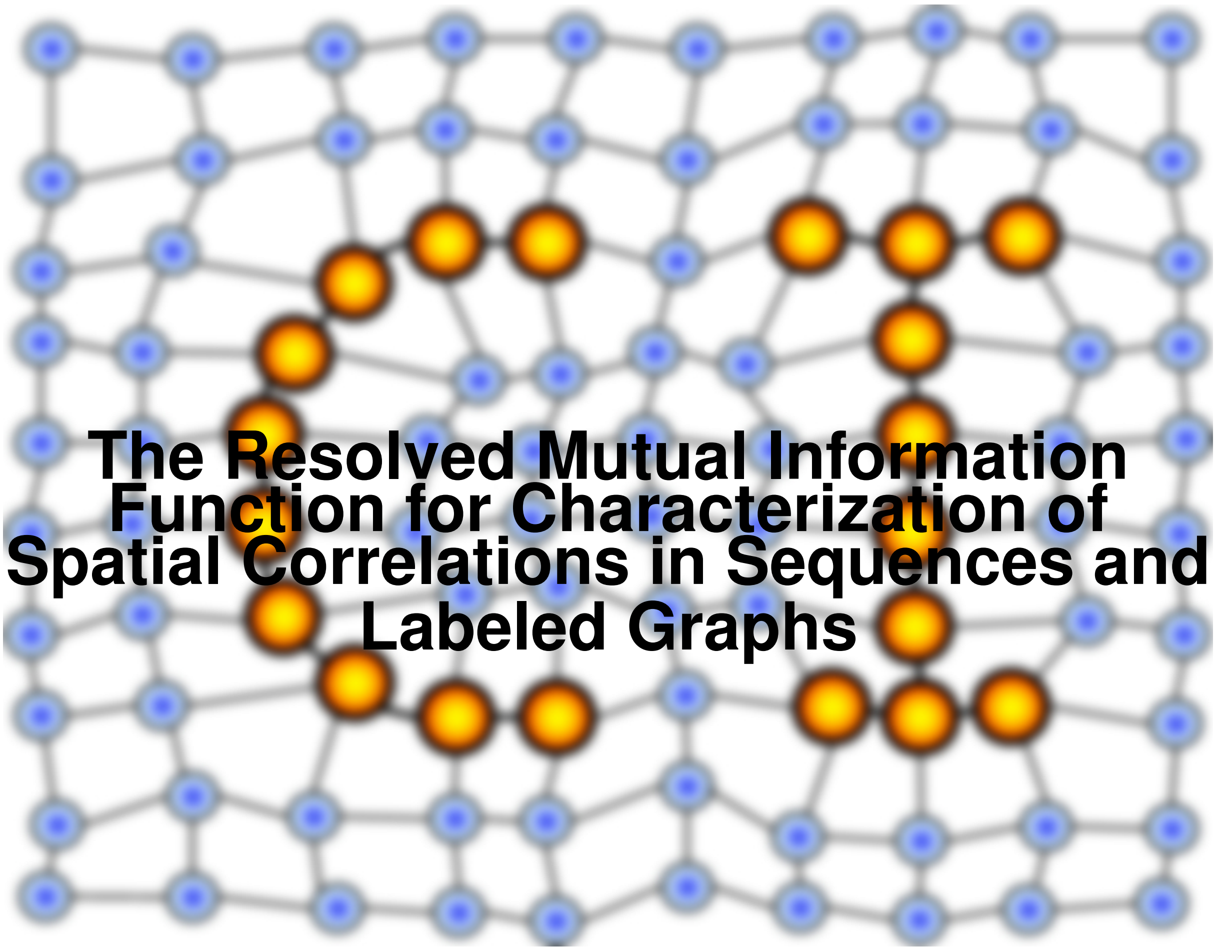
MACHINE LEARNING REPORTS



The Resolved Mutual Information Function for Characterization of Spatial Correlations in Sequences and Labeled Graphs

Report 02/2021
Submitted: 20.06.2021
Published: 23.06.2021

T. Villmann [1,*], K. S. Bohnsack [1], J. Abel [1], M. Kaden [1]
(1) Saxon Institute for Computational Intelligence and Machine Learning (SICIM)
Hochschule Mittweida
Mittweida, Germany

Impressum

Publisher:	University of Applied Sciences Mittweida Technikumplatz 17, 09648 Mittweida, Germany
Editor:	Prof. Dr. Thomas Villmann Dr. Frank-Michael Schleif
Techical-Editor:	Dr. Frank-Michael Schleif
Contact:	fschleif@techfak.uni-bielefeld.de
URL:	http://techfak.uni-bielefeld.de/∼ fschleif/mlr/mlr.html
ISSN:	1865-3960

The Resolved Mutual Information Function for Characterization of Spatial Correlations in Sequences and Labeled Graphs

T. Villmann[*], K. S. Bohnsack, J. Abel, and M. Kaden

Saxon Institute for Computational Intelligence and Machine Learning (SICIM),

University of Applied Sciences Mittweida, 09648 Mittweida, Technikumplatz 17, Germany

Abstract

In this technical report we collect and propose some useful quantities, notations, and approaches regarding information theoretic concepts, which might be helpful in machine learning when analyzing sequences and labeled graphs. In particular, we introduce resolved mutual information functions, which can be seen as a determining feature (profile) to characterize the spatial correlations in those data objects.

---

[*] corresponding author – email: villmann@hs-mittweida.de

1 Introduction

Information theoretic measures become more and more popular in applications far away from signal processing, information science and physics [33]. Among them, molecular-biological topics are one of the most promising areas as well as natural language processing [1, 45]. For example, sequence processing in biology and text analysis in linguistics are treated with very similar mathematical methods of information theory [29]. Those approaches comprise also methods which generate characteristic features of sequences or texts allowing a further processing by machine learning methods for clustering, classification and novelty detection, to name just a few application areas.

In this report, we consider some information theoretic quantities, which gained interest and generalize respective concepts to motivate further applications and the transfer to other application areas [6]. We will use these quantities to characterize spatial correlations and dependencies in sequences as well as in labeled graphs. Particularly, we will focus on quantities derived from the mutual information and related divergence measures [44]. Thereby, we will not restrict the approaches to the Shannon paradigm for the quantities but give also attention to computationally and theoretically interesting alternatives like Rényi divergences [35, 37]. Yet, we will start with the most prominent Shannon concepts.

2 Information Theoretic Quantities Related to the Shannon Concept of Information

2.1 Concepts, Basic Definitions and Mutual Information Functions

The Shannon entropy $H(X)$ of a a random quantity $X \subseteq \mathcal{X}$ with the density measure $p(x)$ is the expectation value

$$\mathbb{E}\left[\log\left(\frac{1}{p(x)}\right)\right] = \int_{\mathcal{X}} p(x) \cdot \log\left(\frac{1}{p(x)}\right) dx \tag{1}$$

of the information $\log\left(\frac{1}{p(x)}\right)$, such that we have

$$H(X) = \mathbb{E}\left[\log\left(\frac{1}{p(x)}\right)\right] \tag{2}$$

for this entropy. The maximum value of $H(X)$ is obtained for an uniform density $p(x)$ and, hence, $H(X)$ serves as a measure of uncertainty [29]. The non-symmetric Kullback-Leibler divergence of $p(x)$ to a random quantity $Y \subseteq \mathcal{Y} = \mathcal{X}$ with density $p(y)$ is given as

$$D_{KL}(X \parallel Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(x) \cdot \log\left(\frac{p(x)}{p(y)}\right) dydx \tag{3}$$

yielding a zero value for $p(x) \equiv p(y)$ [23]. Thus, the Kullback-Leibler divergence $D_{KL}(X \parallel Y)$ can be taken as a dissimilarity measure between $p(x)$ and $p(y)$ [31]. The mutual information

$$I(X,Y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} p(x,y) \cdot \log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right) dydx \tag{4}$$

quantifies the joint information of $p(x)$ and $p(y)$ with $p(x,y)$ being the joint density [23, 37]. Hence, the mutual information is the Kullback-Leibler divergence $D_{KL}(p(x,y) \parallel p(x) \cdot p(y))$ between the joint probability and the product of the marginal probabilities. Equivalently, we can write the mutual information as

$$I(X,Y) = \int_{\mathcal{X}} F(x,Y)\, dx$$

with

$$F\left(x,Y\right)=\int_{\mathcal{Y}}p\left(x,y\right)\cdot\log\left(\frac{p\left(x,y\right)}{p\left(x\right)\cdot p\left(y\right)}\right)dy \tag{5}$$

describing a *mutual information relation* of a particular object (value) x of the random quantity X with respect to the random quantity Y. We denote $F\left(x,Y\right)$ as the (feature) *resolved mutual information* (rMI).

The mutual information $I\left(X,Y\right)$ can be written in terms of the Shannon entropies as

$$\begin{aligned}I\left(X,Y\right) &= H\left(X\right)+H\left(Y\right)-H\left(X,Y\right) \\ &= H\left(X\right)-H\left(X|Y\right)\end{aligned} \tag{6}$$

with the conditional entropy

$$\begin{aligned}H\left(X|Y\right) &= H\left(X,Y\right)-H\left(Y\right) \\ &= -\int_{\mathcal{X}}\int_{\mathcal{Y}}p\left(x,y\right)\cdot\log\left(\frac{p\left(x,y\right)}{p\left(y\right)}\right)dydx\,.\end{aligned} \tag{7}$$

Combining (6) and (7), the mutual information $I\left(X,Y\right)$ is obtained to be

$$\begin{aligned}I\left(X,Y\right) &= -\int_{\mathcal{X}}p\left(x\right)\cdot\log\left(p\left(x\right)\right)dx+\int_{\mathcal{X}}\int_{\mathcal{Y}}p\left(x,y\right)\cdot\log\left(\frac{p\left(x,y\right)}{p\left(y\right)}\right)dydx \\ &= -\int_{\mathcal{X}}\left(\int_{\mathcal{Y}}p\left(x,y\right)\cdot\log\left(\frac{p\left(x,y\right)}{p\left(y\right)}\right)dy-p\left(x\right)\cdot\log\left(p\left(x\right)\right)\right)dx\end{aligned}$$

as an alternative formulation. From this we can conclude that for the rMI in eq. (5) the relation

$$F\left(x,Y\right)=-p\left(x\right)\cdot\log\left(p\left(x\right)\right)+\int_{\mathcal{Y}}p\left(x,y\right)\cdot\log\left(\frac{p\left(x,y\right)}{p\left(y\right)}\right)dy$$

holds.

The *cross mutual information* for sequences $X\left(t\right)$ and $Y\left(t+\tau\right)$ at time t with shift $\tau\geq0$ is defined as

$$I\left(X\left(t\right),Y\left(t+\tau\right)\right)=\int_{\mathcal{X}}\int_{\mathcal{Y}}p\left(x\left(t\right),y\left(t+\tau\right)\right)\cdot\log\left(\frac{p\left(x\left(t\right),y\left(t+\tau\right)\right)}{p\left(x\left(t\right)\right)\cdot p\left(y\left(t+\tau\right)\right)}\right)dydx \tag{8}$$

which yields by setting $Y\left(t+\tau\right)=X\left(t+\tau\right)$

$$I\left(X\left(t\right),X\left(t+\tau\right)\right)=\int_{\mathcal{X}}\int_{\mathcal{X}}p\left(x\left(t\right),x\left(t+\tau\right)\right)\cdot\log\left(\frac{p\left(x\left(t\right),x\left(t+\tau\right)\right)}{p\left(x\left(t\right)\right)\cdot p\left(x\left(t+\tau\right)\right)}\right)dx\left(t+\tau\right)dx$$

as the *auto mutual information* at time t with shift (delay) τ [22, 12]. If $p\left(x\left(t\right)\right)$ is independent from t, only the joint probability $p\left(x\left(t\right),x\left(t+\tau\right)\right)$ remains t-dependent or, more precisely, it becomes dependent only on the shift τ such that we simply write $p\left(x,x\left(\tau\right)\right)$ for this. Thus, the auto mutual information in dependence on the shift τ is obtained as

$$I\left(X,\tau\right)=\int_{\mathcal{X}}\int_{\mathcal{X}}p\left(x,x\left(\tau\right)\right)\cdot\log\left(\frac{p\left(x,x\left(\tau\right)\right)}{p\left(x\right)\cdot p\left(x\left(\tau\right)\right)}\right)dx\left(\tau\right)dx \tag{9}$$

as an information theoretic analagon to the auto-correlation function. In [25, 26] this shift-dependent auto mutual information is denoted as the *mutual information function* (MIF). Adapting the rMI from eq. (5) to the auto mutual information $I\left(X,\tau\right)$ we result the function

$$\begin{aligned} F\left(x,\tau\right) &= \int_{\mathcal{X}} p\left(x,x\left(\tau\right)\right)\cdot\log\left(\frac{p\left(x,x\left(\tau\right)\right)}{p\left(x\right)\cdot p\left(x\left(\tau\right)\right)}\right)dx\left(\tau\right) \qquad (10)\\ &= \int_{\mathcal{X}} p\left(x,x\left(\tau\right)\right)\cdot\log\left(\frac{p\left(x,x\left(\tau\right)\right)}{p\left(x\left(\tau\right)\right)}\right)dx\left(\tau\right)-p\left(x\right)\cdot\log\left(p\left(x\right)\right)\end{aligned}$$

which can be seen as a quantity characterizing the inherent correlations of the sequence values $x\left(t\right)$. We denote $F\left(x,\tau\right)$ as the (feature) *resolved mutual information function* (rMIF), which trivially fulfills

$$I\left(X,\tau\right)=\int_{\mathcal{X}} F\left(x,\tau\right)dx \qquad (11)$$

according to its definition. For (finite) discrete distributions it becomes simply a matrix $\mathbf{F}$. Hence, we can compare those matrices by an arbitrary matrix norm, e.g. by the Frobenius norm (metric) [14, 18].

2.2 Applications of Mutual Information Functions for Sequence Analysis in Bioinformatics

The consideration of organization processes, structuring and self-organization as well as information transfer belong to the key aspects of living systems [36, 9, 15, 16]. Thus information theoretic concepts play an important role in sequence analysis for understanding and analyzing RNA- and DNA- as well as protein sequences to explain biological systems [1].

Long-range correlations in sequences are well-known and intensively studied also in alignment free sequence comparison [28, 27, 32, 46]. The use of MIF as alternative to correlations was firstly investigated in [26] and this idea was renewed in [3] and [40]. It is followed by several studies using MIF for different topics like *in silico* comparison of bacterial strains using mutual information [41]. In [2], the MIF was renamed to be the *average mutual information profile* (AMI-profile) and proposed to serve as a genomic signature. This notation was taken over in the review [45] about information theoretic methods for sequence analysis. Following this notation, we denote the feature resolved mutual information function (rMIF) $F\left(x,\tau\right)$ from eq. (10) in this context also as the (feature) *resolved average mutual information profile* (rAMI-profile).

3 Rényi Information Concepts and α-scaled Mutual Information Functions

The Rényi-entropy

$$H_{\alpha}\left(X\right)=\frac{1}{1-\alpha}\log\left(\int_{\mathcal{X}}\left(p\left(x\right)\right)^{\alpha}dx\right)$$

is a generalization of the Shannon-entropy, where $\alpha>0$ and $\alpha\neq1$ is a parameter [34]. Depending on the context it is also denoted as α-entropy. In the limit $\alpha\to1$, the Shannon entropy is obtained. The Rényi-mutual-information (RMI) or α-mutual information (α-MI) is defined as

$$I_{\alpha}\left(X,Y\right)=\frac{1}{\alpha-1}\log\left(\int_{\mathcal{X}}\int_{\mathcal{Y}}\frac{\left(p\left(x,y\right)\right)^{\alpha}}{\left(p\left(x\right)\right)^{\alpha-1}\cdot\left(p\left(y\right)\right)^{\alpha-1}}dydx\right) \qquad (12)$$

as outlined in [35]. This mutual information is widely applied in data analysis and pattern recognition as well as in information theoretic machine learning [10, 17, 21, 24, 33, 38, 44, 30, 4]. In analogy to the resolved mutual information rMI $F(x,Y)$ from eq. (10), we denote

$$F_{\alpha}(x,Y)=\int_{\mathcal{Y}}\frac{(p(x,y))^{\alpha}}{(p(x))^{\alpha-1}\cdot(p(y))^{\alpha-1}}dy \tag{13}$$

as the *α-scaled resolved mutual information* (α-rMI) or as *α-scaled object dependent average mutual information profile* (α-AMI-profile) in bioinformatics context.

The Rényi variant of the *cross mutual information* for sequences $X(t)$ and $Y(t+\tau)$ at time t with shift $\tau\geq 0$ is defined as

$$I_{\alpha}(X(t),Y(t+\tau))=\frac{1}{\alpha-1}\log\left(\int_{\mathcal{X}}\int_{\mathcal{Y}}\frac{(p(x(t),y(t+\tau)))^{\alpha}}{(p(x(t)))^{\alpha-1}\cdot(p(y(t+\tau)))^{\alpha-1}}dy(t+\tau)\,dx(t)\right) \tag{14}$$

which gives by setting $Y(t+\tau)=X(t+\tau)$

$$I_{\alpha}(X(t),X(t+\tau))=\frac{1}{\alpha-1}\log\left(\int_{\mathcal{X}}\int_{\mathcal{X}}\frac{(p(x(t),x(t+\tau)))^{\alpha}}{(p(x(t)))^{\alpha-1}\cdot(p(x(t+\tau)))^{\alpha-1}}dx(t+\tau)\,dx(t)\right)$$

as the Rényi variant of the *auto mutual information* at time t with shift (delay) τ. Again, if $p(x(t))$ is independent from t, only the joint probability $p(x(t),x(t+\tau))$ remains t-dependent such that it becomes dependent only on the shift τ and we simply write $p(x,x(\tau))$ for this. Hence, the Rényi auto mutual information in dependence on the shift τ is obtained as

$$I_{\alpha}(X,\tau)=\frac{1}{\alpha-1}\log\left(\int_{\mathcal{X}}\int_{\mathcal{X}}\frac{(p(x,x(\tau)))^{\alpha}}{(p(x))^{\alpha-1}\cdot(p(x(\tau)))^{\alpha-1}}dx(\tau)\,dx\right)$$

and denoted as Rényi variant of, or *α-scaled mutual information function* (α-MIF). Accordingly, the α-scaled resolved version of α-MIF $I_{\alpha}(X,\tau)$ is

$$F_{\alpha}(x,\tau,X)=\int_{\mathcal{X}}\frac{(p(x,x(\tau)))^{\alpha}}{(p(x))^{\alpha-1}\cdot(p(x(\tau)))^{\alpha-1}}dx(\tau) \tag{15}$$

describing again the inherent correlations of the sequence and, hence, can serve as a characterizing quantity. We denote this function as the *α-scaled resolved mutual information function* (α-rMIF). In bioinformatics context it is denoted as *α-scaled object dependent average mutual information profile* (α-rAMI-profile). We immediately have

$$I_{\alpha}(X,\tau)=\frac{1}{\alpha-1}\log\left(\int_{\mathcal{X}}F_{\alpha}(x,\tau)\,dx\right)$$

in complete analogy to eq. (11).

In contrast to eq. (6) being valid for the (Shannon) mutual information, now the inequality

$$I_{\alpha}(X,Y)\neq H_{\alpha}(X)+H_{\alpha}(Y)-H_{\alpha}(X,Y) \tag{16}$$

holds for RMI. Thus, the RMI is not based on the Rényi-entropy in a straightforward way as the Shannon mutual information is based on the Shannon entropy [17]. This problem arises from the difficulty to define a conditional Rényi entropy to be consistent with the setting in the Shannon case [8, 11, 42]. According to [20, 19] we can distinguish at least: [5, 43]

- Jizba-Arimitsu conditional Rényi-entropy

$$H_\alpha^{\mathrm{JA}}(X|Y) = H_\alpha(X,Y) - H_\alpha(Y) \tag{17}$$

- Arimoto conditional Rényi-entropy

$$H_\alpha^{\mathrm{A}}(X|Y) = \frac{\alpha}{1-\alpha} \log\left(\int_{\mathcal{X}} p(x) \left(\int_{\mathcal{Y}} (p(x|y))^\alpha \, dy \right)^{1/\alpha} dx \right) \tag{18}$$

- Hayashi conditional Rényi-entropy

$$H_\alpha^{\mathrm{H}}(X|Y) = \frac{1}{1-\alpha} \log\left(\int_{\mathcal{X}} p(x) \left(\int_{\mathcal{Y}} (p(x|y))^\alpha \, dy \right) dx \right) \tag{19}$$

It can be shown that $H_\alpha^{\mathrm{H}}(X|Y) \leq H_\alpha^{\mathrm{A}}(X|Y)$ is valid using the Jensen's inequality [20]. Obviously, $H_\alpha^{\mathrm{JA}}(X|Y)$ can be interpreted as an extension of the conditional Shannon entropy $H(X|Y)$ because the definition (17) precisely coincides with Shannon chain rule (7).

The non-negative Rényi-α-divergence is defined as

$$D_\alpha(X \parallel Y) = \frac{1}{\alpha-1} \log\left(\int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{(p(x))^\alpha}{(p(y))^{\alpha-1}} dy dx \right) \tag{20}$$

with the limit $\lim_{\alpha\to1} D_\alpha(X \parallel Y) = D_{KL}(X \parallel Y)$ being valid.

4 Resolved Mutual Information Functions for Characterization of Spatial Correlations between Nodes in Labeled Graphs

The above concepts for the (resolved) mutual information function for sequences can be easily transferred to labeled graphs. We assume a set $\mathcal{G}$ of graphs G_k, each of which consists of N_k nodes n_j with node labels $l_j \in \mathcal{L}$ determining the class/type of the node regarding some given classification scheme $\mathcal{L}$. Further, let $\mathbf{D}^{(k)} \in \mathbb{R}^{N_k \times N_k}$ be the matrix of the, possibly directed and weighted, minimum graph distances (shortest path) between the node pairs of the graph G_k and $d_{\max} = \max_{i,j,k}\left(D_{i,j}^{(k)}\right)$ the maximum node distance for all graphs to be considered. For undirected graphs, $\mathbf{D}^{(k)}$ is symmetric.

Let, $\mathcal{I}_{T,R} = [0,T]$ be an interval with a partition $P_T = \{\tau_0 = 0, \tau_1, \ldots, \tau_R = T\}$ such that $\tau_i < \tau_j$ holds for $i < j$. The *dominating distance* $\tau_i(z)$ for a given value z is defined by the inequalities $\tau_{i-1} < z \leq \tau_i$ to be valid. Let $p(l)$ be the probability of the label $l \in \mathcal{L}$ in a graph G and $p(l, l')$ the respective joint probabilities.

Using these definitions we can determine in analogy to (9) the Shannon mutual information function (MIF) for graph as

$$I(G_k, \tau_r) = \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \delta\left(\tau_r, \tau_i\left(D_{i,j}^{(k)}\right)\right) \cdot p(l_i, l_j) \cdot \log\left(\frac{p(l_i, l_j)}{p(l_i) \cdot p(l_j)} \right) \tag{21}$$

where

$$\delta(z, z') = \begin{cases} 1 & \text{for } z = z' \\ \\ 0 & \text{for } z \neq z' \end{cases}$$

is the Kronecker-symbol and $\tau_r \in P_T$. Accordingly, the resolved MIF (rMIF) for graphs reads as

$$F\left(l_i, \tau, G_k\right) = \sum_{j=1}^{N_k} \delta\left(\tau_r, \tau_i\left(D_{i,j}^{(k)}\right)\right) \cdot p\left(l_i, l_j\right) \cdot \log\left(\frac{p\left(l_i, l_j\right)}{p\left(l_i\right) \cdot p\left(l_j\right)}\right) \tag{22}$$

with $l_i \in \mathcal{L}$.

Applying the Rényi-variants α-MIF and α-rMIF to graphs, we obtain

$$I_\alpha\left(G_k, \tau\right) = \frac{1}{\alpha - 1} \log\left(\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} \delta\left(\tau_r, \tau_i\left(D_{i,j}^{(k)}\right)\right) \cdot p\left(l_i, l_j\right) \cdot \frac{\left(p\left(l_i, l_j\right)\right)^\alpha}{p\left(l_i\right)^{\alpha-1} \cdot p\left(l_j\right)^{\alpha-1}}\right)$$

and

$$F_\alpha\left(l_i, \tau, G_k\right) = \sum_{j=1}^{N_k} \delta\left(\tau_r, \tau_i\left(D_{i,j}^{(k)}\right)\right) \cdot p\left(l_i, l_j\right) \cdot \frac{\left(p\left(l_i, l_j\right)\right)^\alpha}{p\left(l_i\right)^{\alpha-1} \cdot p\left(l_j\right)^{\alpha-1}}$$

respectively.

5 Conclusion

In this technical report we provide information theoretic concepts and quantities to characterize spatial correlations in sequences and labeled graphs. In particular, we introduced several types of mutual information functions for both, Shannon and Rényi information theoretic approaches. These functions/quantities can be used to compare those objects (sequences/graphs) in machine learning approaches regarding their inherent spatial correlations. Respective applications we see for sequence analysis in bioinformatics and text analysis as well as for social relation analysis by means of graphs as well as for other relational graphs, e.g. chemical structural formulae.

In the future work we will extend these approaches to further mutual information concept related to other widely considered entropy measures and information theoretic quantities, e.g. Tsallis-entropies and divergences or more general α-, β- and γ-divergences with related mutual informatin concepts .

Acknowledgement

K.S.B., J.A., and M.K. acknowldege funding from the European Social Fund (ESF) given to the junior researche group MaleKITA - Machinelles Lernen und KI in Theorie und Anwendung.

References

[1] P. Baldi and S. Brunak. *Bioinformatics.* Bradford Books, 2nd edition, 2001.

[2] M. Bauer, S. Schuster, and K. Sayood. The average mutual information profile as a genomic signature. *BMC Bioinformatics*, 9(48):1–11, 2008.

[3] M. Berryman, A. Allison, and D. Abbott. Mutual information for examining correlataions in DNA. *Fluctuation and Noise Letters*, 4(2):237–246, 2004.

[4] K. Bunte, S. Haase, M. Biehl, and T. Villmann. Stochastic neighbor embedding (SNE) for dimension reduction and visualization using arbitrary divergences. *Neurocomputing*, 90(9):23–45, 2012.

[5] C. Cai and S. Verdú. Conditional Rényi divergence saddlepoint and the maximization of α-mutual information. *entropy*, 21(969):1–25, 2020.

[6] A. Cichocki and S. Amari. Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12:1532–1568, 2010.

[7] I. Csiszár. Information-type measures of differences of probability distributions and indirect observations. *Studia Sci. Math. Hungaria*, 2:299–318, 1967.

[8] I. Csiszár. Axiomatic characterization of information measures. *Entropy*, 10:261–273, 2008.

[9] M. Eigen and P. Schuster. Stages of emerging life —five principles of early organization. *Journal of Molecular Evolution*, 19(1):47–61, Dec 1982.

[10] D. Erdogmus, J. Principe, and K. H. II. Beyond second-order statistics for learning: A pairwise interaction model for entropy estimation. *Natural Computing*, 1(1):85–108, 2002.

[11] S. Fehr and S. Berens. On the conditional Rényi entropy. *IEEE Transactions on Information Theory*, 60(11):6801–6810, 2014.

[12] A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2):1134–1140, 1986.

[13] S. Furuichi. Information theoretical properties of Tsallis entropies. *Journal of Mathematical Physics*, 2006.

[14] G. Golub and C. V. Loan. *Matrix Computations.* Johns Hopkins Studies in the Mathematical Sciences. John Hopkins University Press, Baltimore, Maryland, 4th edition, 2013.

[15] H. Haken. *Synergetics – An Introduction Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology.* Springer Berlin Heidelberg, Heidelberg, 1983.

[16] H. Haken. *Information and Self-Organization.* Springer Berlin Heidelberg, Heidelberg, 1988.

[17] K. Hild, D. Erdogmus, and J. Principe. Blind source separation using Rényi's mutual information. *IEEE Signal Processing Letters*, 8(6):174–176, 2001.

[18] R. Horn and C. Johnson. *Matrix Analysis.* Cambridge University Press, 2nd edition, 2013.

[19] V. Ilić, I. Djordjević, and M. Stanković. On a general definition of conditional Rényi entropies. In *Proceedings of the 4th International Electronic Conference on Entropy and Its Application (ECEA 2017)*, volume 2, pages 1–6. MDPI open access, 2018.

[20] M. Iwamoto and J. Shikata. Revisiting conditional Rényi entropies and generalizing Shannons bounds in information theoretically secure encryption. Cryptology ePrint Archive 440/2013, 2013.

[21] R. Jenssen, J. Principe, D. Erdogmus, and T. Eltoft. The Cauchy-Schwarz divergence and Parzen windowing: Connections to graph theory and Mercer kernels. *Journal of the Franklin Institute*, 343(6):614–629, 2006.

[22] H. Kantz and T. Schreiber. *Nonlinear time series analysis*, volume 7 of *Cambridge Nonlinear Science series*. Cambridge University Press, 1997.

[23] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[24] T. Lehn-Schiøler, A. Hegde, D. Erdogmus, and J. Principe. Vector quantization using information theoretic concepts. *Natural Computing*, 4(1):39–51, 2005.

[25] W. Li. Mutual information functions of natural language texts. Technical Report SFI-89-10-008, Santa Fe Institute, 1989.

[26] W. Li. Mutual information functions versus correlation function. *Journal of Statistical Physics*, 60(5/6):823–837, 1990.

[27] W. Li. The study of correlation structures of DNA sequences: a critical review. *Computers and Chemistry*, 21(4):257–271, 1997.

[28] J. Lin, D. Adjeroh, B.-H. Jiang, and Y. Jiang. k_2 and k_2^* : efficient alignment-free sequence similarity measurement based on kendall statistics. *Bioinformatics*, 34(10):1682–1689, 2018.

[29] D. Mackay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.

[30] E. Mwebaze, P. Schneider, F.-M. Schleif, J. Aduwo, J. Quinn, S. Haase, T. Villmann, and M. Biehl. Divergence based classification in learning vector quantization. *Neurocomputing*, 74(9):1429–1435, 2011.

[31] D. Nebel, M. Kaden, A. Villmann, and T. Villmann. Types of (dis−)similarities and adaptive mixtures thereof for improved classification learning. *Neurocomputing*, 268:42–54, 2017.

[32] C.-K. Peng, S. Buldyrev, A. Goldbergert, S. Havlin, F. Sciortino, M. Simonst, and H. Stanley. Long-range correlations in nucleotide sequences. *Nature*, 356:168–170, 1992.

[33] J. Principe. *Information Theoretic Learning.* Springer, Heidelberg, 2010.

[34] A. Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 1961. University of California Press.

[35] A. Rényi. *Probability Theory.* North-Holland Publishing Company, Amsterdam, 1970.

[36] E. Schrödinger. *What is life?* Cambridge University Press, 1944.

[37] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–432, 1948.

[38] A. Singh and J. Principe. Information theoretic learning with adaptive kernels. *Signal Processing*, 91(2):203–213, 2011.

[39] A. Sparavigna. Mutual information and nonadditive entropies: The case of Tsallis entropy. *International Journal of Sciences*, 4(10), 2015.

[40] D. Swati. Use of mutual information function and power spectra for analyzing the structure of some prokaryotic genomes. *American Journal of Mathematical and Management Sciences*, 27(1–2):179–198, 2007.

[41] D. Swati. In *silico* comparison of bacterial strains using mutual information. *Journal of Bio-Sciences*, 32(6):1169–1184, 2009.

[42] A. Teixeira, A. Matos, and L. Antunes. Conditional Rényi entropies. *IEEE Transactions on Information Theory*, 58(7):4273–4277, 2012.

[43] S. Verdú. α-mutual information. In *Information Theory and Applications Workshop (ITA), San Diego*, pages 1–6. IEEE Press, 2015.

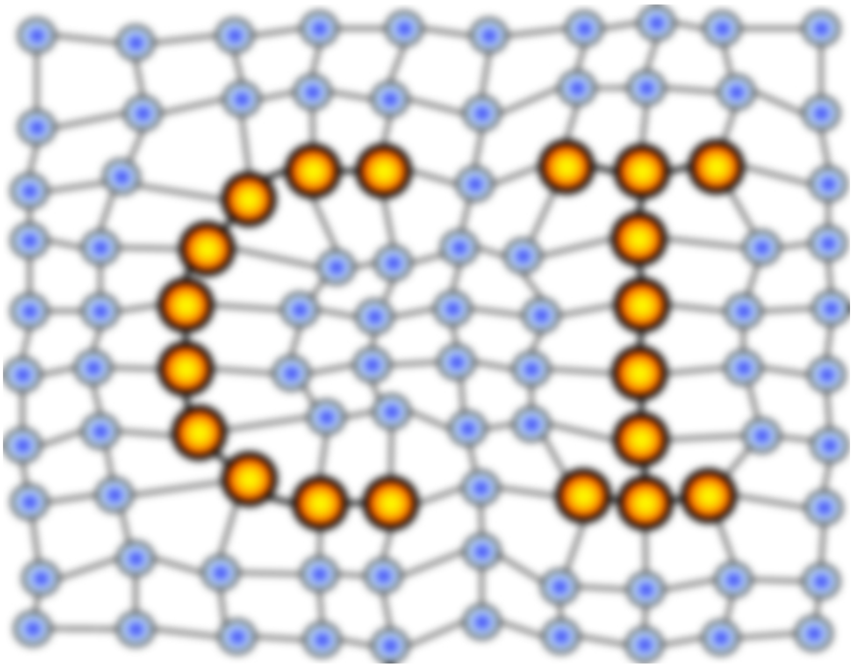
[44] T. Villmann and S. Haase. Divergence based vector quantization. *Neural Computation*, 23(5):1343–1392, 2011.

[45] S. Vinga. Information theory applictions for biological sequence analysis. *Bioinformatics*, 15(3):376–389, 2014.

[46] R. Voss. Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences. *Physical Review A*, 68(25):3805–3808, 1992.

MACHINE LEARNING REPORTS

Report 02/2021



Impressum

Machine Learning Reports ISSN: 1865-3960

▽ Publisher/Editors

Prof. Dr. rer. nat. Thomas Villmann
University of Applied Sciences Mittweida
Technikumplatz 17, 09648 Mittweida, Germany
• http://www.mni.hs-mittweida.de/

Dr. rer. nat. Frank-Michael Schleif
University of Bielefeld
Universitätsstrasse 21-23, 33615 Bielefeld, Germany
• http://www.cit-ec.de/tcs/about

▽ Acknowledgments

We would like to thank the reviewers for their time and patience.