Identifiable Feature Learning for Spatial Data with Nonlinear ICA

Hermanni Hälvä[1], Jonathan So[2], Richard E. Turner[2], Aapo Hyvärinen[1]

[1]Department of Computer Science, University of Helsinki & [2]Department of Engineering, University of Cambridge

Summary

- Nonlinear ICA: identifiable feature learning
- Previous works: latent dependencies guarantee identifiability of components
- Limitation: mainly exploited temporal dependencies
- Our work: nonlinear ICA for high-dimensional dependencies e.g. spatial and spatio-temporal data

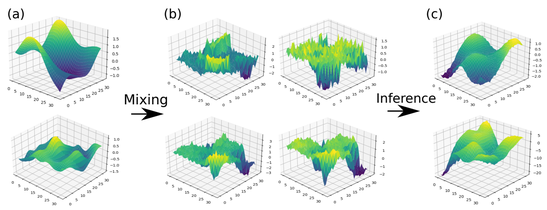

Figure 1. Example simulated spatial data with mixing and demixing with tp-NICA.

Background

Identifiability problem

Model is identifiable if:

$$p_\theta(x) = p_{\hat{\theta}}(x) \Rightarrow \theta = \widehat{\theta}$$

Deep generative models with factorial priors e.g. VAEs...:

$$p(s) = \prod_{i=1}^{N} p(s_i)$$
$$x = f(s)$$

... are unidentifiable. Thus the ground-truth mixing function f and the latent components s can *not* be recovered.

$$p_f(x) = p_{\hat{f}}(x) \;\not\Rightarrow\; f = \hat{f}$$

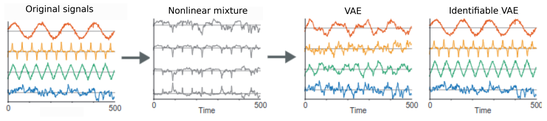

Figure 2. Latent variable reconstruction for identifiable vs. unidentifiable VAE

Identifiability from latent dependencies

Structured Nonlinear ICA (SNICA) [2]: latent dependencies guarantee identifiability if sufficient conditions satisfied:

- (A1) "Sufficiently strong" dependency between "nearby" data points
- (A2) Distribution of each independent component is non-quasi-Gaussian

$\Rightarrow f^{-1}$ and the ground-truth latent components s are identifiable.

Limitations of previous works

SNICA only provides general identifiability theorem and model class

Specific model and algorithms for temporal data but not for general latent dependencies

Unable to exploit dependencies e.g. in spatial & spatio-temporal data

Proposed approach

Identifiable nonlinear ICA model that:

- Exploits dependencies of any order e.g. spatial and spatio-temporal dependencies
- Utilizes $t-$process latent components $\Rightarrow$ non-Gaussian $\Rightarrow$ identifiable
- Scalable variational inference algorithm
- Prove necessity of non-Gaussianity for general identifiability

$t-$process nonlinear ICA

- Assume indexing set $\mathbb{L}$ of arbitrary dimension e.g. 2D for spatial data
- Each independent component follows t-process over this set:

$$(s_l^{(i)})_{l\in\mathbb{L}} \sim \mathcal{TP}_{\nu^{(i)}}(h^{(i)}, \kappa^{(i)}),$$

  for mean function $h(\cdot)$, covariance kernel $\kappa(\cdot,\cdot)$, degrees-of-freedom ν
- At each index location $l \in \mathbb{L}$, nonlinear mixing produces the observations:

$$x_l = f(s_l) + \varepsilon_l,$$

  with observation noise ε_l
- We call this model **tp-NICA**.
- In the limit $\nu \to \infty$, the t-process becomes a Gaussian process (GP) and we get **gp-NICA** as a special case

Learning and Inference

Problem

Marginal likelihood for tp-NICA...:

$$\log p(x) = \log \int_s \prod_{j=1}^{m} p(x_{l_j} \mid s_{l_j}) \prod_{i=1}^{N} p(s^{(i)})\, ds,$$

...is problematic:

1. $p(x_l \mid s_l)$ is a nonlinear observation likelihood
2. $p(s^{(i)})$ is non-exponential family and thus non-conjugate
3. $t-$process computational complexity $\mathcal{O}((mN)^3)$

Solution

Variational inference algorithm:

1. Approximate $q(s_{l_j} \mid x_{l_j})$ as Gaussian factor with free-form variational parameters [3]
2. Represent $t-$process as infinite mixture of gamma-scaled GPs
   $\Rightarrow$ Conjugacy
3. Pseudo-latents from sparse VAE literature [1]

Resulting variational lower bound:

$$\mathbb{E}_{q(\tau)} \underbrace{\mathbb{E}_{\tilde{q}(s|\tau)}\left[\log p(x \mid s)\right]}_{\text{From 1.}} - \underbrace{\mathrm{KL}[q(u \mid \tau) \,\|\, p(u \mid \tau)]]}_{\text{From 2.,3.}} - \underbrace{\mathrm{KL}[q(\tau) \,\|\, p(\tau)]}_{\text{From 2.}},$$

where u are pseudo-latents whose locations are optimized and τ is gamma random variable.

Computational complexity: max $\mathcal{O}((kN)^3)$ where k num. of pseudo-points

Identifiability of tp-NICA and gp-NICA

Theorem 1 – Identifiability of tp-NICA

$t-$processes are proven to be non-quasi-Gaussian $\Rightarrow$ assumptions (A1) and (A2) (see left) are satisfied $\Rightarrow$ f^{-1} is identifiable

Theorem 2 – Necessity of distinct covariance kernels in gp-NICA

[2]: *If* GP components have distinct covariance kernels $\Rightarrow$gp-NICA is identifiable

This work: *if* GP components have the same covariance kernel $\Rightarrow$gp-NICA is not identifiable

Combined: gp-NICA is identifiable if and only if all components have distinct covariance kernels

Corollary: non-Gaussianity is necessary for general identifiability

Experiments

Simulated data

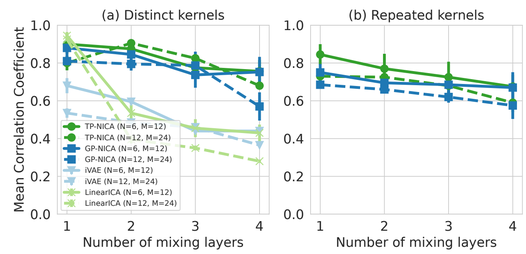

Figure 3. Mean absolute correlation coefficients between ground-truth independent components and their estimates for (a) distinct covariance kernels (b) components with equivalent kernels.

Real data: crop growth cycles

Multi-spectral 16 dimensional 32×32 satellite imagery of 4000 crop fields over 6 month growth cycle

tp-NICA and gp-NICA trained on the data to perform feature extraction

Task: Shuffle the data temporally. Multi-class classification (Random forest; RF) on the learned features to put data in correct temporal order.

	Random baseline	RF only	gp-NICA (l=2) + RF	gp-NICA (l=3) + RF	tp-NICA (l=2, nu=4) + RF	tp-NICA (l=3, nu=4) + RF
Avg. cross-entropy	1.81 ± 0.01	1.33 ± 0.02	1.19 ± 0.01	1.3 ± 0.01	**0.97** ± 0.01	1.08 ± 0.02
Avg. accuracy	0.17 ± 0.01	0.27 ± 0.01	0.5 ± 0.01	0.47 ± 0.02	**0.58** ± 0.01	0.52 ± 0.02

References

[1] Ashman, M., So, J., Tebbutt, W., Fortuin, V., Pearce, M., and Turner, R. E. (2020). Sparse gaussian process variational autoencoders. *arXiv preprint arXiv:2010.10177*.

[2] Hälvä, H., Le Corff, S., Lehéricy, L., So, J., Zhu, Y., Gassiat, E., and Hyvarinen, A. (2021). Disentangling identifiable features from noisy data with structured nonlinear ica. *Advances in Neural Information Processing Systems*, 34:1624–1633.

[3] Ranganath, R., Gerrish, S., and Blei, D. (2014). Black Box Variational Inference. In Kaski, S. and Corander, J., editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 814–822, Reykjavik, Iceland. PMLR.