Bayesian Domain Adaptation with Gaussian Mixture Domain-Indexing

Yanfang Ling, Jiyong Li, Lingbo Li, Shangsong Liang*

Speaker**: Yanfang Ling**
Sun Yat-sen University, China
lingyf3@mail2.sysu.edu.cn

Outlines

- Background and Motivation
- Methodology
    - Mixture of Domain Index Distributions
    - Generative Process of GMDI
    - Variational Inference
- Theory
- Experimental Results
- Conclusion

Outlines

- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

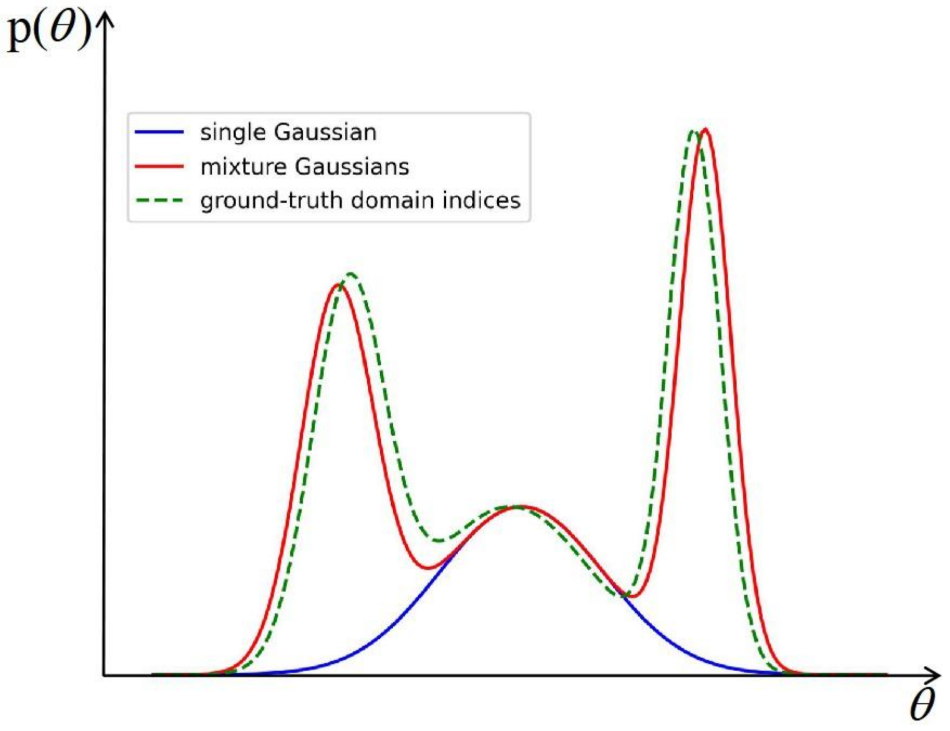
Background and Motivation

- **Limitation of the domain indices space.**

  - A single Gaussian distribution struggles to adequately fit the domain indices, neglecting the inherent structures among different domains.

  - Motivate us to model domain indices as a **Gaussian mixture distribution**, with the number of mixture components dynamically determined by a **Chinese Restaurant Process**.

  - The mixtures of distributions provide a higher level of flexibility in a larger latent space, thereby increasing the capability to adapt to various target domains with domain shift.



Outlines

- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

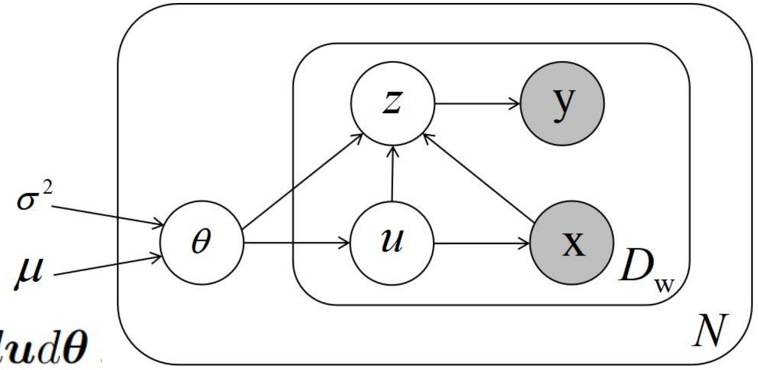
Methodology

- **Problem Formulation**
  - Unsupervised domain adaptation:
    - Given N domains with domain identity, and each domain contains D_w data points;
    - N domains are divided into source domains with labeled data and target domains with unlabeled data.
  - We aim to:
    - Predict the label $\mathbf{y}$ of target domain data;
    - Infer local domain index u and global domain index θ.

Outlines

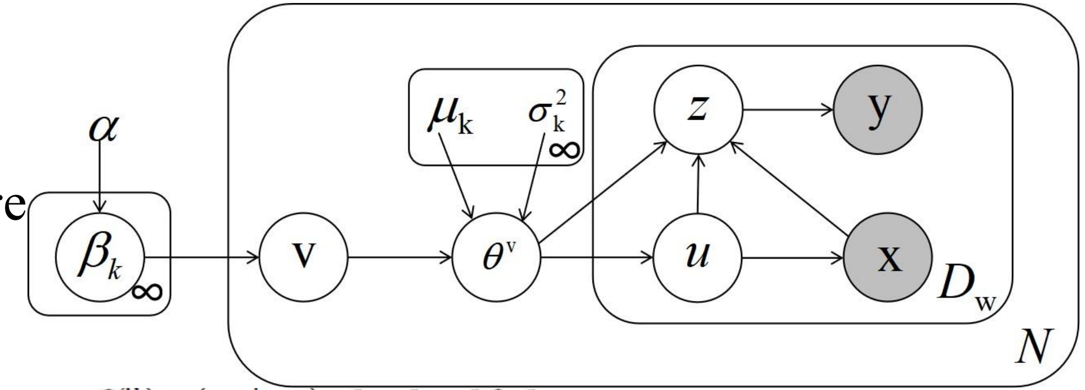
Methodology

- **Mixture of Domain Index Distributions**
  - VDI:
    - Global domain index θ follow certain distribution.



$$p(y, \boldsymbol{x} \mid \boldsymbol{\varepsilon}) = \int p(\boldsymbol{\theta} \mid \boldsymbol{\varepsilon}) p(\boldsymbol{u} \mid \boldsymbol{\theta}) p(\boldsymbol{x} \mid \boldsymbol{u}) p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}) p(y \mid \boldsymbol{z}) \, d\boldsymbol{z} d\boldsymbol{u} d\boldsymbol{\theta}$$

  - GMDI:
    - Global domain index θ follow a GaussianMixture Model (GMM).



$$p(y, \boldsymbol{x} \mid \boldsymbol{\varepsilon}) = \int p(v) p(\boldsymbol{\theta}^v \mid \boldsymbol{\varepsilon}) p(\boldsymbol{u} \mid \boldsymbol{\theta}^v) p(\boldsymbol{x} \mid \boldsymbol{u}) p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v) p(y \mid \boldsymbol{z}) \, d\boldsymbol{z} d\boldsymbol{u} d\boldsymbol{\theta} dv$$

Outlines

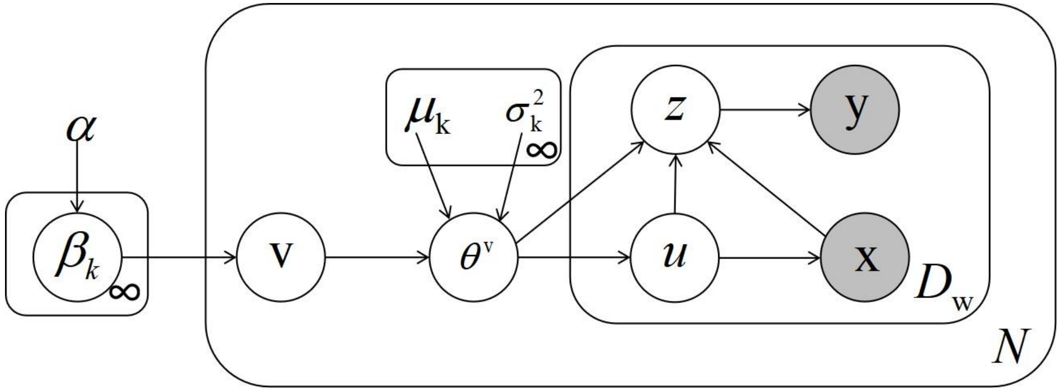
- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

Methodology

- **Generative Process of GMDI**

    - Chinese Restaurant Process (CRP) prior on **v**:

$$P(v=k)=\begin{cases}\dfrac{n_k}{N-1+\alpha} & \text{if the cluster } k \text{ exists}\,,\\ \dfrac{\alpha}{N-1+\alpha} & \text{if cluster } k \text{ is a new cluster}\,,\end{cases}$$



    - Generative process of GMDI:

$$\begin{aligned} v \mid \boldsymbol{\pi} &\sim \mathrm{Categorical}_{\infty}(\boldsymbol{\pi})\,,\\ \boldsymbol{\theta}^{v=k} &\sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\sigma}_k^2)\,,\\ \boldsymbol{u} \mid \boldsymbol{\theta}^{v=k} &\sim p(\boldsymbol{u} \mid \boldsymbol{\theta}^{v=k})\,,\\ \boldsymbol{x} \mid \boldsymbol{u} &\sim p(\boldsymbol{x} \mid \boldsymbol{u})\,,\\ \boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v &\sim p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v)\,, \end{aligned}$$

Outlines

- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

Methodology

- **Variational Inference**

  - Variational Posterior:

$$q(\boldsymbol{u}, \boldsymbol{\theta}, \boldsymbol{z}, v, \boldsymbol{\beta} \mid \boldsymbol{x}) = q(\boldsymbol{\beta}; \gamma) q(v; \boldsymbol{\eta}) q(\boldsymbol{u} \mid \boldsymbol{x}; \boldsymbol{\psi}_u) q(\boldsymbol{\theta}^v \mid \boldsymbol{u}; \boldsymbol{\psi}_\theta) q(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v; \boldsymbol{\psi}_z)$$

  - Evidence Lower Bound (ELBO):

$$\begin{aligned}
\mathcal{L}_{\text{ELBO}} = & \ \mathbb{E}_{q(\boldsymbol{u}, \boldsymbol{\theta}^v, \boldsymbol{z} | \boldsymbol{x}; \boldsymbol{\phi}) q(v; \boldsymbol{\eta})} [\log p(y | \boldsymbol{z})] + \mathbb{E}_{q(\boldsymbol{u} | \boldsymbol{x}; \boldsymbol{\psi}_u)} [\log p(\boldsymbol{x} | \boldsymbol{u})] \\
& + \mathbb{E}_{q(v; \boldsymbol{\eta}) q(\boldsymbol{\beta}; \boldsymbol{\gamma}) q(\boldsymbol{u} | \boldsymbol{x}; \boldsymbol{\psi}_u) q(\boldsymbol{\theta}^v | \boldsymbol{u}; \boldsymbol{\psi}_\theta)} [\log p(\boldsymbol{u} | \boldsymbol{\theta}^v)] - \mathrm{KL}[q(\boldsymbol{\beta}; \boldsymbol{\gamma}) || p(\boldsymbol{\beta})] \\
& - \mathbb{E}_{q(\boldsymbol{\beta}; \boldsymbol{\gamma})} [\mathrm{KL}[q(v; \boldsymbol{\eta}) || p(v | \boldsymbol{\beta}; \boldsymbol{\psi}_v)]] - \mathbb{E}_{q(\boldsymbol{u} | \boldsymbol{x}; \boldsymbol{\psi}_u) q(v; \boldsymbol{\eta})} [\mathrm{KL}[q(\boldsymbol{\theta}^v | \boldsymbol{u}; \boldsymbol{\psi}_\theta) || p(\boldsymbol{\theta}^v)]] \\
& - \mathbb{E}_{q(v; \boldsymbol{\eta}) q(\boldsymbol{u}, \boldsymbol{\theta}^v | \boldsymbol{x}; \boldsymbol{\xi})} [\mathrm{KL}[q(\boldsymbol{z} | \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v; \boldsymbol{\psi}_z) || p(\boldsymbol{z} | \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{\theta}^v)]] \\
& - \mathbb{E}_{q(\boldsymbol{u} | \boldsymbol{x}; \boldsymbol{\psi}_u)} [\log q(\boldsymbol{u} | \boldsymbol{x}; \boldsymbol{\psi}_u)] ,
\end{aligned}$$

Outlines

- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

Theory

- **Final Objective Function:**

$$\mathcal{L}_{\mathrm{GMDI}} = \max \min_{D} \mathcal{L}_{\mathrm{ELBO}} - \lambda * \mathcal{L}_{\mathrm{D}}$$

- **Theorem(optimum):**

$$\textit{(1)} I(\boldsymbol{z};\boldsymbol{\theta}) \;=\; I(\boldsymbol{z};w|\boldsymbol{\theta}) \;=\; 0,$$

$$\textit{(2)} I(y;\boldsymbol{z}) \;\; \textit{and} \;\; I(\boldsymbol{x};\boldsymbol{u},\boldsymbol{\theta},\boldsymbol{z},v) \;\; \textit{are maximized,}$$

$$\textit{(3)} \mathrm{KL}[q(\boldsymbol{u},\boldsymbol{\theta},v,\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{u},\boldsymbol{\theta},v,\boldsymbol{z})] \;=\; 0 \;\; \textit{and} \;\; \mathrm{KL}[q(\boldsymbol{x}|\boldsymbol{u},\boldsymbol{\theta},v,\boldsymbol{z})||p(\boldsymbol{x}|\boldsymbol{u},\boldsymbol{\theta},v,\boldsymbol{z})] = 0.$$

- **Tighter Upper Bound of the Objective:**

$$\mathcal{L}_{\mathrm{VDI-ELBO}} \le \mathcal{L}_{\mathrm{ELBO}} \le \log p(\boldsymbol{x},y) \;\textit{and}\; \mathcal{L}_{\mathrm{VDI}} \le \mathcal{L}_{\mathrm{GMDI}}.$$

Outlines

- Background and Motivation
- Methodology
  - Mixture of Domain Index Distributions
  - Generative Process of GMDI
  - Variational Inference
- Theory
- Experimental Results
- Conclusion

Experimental Results

- Accuracy on binary classification tasks (Circle, DG-15, and DG-60) and 4-way classification task (CompCars):

	Method								
Dataset	Source-only	DANN	ADDA	CDANN	MDD	SENTRY	D2V	VDI	GMDI (Ours)
Circle	55.5	53.4	56.2	54.9	53.4	59.5	60.1	94.3	**96.9**
DG-15	39.7	43.3	33.5	38.8	37.2	42.6	79.9	94.7	**96.5**
DG-60	55.0	66.3	60.8	65.3	54.6	51.3	82.1	95.9	**99.3**
CompCars	39.1	38.9	42.8	41.8	41.4	41.8	40.7	42.5[2]	**44.4**

2: Reproduced result from VDI.

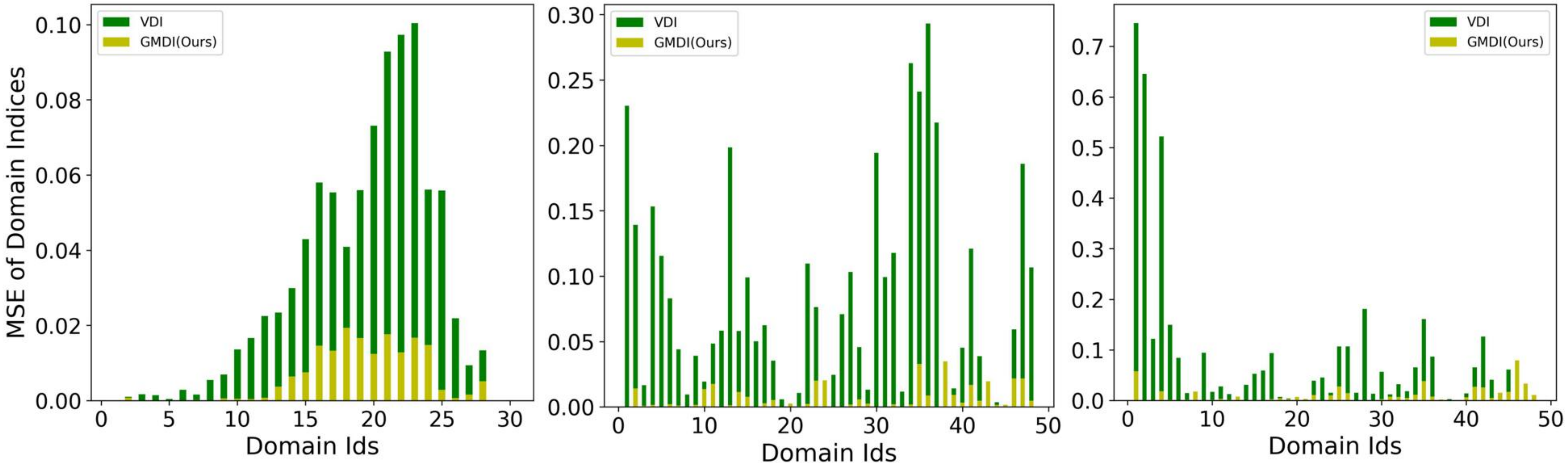
Experimental Results

- MSE for various DA methods in both tasks W (6) → E (42) and N (24) → S (24) on TPT-48. We report the average MSE of all domains as well as more detailed average MSE of level-1, level-2, level-3 target domains, respectively:

Task	Domain	Source-only	DANN	ADDA	CDANN	MDD	SENTRY	VDI	GMDI(Ours)
W (6) → E (42)	Average of 4 level-1 domains	**1.184**	1.984	5.448	6.168	5.544	2.515	2.160	1.346
	Average of 6 level-2 domains	3.128	5.112	7.624	7.016	7.912	5.136	3.000	**2.393**
	Average of 32 level-3 domains	5.272	5.880	7.256	6.986	8.008	5.872	2.448	**2.122**
	Average of all 42 domains	4.576	5.400	7.136	6.896	7.76	5.456	2.496	**2.087**
N (24) → S (24)	Average of 10 level-1 domains	1.648	1.832	5.872	1.832	2.736	3.976	1.536	**1.479**
	Average of 6 level-2 domains	3.128	3.296	6.888	2.856	6.144	3.760	2.584	**2.119**
	Average of 8 level-3 domains	9.280	6.744	7.088	7.688	10.608	**3.672**	5.624	3.942
	Average of all 24 domains	4.560	3.840	6.528	4.040	6.216	3.816	3.160	**2.493**

Experimental Results

- MSE of domain indices on Circle and TPT-48(N→S, W→E) datasets:



Outlines

- Background and Motivation
- Methodology
    - Mixture of Domain Index Distributions
    - Generative Process of GMDI
    - Variational Inference
- Theory
- Experimental Results
- Conclusion

Conclusion

- GMDI, a novel Gaussian Mixture Domain-Indexing algorithm, to address the challenge of **inferring domain indices when they are unavailable**.

- GMDI is the first one to utilize a **mixture of dynamic Gaussians**. The number of mixture components is determined adaptively by the **Chinese Restaurant Process**, enhancing the flexibility and effectiveness of domain adaptation.

- Our theoretical analysis confirms that GMDI achieves a more stringent evidence lower bound, closer to the log-likelihood.

- Extensive experiments validate the effectiveness of GMDI in inferring domain indices and highlight its potential practical applications

Thank you!

Code