Available online at http://scik.org
J. Math. Comput. Sci. 2022, 12:96
https://doi.org/10.28919/jmcs/6951
ISSN: 1927-5307

ASYMMETRY QUANTIFICATION IN CROSS MODAL RETRIEVAL USING COPULAS

LOUBNA KARBIL[1,*], MOHAMED EL MAAZOUZ[2], AHMED SANI[2], IMANE DAOUDI[1]

[1]Engineering Research Laboratory, ENSEM, Casablanca, Morocco

[2]Department of Mathematics, University Ibn Zohr, Agadir, Morocco

Abstract. Copulas are used to describe and explain the asymmetry between image-text and text-image retrieval observed in different values of the mean average precision (MAP). We use empirical copulas to quantify the asymmetry in a general framework of cross-modal retrieval via suitable asymmetry measures. Several experiments are done on real world dataset feautures to prove the relevance of our analysis.

Keywords: multimodal; copulas; asymmetry; database.

2010 AMS Subject Classification: 68P15.

1. MULTIMODAL HASHING

During recent years, Cross-modal retrieval has already gained wide attention due to the enormous growth of multimedia data on the Internet and social networks. Multimodal search is used in many domains such as pattern recognition, video surveillance and audio-text recognition. The task of cross modal retrieval is to submit a request from one modality and have results from another one [1, 2]. Cross modal retrieval can have advantages of high efficiency and good results with information from different modalities compared to unimodal retrieval methods [3, 4] or more recently [5]. Transforming high multimodal data into a set of a small length

*Corresponding author
E-mail address: l.karbil@gmail.com
Received October 30, 2021

compact code is the main idea of multimodal hashing. This code needs to preserve the cross-modality similarities of the original data. Then, using a convenient metric between obtained codes, we can compute the multimodal similarity distance between the original data. Actually, the limited storage resource and the semantic hole between the representation and original data, makes performing accurate cross-modal retrieval on large scale datasets very challenging. Many approaches have been proposed to address the cross-modality retrieval problem and can be grouped into two categories: Unsupervised methods that learn the cross-modal feature representation space by maximizing the correlation between pairwise data from different modalities such as cross-view hashing (CVH) proposed by Kumar and Udupa [6], is a relevant work that extends the spectral hashing (SH) method [7]. This method aims to minimize the weighted average distance between hash codes to preserve the similarity between data modalities. Co-Regularized Hashing (CRH)[8], the hash functions for each bit of the hash codes are learned by solving difference of convex function programms. Inter-media hashing (IMH) [9] using a similarity graph, the method explores the correlations within each single modality and then keeps the binary codes of the paired data points with different modalities consistent. Multimodal latent binary embedding (MLBE) [10] directly learns the binary hash codes with latent variable models. Multi-View Spectral Hashing (MVSH) [11] integrates Multiview information into binary codes, and uses product of code words to avoid undesirable embedding. [12] proposes a novel Latent Semantic Sparse Hashing (LSSH) employ sparse coding to capture the salient structures of images and Matrix Factorization to learn the latent concepts from text to perform cross-modal similarity search. Semantic Topic Multimodal Hashing (STMH) [13] discovers clustering patterns of texts and robust factorizes the matrix of images to obtain multiple semantic topics of texts and concepts of images. [14] proposes the RFDH model where binary codes are directly learned based on discrete matrix decomposition. Supervised multimodal hashing methods learn hash codes by approximating the pairwise similarity matrices that are constructed by available supervised information such as semantic correlation maximization (SCM) [15] which, based on seamlessly, integrate semantic labels into the hashing learning procedure for large-scale data modeling. Semantics-Preserving Hashing (SePH) method [16] proposed to transform supervised information into a probability distribution learned hash codes in Hamming space using the Kullback-Leibler divergence. Multimodal discriminative binary embedding (MDBE) [10] learns discriminative hash codes creating a classification to learn the hash function. Cross Modal Hashing (CMDH) [17] learned a set of shared binary codes from different modalities, to remove the modality effectively in cross-modal multimedia retrieval. Supervised hierarchical cross-modal hashing (HiCHNet) [18] exploits the hierarchical labels of instances by the pre-established label hierarchy and characterizes each modality of the instance with a set of layer-wise hash representations. Supervised robust discrete multimodal hashing (SRDMH) [19] incorporates full label information into the hash functions learning to preserve the similarity in the original space. Recently, thanks to using deep neural networks while constructing the hash code, high performance can be achieved using deep supervised multimodal hashing methods. Those methods efficiently capture nonlinear correlations between multimodal data such as deep semantic multimodal hashing network (DSMHN) [20], Cross-modal hashing with semantic deep embedding (CMSDE) [21] and Pairwise Correlation Discrete Hashing (PCDH)[22]. Supervised information of the entire training data set are required in supervised methods. This makes performing accurate cross-modal retrieval on large scale datasets very challenging.

In the current work, we are interested to quantify the asymmetry observed between the request/result: text-image and image-text. We recall some obvious results stating this asymmetry in cross modal retrieval. We introduce *Copulas* as a mathematical tool to describe this asymmetry which is, in the best of our knowledge, never treated before. To this end, we use results on asymmetric copulas stated in [23]. In fact, there are some scattered papers where machine learning was related to copulas such as [5], but without evoking asymmetry questions. So in the first section, mathematical ingredients are presented mainly classical results on copulas and some facts on cross modal retrieval. The second section is devoted to describe, via examples, the impact of eventual non linear correlation between image and text descriptors using a main real world database which is in the best of our knowledge not treated before.

2. BIVARIATE COPULAS

It is worth to mention that for multimodal researches, multivariate copulas are more convenient but simultaneously more difficult to handle. For the sake of simplicity and clarity, restriction to the bivariate case is more convenient as it will be seen at the experimental apparatus. In the following, $\mathbb{I}$ denotes the interval $[0,1]$.

Definition 2.1. *A copula C is a bifunction on $\mathbb{I}^2$ into $\mathbb{I}$ which satisfies the following conditions for all u,v,u_1,v_1,u_2,v_2 in $\mathbb{I}$*

1. *Border conditions: $C(0,v) = C(u,0) = 0$.*
2. *Uniform margins: $C(1,v) = v$ and $C(u,1) = u$.*
3. *the $C-$volume property: $V_C(R) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$ for all rectangle $R = [u_1,u_2] \times [v_1,v_2] \subset \mathbb{I}^2$ with $u_1 < u_2$ and $v_1 < v_2$.*

Any copula C is framed between Fréchet-Hoeftding bounds W and M given by $W(u,v) = \max(u+v-1,0)$ and $M(u,v) = \min(u,v)$. Precisely, we have

$$\forall (x,y) \in \mathbb{I}^2 : \quad W(x,y) \leq C(x,y) \leq M(x,y).$$

Definition 2.2. *A bivariate copula C is said symmetric if it satisfies*

$$\forall u,v \in \mathbb{I} : \quad C(u,v) = C(v,u).$$

Exemples 1. *(1) The Fréchet-Hoeftding bounds W and M and the independence copula $\Pi(u,v) = uv$ are all symmetric.*

(2) All archimedean *copulas are symmetric. By archimedean copula, we insinuate a bifunction C satisfying hypotheses (1), (2) and (3) in Definition (2.2) above and for which there exists a decreasing convex function ϕ called* generator *such that*

$$\forall (u,v) \in \mathbb{I}^2 : \quad C(u,v) = \phi^{(-1)}(\phi(u) + \phi(v)).$$

Here, $\phi^{(-1)}$ denotes the generalized inverse of a monotonic function ϕ (not necessarily strictly monotonic). Precisely $\phi^{(-1)}(v) = \inf_{u \in \mathbb{I}}\{u/\phi(u) \geq v\}$. It is clear that such copula is manifestly symmetric.

(3) Let α and β be in $(0,1)$. The Marshall Olkin family of copulas defined by

$$C_{\alpha,\beta}(u,v) = min(u^{1-\alpha}v, uv^{1-\beta})$$

is asymmetric if $\alpha \neq \beta$.

For $\alpha = \frac{1}{2}$ and $\beta = \frac{3}{4}$ the scatter plots of the copula $C_{\alpha,\beta}$ are given by the following figure:

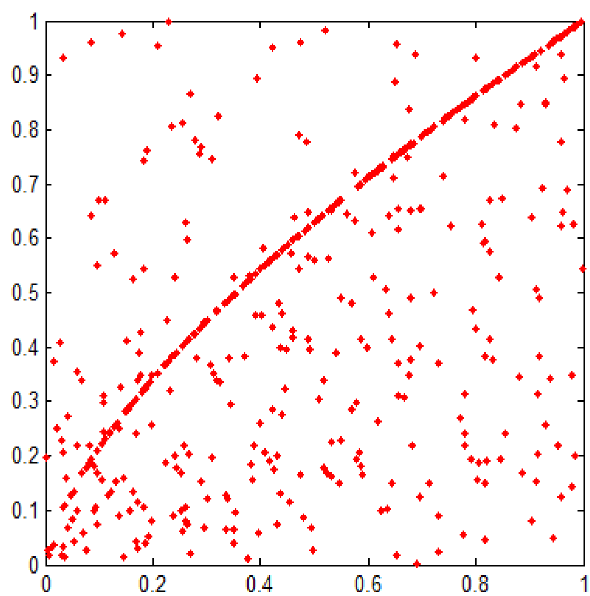

FIGURE 1. 500 data scatter plots of $C_{\frac{1}{2},\frac{3}{4}}$

Comment: This copula is introduced to illustrate the asymmetry in terms of copulas. Observe mainly the density of scatter plot under the curve $v = \sqrt{u}$ which supports mostly the copula. Note that this curve is obtained by $u^{1-\alpha}v = uv^{1-\beta}$ and precise that the plot was realized by choosing $\alpha = \frac{1}{2}$, $\beta = \frac{3}{4}$.

Copulas allow also the calculus of concordance parameters like Spearman's ρ and Kendall's τ. Let us recall briefly the important notions on concordance. To do so, let $\mathcal{Q}$ denote the difference of two probabilities $\mathcal{Q} = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$ for a pair of continuous random vectors (X_1, X_2) and (Y_1, Y_2). If the corresponding copulas are C_1

and C_2, then we have

$$\mathcal{Q}(C_1,C_2) = 4\int_{[0,1]^2} C_1(u,v)\, dC_2(u,v) - 1$$

The usual measures of concordance may be defined in terms of the concordance function $\mathcal{Q}$. Kendall's tau of C is defined by

$$\tau(C) = \mathcal{Q}(C,C)$$

Spearman's rho by

$$\rho(C) = 3\mathcal{Q}(C,\Pi)$$

The undeniable asset of copulas is certainly the seminal Sklar theorem which gives a bridge between joint distributions and marginal ones. It is so inevitable to recall it.

Theorem 2.3 (Sklar's theorem)**.** *Let H be two-dimensional distribution function on a probability space (Ω, p) with marginal distribution functions F and G. Then there exists copula C such that*

$$\forall (x,y) \in \mathbb{I}^2: \quad H(x,y) = C(F(x),G(y)). \tag{2.1}$$

The copula C is unique if F and G are continuous.

When the empiric problem is asymmetric, archimedean copulas, class of copulas that are easy to handle, are not adequate for a precise description of data dependence. It is the case when, for instance, we request an image basing on a given text and vice versa. Indeed all experimental results prove that the cross retrieval process is far to be symmetric (see section below). Whereas it seems difficult presently to overcome this impurity, we suggest first to highlight the asymmetry by measuring it using a fitting copula measure. In fact, this was the real catalyst that prompts us for an investment in asymmetry measures of copulas (see [23]). For the best of our knowledge, the first use of copulas in Machine learning was the paper of G.Elidan in [24]. Therein, the premises of copula use have appeared and were related to artificial intelligence using Bayesian mixture of copula trees and copula Bayesian networks.

3. QUANTIFICATION OF THE ASYMMETRY IN CROSS MODAL RETRIEVAL

3.1. Experimental apparatus: Let us first present the experimental apparatus that we will adopt to quantify the asymmetry in cross modal retrieval. To be concise and clear, we describe roughly the important steps of the experiment.

(1) From a given image *Im*, we precise its descriptors $(d_1^i, d_2^i, id_3 \ldots d_n^i)$.
(2) As result of our request, we obtain a list of texts paired to the image *Im* which fits with the research.
(3) Only the head of the list T_1 is considered as far automatically as the most *similar* to the *Im*. For more information on similarity measures, we refer for instance to [25].
(4) Descriptors $(d_1^t, d_2^t, d_3^t \ldots d_n^t)$ of the T_1 are then given precisely.
(5) At last, we summarize results in contingency table that we analyze in terms of copulas. More precise, we dress a two entries table where the main variables are T and *Im* materialized with their vectors of descriptors:
   a: On the principal row put T descriptors.
   b: On the principal column put *Im* descriptors.
   c: On the cell intersection of a row k and column l, we quantify the jointness by

$$\frac{d_k^i}{d_k^i + d_j^t} \tag{3.2}$$

Data are extracted from Wiki database [26]: it is explored from featured Wikipedia articles. It consists on 2,866 documents which are image-text pairs and annotated with semantic tags of 10 classes. A 128-dimensional SIFT feature vector represents each image and each text is represented by a 10-dimensional topic vector.

It is clear that the formula (3.2) is manifestly asymmetric. We recall that the lack of symmetry is obvious. The goal is to quantify it by a fitting use of asymmetry copula measure.

Now the problem consists in searching of a suitable copula which describes the dependence between *Im* and T. It is used in the literature to consider, with or without any specific awareness or knowledge, the Gaussian copulas as automatic tool for modeling dependence, mainly in applications to risk management. In fact they are the most commonly used also in chemical and biological studies. By the Gaussian copula we mean the following one

$$C(F(x),G(y)) = \Phi(\Phi^{-1}((F(x)),\Phi^{-1}(G(x))), \tag{3.3}$$

where Φ denotes the standard Gaussian distribution function. The common use of this copula is surely explained by the the law of large numbers mainly the limit central theorem which may be found in any elementary or advanced course of probability, e.g [27].

3.2. The empiric fitting copula. We start by explaining estimation procedure of empiric copulas. For our concern, we have at disposal a random sample of $((d_k^i,d_j^l))_{k,l}$ extracted from an unknown joint distribution H. The question now is to determine the underlying copula C. In fact, as the problem is nonparametric and data are discrete, we will estimate the copula C at nods (d_k^i,d_j^l) to obtain the estimated copula $\widehat{C}$. We follow the same way as in [24] to give a fit punctual estimation $\widehat{C}$ of C.
To this end, as usual in this type of situations, we define the joint distribution $H(Im,T)$ as follows

$$H(u,v) = \frac{1}{N+1}\sum_{k=1}^{N} \mathbb{1}_{(Im\leq u;T\leq v)}. \tag{3.4}$$

The function $\mathbb{1}_{(Im\leq u;T\leq v)}$ denotes here the indicator mapping which maps every couple (x,y) to 1 if $x \leq u$ and $y \leq v$, and 0 otherwise.
The formula (3.4) is often used when the entire empiric joint distribution is not large enough and when a direct use of Gaussian copula is adopted . This explains partially why we adopt here the alternative formula (3.2). The mathematical background of this result is the famous Glivenko-Cantelli theorem and its precision on rate of convergence Donsker theorem [27].

To pursue the experiment process, we normalize data from the table. It is a classical technique by adding up all values in the table and dividing each cell on the obtained sum.
To explain the manner in which the formula allows to fill in the aforementioned table, we give in details the algorithm showing how to do so. For numerical results, one may flip the attached txt-files

We consider now the bivariate uniform joint distribution (U,V) defined by

$$(\widehat{U_k},\widehat{V_l}) = (F(d_k^i)), G(v_l^t)).$$

This latter value is considered as an estimation of the empiric copula $\widehat{C}$. For the sack of streamlining, we treat the diagonal data $(\widehat{U_k},\widehat{V_k}) = (F(d_k^i), G(v_k^t))$.
At this stage, we may estimate the empiric copula as follows:

$$\widehat{C}(u,v) = \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{\{\widehat{U_k}\le u;\widehat{V_k}\le v\}}.$$

3.3. The asymmetry measurement. As treated by Siburg and resumed and developed in [23], there are many ways to measure the asymmetry of a given copula C and hence of the joint underlying data that it describes. The most common are $p-$measures ($p \geq 1$) which are defined by the following formula

$$\mu_p(C) = \left(\iint |C(u,v) - C(v,u)|^p \, dudv\right)^{\frac{1}{p}}.$$

The double integral is taken on $\mathbb{I}^2 = [0,1]^2$ and for a suitable presentation, it is used to denote $\overline{C}(u,v) = C(v,u)$ the well known transpose copula.
Mathematically speaking, it is a classical exercise to see that

$$\lim_{p\to\infty} \mu_p = \mu_\infty = \sup_{(u,v)\in\mathbb{I}^2} |C(u,v) - \overline{C}(u,v)|.$$

In the statistical point of view, the measure μ_p also tends to the same limit when the sample size is large enough, say when $p \to +\infty$. This explains the frequent use of μ_∞ instead of general other measures. For curious readers, we refer to [23] for more details and developments on measures of asymmetry and their topological and mathematical treatment.

3.4. Data treatment. To measure the asymmetry , we take an image descriptor and its coresponding text one from the wiki dataset. Then , we measure the jointness between the image/text features and vise versa using the formula 3.2 . We obtain results where part of them are shown in table I and table II.

Image / Text	**037323039,**	**0,003861004**	**0,052767053**	**0**	**0,002574003**	**...**
0,072571837	0,339624924	0,050514984	0,420995055	0	0,034253428	...
0,077900802	0,323917677	0,047222584	0,403825815	0	0,031985198	...
0,032538696	0,534241512	0,106072407	0,618563855	0	0,073306886	...
0,085489399	0,303902762	0,043211935	0,381660693	0	0,029228971	...
0,054061688	0,408416595	0,066657879	0,493940607	0	0,045448419	...
0,069954431	0,347911254	0,052306185	0,429974048	0	0,035489566	...
0,054428878	0,406782116	0,066237976	0,492248657	0	0,045155658	...
0,413522013	0,082784626	0,009250505	0,113163823	0	0,006186078	...
0,094706813	0,282686366	0,039171039	0,35780613	0	0,026459509	...
0,044825443	0,454336318	0,079303461	0,540687604	0	0,054304487	...

TABLE 1. The jointness between image and text in a wiki data raw.

Im / Txt	**0,037323039**	**0,003861004**	**0,052767053**	**0**	**0,002574003**	**...**
0,072571837	0,660375076	0,949485016	0,579004945	1	0,965746572	...
0,077900802	0,676082323	0,952777416	0,596174185	1	0,968014802	...
0,032538696	0,465758488	0,893927593	0,381436145	1	0,926693114	...
0,085489399	0,696097238	0,956788065	0,618339307	1	0,970771029	...
0,054061688	0,591583405	0,933342121	0,506059393	1	0,954551581	...
0,069954431	0,652088746	0,947693815	0,570025952	1	0,964510434	...
0,054428878	0,593217884	0,933762024	0,507751343	1	0,954844342	...
0,413522013	0,917215374	0,990749495	0,886836177	1	0,993813922	...
0,094706813	0,717313634	0,960828961	0,64219387	1	0,973540491	...
0,044825443	0,545663682	0,920696539	0,459312396	1	0,945695513	...

TABLE 2. The jointness between text and Image in a wiki data raw.

As mentioned above, we construct the empiric copula C_1 associated with the first table. For all $(i, j) \in \{0, ..., 9\}^2$ The matrix obtained is given by:

$$\widehat{C_1}\left(\frac{i}{10},\frac{j}{10}\right)=\frac{1}{1280}\begin{pmatrix}
6. & 12. & 12. & 18. & 24. & 30. & 36. & 48. & 54. & 60. \\
18. & 36. & 36. & 54. & 72. & 90. & 108. & 144. & 162. & 180. \\
36. & 72. & 72. & 108. & 144. & 180. & 216. & 288. & 324. & 360. \\
40. & 80. & 80. & 120. & 160. & 200. & 240. & 320. & 360. & 400. \\
56. & 112. & 112. & 168. & 224. & 280. & 336. & 448. & 504. & 560. \\
70. & 140. & 140. & 210. & 280. & 350. & 420. & 560. & 630. & 700. \\
80. & 160. & 160. & 240. & 320. & 400. & 480. & 640. & 720. & 800. \\
96. & 192. & 192. & 288. & 384. & 480. & 576. & 768. & 864. & 960. \\
106. & 212. & 212. & 318. & 424. & 530. & 636. & 848. & 954. & 1060. \\
128. & 256. & 256. & 384. & 512. & 640. & 768. & 1024. & 1152. & 1280.
\end{pmatrix}$$

The Siburg measure of asymmetry of this copula is $\mu_\infty(C_1)=0.0875$.

The same calculus leads to the copula matrix C_2.
For all $(i,j)\in\{0,....,9\}^2$, the matrix obtained is given by:

$$\widehat{C_2}\left(\frac{i}{10},\frac{j}{10}\right)=\frac{1}{1280}\begin{pmatrix}
0. & 13. & 39. & 52. & 65. & 78. & 91. & 91. & 104. & 130. \\
0. & 26. & 78. & 104. & 130. & 156. & 182. & 182. & 208. & 260. \\
0. & 39. & 117. & 156. & 195. & 234. & 273. & 273. & 312. & 390. \\
0. & 51. & 153. & 204. & 255. & 306. & 357. & 357. & 408. & 510. \\
0. & 64. & 192. & 256. & 320. & 384. & 448. & 448. & 512. & 640. \\
0. & 77. & 231. & 308. & 385. & 462. & 539. & 539. & 616. & 770. \\
0. & 90. & 270. & 360. & 450. & 540. & 630. & 630. & 720. & 900. \\
0. & 103. & 309. & 412. & 515. & 618. & 721. & 721. & 824. & 1030. \\
0. & 115. & 345. & 460. & 575. & 690. & 805. & 805. & 920. & 1150. \\
0. & 128. & 384. & 512. & 640. & 768. & 896. & 896. & 1024. & 1280.
\end{pmatrix}$$

The Siburg measure of asymmetry of this copula is $\mu_\infty(C_2)=0.1046875$.

The integrals involving above in concordance parameters formula can be calculated with approximation methods of integral calculus. Here we expose which one is used
For $(i,j)\in\{0,....,9\}$ we denote by $R_{i,j}$ the square $\left[\frac{i}{10},\frac{i+1}{10}\right]\times\left[\frac{j}{10},\frac{j+1}{10}\right]$ Then

$$\begin{aligned}
\iint_{\mathbb{I}^2} C(u,v)\ dC(u,v) &= \sum_{i=0}^{9}\sum_{j=0}^{9}\iint_{\left[\frac{i}{10},\frac{i+1}{10}\right]\times\left[\frac{j}{10},\frac{j+1}{10}\right]} C(u,v)\ dC(u,v) \\
&\simeq \frac{1}{2}\sum_{i=0}^{9}\sum_{j=0}^{9}\left(C\left(\frac{i}{10},\frac{j}{10}\right)+C\left(\frac{i+1}{10},\frac{j+1}{10}\right)\right)V_C\left(R_{i,j}\right)
\end{aligned}$$

After all computations, we obtain:

$$\tau(C_1) = 0.01400$$

and

$$\tau(C_2) = 0.01401$$

For Spearman rho calculus, since $V_\Pi(R_{i,j}) = \frac{1}{100}$, the integral :

$$\iint_{\mathbb{I}^2} C(u,v)\, dudv$$

is approximated similary by:

$$\frac{1}{200}\sum_{i=0}^{9}\sum_{j=0}^{9}\left(C\left(\frac{i}{10},\frac{j}{10}\right)+C\left(\frac{i+1}{10},\frac{j+1}{10}\right)\right)$$

After computations theses results are obtained:

$$\rho(C_1) = -0.55687.$$

and

$$\rho(C_2) = -0.201374999.$$

4. RESULTS AND INTERPRETATIONS

It is possible, without loss of generality, to restrict calculus to $\mu_\infty(Txt - Image)$ but for a deep understanding, the inverse retrieval is also so fundamental. So, on one hand the asymmetry is manifestly made clearer since $\mu_\infty(Image - Txt) - \mu_\infty(Txt - Image) > 0$ and on the other hand, the experience has allowed to deal with the problem of concordance and correlation by estimation of some classical statistic parameters (τ and ρ.) The weak value of $\mu_\infty(Image - Txt)$ compared to $\mu_\infty(Txt - Image)$ may be interpreted in terms of semantic parameters that characterize the text. To express the result in a pith and lapidary language, one will say: the retrieval Text-Image is less asymmetric than the research Image-Text. While the choice of dependence quantification tool explains the strict positive behavior of the both measures, the difference $|\mu_\infty(Image - Txt) - \mu_\infty(Txt - Image)|$ gives a new way to measure the asymmetry in the bimodal-cross retrieval.

Moreover, we observed that the siburg asymmetry of the two copulas is invariant to the the change of the copula matrix order when that one is greater than or equals to 10 which is simultaneously impressing and unpredictable. This expresses the convergence velocity of $(\hat{C}_n)$ to theoretical copula C that joint really the raw data is so ”fast” in the sense given by Donsker theorem.

A simple calculus leads to $\mu_\infty(Txt - Image) = 0.0875$. Let us now discuss predictability of this empirical result. It is will known that for every copula C, $\mu_\infty \leq \frac{1}{3}$. This is a direct consequence of Fréchet-Hoffding bounds which stipulate that the double inequality $\max(u+v-1,0) \leq C(u,v) \leq \min(u,v)$ holds for all copulas. Therefore, the value obtained is so plausible. However, $\mu_\infty = 0.0875$ is of order 10^{-2} which is, statistically speaking, low but considerable. A suitable interpretation is that the value confirms the fact that the asymmetry in research of different modalities is consistent but maybe avoidable if descriptors are improved to obtain more faithful description for documents.

For an adequate interpretation of negative values of the coefficient ρ, a deeper comprehension of the manner to construct descriptors of each document is probably the best way to explain the weak correlation of the two documents in an indirect sense. We postpone the treatment of this interesting question to an other work

5. Conclusion

It is certainly better to take in account all document modalities or at least a lot of them, but this implies an avoidable use of multivariate copulas. This will be our next investment in the future work.

Conflict of Interests

The author(s) declare that there is no conflict of interests.

References

[1] J. Costa Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G.R.G. Lanckriet, R. Levy, N. Vasconcelos, On the Role of Correlation and Abstraction in Cross-Modal Multimedia Retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014), 521–535.

[2] W. Wang, X. Yang, B.C. Ooi, D. Zhang, Y. Zhuang, Effective deep learning-based multi-modal retrieval, VLDB J. 25 (2016), 79–101.

[3] B. Kulis, K. Grauman, Kernelized locality-sensitive hashing for scalable image search, in: 2009 IEEE 12th international conference on computer vision, IEEE, Kyoto, 2009: pp. 2130–2137.

[4] W. Liu, J. Wang, R. Ji, Y.G. Jiang, S.F. Chang, Supervised hashing with kernels, in: 2012 IEEE conference on computer vision and pattern recognition, IEEE, Providence, RI, 2012: pp. 2074–2081.

[5] D. Zhang, X.-J. Wu, H.-F. Yin, J. Kittler, MOON: Multi-hash codes joint learning for cross-media retrieval, Pattern Recogn. Lett. 151 (2021), 19–25.

[6] S. Kumar, R. Udupa, Learning hash functions for cross-view similarity search. In: Proceedings of the 25th international joint conference on artificial intelligence, New York, 2011. 1360–1365.

[7] Y. Weiss, A. Torralba, R. Fergus, et al., Spectral hashing., in: Nips, Vol. 1, Citeseer, 2008, p. 4.

[8] Y. Zhen, D. Y. Yeung, Co-regularized hashing for multimodal data, Adv. Neural Inform. Proc. Syst. 2 (2012), 1376.

[9] J. Song, Y. Yang, Y. Yang, Z. Huang, H. T. Shen, Inter-media hashing for large-scale retrieval from heterogeneous data sources, in: Proceedings of the 2013 ACM SIGMOD international conference on management of data, 2013, pp. 785–796.

[10] D. Wang, X. Gao, X. Wang, L. He, B. Yuan, Multimodal discriminative binary embedding for large-scale cross-modal retrieval, IEEE Trans. Image Process. 25 (2016), 4540–4554.

[11] X. Shen, F. Shen, Q.-S. Sun, Y.-H. Yuan, Multi-view latent hashing for efficient multimedia search, in: Proceedings of the 23rd ACM international conference on Multimedia, 2015, pp. 831–834.

[12] J. Zhou, G. Ding, Y. Guo, Latent semantic sparse hashing for cross-modal similarity search, in: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014, pp. 415–424.

[13] D. Wang, X. Gao, X. Wang, L. He, Semantic topic multimodal hashing for cross-media retrieval, in: Twenty-fourth international joint conference on artificial intelligence, 2015.

[14] D. Wang, Q. Wang, X. Gao, Robust and flexible discrete hashing for cross-modal similarity search, IEEE transactions on circuits and systems for video technology 28 (10) (2017) 2703–2715.

[15] D. Zhang, W.-J. Li, Large-scale supervised multimodal hashing with semantic correlation maximization, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 28, 2014.

[16] D. Zhang, W.-J. Li, Large-scale supervised multimodal hashing with semantic correlation maximization, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 28, 2014.

[17] Q.-Y. Jiang, W.-J. Li, Deep cross-modal hashing, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3232–3240.

[18] C. Sun, X. Song, F. Feng, W. X. Zhao, H. Zhang, L. Nie, Supervised hierarchical cross-modal hashing, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 725–734.

[19] T.-K. Yan, X.-S. Xu, S. Guo, Z. Huang, X.-L. Wang, Supervised robust discrete multimodal hashing for cross-media retrieval, in: Proceedings of the 25th ACM international on conference on information and knowledge management, 2016, pp. 1271–1280.

[20] L. Jin, Z. Li, J. Tang, Deep semantic multimodal hashing network for scalable image-text and video-text retrievals, IEEE Transactions on Neural Networks and Learning Systems (2020), 1-14.

[21] C. Yan, X. Bai, S. Wang, J. Zhou, E. R. Hancock, Cross-modal hashing with semantic deep embedding, Neurocomputing, 337 (2019), 58–66.

[22] Y. Chen, X. Lu, Deep discrete hashing with pairwise correlation learning, Neurocomputing 385 (2020) 111–121.

[23] L. K. Ahmed Sani, A functional treatment of asymmetric copulas, Electron. J. Math. Anal. Appl. 8 (2020), 17–26.

[24] G. Elidan, Lecture notes in statistics,proceedings of the workshop held in cracow,, in: Copulae in Mathematical and Quantitative Finance, Springer, July 2012, pp. 39–60.

[25] A. Idarrou, D. Mammas, A new structural similarity measure for clustering multi-structured documents, J. Theor. Appl. Inform. Technol. 86 (2016), 34–43.

[26] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, Y. Zheng, Nus-wide: a real-world web image database from national university of singapore, in: Proceedings of the ACM international conference on image and video retrieval, 2009, pp. 1–9.

[27] P. Billingsley, Probability and Measure (3rd ed.), John Wiley and Sons, 1995.