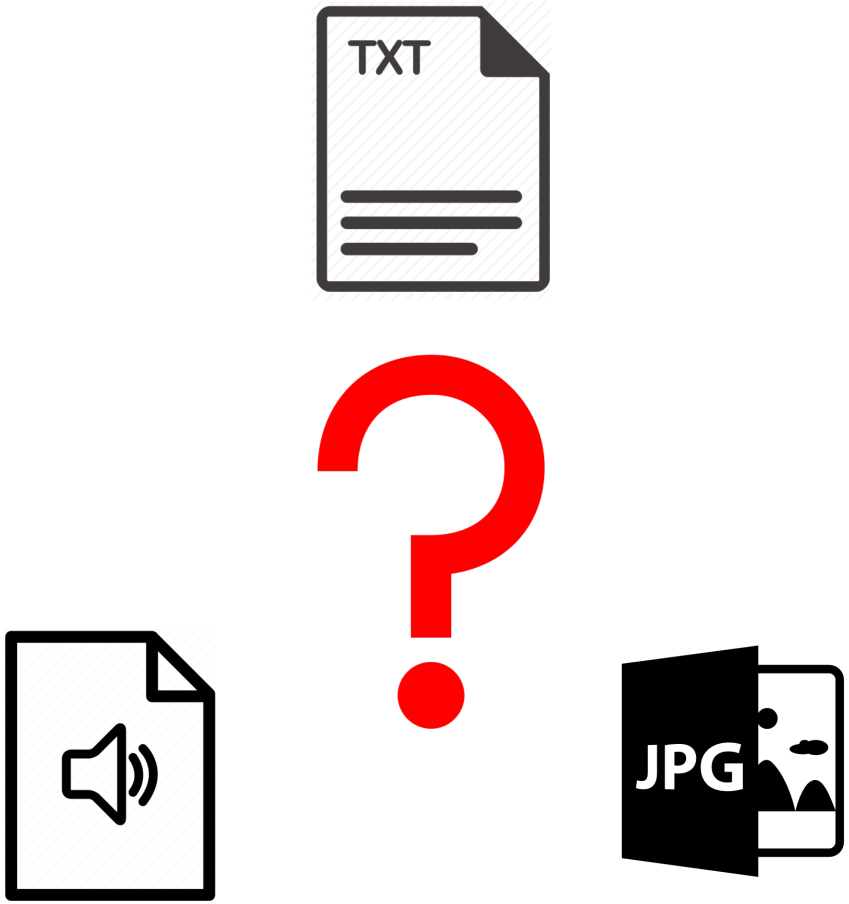
ETH zürich

medical data science

Generalized Multimodal ELBO

Thomas M. Sutter, Imant Daunhawer,
Prof. Dr. Julia E. Vogt

ICLR 2021


TXT
JPG

Divergence between true and approximate posterior

- Multimodal ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z\,|X)}[\log p_\theta(X \mid z)] - KLD\left(q_\phi(z \mid X) \middle| \, p_\theta(z)\right)$$

- Maximizing ELBO is equal to minimizing KL-divergence:

$$\arg\min_\phi KLD\left(q_\phi(z\,|X) \middle| p(z|X)\right)$$

- What if not all data types are present?

Divergence between true and approximate posterior

medical data science

- Multimodal ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z\,|X)}[\log p_\theta(X \mid z)] - KLD\left(q_\phi(z \mid X) \middle| \, p_\theta(z)\right)$$

- Maximizing ELBO is equal to minimizing KL-divergence:

$$\arg\min_\phi KLD\left(q_\phi(z\,|X) \middle| p(z|X)\right)$$

- What if not all data types are present?

Divergence between true and approximate posterior II

- Optimization given all data types:

$$\arg\min_{\phi} KLD\big(q_{\phi}(z\,|X)\big|p(z|X)\big)$$

- Optimization given a subset X_k of all data types X:

$$\arg\min_{\phi} KLD\big(q_{\phi}(z \mid X_k)\big|p(z|X)\big)$$

Divergence between true and approximate posterior II

medical data science

- Optimization given all data types:

$$\arg\min_{\phi} KLD\big(q_{\phi}(z\,|X)\big|p(z|X)\big)$$

- Optimization given a subset X_k of all data types X:

$$\arg\min_{\phi} KLD\big(q_{\phi}(z\mid X_k)\big|p(z|X)\big)$$

Convex Combination of KLDs

medical data science

- Optimization given all data types:

$$\arg\min_{\phi} KLD\left(q_{\phi}(z \mid X) \middle| p(z|X)\right)$$

- Joint optimization of the powerset $\mathcal{P}(X)$ of all subsets X_k:

$$\arg\min_{\phi} \sum_{X_k \in \mathcal{P}(X)} KLD\left(q_{\phi}(z \mid X_k) \middle| p(z|X)\right)$$

- **Lemma 1:**
  The convex combination of KLDs over the powerset $\mathcal{P}(X)$ describes the joint log-probability $\log p_{\theta}(X)$.

MoPoE: Mixture-of-Products-of-Experts

medical data science

- ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z\,|X)}[\log p_\theta(X \mid z)] - KLD\left(\frac{1}{2^M}\sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) \,\middle|\, p_\theta(z)\right)$$

where $\tilde{q}_\phi(z|X_k) = \mathrm{PoE}\left(\{q_j(z \mid x_j)\ \forall\, x_j\ \in\ X_k\}\right)$ [1]

and $q_\phi(z|X) = \frac{1}{2^M}\sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) = MoE\left(\tilde{q}_\phi(z|X_k)\ \forall\, X_k\ \in\ \mathcal{P}(X)\right)$ [2]

[1] Wu and Goodman: Multimodal Generative Models for Scalable Weakly-Supervised Learning, NeurIPS 2018

[2] Shi et al: Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models, NeurIPS 2019

MoPoE: Mixture-of-Products-of-Experts

- ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z\,|X)}[\log p_\theta(X \mid z)] - KLD\left(\frac{1}{2^M}\sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) \,\middle|\, p_\theta(z)\right)$$

where $\tilde{q}_\phi(z|X_k) = \mathrm{PoE}(\{q_j(z \mid x_j)\ \forall\, x_j \ \in\ X_k\})$ [1]

and $q_\phi(z|X) = \frac{1}{2^M}\sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) = MoE\left(\tilde{q}_\phi(z|X_k)\ \forall\, X_k \ \in\ \mathcal{P}(X)\right)$ [2]

[1] Wu and Goodman: Multimodal Generative Models for Scalable Weakly-Supervised Learning, NeurIPS 2018

[2] Shi et al: Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models, NeurIPS 2019

MoPoE: Mixture-of-Products-of-Experts

medical data science

- ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z \mid X)}[\log p_\theta(X \mid z)] - KLD\left(\frac{1}{2^M} \sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) \,\middle|\, p_\theta(z)\right)$$

where $\tilde{q}_\phi(z|X_k) = \mathrm{PoE}\left(\{q_j(z \mid x_j) \; \forall \, x_j \; \in \; X_k\}\right)$ [1]

and $q_\phi(z|X) = \frac{1}{2^M} \sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) = MoE\left(\tilde{q}_\phi(z|X_k) \; \forall \, X_k \; \in \; \mathcal{P}(X)\right)$ [2]

[1] Wu and Goodman: Multimodal Generative Models for Scalable Weakly-Supervised Learning, NeurIPS 2018

[2] Shi et al: Variational Mixture-of-Experts Autoencoders for Multi-Modal Deep Generative Models, NeurIPS 2019

MoPoE: Generalized ELBO

- ELBO:

$$\log p_\theta(X) \geq E_{q_\phi(z\,|X)}[\log p_\theta(X \mid z)] - KLD\left(\frac{1}{2^M} \sum_{X_k \in \mathcal{P}(X)} \tilde{q}_\phi(z|X_k) \,\middle|\, p_\theta(z)\right)$$

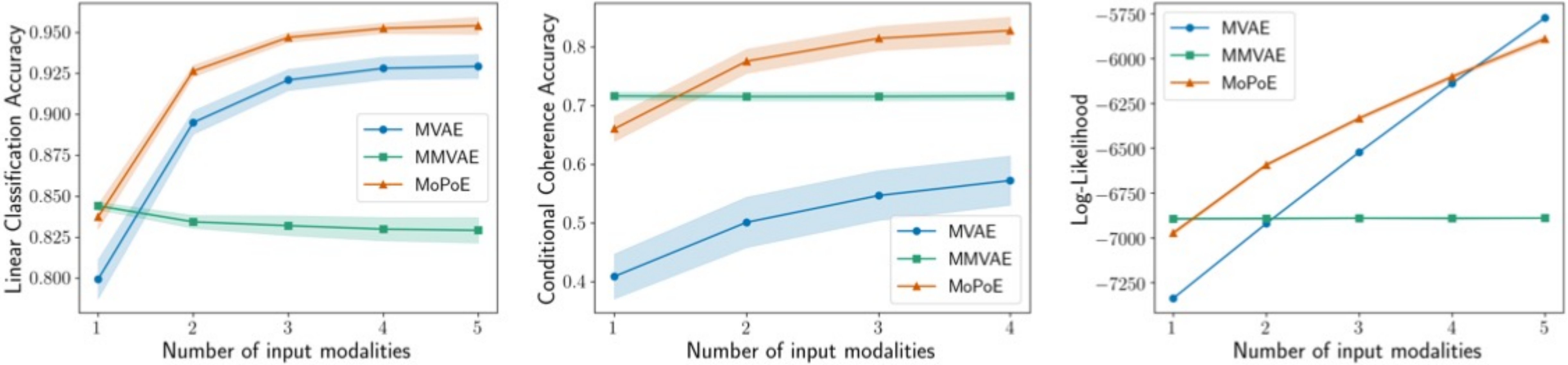
- **Lemma 2:**
  The Generalized ELBO is a multimodal ELBO which approximates the log joint probability $\boldsymbol{\log p_\theta(X)}$.

- **Lemma 3:**
  Maximizing the Generalized ELBO minimizes the convex combination of KL-divergences shown before.

Experiments

- MNIST-SVHN-Text: 2 image modalities and 1 text modality
- **polyMNIST**: 5 image modalities
- Bimodal CelebA: 1 image modality and 1 text modality

Experiments: polyMNIST



- X-axis: number of input modalities
- Y-axis: Evaluation metric
- Experimental results provide empirical evidence for the insights from analysis of ELBO

Conclusion

- New ELBO formulation: based on a convex combination of KL-divergences
- New formulation generalizes previous works
- Superior experimental results: proposed method outperforms or reaches state-of-the-art performance
- Experimental results can be related to objective of models and their choice of posterior approximation function

Come visit us in Poster Session 4!