Gaussian Mixture Variational Autoencoder with Contrastive Learning for Multi-Label Classification

Junwen Bai, Shufeng Kong, Carla Gomes

Cornell University

Motivations

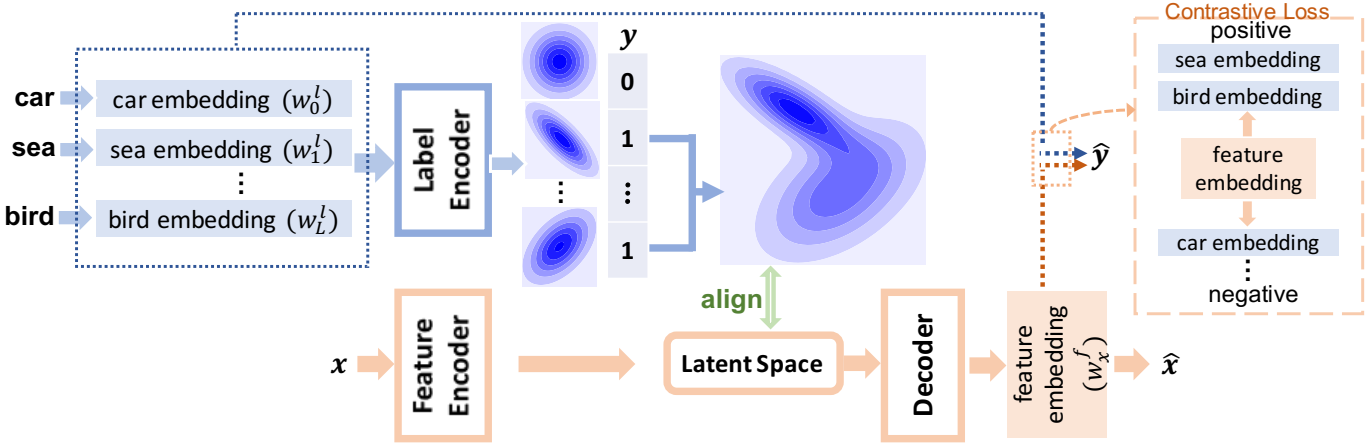
- Learn a deep latent space shared by features and labels
  - Deterministic
  - Unimodal Gaussian
  - Gaussian mixture (ours)

- Label correlation modeling
  - Pairwise ranking loss
  - Covariance matrix
  - Graph Neural Nets
  - Contrastive loss (ours)

Background

- Given a dataset of N samples (x, y)
  - x is feature
  - y is binary coding indicating labels
  - Goal: find a mapping from x to y

C-GMVAE



- Contrastive learning boosted Gaussian mixture variational autoencoder
  - A Gaussian mixture latent space
  - Contrastive learning for feature and label embeddings

- w_i^l is the label embedding
- w_i^f is the feature embedding

C-GMVAE

- Gaussian mixture latent space
  - Every label category is mapped to a learnable embedding
  - The label set selects the positive latent spaces and forms a Gaussian mixture prior

- Contrastive learning module
  - Anchor: feature embedding w_i^f
  - Pos/Neg samples: label embeddings w_i^l
  - If two label embeddings co-appear often as positive samples, they would implicitly become similar

C-GMVAE

- Supervised cross-entropy loss
  - $L_{CE} = \sum_{i=1}^{L} y_i \log s\left(w_x^f w_i^l\right) + (1 - y_i) \log(1 - s\left(w_x^f w_i^l\right))$

- Objective function
  - $L = L_{KL} + L_{recon} + \alpha L_{CL} + \beta L_{CE}$

Experiments

- Nine Datasets
  - Image: mirflickr, nuswide, scene
  - Biology: sider, yeast, eBird
  - Text: reuters, bookmarks, delicious
- # samples: 1427 ~ 270k
- # labels: 6 ~ 983

- Metrics
  - example-F1, micro-F1, macro-F1, hamming acc, precision@1

Experiments

Metric	example-F1									micro-F1								
Dataset	*eBird*	*mirflickr*	*nus-vec*	*yeast*	*scene*	*sider*	*reuters*	*bkms*	*delicious*	*eBird*	*mirflickr*	*nus-vec*	*yeast*	*scene*	*sider*	*reuters*	*bkms*	*delicious*
BR	0.365	0.325	0.343	0.630	0.606	0.766	0.733	0.171	0.174	0.384	0.371	0.371	0.655	0.706	0.796	0.767	0.125	0.197
MLKNN	0.510	0.383	0.342	0.618	0.691	0.738	0.703	0.213	0.259	0.557	0.415	0.368	0.625	0.667	0.772	0.680	0.181	0.264
HARAM	0.510	0.432	0.396	0.629	0.717	0.722	0.711	0.216	0.267	0.573	0.447	0.415	0.635	0.693	0.754	0.695	0.230	0.273
SLEEC	0.258	0.416	0.431	0.643	0.718	0.581	0.885	0.363	0.308	0.412	0.413	0.428	0.653	0.699	0.697	0.845	0.300	0.333
C2AE	0.501	0.501	0.435	0.614	0.698	0.768	0.818	0.309	0.326	0.546	0.545	0.472	0.626	0.713	0.798	0.799	0.316	0.348
LaMP	0.477	0.492	0.376	0.624	0.728	0.766	0.906	0.389	0.372	0.517	0.535	0.472	0.641	0.716	0.797	0.886	0.373	0.386
MPVAE	0.551	0.514	0.468	0.648	0.751	0.769	0.893	0.382	0.373	0.593	0.552	0.492	0.655	0.742	0.800	0.881	0.375	0.393
ASL	0.528	0.477	0.468	0.613	0.770	0.752	0.880	0.373	0.359	0.580	0.525	0.495	0.637	0.753	0.795	0.869	0.354	0.387
RBCC	0.503	0.468	0.466	0.605	0.758	0.733	0.857	-	-	0.558	0.513	0.490	0.623	0.749	0.784	0.825	-	-
C-GMVAE	**0.576**	**0.534**	**0.481**	**0.656**	**0.777**	**0.771**	**0.917**	**0.392**	**0.381**	**0.633**	**0.575**	**0.510**	**0.665**	**0.762**	**0.803**	**0.890**	**0.377**	**0.403**
std (±)	0.001	0.002	0.000	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.000	0.002	0.002	0.000	0.001	0.001	0.002

- On ex-F1, C-GMVAE improves over ASL by 5.3%, RBCC by 7.7%, MPVAE by 2.5%, and LaMP by 8.8%
- On mi-F1, C-GMVAE improves over ASL by 4.4%, RBCC by 6.7%, MPVAE by 2.4% and LaMP by 6.1%

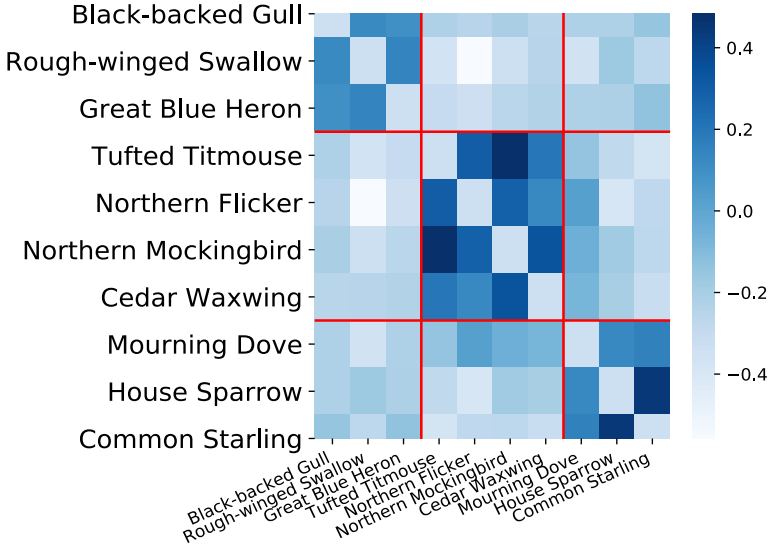
Experiments

Metric	macro-F1									Hamming Accuracy								
Dataset	*eBird*	*mirflickr*	*nus-vec*	*yeast*	*scene*	*sider*	*reuters*	*bkms*	*delicious*	*eBird*	*mirflickr*	*nus-vec*	*yeast*	*scene*	*sider*	*reuters*	*bkms*	*delicious*
BR	0.116	0.182	0.083	0.373	0.704	0.588	0.137	0.038	0.066	0.816	0.886	0.971	0.782	0.901	0.747	0.994	0.990	0.982
MLKNN	0.338	0.266	0.086	0.472	0.693	0.667	0.066	0.041	0.053	0.827	0.877	0.971	0.784	0.863	0.715	0.992	0.991	0.981
HARAM	0.474	0.284	0.157	0.448	0.713	0.649	0.100	0.140	0.074	0.819	0.634	0.971	0.744	0.902	0.650	0.905	0.990	0.981
SLEEC	0.363	0.364	0.135	0.425	0.699	0.592	0.403	0.195	0.142	0.816	0.870	0.971	0.782	0.894	0.675	0.996	0.989	0.982
C2AE	0.426	0.393	0.174	0.427	0.728	0.667	0.363	0.232	0.102	0.771	0.897	0.973	0.764	0.893	0.749	0.995	0.991	0.981
LaMP	0.381	0.387	0.203	0.480	0.745	0.668	0.520	0.286	0.196	0.811	0.897	0.980	0.786	0.903	0.751	0.997	0.992	0.982
MPVAE	0.494	0.422	0.211	0.482	0.750	0.690	0.545	0.285	0.181	0.829	0.898	0.980	0.792	0.909	0.755	0.997	0.991	0.982
ASL	0.467	0.410	0.208	0.484	0.765	0.668	0.563	0.264	0.183	0.831	0.893	0.975	0.796	0.912	0.759	0.997	0.991	0.982
RBCC	0.443	0.409	0.202	0.480	0.753	0.654	0.503	-	-	0.815	0.888	0.975	0.793	0.904	0.753	0.997	-	-
C-GMVAE	**0.538**	**0.440**	**0.226**	**0.487**	**0.769**	**0.691**	**0.582**	**0.291**	**0.197**	**0.847**	**0.903**	**0.984**	**0.796**	**0.915**	**0.767**	**0.997**	**0.992**	**0.983**
std (±)	0.000	0.001	0.001	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.002	0.001	0.003	0.000	0.000	0.000

- On ma-F1, the improvements are as large as 6.1%, 9.4%, 4.1% and 11%

Interpretability

- C-GMVAE also facilitates the model interpretability
- The heatmap matrix clearly forms three blocks on the diagonal
  - 1st block: water birds living near sea or lake
  - 2nd block: forest birds
  - 3rd block: commonly seen residential birds



Take-aways

- C-GMVAE is a novel method for multi-label prediction
- Combine Gaussian mixture latent space and contrastive learning
- Provide interpretable insights

Q & A

- Thanks for listening!