Label Embedding with Partial Heterogeneous Contexts

Yaxin Shi[†], Donna Xu[†], Yuangang Pan[†], Ivor W. Tsang[†], Shirui Pan[†]

[†]Centre for Artificial Intelligence(CAI),
University of Technology Sydney, Australia

March, 2019

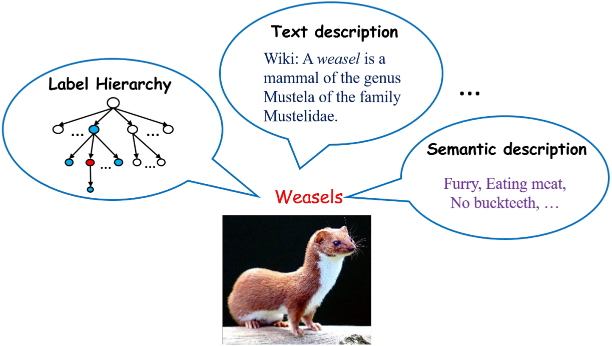
Outline

Introduction

Background: Label Embedding with Multiple Contexts

Background

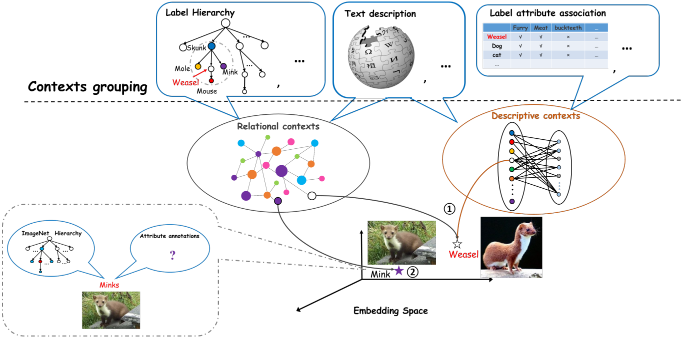
- Label embedding plays a key role in many real-world applications.
- Multiple contexts were adopted to learn label embeddings.



To learn label embeddings that capture the label relatedness in multiple aspects.

Challenges:

- The incorporated contexts are heterogeneous.
- Some contexts are partially observed in practical tasks.



How to exploit the label relatedness conveyed in partial heterogeneous contexts?

Motivation:Partial heterogeneous Contexts Label Embedding

Key problems:

- How to effectively exploit the label relatedness in each context?
- How to align the label relatedness conveyed in different contexts?
- How to overcome the partial context problem?

Related works:

Table: Comparison between existing methods.

Methods	Problem1	Problem2	Problem3
CNC [1]	✗	✗	✓
CCA [2]	✗	✓	✗
SNE [3]	✗	✓	✗
TADW [4]	✓	✓	✗
PHCLE	✓	✓	✓

Partial Heterogeneous Context Label Embedding (PHCLE)

Partial Heterogeneous Context Label Embedding (PHCLE)

Technique1: Tailor-make formulas for heterogeneous contexts.

- **Relational context:** D
  - Property: Convey direct information on label relatedness;
  - Formulation: SGNS as Explicit Matrix Factorization (EMF) [5];

$$\begin{aligned}&\min_{C,W} EMF(D, C^T W)\\ &\qquad = -tr(D^T C^T W) + \sum_{w \in V_W} \log\Big(\sum_{d'_w \in \mathcal{S}_w} e^{d'^T_w C^T w}\Big),\end{aligned}$$

  - Specification: Label hierarchy, which conveys the intrinsic structure of the labels, are exploited to construct D.
- **Descriptive context:** A
  - Property: Convey the label-description associations;
  - Formulation: Traditional matrix factorization;

$$\|A - W^T U\|_F^2$$

Technique2: Align partially observed contexts via shared embedding.

- **Formulation**

$$\min_{C,W,U} EMF(D, C^T W) + \frac{\lambda_1}{2} \| I \odot (A - W^T U) \|_F^2$$

where W is the shared label embedding matrix, C denotes the context embedding and U is the description embedding.

  - Remark:

    - Context alignment: the shared label embedding variable W;

    - Partial contexts: the indicator matrix I for the missing entries;

Technique3: Enhance alignment via discriminative contexts selection.

- **Formulation**

$$\min_{C,W,U} EMF(D, C^T W) + \frac{\lambda_1}{2} \| I \odot (A - W^T U) \|_F^2 + \lambda_2 \|U\|_1 + \frac{\lambda_3}{2} (\|W\|_F^2 + \|U\|_F^2)$$

PHCLE addresses the aforementioned key problems by:

1) Tailor-making formulas for these contexts.
2) Aligning different contexts via the shared embedding.
3) Enhancing the alignment via discriminative contexts selection.

Model Comparison and Generalization

Model Comparison

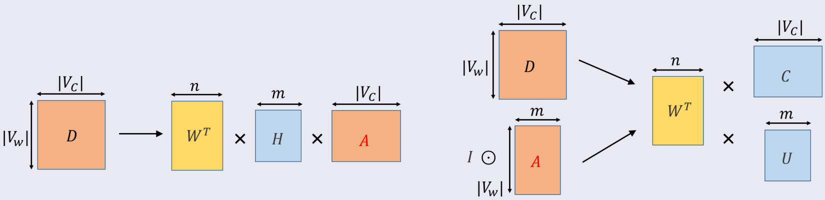
Attributed graph embedding methods

- **TADW**

$$\min_{W,H} \|D - W^T H A\|_F^2 + \frac{\lambda}{2}(\|W\|_F^2 + \|H\|_F^2)$$

The label embedding is obtained with $[W, HA]$.

- **Comparison**



Model Generalization

Generalized PHCLE

- **Reason:** Additive property of the matrix factorization formulation.
- **Formulation**

$$\min_{C^{(i)},W,U^{(j)}} \sum_{i=1}^{n} \alpha^{(i)} EMF(D^{(i)}, C^{(i)T} W) + \sum_{j=1}^{m} \beta^{(j)} ||A^{(j)} - W^T U^{(j)}||_F^2 + \Omega(W, C^{(i)}, U^{(j)})$$

$$s.t. \quad \alpha^{(i)} \geq 0, \alpha^{\mathbf{T}} \mathbf{1}_m = 1, \beta^{\mathbf{T}} \mathbf{1}_m = 1$$

where $\alpha = [\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(i)}]$ controls the weight for relational contexts, $\beta = [\beta^{(1)}, \beta^{(2)}, ..., \beta^{(j)}]$ controls the weight for descriptive contexts, $\Omega(W, C^{(i)}, U^{(j)})$ represents the regularizers..

Optimization

Algorithms

- Gradient descent based alternating minimization.
  - Due to the sparsity constraint, U is updated via FISTA.

Algorithm 1: Alternating Minimization for PHCLE

Input: co-occurrence matrix D, label-attribute association matrix A, step-size η, iterations number K, trade-off factors λ_1, λ_2, λ_3 and λ_4

Output: C_K, W_K, U_K

Initialize C_0, W_0, V_0 to the matrix with all elements equal to 1;

while $i < K$ **do**

  repeat $C_i = C_{i-1} - \eta \frac{\partial L}{\partial C}$ (See Eq. (7))

  repeat $W_i = W_{i-1} - \eta \frac{\partial L}{\partial W}$ (See Eq. (8))

  Update U using FISTA (Algorithm 2)

  $i = i + 1$;

end

Algorithm 2: FISTA with fixed step size for updating U

Input: A: the label-description association matrix;

  $I \odot W_{i-1}$: label embedding obtained in last iteration;

  U_{i-1}: description embedding obtained in last iteration;

  L: the Lipschitz constant of $\bigtriangledown f$;

  $MaxIter$: maximum iteration of the algorithm

Output: The optimal solution of $\hat{U}$

Initialize $\hat{U}_0 = U_{i-1}, Z_1 = \hat{U}_0, t_1 = 1$;

while $j < MaxIter$ **do**

  $\hat{U}_j = p_\tau(Z_j)$;

  $t_{j+1} = \frac{1+\sqrt{1+4t_j^2}}{2}$;

  $Z_{j+1} = \hat{U}_j + (\frac{t_j - 1}{t_{j+1}})(\hat{U}_j - \hat{U}_{j-1})$;

  $j = j + 1$;

  if $\frac{F(\hat{U}) - F(\hat{U})}{F(\hat{U})} < \epsilon$ break;

end

Experiments

Experimental Setups

Setup for PHCLE - partial heterogeneous contexts

- **Labels:** the 1000 labels of ImageNet 2012 dataset;
- **Relational contexts:** the label hierarchy in WordNet;
- **Descriptive contexts:** available attributes in AWA and aPY datasets.

Baselines

- **Single context embeddings:** ALE, WLE, HLE;
- **Multi-contexts embeddings:** CNC, CCA, TADW;
- **PHCLE variants:** PHCLE_FC, PHCLE_NoSp;

Tasks

- Zero-shot image classification;
- Label similarity and interpretability analysis;
- Novel image understanding.

Zero-shot image classification

Table: The results of classification accuracy on AWA and aPY.

Methods	Datasets					
	AWA			aPY		
	ESZSL	ConSE	SJE	ESZSL	ConSE	SJE
ALE	63.66	50.83	61.48	52.60	**52.87**	49.75
WLE	56.98	45.92	38.63	45.53	39.13	38.04
HLE	53.76	58.08	54.18	49.79	43.35	**52.18**
CNC	63.66	57.97	74.27	48.71	50.78	33.34
CCA	47.05	35.46	47.05	41.94	34.34	39.45
PHCLE	**69.11**	**58.43**	**77.47**	**54.61**	**52.87**	50.91

Table: Result comparison for label embeddings with full contexts.

Methods	Datasets					
	AWA			aPY		
	ESZSL	ConSE	SJE	ESZSL	ConSE	SJE
TADW	50.18	56.62	40.84	28.13	25.21	23.48
PHCLE_FC	**69.08**	**56.77**	**53.33**	**50.56**	**53.04**	**42.52**

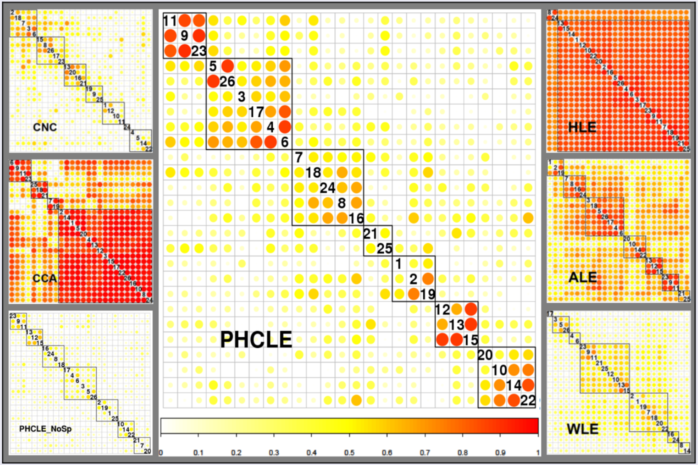
Label similarity and interpretability analysis

Label retrieval

Table: Label retrieval results. The labels are listed in descending order. Highly relevant labels are highlighted in bold, weakly relevant labels are in normal font, and irrelevant labels are in italics.

Query label	PHCLE	WLE	HLE
coffeepot	**teapot**	chiffonier	**cauldron**
	cauldron	*fire screen*	**teapot**
	beaker	washbasin	barrel
	vase	chocolate sauce	bathtub
	coffee mug	*window shade*	bucket

Clustering visualization

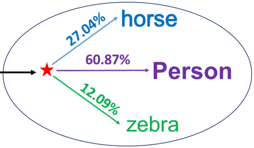


1:killer whale
2: beaver
3: dalmatian
4: persian cat
5: german shepherd
6: siamese cat
7: skunk
8: mole
9: tiger
10: hippopotamus
11: leopard
12: spider monkey
13: gorilla
14: ox
15: chimpanzee
16: hamster
17: chihuahua
18: weasel
19: otter
20: zebra
21: giant panda
22: pig
23: lion
24: mouse
25: polar bear
26: collie

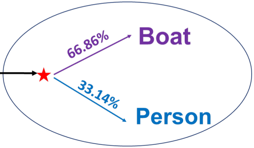
Figure: Cluster visualization of the correlation matrix of the label embeddings.

- **Task:** Describe an image with relative labels and specific descriptions.
- **Reason:** Multi-contexts aligned via the shared embedding.

Attribute	Value
Skin	0.37
Tail	0.31
Torso	0.24
Head	0.23
Eye	0.22
Arm	0.13



Attribute	Value
Shiny	0.67
Sail	0.51
Plastic	0.34
Skin	0.33
Head	0.23
Cloth	0.17



Descriptive Contexts **Relational Contexts**

Figure: For relational contexts, the digits represent the images' similarity with the related labels. For descriptive contexts, attributes marked in green are creatively exploited, while those marked in red are wrongly predicted.

Conclusion

Conclusion

- **Problem:** Label embedding with Partial Heterogeneous Contexts
- **Model:** Partial Heterogeneous Context Label Embedding (PHCLE)
  - Tailor-make formula for the heterogeneous contexts;
  - Align partial heterogeneous contexts via shared label embedding;
  - Enhance alignment with discriminative contexts selection.
- **Model Speciality**
  - Can handle partial context with an matrix to indicate the missing entries;
  - Can be easily generalized to incorporate more contexts.
- **Empirical results**
  - Superb performance in image classification task;
  - Exhibits good human interpretability;
    - label similarity analysis;
    - novel image understanding;

References

[1] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele. Evaluation of output embeddings for fine-grained image classification. In *CVPR*, pages 2927–2936, 2015.

[2] Y. Fu, T.M. Hospedales, T. Xiang, Z. Fu, and S. Gong. Transductive multi-view embedding for zero-shot recognition and annotation. In *ECCV*, pages 584–599, 2014.

[3] L. Liao, X. He, H. Zhang, and T. Chua. Attributed social network embedding. *arXiv:1705.04969*, 2017.

[4] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Y. Chang. Network representation learning with rich text information. In *IJCAI*, pages 2111–2117, 2015.

[5] Y. Li, L. Xu, F. Tian, L. Jiang, X. Zhong, and E. Chen. Word embedding revisited: A new representation learning and explicit matrix factorization perspective. In *IJCAI*, pages 3650–3656, 2015.

Thank you!
Q & A