang and Mao researched information entropy-based metrics for measuring emergences in artificial societies [20].

In recent years, many scholars have studied entropy-based methods in recognition and classification. Soares and Knobbe studied entropy-based discretization methods for ranking data [21]. Li et al. proposed a method to solve the problem of molecular docking using information entropy and the ant colony genetic algorithm [22]. Ma et al. used information entropy to analyze the changes of substance in the processes of chemical changes [23]. Kö et al. researched operational meaning of min- and max-entropy [24]. In addition, Müller and Pastena studied a generalization of majorization based on Shannon entropy [25]. Zhang et al. proposed a feature selection algorithm for fuzzy rough sets on the basis of information entropy [26]. Guariglia et al. studied some fractal properties of entropy [27]. Ebrahimzadeh et al. proposed the concept of logical entropy based on entropy, and applied it to a quantum dynamical system [28]. Lopez-Garcia et al. proposed a method to make a prediction for a traffic jam in a short period of time; they combined the genetic algorithm with cross entropy [29]. Sutter et al. studied the monotonicity of cross entropy [30]. Opper provided an estimator for the relative entropy rate of path measures for stochastic differential equations [31]. Tang et al. studied an EEMD-based multi-scale fuzzy entropy approach for complexity analysis in clean energy markets [32].

3. Generalized Relative Entropy

3.1. Structure of Generalized Relative Entropy

Nowadays, relative entropy is becoming one of the most important dissimilarity measures between two multidimensional vectors. Let $X(x_1, \dots, x_s)$ and $Y(y_1, \dots, y_s)$ be two multidimensional vectors, which are constituted by s components with different counts. The count of i -th component is x_i in vector X and y_i in vector Y . Therefore, the relative entropy $RE(X, Y)$, which denotes relation from X to Y , is defined in Equation (5), where $p_x(i) \stackrel{\text{def}}{=} \frac{x_i}{\sum_{i=1}^s x_i}$ means the probability of x_i in X for each i . Herein, we define $p_x(i) \cdot \log \frac{p_x(i)}{p_y(i)} = 0$ when $p_x(i) = 0$ in order to avoid form of equation $0 \cdot \log 0$. In real application, $\varepsilon > 0$ is added in the denominator of $\log()$, which is a very small positive number to avoid form of equation $\log \infty$.

$$RE(X, Y) = \sum_{i=1}^s p_x(i) \cdot \log \frac{p_x(i)}{p_y(i)} \quad (5)$$

It is known that $RE(X, Y)$ is not a distance metric because usually $RE(X, Y) \neq RE(Y, X)$ when $X \neq Y$. However, relative entropy does not have a finite upper bound, which means it can not be easily used to measure difference between high dimensional vectors in real application. So in this paper, based on definition of relative entropy, we present a generalized relative entropy $d(X, Y)$ by Equation (6), where s denotes the number of components and $k \geq 1$ denotes the control parameter of function $d()$, $r = 0$ when $X = Y$; $r = 1$ when $X \neq Y$. We believe the generalized relative entropy has better properties than relative entropy. Moreover, it is a distance metric.

$$d(X, Y) = \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} + p_y(i) \cdot \log \frac{k \cdot p_y(i)}{p_x(i) + (k-1)p_y(i)} \right) + r \cdot \log \left(1 + \frac{1}{k-1} \right)^2 \quad (6)$$

3.2. Properties of Generalized Relative Entropy

Theorem 1 will be presented to prove that the generalized relative entropy $d()$ is a distance metric. However, Lemmas 1 and 2 and Inferences 1 and 2 are presented first.

Lemma 1. $\sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} \right)$ is constant nonnegative if $p_x(i) \geq 0, p_y(i) \geq 0, \sum_{i=1}^s p_x(i) = 1, k \geq 1$.

Proof. Because $p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} = -p_x(i) \cdot \log \frac{(k-1)p_x(i) + p_y(i)}{k \cdot p_x(i)}, \frac{(k-1)p_x(i) + p_y(i)}{k \cdot p_x(i)} \geq 0$, we have following Equation (7).

$$\begin{aligned} \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{(k-1) \cdot p_x(i) + p_y(i)}{k \cdot p_x(i)} \right) &\leq \log \sum_{i=1}^s \frac{(k-1) \cdot p_x(i) + p_y(i)}{k \cdot p_x(i)} \cdot p_x(i) \\ &= \log \frac{(k-1) \cdot \sum_{i=1}^s p_x(i) + \sum_{i=1}^s p_y(i)}{k} = \log 1 = 0 \end{aligned} \quad (7)$$

So $\sum_{i=1}^s p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} = -\sum_{i=1}^s p_x(i) \cdot \log \frac{(k-1)p_x(i) + p_y(i)}{k \cdot p_x(i)} \geq 0$
Lemma 1 is proved. \square

Then, with consideration of condition that sign “=” appeared, we have Inference 1.

Inference 1. $\sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} \right)$ is zero if and only if $p_x(i) = p_y(i)$ for all i where $p_x(i) \geq 0, p_y(i) \geq 0, \sum_{i=1}^s p_x(i) = \sum_{i=1}^s p_y(i) = 1$ and $k \geq 1$.

Then, we have Lemma 2 and Inference 2 based on Lemma 1 and Inference 1 to prove upper bound of $d(X, Y)$.

Lemma 2. $\sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} \right) \leq \log \frac{k}{k-1}$ if $p_x(i) \geq 0, p_y(i) \geq 0, k \geq 1, \sum_{i=1}^s p_x(i) = 1$.

Proof. We have Equation (8) to prove Lemma 2 based on Equation (9).

$$\sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} \right) \leq \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k}{k-1} \right) = \log \frac{k}{k-1} \quad (8)$$

$$\log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} \leq \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)} = \log \frac{k}{k-1} \quad (9)$$

Lemma 2 is proved. \square

Inference 2. Upper bound of $d(X, Y)$ is $4 \cdot \log \frac{k}{k-1}$ where $\sum_{i=1}^s p_x(i) = \sum_{i=1}^s p_y(i) = 1$ and $k \geq 1$.

Proof. We have Equation (10) to prove Inference 2 based on Lemma 2.

$$\begin{aligned} d(X, Y) &= \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i) + p_y(i)} + p_y(i) \cdot \log \frac{k \cdot p_y(i)}{p_x(i) + (k-1)p_y(i)} \right) + r \\ &\cdot \log \left(1 + \frac{1}{k-1} \right)^2 \leq \log \frac{k}{k-1} + \log \frac{k}{k-1} + 2r \cdot \log \frac{k}{k-1} \end{aligned} \quad (10)$$

Inference 2 is proved. \square

After that, Theorem 1 is presented to prove $d(X, Y)$ is a distance metric between two elements X and Y with same diversity s and length n .

Theorem 1. Function $d()$ is a distance metric of elemental set $E\{\}$ in space $S(E\{\}, d())$ where all elements in E have same diversity s .

Proof. Let X and Y be two elements in E , $p_x(i)$ and $p_y(i)$ denote the frequency of the i -th component in X or Y , $k > 1$ is a control parameter, s is the number of components in X and Y , $r = 0$ when $X = Y$ and $r = 1$ when $X \neq Y$, we have $d(X, Y) = \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_y(i)} + p_y(i) \cdot \log \frac{k \cdot p_y(i)}{p_x(i)+(k-1)p_y(i)} \right) + r \cdot \log \left(1 + \frac{1}{k-1} \right)^2$ from Equation (2). Then, we use following properties to prove Theorem 1.

Property 1. $d(X, Y) \geq 0$ for every X and Y , $d(X, Y) = 0$ if and only if $X = Y$.

First, we know $\sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_y(i)} \right)$ is nonnegative from Lemma 1. It implies $\sum_{i=1}^s \left(p_y(i) \cdot \log \frac{k \cdot p_y(i)}{(k-1)p_y(i)+p_x(i)} \right) \geq 0$. Then, we know $r \cdot \log \left(1 + \frac{1}{k-1} \right)^2 \geq 0$. So $d(X, Y) = \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_y(i)} \right) + \sum_{i=1}^s \left(p_y(i) \cdot \log \frac{k \cdot p_y(i)}{p_x(i)+(k-1)p_y(i)} \right) + r \cdot \log \left(1 + \frac{1}{k-1} \right)^2 \geq 0$, which means Property 1 is proved.

Property 2. $d(X, Y) = d(Y, X)$ for every X and Y .

It is known that the formation of $d(X, Y)$ is symmetrical to $d(Y, X)$ for every pair of vectors X and Y , which means Property 2 is proved.

Property 3. $D(X, Y) + d(Y, Z) \geq d(X, Z)$.

First, if there are at least two elements in $\{X, Y, Z\}$ that are equal, it is known that $d(X, Y) + d(Y, Z) \geq d(X, Z)$ because in the three functions $d()$, one value is zero and other two values are the same and nonnegative.

So, Equation (11) is used to describe $d(X, Y) + d(Y, Z) - d(X, Z)$ when $X \neq Y \neq Z$.

$$\begin{aligned}
 d(X, Y) + d(Y, Z) - d(X, Z) &= \sum_{i=1}^s \left(p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_y(i)} + p_y(i) \cdot \log \frac{k \cdot p_y(i)}{p_x(i)+(k-1)p_y(i)} + p_z(i) \cdot \log \frac{k \cdot p_z(i)}{(k-1)p_y(i)+p_z(i)} - p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_z(i)} - p_z(i) \cdot \log \frac{k \cdot p_z(i)}{p_x(i)+(k-1)p_z(i)} \right) + \log \left(1 + \frac{1}{k-1} \right)^2
 \end{aligned} \tag{11}$$

Then, we have Equation (12) to prove Property 3 based on Lemmas 1 and 2.

$$\begin{aligned}
 d(X, Y) + d(Y, Z) - d(X, Z) &\geq \log \left(1 + \frac{1}{k-1} \right)^2 - \sum_{i=1}^s p_x(i) \cdot \log \frac{k \cdot p_x(i)}{(k-1)p_x(i)+p_z(i)} - \sum_{i=1}^s p_z(i) \cdot \log \frac{k \cdot p_z(i)}{p_x(i)+(k-1)p_z(i)} \geq 2 \cdot \log \frac{k}{k-1} - 2 \cdot \log \frac{k}{k-1} \geq 0
 \end{aligned} \tag{12}$$

To summarize Properties 1–3, Theorem 1 is proved. \square

Then, we use Theorem 2 to provide the range of $d(X, Y)$ for all elements X and Y which were combined with s components.

Theorem 2. Range of $d(X, Y)$ is $\{0\} \cup [2 \cdot \log \frac{k}{k-1}, 4 \cdot \log \frac{k}{k-1}]$.

Proof.

Case (1) $X = Y$

When $X = Y$, we have $d(X, Y) = d(X, X) = 0$ by Inference 1.

Case (2) $X \neq Y$

When $X \neq Y$, we have $d(X, Y) \leq 4 \cdot \log \frac{k}{k-1}$ by Inference 2, and $d(X, Y) \geq 2 \cdot \log \frac{k}{k-1}$ by Lemma 1.

To summarize Cases 1 and 2, Theorem 2 is proved. \square

In this way, the generalized relative entropy is provided and some properties are proved.

4. Experiment

4.1. Model Predicting Nucleosome Positioning

In this paper, we undertake some experiments in order to prove that generalized relative entropy has better properties than relative entropy. We consider the following two species: fly and yeast, and we use those datasets to predict nucleosome positioning. The datasets of fly are downloaded from Supplementary data in [33], including 2900 core DNA sequences and 2850 linker DNA sequences of 147 bp, and the datasets of yeast are downloaded from Supporting Information S1 in [34], including 1880 core DNA sequences and 1740 linker DNA sequences of 150 bp.

The following describes the processes of the experiments. Firstly, we introduce the definition of k -nucleotide sequences combinations. This means the combination of four nucleotides (A or G or C or T). Thus, di-nucleotide sequences have 16 combinations such as AA or AT. Next, we introduce the processes of statistics taking fly datasets for an example. Firstly, we count the frequencies of di-nucleotide sequences in core DNA sequences and linker DNA sequences, respectively, which construct two real distributions $p_{1x}(i)$, $p_{2x}(i)$, where i represents the i -th di-nucleotide sequences. Secondly, we count the frequencies of all di-nucleotide sequences for each DNA sequence and we construct the unreal distribution $p_y(i)$. Thirdly, we count the relative entropy between each DNA sequence and the core sequences and the linker DNA sequences, respectively, which constructs two dimensions feature vectors $R(RE_1, RE_2)$. Then, we put them into the back propagation neural network (BP neural network) to train a classification model to predict nucleosome positioning and use 10-fold cross-validation to examine the quality of the model. Then, we count generalized relative entropy between each DNA sequence and the core sequences and the linker DNA sequences, respectively, which constructs two dimensions feature vectors $R(d_1, d_2)$. The k ranges from 1.1 to 5. Then, we put them into a BP neural network to train a classification model to predict nucleosome positioning and use 10-fold cross-validation to examine the quality of the model. Next, we predict nucleosome positioning of yeast datasets using the same methodology as for predicting nucleosome positioning of the fly datasets.

4.2. Evaluations of the Equality of Prediction

In this paper, four variables TP, FP, FN, TN are defined. TP represents the situation such that both the prediction and the fact are the core DNA sequences. FP presents the situation that the linker DNA sequences incorrectly predicted the core DNA sequences. FN represents the situation that the core DNA sequences incorrectly predicted the linker DNA sequences. TN represents the situation that both the prediction and fact are the linker DNA sequences. We define the following standard to examine the quality of the prediction of a model [35].

$$S_n = \frac{TP}{TP + FN}$$

$$S_p = \frac{TN}{TN + FP}$$

$$Acc = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}$$

where Sn represents sensitivity, Sp represents specificity, Acc represents accuracy, and Mcc represents Mathew correlation coefficient.

4.3. Results and Analysis

We use 10-fold cross-validation to examine the quality of the model for fly datasets and yeast datasets. From the following Table and Figures (Tables 1 and 2, Figures 1–4), we can come to the conclusion that the results obtained by generalized relative entropy are better than relative entropy. Besides, it is obvious that the values obtained by generalized relative entropy are higher than the values obtained by relative entropy when k equals 2, 3.1 and 4.1 (Table 1). Meanwhile, we can see that the values of Acc for yeast datasets are higher than fly datasets (Figure 1, Table 2), which illustrates that nucleosome positioning is more easily obtained in yeast than fly.

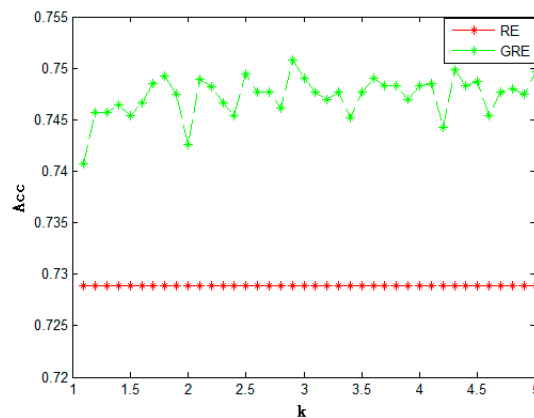


Figure 1. Accuracy of fly datasets (values marked in red are obtained by relative entropy, RE for short in Figures 1–4, values marked in green are obtained by generalized relative entropy, GRE for short in Figures 1–4).

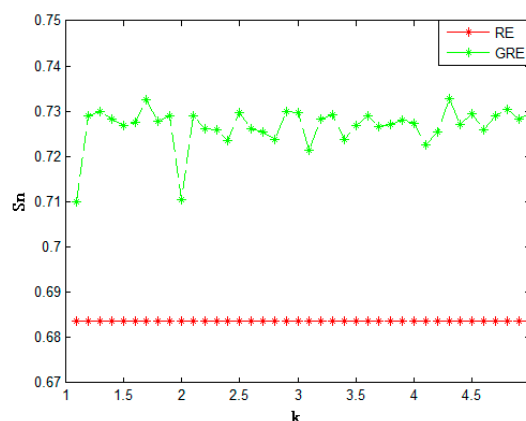


Figure 2. Sensitivity of fly datasets.

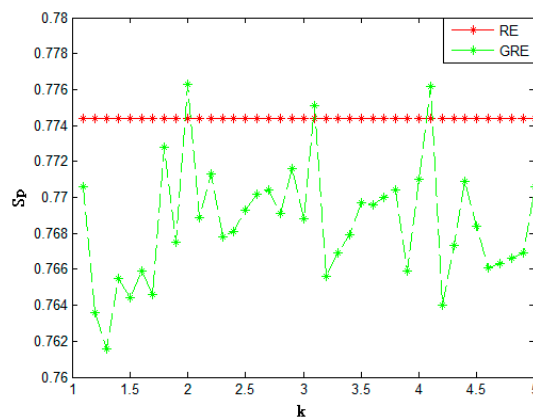


Figure 3. Specificity of fly datasets.

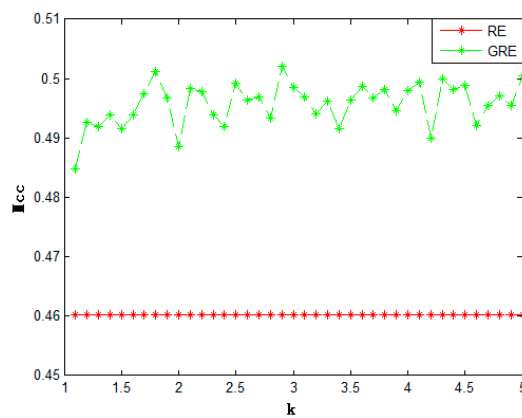


Figure 4. Mathew correlation coefficient of fly.

Table 1. The prediction results of fly datasets.

Method	Acc	Sn	Sp	Mcc
Relative entropy	0.7289	0.6837	0.7744	0.4603
Generalized relative entropy ($k = 2$)	0.7426	0.7105	0.7763	0.4885
relative entropy ($k = 3.1$)	0.7477	0.7215	0.7751	0.4970
entropy ($k = 4.1$)	0.7485	0.7225	0.7762	0.4994

Table 2. The prediction results of yeast datasets.

Method	Acc	Sn	Sp	Mcc
Relative entropy	0.9843	0.9875	0.9809	0.9684
Generalized relative entropy ($k = 2$)	0.9901	0.9937	0.9860	0.9801

5. Conclusions

In this paper, we provided a novel distance metric based on relative entropy, which was called generalized relative entropy. The generalized relative entropy surmounted the disadvantage of relative entropy because it had an upper bound and satisfies the triangle inequality of distance. The properties of the distance metric and upper bound were proved in this paper. Then, the range of the provided generalized relative entropy was computed, and k ranges from 1.1 to 5. In order to validate the advantages of generalized relative entropy, we predict nucleosome positioning of fly and yeast based on generalized relative entropy and relative entropy, respectively. The experimental results show that generalized relative entropy is better than relative entropy in nucleosome positioning. Finally, since

there was a parameter k to control the generalized relative entropy, we believe that this metric can be used in a variety of real applications by adjusted k .

Acknowledgments: The authors wish to thank the anonymous editors and reviewers for their helpful comments in this paper. This work is supported by Grants Programs of National Natural Science Foundation of China (No. 61502254), the Program for New Century Excellent Talents in University (NCET-12-1016), the program for Young Talents of Science and Technology in the Universities of Inner Mongolia Autonomous Region (NJYT-12-B04). We also want to thank Prof. Guo for his help to this paper.

Author Contributions: Shuai Liu conceived the method and finished some proofs; Mengye Lu, Gaocheng Liu and Zheng Pan performed and analyzed the experiments; Gaocheng Liu cleaned the data; Shuai Liu and Mengye Lu wrote the paper. All authors have read and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Biaynicki-Birula, I.; Mycielski, J. Uncertainty relations for information entropy in wave mechanics. *Commun. Math. Phys.* **1975**, *44*, 129–132. [[CrossRef](#)]
2. Uhlmann, A. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Commun. Math. Phys.* **1977**, *54*, 21–32. [[CrossRef](#)]
3. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37. [[CrossRef](#)]
4. Fraser, A.M.; Swinney, H.L. Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **1986**, *33*, 1134–1140. [[CrossRef](#)]
5. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
6. Hyvärinen, A. New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit. 1998. Available online: <https://papers.nips.cc/paper/1408-new-approximations-of-differential-entropy-for-independent-component-analysis-and-projection-pursuit.pdf> (accessed on 12 June 2017).
7. Petersen, I.R.; James, M.R.; Dupuis, P. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Trans. Autom. Control* **2000**, *45*, 398–412. [[CrossRef](#)]
8. Kwak, N.; Choi, C.-H. Input feature selection by mutual information based on Parzen window. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 1667–1671. [[CrossRef](#)]
9. Pluim, J.P.W.; Maintz, J.B.A.; Viergever, M.A. Mutual-information-based registration of medical images: A survey. *IEEE Trans. Med. Imaging* **2003**, *22*, 986–1004. [[CrossRef](#)] [[PubMed](#)]
10. Arif, M.; Ohtaki, Y.; Nagatomi, R.; Inooka, H. Estimation of the Effect of Cadence on Gait Stability in Young and Elderly People using Approximate Entropy Technique. *Meas. Sci. Rev.* **2004**, *4*, 29–40.
11. Phillips, S.J.; Anderson, R.P.; Schapire, R.E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259. [[CrossRef](#)]
12. Krishnaveni, V.; Jayaraman, S.; Ramadoss, K. Application of Mutual Information based Least dependent Component Analysis (MILCA) for Removal of Ocular Artifacts from Electroencephalogram. *Int. J. Biomed. Sci.* **2006**, *1*, 63–74.
13. Wolf, M.M.; Verstraete, F.; Hastings, M.B.; Cirac, J.I. Area laws in quantum systems: Mutual information and correlations. *Phys. Rev. Lett.* **2008**, *100*, 070502. [[CrossRef](#)] [[PubMed](#)]
14. Baldwin, R.A. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy* **2009**, *11*, 854–866. [[CrossRef](#)]
15. Verdu, S. Mismatched Estimation and Relative Entropy. *IEEE Trans. Inf. Theory* **2010**, *56*, 3712–3720. [[CrossRef](#)]
16. Batina, L.; Gierlichs, B.; Prouff, E.; Rivain, M.; Standaert, F.X.; Veyrat-Charvillon, N. Mutual Information Analysis: A Comprehensive Study. *J. Cryptol.* **2011**, *24*, 269–291. [[CrossRef](#)]
17. Audenaert, K.M.R. On the asymmetry of the relative entropy. *J. Math. Phys.* **2013**, *54*, 073506. [[CrossRef](#)]
18. Gong, M.; Zhao, S.; Jiao, L.; Tian, D.; Wang, S. A Novel Coarse-to-Fine Scheme for Automatic Image Registration Based on SIFT and Mutual Information. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4328–4338. [[CrossRef](#)]

19. Giagkiozis, I.; Purshouse, R.C.; Fleming, P.J. Generalized decomposition and cross entropy methods for many-objective optimization. *Inf. Sci.* **2014**, *282*, 363–387. [[CrossRef](#)]
20. Tang, M.; Mao, X. Information Entropy-Based Metrics for Measuring Emergences in Artificial Societies. *Entropy* **2014**, *16*, 4583–4602. [[CrossRef](#)]
21. De Sá, C.R.; Soares, C.; Knobbe, A. Entropy-based discretization methods for ranking data. *Inf. Sci.* **2015**, *329*, 921–936. [[CrossRef](#)]
22. Li, Z.; Gu, J.; Zhuang, H.; Kang, L.; Zhao, X.; Guo, Q. Adaptive molecular docking method based on information entropy genetic algorithm. *Appl. Soft Comput.* **2015**, *26*, 299–302. [[CrossRef](#)]
23. Ma, C.W.; Wei, H.L.; Wang, S.S.; Ma, Y.G.; Wada, R.; Zhang, Y.L. Isobaric yield ratio difference and Shannon information entropy. *Phys. Lett. B* **2015**, *742*, 19–22. [[CrossRef](#)]
24. König, R.; Renner, R.; Schaffner, C. The operational meaning of min- and max-entropy. *IEEE Trans. Inf. Theory* **2015**, *55*, 4337–4347. [[CrossRef](#)]
25. Müller, M.P.; Pastena, M. A Generalization of Majorization that Characterizes Shannon Entropy. *IEEE Trans. Inf. Theory* **2016**, *62*, 1711–1720. [[CrossRef](#)]
26. Zhang, X.; Mei, C.; Chen, D.; Li, J. Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy. *Pattern Recognit.* **2016**, *56*, 1–15. [[CrossRef](#)]
27. Guariglia, E. Entropy and Fractal Antennas. *Entropy* **2016**, *18*, 84. [[CrossRef](#)]
28. Ebrahimzadeh, A. Logical entropy of quantum dynamical systems. *Open Phys.* **2016**, *14*, 1–5. [[CrossRef](#)]
29. Lopez-Garcia, P.; Onieva, E.; Osaba, E.; Masegosa, A.D.; Perallos, A. A Hybrid Method for Short-Term Traffic Congestion Forecasting Using Genetic Algorithms and Cross Entropy. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 557–569. [[CrossRef](#)]
30. Sutter, D.; Tomamichel, M.; Harrow, A.W. Strengthened Monotonicity of Relative Entropy via Pinched Petz Recovery Map. *IEEE Trans. Inf. Theory* **2016**, *62*, 2907–2913. [[CrossRef](#)]
31. Opper, M. An estimator for the relative entropy rate of path measures for stochastic differential equations. *J. Comput. Phys.* **2017**, *330*, 127–133. [[CrossRef](#)]
32. Tang, L.; Lv, H.; Yu, L. An EEMD-based multi-scale fuzzy entropy approach for complexity analysis in clean energy markets. *Appl. Soft Comput.* **2017**, *56*, 124–133. [[CrossRef](#)]
33. Guo, S.-H.; Deng, E.-Z.; Xu, L.-Q.; Ding, H.; Lin, H.; Chen, W.; Chou, K.-C. iNuc-PseKNC: A sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. *Bioinformatics* **2014**, *30*, 1522–1529. [[CrossRef](#)] [[PubMed](#)]
34. Chen, W.; Feng, P.; Ding, H.; Lin, H.; Chou, K.-C. Using deformation energy to analyze nucleosome positioning in genomes. *Genomics* **2016**, *107*, 69–75. [[CrossRef](#)] [[PubMed](#)]
35. Awazu, A. Prediction of nucleosome positioning by the incorporation of frequencies and distributions of three different nucleotide segment lengths into a general pseudo k-tuple nucleotide composition. *Bioinformatics* **2017**, *33*, 42–48. [[CrossRef](#)] [[PubMed](#)]

