Copula Entropy

Theory and Applications

MA Jian
majian.2003@tsinghua.org.cn

Contents

Lecture 3: Theoretical applications (1)

Copula Entropy: Application 1

Association Discovery[1]

[1] Jian Ma. "Discovering Association with Copula Entropy". In: *arXiv preprint arXiv:1907.12268* (2019).

Copula Entropy: Association Discovery

- Problem
  - To discover association relations between random variables from data
- History
  - An old and fundamental problem date back to the early days of statistics
- Related Methods
  - Pearson Correlation Coefficient r
  - Spearman's ρ and Kendall's τ
  - Regression

Copula Entropy: Association Discovery

Traditional association measures

- Pearson Correlation Coefficient[2]

$$r_{XY} = corr(X, Y) = \frac{cov(X, Y)}{\delta_X \delta_Y} \tag{1}$$

- Spearman's ρ[3] and Kendall's τ[4]

$$\rho_{XY} = 12 \int_u \int_v C(u, v) du dv - 3 \tag{2}$$

$$\tau_{XY} = 4 \int_u \int_v C(u, v) dC(u, v) - 1 \tag{3}$$

[2] Karl Pearson. "Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs". In: *Proceedings of The Royal Society of London* 60.1 (1896), pp. 489–498.

[3] C. Spearman. "The Proof and Measurement of Association between Two Things". In: *The American Journal of Psychology* 100.3/4 (1987), pp. 441–471.

[4] M. G. Kendall. "A New Measure of Rank Correlation". In: *Biometrika* 30.1/2 (1938), pp. 81–93. ISSN: 00063444. URL: http://www.jstor.org/stable/2332226.

Copula Entropy: Association Discovery

Why Copula Entropy?

Table: Theoretical comparison between CE and CC.

	CC	CE
linearity	linear	nonlinear
Order	2	≥ 2
Assumption	Gaussianity	None
variate	bivariate	multivariate

Copula Entropy: Association Discovery

CE between random vectors

- Theorem

$$H_c(\mathbf{x}_1; \ldots; \mathbf{x}_n) = H_c(\mathbf{x}) - \sum_{i=1}^{n} H_c(\mathbf{x}_i). \tag{4}$$

- Merits
  - rigorous and provable
  - intuitive and natural
  - easy to implement

Copula Entropy: Association Discovery

Experiments on the NHANES data

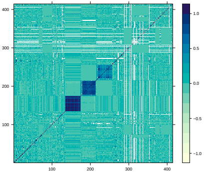
- Objectives of NHANES
  - to monitor trends and emerging issues of population health
  - to investigate its relationship with risk factors, nutrition and environmental exposures, etc.
- NHANES (2013-2014)
  - 14,332 persons from 30 different survey locations were selected
  - Of those selected, 10,175 interviewed and 9,813 examined
  - 5 groups of data: demographics, dietary, examination, laboratory, and questionnaire
- Experimental data

  The laboratory data, which includes 423 variables from blood, urine, oral rinse and vaginal/Penile swabs.
- Missing values

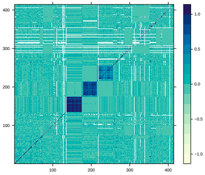
  The missing values were filled with the mean of their corresponding variables.

Copula Entropy: Association Discovery

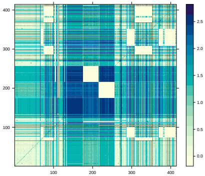
- Results - Correlation matrices



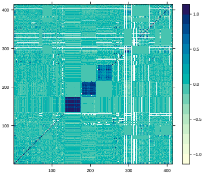
Pearson's r



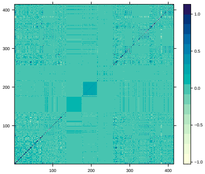
Kendall's τ



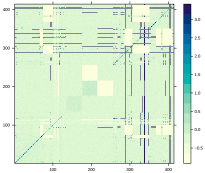
Schweizer & Wolff's σ



Spearman's ρ



Gini's γ



CE

Copula Entropy: Application 2

Structure Learning[5]

[5] **Ma2009**, Jian Ma and Zengqi Sun. "Dependence Structure Estimation via Copula". In: *arXiv preprint arXiv:0804.4451* (2008).

Copula Entropy: Structure Learning

- Problem
  - To learn statistical structure among random variables from data
- Graph Representation
  - A probability density is represented with a directed or indirected graph, of which each node represents a random variable, and each edge represents a (conditional) dependence relation between two random variables
- Related Methods
  - Chow-Liu Algorithm
  - Vine Copula

Copula Entropy: Structure Learning

Given random variables $\mathbf{X} = (X_1, \ldots, X_m)$ and their corresponding graph structure G, then their joint density function can be represented as the following product form:

$$p(\mathbf{x}|G) = \prod_{i=1}^{m} p(x_i|x_{\pi_i}), \tag{5}$$

where π_i is the predecessor nodes of x_i in G.
Graph representation has its corresponding copula, since graph represents dependence relationships and so does copula. In this sense, (5) can be transformed into the following copula form:

$$p(\mathbf{x}|G) = \prod_{i=1}^{m} c_i(x_i, x_{\pi_i}) \prod_{i=1}^{m} p(x_i). \tag{6}$$

Copula Entropy: Structure Learning

Chow-Liu method[6] generates optimal tree structure to approximate density function by maximizing the following MI sum of tree structure with minimal-spanning-tree (MST) algorithm:

$$\max \sum_{i=1}^{m} I(x_i, x_{\pi_i}). \quad (7)$$

[6] C.K. Chow and C.N. Liu. "Approximating discrete probability distributions with dependence trees". In: *IEEE Transactions on Information Theory* 14.3 (1968), pp. 462–467.

Copula Entropy: Structure Learning

Given a sample $\mathbf{X} = X_1, \ldots, X_T$ from $p(\mathbf{x})$ associated with copula function $c(\mathbf{u})$, then

- Likelihood

$$\mathcal{L}(\theta; \mathbf{X}) = \sum_{t=1}^{T} \log p(\mathbf{x}_t; \theta) \tag{8}$$

- **Copula Likelihood**[7]

$$\mathcal{L}_c(\theta_c; \mathbf{X}) = \sum_{t=1}^{T} \log c(\mathbf{u}_t; \theta_c). \tag{9}$$

  - equivalent to copula entropy
  - guide principle of copula based structure learning

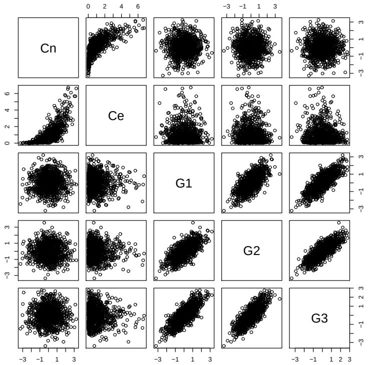
[7] 马健. “Copula 结构学习”. 博士学位论文. 清华大学, 2009.

Copula Entropy: Structure Learning

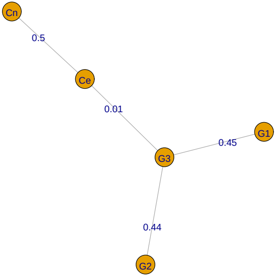
- **Copula Entropy based Chow-Liu Algorithm**
  1. computing dependence matrix $\mathbf{W}_x$ of data x with CE estimation
  2. constructing dependence structure T from $\mathbf{W}_x$ with MST algorithm
- Advantages
  - distribution-free, non-parametric
  - tuning-free, insensitive to parameters
  - easy to implement
  - low computation burden

Copula Entropy: Structure Learning

- Simulation Experiment
  5 random variables: the first three are Gaussian and the others two are governed by Gaussian copula with margins as normal distribution and exponential distribution respectively



Simulated data



Learned graph

Copula Entropy: Structure Learning

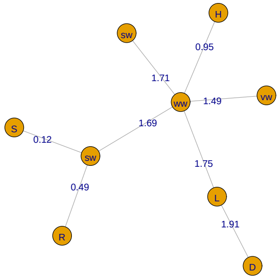
Experiment on real data

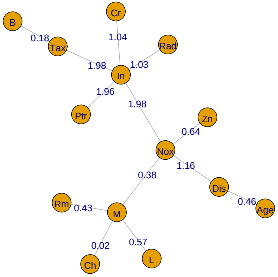
- Abalone data

  Predicting the age of abalone from physical measurements

- Boston housing data

  Concerns housing values in suburbs of Boston





Copula Entropy: Application 3

Variable Selection[8]

[8] Jian Ma. "Variable Selection with Copula Entropy". In: *Chinese Journal of Applied Probability and Statistics* 37.4 (2021). See also arXiv preprint arXiv:1910.12389, 2019, pp. 405–420.

Copula Entropy: Variable Selection

- Problem
  - To select a 'right' subset of variables from the whole group for building classification or regression models with good predictability and interpretability
- History
  - An old and basic problem in statistics and machine learning
- Related Problems
  - Feature Selection
  - Model Selection

Copula Entropy: Variable Selection

Existing methods - Likelihood with penalty

- Information Criteria
  with penalty on the number of parameters in the models

$$\mathbf{AIC} = -2L + 2p \tag{10}$$

$$\mathbf{BIC} = -2L + p \log N \tag{11}$$

- Penalized GLMs
  with penalty on the nonzero coefficients in the GLMs
  - LASSO
  - Ridge Regression
  - Elastic Net

$$\min_{\beta}\{L(\beta; y, \mathbf{X}) + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2\} \tag{12}$$

  - Adaptive LASSO

$$\min_{\beta}\{L(\beta; y, \mathbf{X}) + \lambda \sum_{j=1}^{p} w_j |\beta_j|\} \tag{13}$$

Copula Entropy: Variable Selection

Existing methods - Statistical independence measures

- Distance Correlation

$$\mathbf{dCor}(X, Y) = \frac{\nu^2(X, Y)}{\sqrt{\nu^2(X)\nu^2(Y)}}, \tag{14}$$

where $\nu^2(X, Y)$ be distance covariance.

- Hilbert-Schmidt Independence Criterion (HSIC)

$$\mathbf{dHSIC}(P(\mathbf{X})) = ||\Pi(P(X_1)\otimes, \ldots, \otimes P(X_d)) - \Pi(P(\mathbf{X}))||, \tag{15}$$

where Π be the mean embedding function associated with kernel functions.

Copula Entropy: Variable Selection

- CE based method
  To select variables X_i according to the rank of their negative CE values with target Y

$$-H_c(X_i, Y) > -H_c(X_j, Y) > \cdots > -H_c(X_k, Y) \quad (16)$$

- Advantages
  - model-free, non-parametric
  - tuning-free, insensitive to parameters
  - interpretable with physical meanings
  - supported by rigorous math
  - science instead of art, compared with existing methods
  - easy to implement, low computation burden

Copula Entropy: Variable Selection

Experiments on the UCI heart disease data[9]

- Overview of the data
  The data set contains 4 databases (899 samples) concerning heart disease diagnosis. All attributes are numeric-valued. The data was collected from the four following locations:
  - Cleveland clinic foundation;
  - Hungarian Institute of Cardiology, Budapest;
  - V.A. medical center, long beach, CA;
  - University hospital, Zurich, Switzerland.
- Attributes
  The data has 76 attributes (#58 'num' for diagnosis). Of them, 13 attributes are recommended by professionals as clinical relevant.

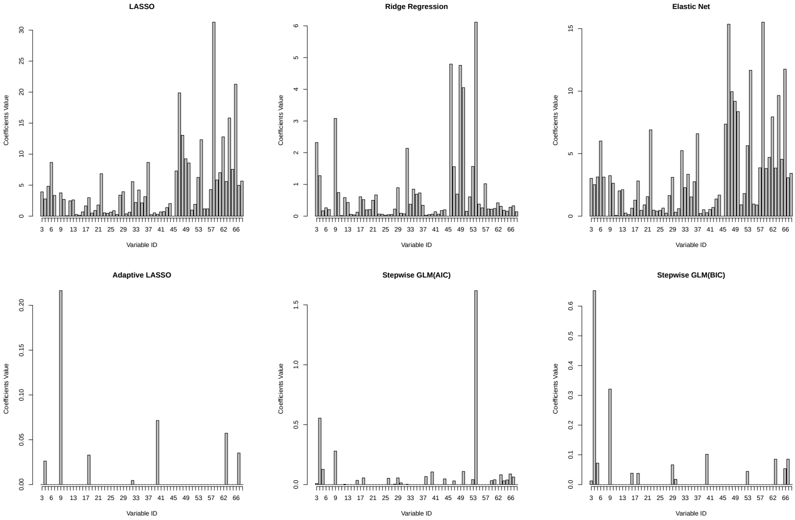
Table: Recommended attributes.

ID	3	4	9	10	12	16	19
Name	age	sex	cp	trestbps	chol	fbs	restecg
ID	32	38	40	41	44	51	58
Name	thalach	exang	oldpeak	slope	ca	thal	num

[9] **Asuncion2007**.

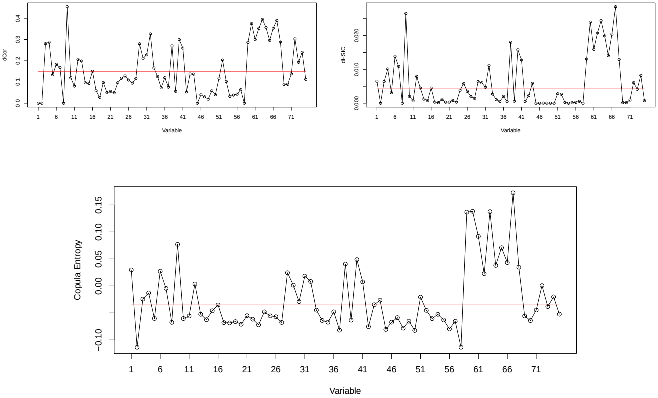
Copula Entropy: Variable Selection

- Results - Coefficients of penalized likelihood based models



Copula Entropy: Variable Selection

- Results - with statistical dependence measures (dCor, dHSIC, CE)



Copula Entropy: Variable Selection

- Results - Prediction accuracy
  the selected variables present the best prediction accuracy.

Model	Accuracy(%)
SVM(Recommended variables)	84.20
SVM(CE)	**84.76**
SVM(dCor)	82.76
SVM(dHSIC)	84.54
Stepwise GLM(AIC)	51.8
Stepwise GLM(BIC)	49.1
LASSO	79.2
Ridge Regression	63.0
Elastic Net	75.9
Adaptive LASSO	35.7

Copula Entropy: Variable Selection

- Results - Selected variables
  Copula Entropy selects more 'right' variables than the other methods do.

Method	Selected Variables' ID	✓
Recommended variables	3,4,9,10,12,16,19,32,38,40,41,44,51	13
CE	3,4,6,7,9,12,16,28-32,38,40,41,44,51,59-68	**11**
dHSIC	3,4,6,7,9,12,13,16,25,29-32,38,40,41,44,59-68	10
dCor	3,4,6,7,9,12,13,16,28-33,38,40,41,52,59-68	9
Stepwise GLM(AIC)	3,4,5,9,12,16,18,20,26,29,30,32,40,44,47,50,53,54,60,61,63,65-67	8
Stepwise GLM(BIC)	3,4,5,9,16,18,29,30,40,53,63,66,67	5
Adaptive LASSO	4,6,9,18,32,40,63,67	4
LASSO		
Ridge Regression	all except 8,45	-
Elastic Net		

Monograph

Jian Ma. "Copula Entropy: Theory and Applications". In: *arXiv preprint arXiv:2025.18168* (2025)

Thanks!

Q&A