SHGR: A Generalized Maximal Correlation Coefficient

Conference on Neural Information Processing Systems

Samuel Stocksieker, Denys Pommeret

04/12/2025

Context

- Traditional correlation measures (e.g., Pearson, Spearman) fail to capture nonlinear or multivariate dependencies.
- The Hirschfeld–Gebelein–Rényi (HGR) maximal correlation addresses this by seeking the strongest possible nonlinear relationship between two variables.
- However, estimating HGR is computationally challenging due to the complexity of its nonlinear optimization.
- We introduce the Spearman HGR (SHGR), a new, rank-based and differentiable, maximal correlation measure inspired by HGR.

Contributions

- We propose SHGR, a Spearman-based extension of the HGR maximal correlation.
- We design the first differentiable and robust estimator that is fast, noise-resistant, and capable of recovering learned transformations while enabling significance testing.
- We develop a stacked cross-encoder architecture to estimate multiple correlations simultaneously in both bivariate and multivariate settings.
- We introduce a comprehensive evaluation protocol, the *Multivariate Power of Correlation Measure*, assessing performance, robustness, and computational efficiency.

Background

HGR maximal correlation coefficient

Let U ($\sim \mathcal{D}_U$) and V ($\sim \mathcal{D}_V$) be two continuous random variables taking values in $\mathcal{U}$ and $\mathcal{V}$, respectively. Let $\mathcal{E}(\mathcal{U})$ (resp. $\mathcal{E}(\mathcal{V})$) denote the set of measurable functions from $\mathcal{U}$ (resp. $\mathcal{V}$) to $\mathbb{R}$. Let $r()$ denote the Pearson correlation coefficient.

The Hirschfeld-Gebelein-Renyi (HGR) maximal correlation coefficient is defined as follows:

$$
\begin{aligned}
HGR(U, V) &:= \sup_{\substack{f_U \in \mathcal{E}(\mathcal{U}) \\ f_V \in \mathcal{E}(\mathcal{V})}} r(f_U(U), f_V(V)) \\
&= \sup_{\substack{f_U \in \mathcal{E}(\mathcal{U}), f_V \in \mathcal{E}(\mathcal{V}) \\ \mathbb{E}(f_U(U))=0, \mathbb{E}(f_V(V))=0 \\ \mathbb{E}(f_U^2(U))=1, \mathbb{E}(f_V^2(V))=1}} \mathbb{E}_{U \sim \mathcal{D}_U, V \sim \mathcal{D}_V}(f_U(U) f_V(V))
\end{aligned}
$$

Background

HGR Neural Estimate

Grari et al. (2021) proposed estimating the HGR transformations using neural networks to capture nonlinear relationships. The algorithm takes u, sample of U and v, sample of V as inputs, and returns as output the estimated Pearson correlation[1] $r(f_u(u), f_v(v))$, where f_u and f_v are parameterized by a compact domain Θ. This measure is then estimated using a neural network by minimizing the following loss function:

$$\mathcal{L}_{NHGR} = -\sup \, r(f_{\theta_u}(u), f_{\theta_v}(v)),$$

where f_{θ_u} (resp. f_{θ_v}) denotes a neural estimator of f_u (resp. f_v). This estimator of the maximal correlation coefficient $HGR(u, v)$, denoted $NHGR$ (for Neural-HGR) is defined as:

$$NHGR_{\Theta}(u, v) = r(f^*_{\theta_u}(u), f^*_{\theta_v}(v)), \text{ with } (f^*_{\theta_u}, f^*_{\theta_v}) = \underset{f_{\theta_u}, f_{\theta_v} \in \Theta}{\arg\max} \, r(f_{\theta_u}(u), f_{\theta_v}(v))$$

[1] Its estimate is also designated by r to simplify notation.

Spearman-HGR (SHGR) coefficient

Let $F_U(u)$ the cumulative distribution function of a random variable U. Using the previous notations, the Spearman-HGR (SHGR) coefficient associated to (U, V) is defined by:

$$SHGR(U,V) := \max_{\substack{f_u \in \mathcal{E}(\mathcal{U}), f_v \in \mathcal{E}(\mathcal{V}) \\ \mathbb{E}(f_u(U))=0, \mathbb{E}(f_v(V))=0 \\ \mathbb{E}(f_u^2(U))=1, \mathbb{E}(f_v^2(V))=1}} r(F_{f_u(U)}(f_u(U)), F_{f_v(V)}(f_v(V))).$$

Using the empirical estimator $\widehat{F}$ of F, we obtain an estimator of the SHGR, $\widehat{SHGR}$, and its neural version, $SHGR_\Theta$, defined as following:

$$\widehat{SHGR} = \max_{\substack{f_u \in \mathcal{E}(\mathcal{U}), f_v \in \mathcal{E}(\mathcal{V}) \\ \mathbb{E}(f_u(U))=0, \mathbb{E}(f_v(V))=0 \\ \mathbb{E}(f_u^2(U))=1, \mathbb{E}(f_v^2(V))=1}} r(\widehat{F}_{f_u(U)}(f_u(U)), \widehat{F}_{f_v(V)}(f_v(V))),$$

$$SHGR_\Theta(u,v) = \max_{f_{\theta_u}, f_{\theta_v} \in \Theta} r(\widehat{F}_{f_{\theta_u}(U)}(f_{\theta_u}(u)), \widehat{F}_{f_{\theta_v}(V)}(f_{\theta_v}(v)))$$

Stacked Cross-Encoder Design

We extend SHGR to handle multivariate dependencies:

i) between pairs of variables (bivariate design)
ii) between one variable and all remaining variables (multivariate design)
iii) between two sets of variables (full design)

To estimate multiple correlations efficiently at once, we design a neural stacked cross-encoder that produces complete SHGR correlation matrices:

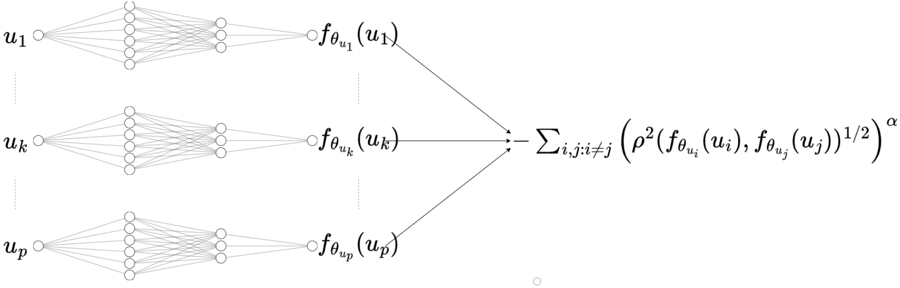


Figure: Stacked Cross-Encoder for bivariate correlations estimation

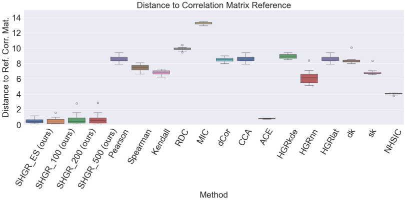
Experiments

We evaluate `SHGR` and several competing methods through a comprehensive benchmark, called the *Multivariate Power of Correlation Measures*, across three designs (bivariate, multivariate, and full).

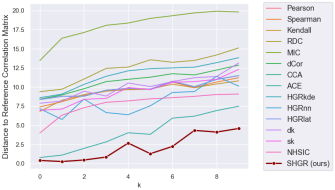
Evaluation criteria:

- **Performance:** ability to detect complex nonlinear dependencies;
- **Noise robustness:** ability to maintain accuracy in the presence of noise;
- **Hallucination robustness:** ensuring null correlation under independence;
- **Outlier robustness:** maintaining stability in the presence of extreme values;
- **Power of dependence measure:** as proposed by Lopez-Paz et al. (2013);
- **Computation time:** efficiency in estimating correlations;
- **Significance testing:** ability to correctly reject the null hypothesis of zero correlation.

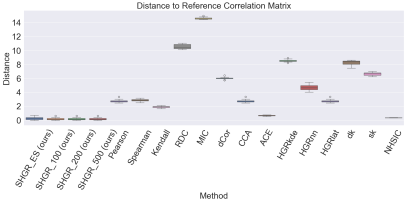
Experiments



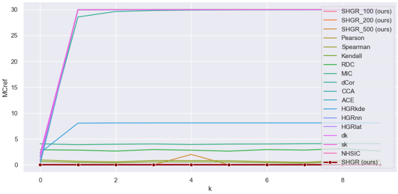
(a) Performance



(b) Robustness to noise

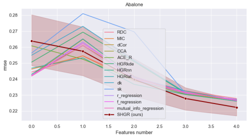


(c) Robustness to hallucinations

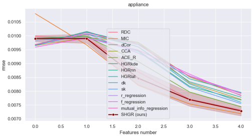


(d) Robustness to extreme values

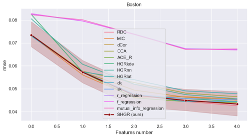
Figure: Some results for pairwise correlation estimation. SHGR (in red)



(a) Abalone dataset



(b) Appliance dataset



(c) Boston dataset

Figure: RMSE on test set prediction with feature selection on real-world datasets (lower is better). Based on maximal (bivariate) correlation — results with SHGR are shown in red.

Conclusion

This work introduces SHGR, a novel maximal correlation coefficient that extends the Hirschfeld–Gebelein–Rényi (HGR) framework through rank-based transformations and a neural architecture specifically designed for its estimation. The proposed method efficiently estimates full correlation matrices, remains robust to noise and overfitting, and consistently outperforms state-of-the-art approaches across multiple settings.

Bibliography

Grari, V., Hajouji, O. E., Lamprier, S., and Detyniecki, M. (2021). Learning unbiased representations via rényi minimization. In Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., and Lozano, J. A., editors, *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 749–764, Cham. Springer International Publishing.

Lopez-Paz, D., Hennig, P., and Schölkopf, B. (2013). The randomized dependence coefficient. *Advances in neural information processing systems*, 26.