Energy-Based Processes
for Exchangeable Data

Mengjiao Yang*, Bo Dai*, Hanjun Dai, Dale Schuurmans

Google Brain

Paper: https://arxiv.org/abs/2003.07521
Code: https://github.com/google-research/google-research/tree/master/ebp
Sets

- Record data
- 3D point clouds
- Images

(x, y, R, G, B)
Sets Properties

• Exchangeability

\[
\begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
7 & 5 & 6 & 3 & 4 & 1 & 2 \\
\end{array} =
\]

Same set

• Varying cardinality

Same chair
Modeling Sets (Unconditional)

- RNNs

\[ p(x_{1:n}) = \prod_{i=1}^{n} p(x_i | x_{1:i-1}) \]

- Varying Cardinality
- Exchangeability

Larochelle, H. and Murray, I. The neural autoregressive distribution estimator. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 29–37, 2011.
Modeling Sets (Unconditional)

• Latent variable models

\( \{ x_i \} \) conditionally i.i.d.

\[
p(x_{1:n}) = \int \prod_{i=1}^{n} p(x_i | \theta)p(\theta)d\theta
\]

Known prior

✓ Varying Cardinality
✓ Exchangeability

Edwards, H. and Storkey, A. Towards a neural statistician. arXiv preprint arXiv:1606.02185, 2016
Korshunova, I., Degrave, J., Huszar, F., Gal, Y., Gretton, A., and Dambre, J. Bruno: A deep recurrent model for exchangeable data. In Advances in Neural Information Processing Systems, 2018.
Pointflow: 3d point cloud generation with continuous normalizing flows.
Modeling Sets (Unconditional)

- Latent variable models

\{x_i\} conditionally i.i.d.

\[ p(x_{1:n}) = \int \prod_{i=1}^{n} p(x_i | \theta)p(\theta)d\theta \]

- Varying Cardinality
- Exchangeability
- Flexibility

Known prior

Tractable

---

Edwards, H. and Storkey, A. Towards a neural statistician. arXiv preprint arXiv:1606.02185, 2016

Korshunova, I., Degrave, J., Huszar, F., Gal, Y., Gretton, A., and Dambre, J. Bruno: A deep recurrent model for exchangeable data. In Advances in Neural Information Processing Systems, 2018.

Pointflow: 3d point cloud generation with continuous normalizing flows.
Modeling Sets (Conditional)

• Stochastic processes
  A set of random variables: \( \{X_t; t \in \mathcal{T}\} \)
  with finite-dimensional marginal distribution: 
  \[ p(x_{t_1:t_n} \mid \{t_i\}_{i=1}^n) \]

Øksendal, B. Stochastic differential equations. In Stochastic differential equations, pp. 65–84. Springer, 2003.
Modeling Sets (Conditional)

- Stochastic processes

A set of random variables: \( \{X_t; \ t \in \mathcal{T}\} \)

with finite-dimensional marginal distribution: \( p(x_{t_1:t_n} \mid \{t_i\}_{i=1}^n) \)

✓ Consistency: \( p(x_{t_1:t_m}) = \int p(x_{t_1:t_n}) dx_{t_{m+1}:t_n} \)

✓ Exchangeability: \( p(x_{t_1:t_n}) = p(\pi(x_{t_1:t_n})) \)

Øksendal, B. Stochastic differential equations. In Stochastic differential equations, pp. 65–84. Springer, 2003.
Rasmussen, C. E. and Williams, C. K. I. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, 2006.
Shah, A., Wilson, A., and Ghahramani, Z. Student-t processes as alternatives to gaussian processes. In Artificial intelligence and statistics, pp. 877–885, 2014.
Modeling Sets (Conditional)

- Stochastic processes
  A set of random variables: \( \{X_t; t \in \mathcal{T}\} \)
  with finite-dimensional marginal distribution: \( p(x_{t_1:t_n} | \{t_i\}_{i=1}^n) \)

  ☑ Consistency: \( p(x_{t_1:t_m}) = \int p(x_{t_1:t_n}) dx_{t_{m+1}:t_n} \)

  ☑ Exchangeability: \( p(x_{t_1:t_n}) = p(\pi(x_{t_1:t_n})) \)

  ❓ Flexibility:
  - Gaussian processes: \( p(x_{t_1:t_n}) = \mathcal{N}(0,K(t_1:n) + \sigma^2 I_n) \)
  - Student-t processes: \( p(x_{t_1:t_n}) = \mathcal{N}(\nu,0,K(t_1:n) + \sigma^2 I_n) \)

Øksendal, B. Stochastic differential equations. In Stochastic differential equations, pp. 65–84. Springer, 2003.
Rasmussen, C. E. and Williams, C. K. I. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, 2006.
Shah, A., Wilson, A., and Ghahramani, Z. Student-t processes as alternatives to gaussian processes. In Artificial intelligence and statistics, pp. 877–885, 2014.
Modeling Sets (Conditional)

- Stochastic processes

Rasmussen, C. E. and Williams, C. K. I. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, 2006.
Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S., and Teh, Y. W. Neural processes. arXiv preprint arXiv:1807.01622, 2018b.
Ma, C., Li, Y., and Hernández-Lobato, J. M. Variational implicit processes. arXiv preprint arXiv:1806.02390, 2018.
Energy-Based Processes

- Stochastic processes as latent variable models

\[
p(x_{t_1:t_n}) = \int \Pi_{i=1}^{n} p(x \mid \theta, t_i)p(\theta)d\theta
\]

- Varying Cardinality
- Exchangeability
Energy-Based Processes

• Stochastic processes as latent variable models

\[ p(x_{t_1:t_n}) = \int \prod_{i=1}^{n} p(x \mid \theta, t_i)p(\theta)d\theta \]

- Varying Cardinality
- Exchangeability

• Deep energy-based models for likelihood

- Flexibility

Deep EBMs
Energy-Based Processes

• Stochastic processes as latent variable models

\[ p(x_{t_1:t_n}) = \int \prod_{i=1}^{n} p(x | \theta, t_i)p(\theta)d\theta \]

- Varying Cardinality
- Exchangeability

• Deep energy-based models for likelihood

- Flexibility

• Neural collapsed inference => unconditional EBPs

\[ p(x_{1:n}) = \int p(x_{1:n} | \theta)p(\theta)d\theta \]

Teh, Y. W., Newman, D., and Welling, M. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In Advances in Neural Information Processing Systems, Volume 19, pp. 1353–1360, 2007. ISBN 9780262195683.
Energy-Based Processes

• Learning EBPs: \( \max_w \mathbb{E}_{x_{1:n} \sim \mathcal{D}}[\log p_w(x_{1:n})] \)
Energy-Based Processes

- Learning EBPs: \( \max_w \mathbb{E}_{x_{1:n} \sim \mathcal{D}}[\log p_w(x_{1:n})] \)

? Intractable integration over \( \theta \)

\[ \log \int p_w(x_{1:n} | \theta)p(\theta)d\theta \]
Energy-Based Processes

• Learning EBPs: \( \max_w \mathbb{E}_{x_1:n \sim D}[\log p_w(x_{1:n})] \)

\( \log \int p_w(x_{1:n} | \theta)p(\theta)d\theta = \max_q \mathbb{E}_q[\log p_w(x_{1:n} | \theta)] - KL(q || p) \)

Intractable integration over \( \theta \)

\( \checkmark \) ELBO

Dai, B., Liu, Z., Dai, H., He, N., Gretton, A., Song, L., and Schuurmans, D. Exponential family estimation via adversarial dynamics embedding. arXiv preprint arXiv:1904.12083, 2019.
Energy-Based Processes

• Learning EBPs: \( \max_w \mathbb{E}_{x_{1:n} \sim \mathcal{D}}[\log p_w(x_{1:n})] \)

? Intractable integration over \( \theta \)
\[
\log \int p_w(x_{1:n} | \theta)p(\theta)d\theta = \max_q \mathbb{E}_q[\log p_w(x_{1:n} | \theta)] - KL(q || p)
\]

✓ ELBO

? Intractable partition function
\[
\log p_w(x_{1:n} | \theta) = f_w(x_{1:n}; \theta) - \log Z(f_w, \theta)
\]
Energy-Based Processes

- **Learning EBPs:** \( \max_w \mathbb{E}_{x_1:n \sim \mathcal{D}}[\log p_w(x_1:n)] \)

  - Intractable integration over \( \theta \)
    \[
    \log \int p_w(x_1:n \mid \theta)p(\theta)d\theta = \max_{q(\theta \mid x_1:n)} \mathbb{E}_q[\log p_w(x_1:n \mid \theta)] - KL(q || p)
    \]

  - Intractable partition function
    \[
    \log p_w(x_1:n \mid \theta) = f_w(x_1:n; \theta) - \log Z(f_w, \theta)
    \]
    \[
    \propto \min_{q(x_1:n, \nu \mid \theta)} f_w(x_1:n; \theta) - \mathbb{E}_q[f_w(x_1:n; \theta) - \frac{\lambda}{2} \nu^\top \nu] - H(q)
    \]

- 

  - Adversarial dynamic embeddings

---

Dai, B., Liu, Z., Dai, H., He, N., Gretton, A., Song, L., and Schuurmans, D. Exponential family estimation via adversarial dynamics embedding. arXiv preprint arXiv:1904.12083, 2019.
Energy-Based Processes

- Parametrizing EBPs:

\[ \mu, \sigma \sim \mathcal{N}(0, I) \]

\[ \epsilon \sim \mathcal{N}(0, I) \]

\[ \theta \sim q(\theta | x_{1:n}) \]

\[ x_{1:n} \]
Energy-Based Processes

• Parametrizing EBPs:

\[ \epsilon \sim \mathcal{N}(0, I) \]

\[ \mu, \sigma \]

\[ x_{1:n} \]

\[ \theta \sim q(\theta | x_{1:n}) \]

+ \hspace{1cm} \times \hspace{1cm}

\[ \hat{x}_{1:n} \sim q(x_{1:n}, \nu | \theta) \]
Energy-Based Processes

- Parametrizing EBPs:

\[ x_{1:n} \]

\[ \mu \quad \sigma \]

\[ \epsilon \sim \mathcal{N}(0, I) \]

\[ \theta \sim q(\theta | x_{1:n}) \]

\[ + \]

\[ \times \]

\[ \text{RNN/Flow} + \text{Langevin} \]

\[ \hat{x}_{1:n} \sim q(\hat{x}_{1:n}, \nu | \theta) \]

\[ f_w(x_{1:n}; \theta) \]

Energy
Applications

• Image completion

LeCun, Y. MNIST handwritten digit database, 1998. URL http://yann.lecun.com/exdb/mnist/.
Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision, pp. 3730–3738, 2015.
Applications

• Point-cloud generation

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. 3d shapenets: A deep representation for volumetric shapes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1912–1920, 2015.
## Applications

- **Point-cloud generation**

| Category | Model     | JSD (↓) | MMD (↓) CD | MMD (↓) EMD | COV (%; ↑) CD | COV (%; ↑) EMD |
|----------|-----------|---------|------------|-------------|---------------|---------------|
| **Airplane** | l-GAN     | 3.61    | 0.239      | 3.29        | 47.90         | 50.62         |
|          | PC-GAN    | 4.63    | 0.287      | 3.57        | 36.46         | 40.94         |
|          | PointFlow | 4.92    | **0.217**  | 3.24        | 46.91         | 48.40         |
|          | EBP (ours)| 3.92    | 0.240      | **3.22**    | **49.38**     | **51.60**     |
| **Chair** | l-GAN     | 2.27    | 2.46       | **7.85**    | 41.39         | 41.69         |
|          | PC-GAN    | 3.90    | 2.75       | 8.20        | 36.50         | 38.98         |
|          | PointFlow | 1.74    | **2.42**   | 7.87        | 46.83         | 46.98         |
|          | EBP (ours)| **1.53**| 2.59       | 7.92        | **47.73**     | **49.84**     |
| **Car**  | l-GAN     | 2.21    | 1.48       | 5.43        | 39.20         | 39.77         |
|          | PC-GAN    | 5.85    | 1.12       | 5.83        | 23.56         | 30.29         |
|          | PointFlow | 0.87    | **0.91**   | **5.22**    | 44.03         | 46.59         |
|          | EBP (ours)| **0.78**| 0.95       | 5.24        | **51.99**     | **51.70**     |

Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L. Learning representations and generative models for 3d point clouds. arXiv preprint arXiv:1707.02392, 2017.

Li, C.-L., Zaheer, M., Zhang, Y., Poczos, B., and Salakhutdinov, R. Point cloud gan. arXiv preprint arXiv:1810.05795, 2018.

Yang, G., Huang, X., Hao, Z., Liu, M.-Y., Belongie, S., and Hariharan, B. Pointflow: 3d point cloud generation with continuous normalizing flows. arXiv preprint arXiv:1906.12320, 2019.
Applications

• Unsupervised representation learning

| Model                                | Accuracy |
|--------------------------------------|----------|
| VConv-DAE (Sharma et al., 2016)       | 75.5     |
| 3D-GAN (Wu et al., 2016)             | 83.3     |
| l-GAN (EMD) (Achlioptas et al., 2017)| 84.0     |
| l-GAN (CD) (Achlioptas et al., 2017) | 84.5     |
| PointGrow (Sun et al., 2018)         | 85.7     |
| MRTNet-VAE (Gadelha et al., 2018)    | 86.4     |
| PointFlow (Yang et al., 2019)        | 86.8     |
| PC-GAN (Li et al., 2018)             | 87.8     |
| FoldingNet (Yang et al., 2018)       | **88.4** |
| EBP (ours)                           | **88.3** |

Sharma, A., Grau, O., and Fritz, M. Vconv-dae: Deep volumetric shape learning without object labels. In European Conference on Computer Vision, pp. 236–250. Springer, 2016.

Wu, J., Zhang, C., Xue, T., Freeman, B., and Tenenbaum, J. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In Advances in neural information processing systems, pp. 82–90, 2016.

Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L. Learning representations and generative models for 3d point shapes. arXiv preprint arXiv:1707.02392, 2017.

Sun, Y., Wang, Y., Liu, Z., Siegel, J. E., and Sarma, S. E. Pointgrow: Autoregressively learned point cloud generation with self-attention. arXiv preprint arXiv:1810.05591, 2018.

Gadelha, M., Wang, R., and Maji, S. Multiresolution tree networks for 3d point cloud processing. In Proceedings of the European Conference on Computer Vision (ECCV), pp. 103–118, 2018.

Yang, Y., Feng, C., Shen, Y., and Tian, D. Foldingnet: Point cloud auto-encoder via deep grid deformation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

Li, C.-L., Zaheer, M., Zhang, Y., Poczos, B., and Salakhutdinov, R. Point cloud gan. arXiv preprint arXiv:1810.05795, 2018.

Yang, G., Huang, X., Hao, Z., Liu, M.-Y., Belongie, S., and Hariharan, B. Pointflow: 3d point cloud generation with continuous normalizing flows. arXiv preprint arXiv:1906.12320, 2019.
Applications

• Point-cloud denoising
Summary

- Energy-based processes for flexibility set modeling
- Unifies stochastic process and latent variable perspectives
- Neural collapsed inference for learning
- State-of-the-art performance on a set of supervised and unsupervised tasks