EMD Metric Learning

***Zizhao Zhang* , Yubo Zhang, Xibin Zhao, and Yue Gao**

School of Software, Tsinghua University, China.

Contents

Contents

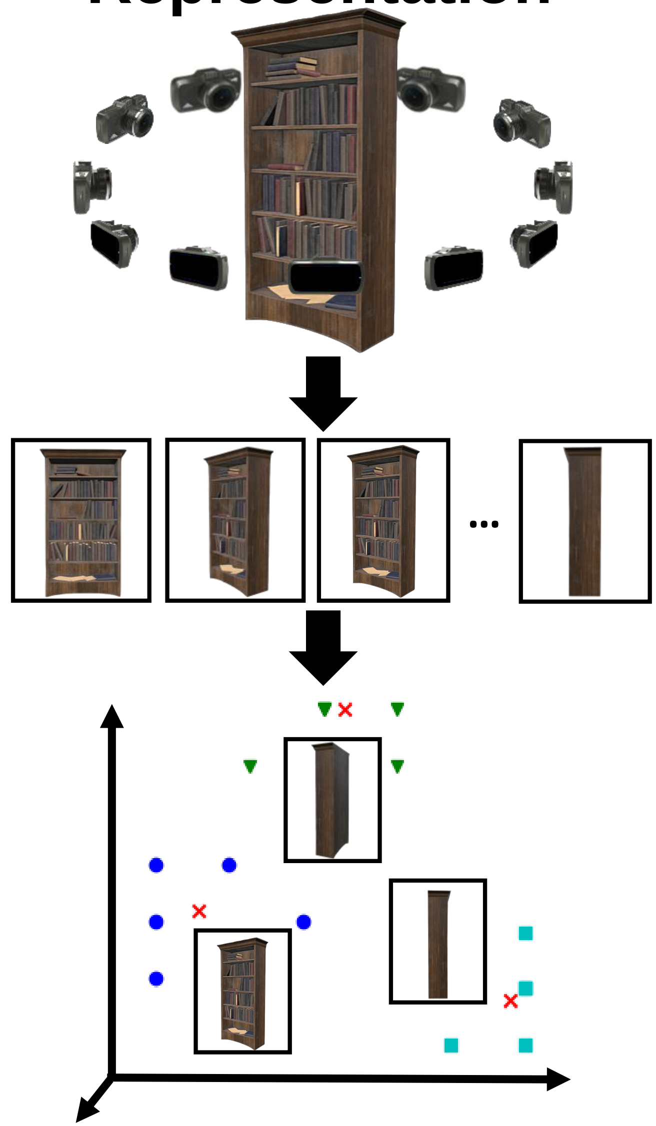
Multivariate Distribution

Subject

Multivariate Distribution

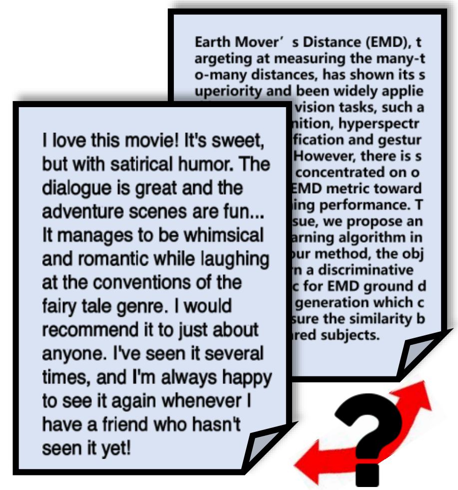
Document Representation

Visual Object Representation



Multivariate Distribution

Bookshelf Door



How to measure the similarity between two multivariate distributions (signatures)?

- K-L Divergence (Goldberger et al. 2003)
- Maximum Mean Discrepancy (Borgwardt et al. 2006)
- **Earth Mover's Distance** [1]

[1] Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40(2): 99-121.

Cited by 3124

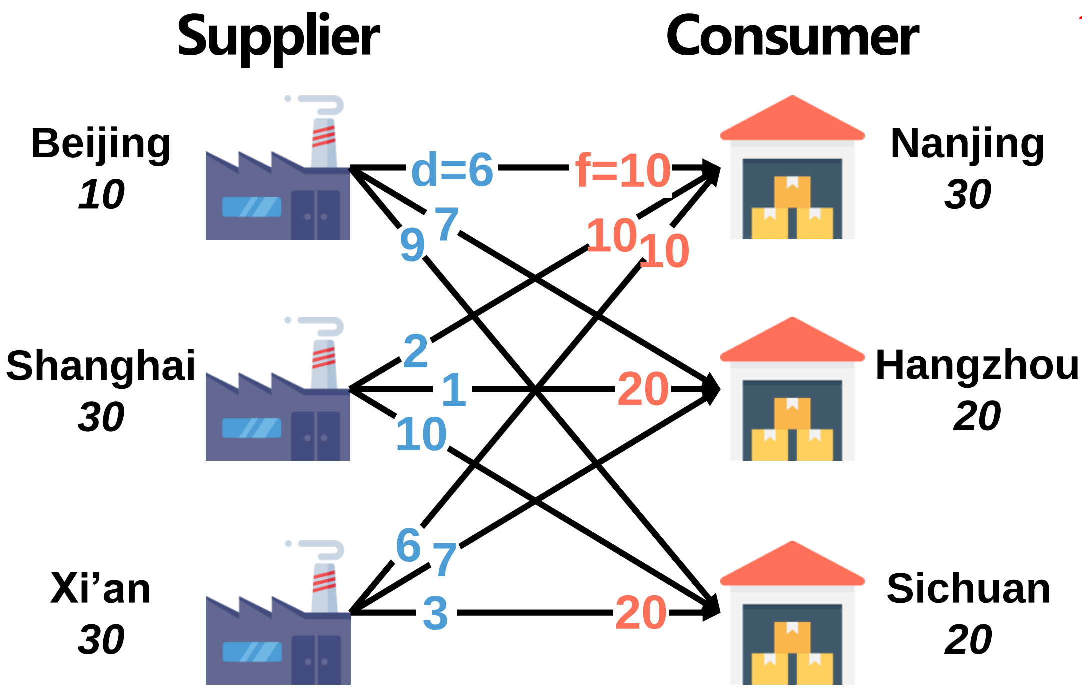
Earth Mover's Distance

EMD is a solution to the old transportation problem.

A set of suppliers: $P = \{(p_1, w_{p_1}), \ldots, (p_m, w_{p_m})\}$

A set of consumers: $Q = \{(q_1, w_{q_1}), \ldots, (q_n, w_{q_n})\}$

$$\boldsymbol{EMD(P,Q) = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} d_{ij} f_{ij}^{*}}{\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}^{*}}}$$



subject to

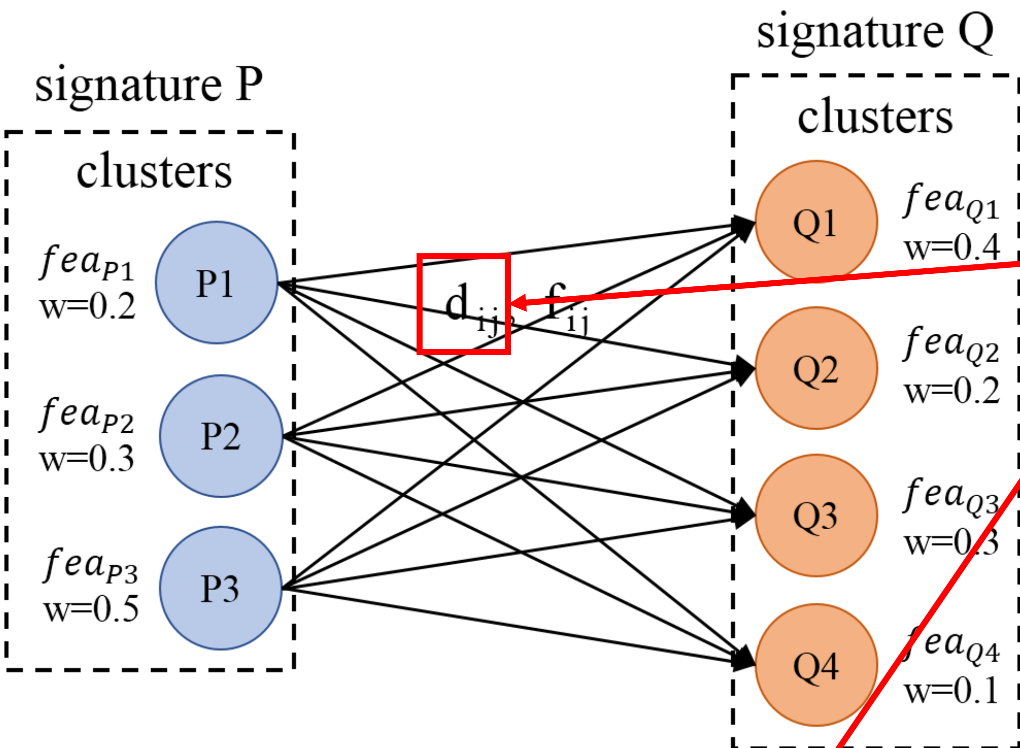
$$f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i} \quad 1 \leq i \leq m$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j} \quad 1 \leq j \leq n$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}\right)$$

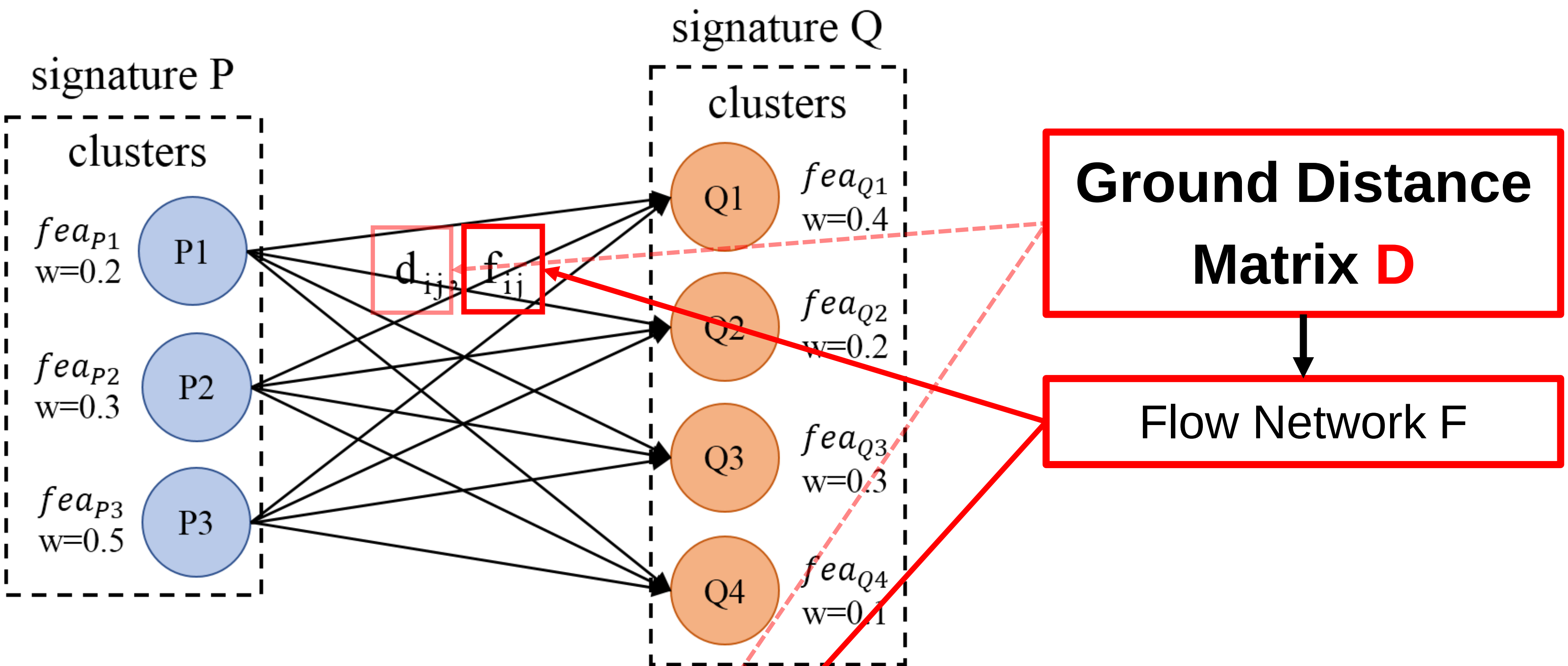
Earth Mover's Distance



Ground Distance Matrix D

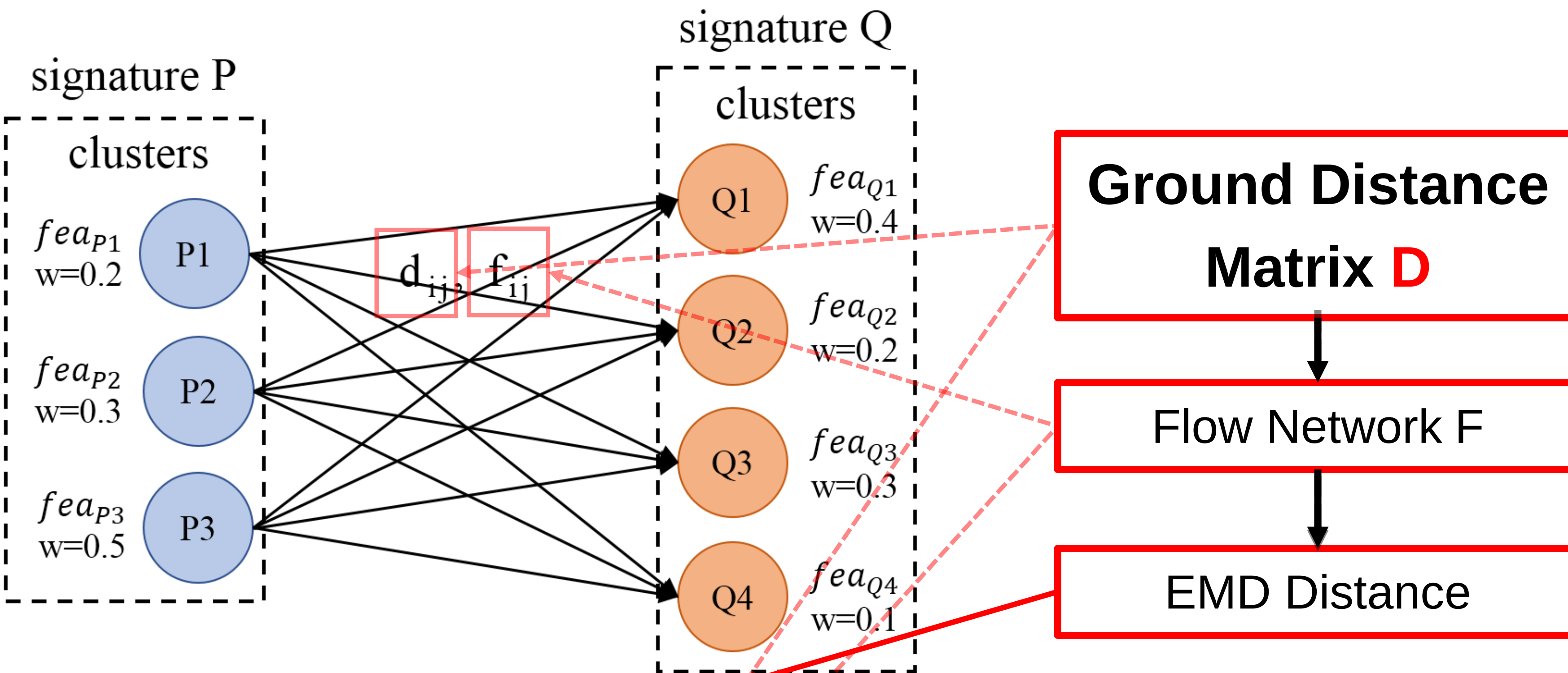
$$EMD(P,Q) = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} \boxed{d_{ij}} f_{ij}^{*}}{\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}^{*}}$$

Earth Mover's Distance



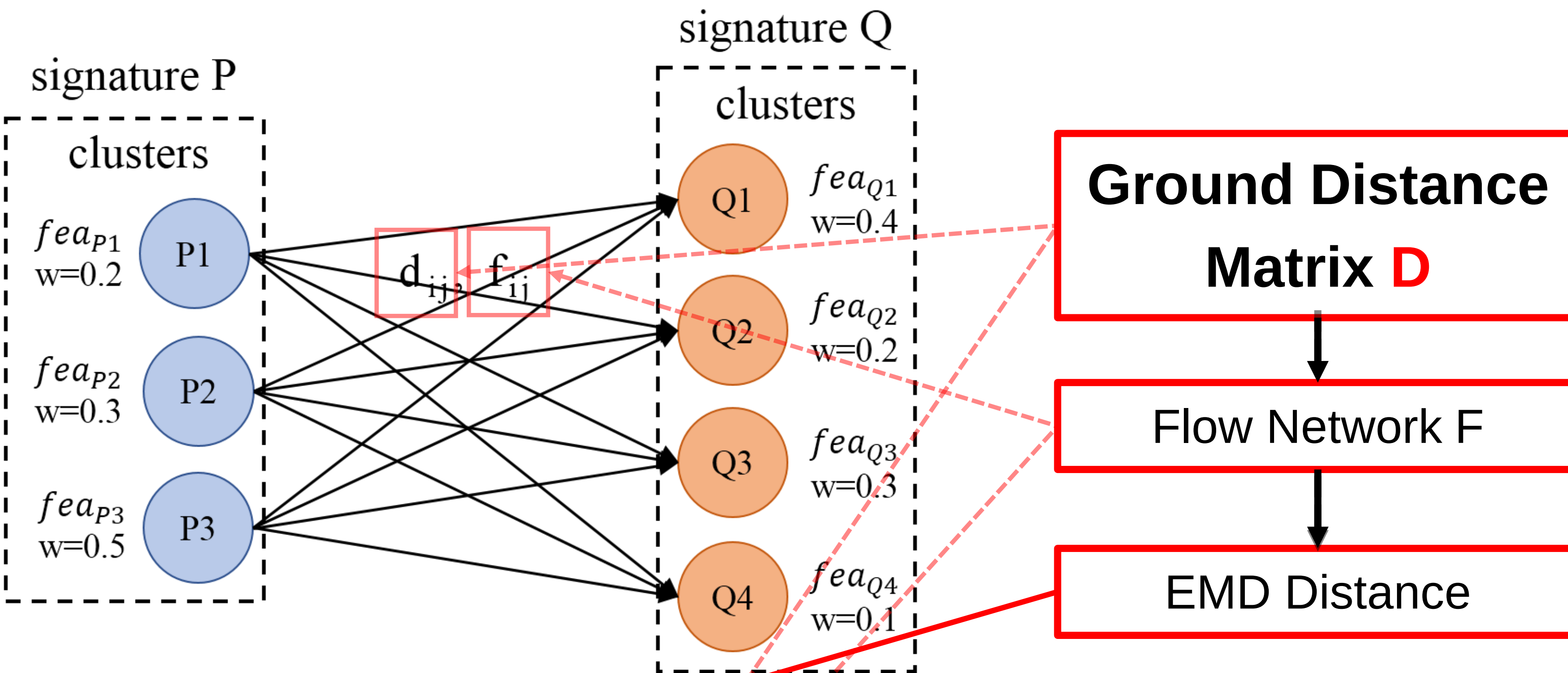
$$EMD(P,Q) = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} \boxed{d_{ij}}\boxed{f_{ij}^{*}}}{\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}^{*}}$$

Earth Mover's Distance



$$\boxed{EMD(P,Q)} = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} \boxed{d_{ij}}\,\boxed{f_{ij}^{*}}}{\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}^{*}}$$

Earth Mover's Distance

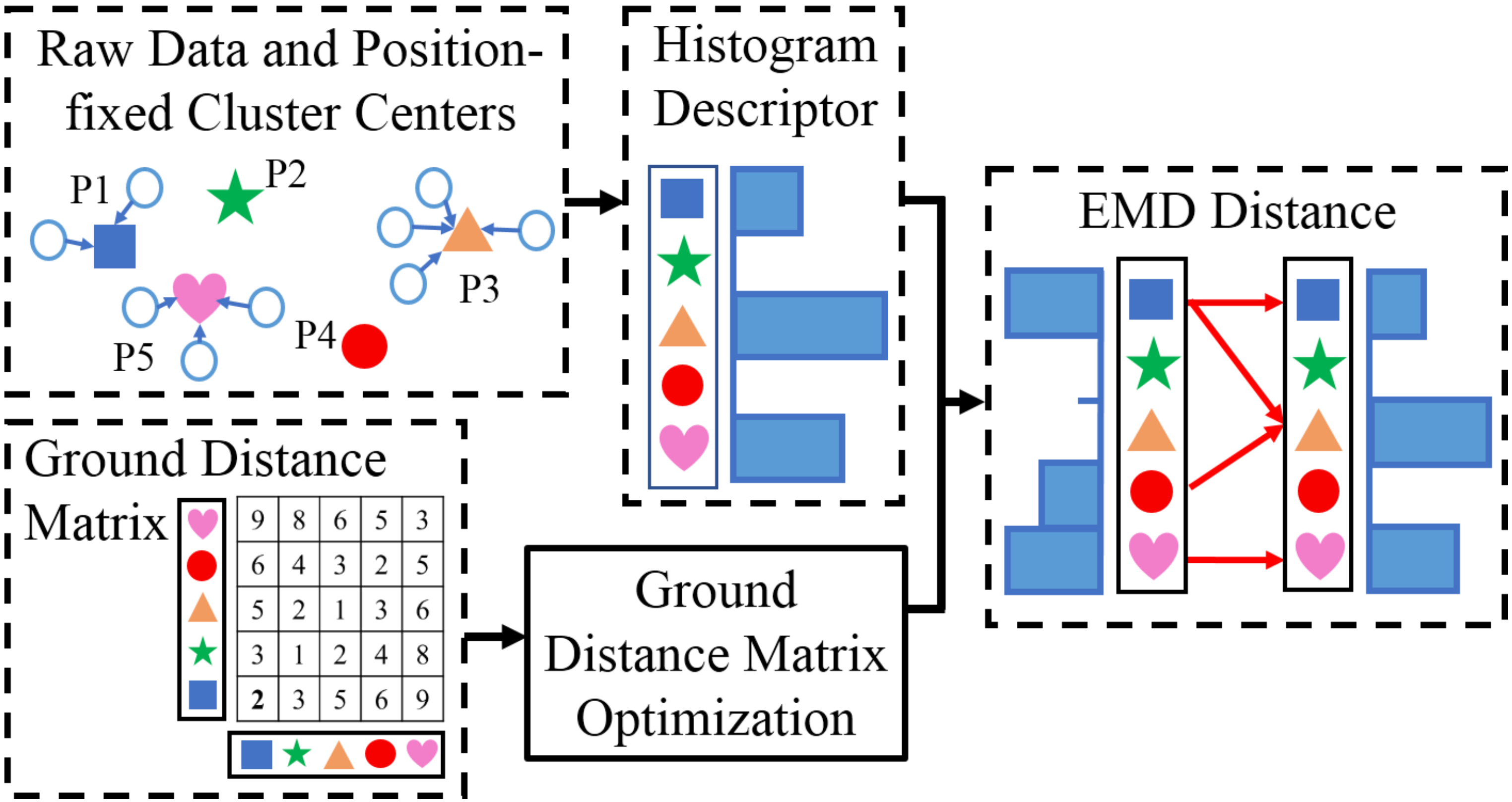


$$\boxed{EMD(P,Q)} = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} \boxed{d_{ij}}\,\boxed{f_{ij}^{*}}}{\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}^{*}}$$

Optimize EMD ⟶ Optimize ground distance matrix

Related Work

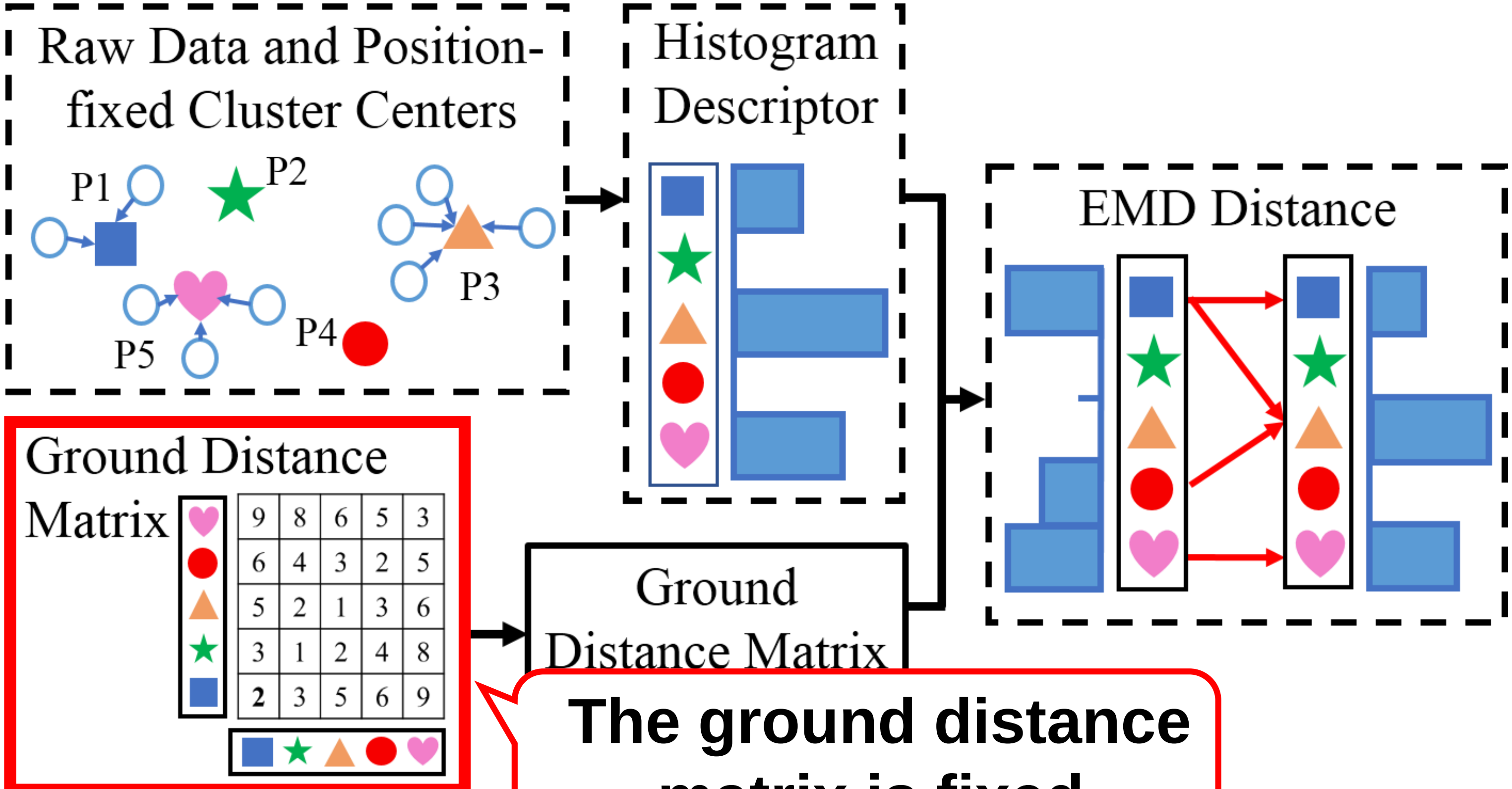
Wang et al. learn an optimal ground distance matrix directly. [2]



[2] Wang, F., and Guibas, L. J. 2012. Supervised earth movers distance learning and its computer vision applications. *In European Conference on Computer Vision*, 442–455. Springer.

Related Work

Wang et al. learn an optimal ground distance matrix directly. [2]



The ground distance matrix is fixed.

A fixed ground distance matrix D is infeasible in many applications.

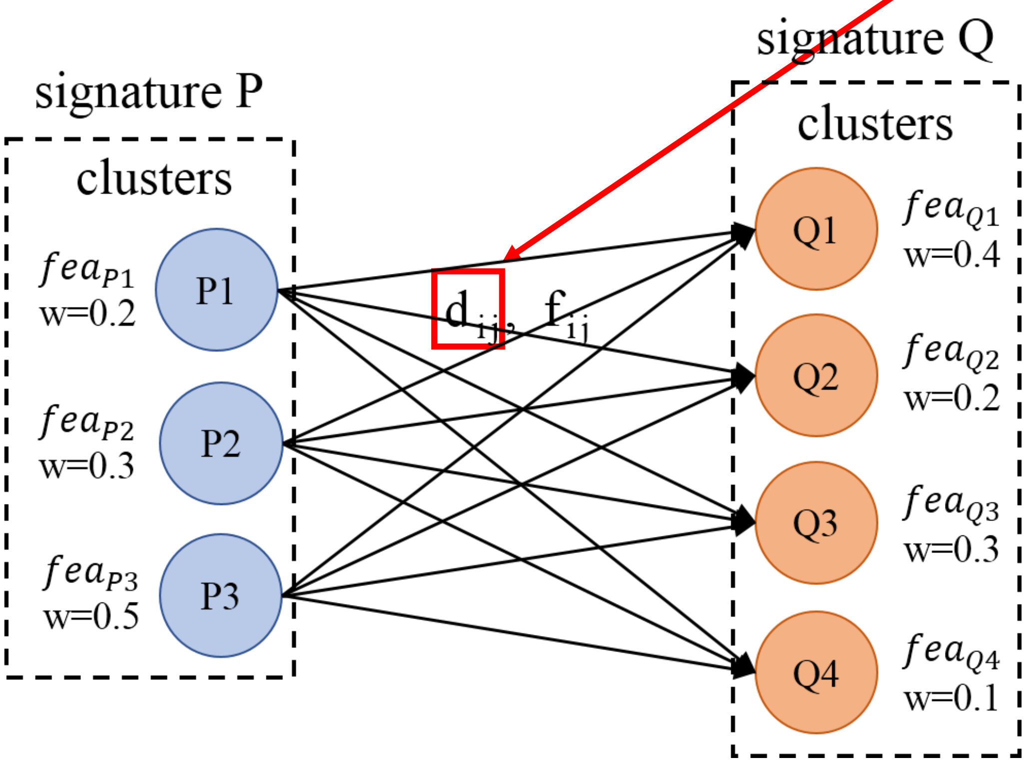
[2] Wang, F., and Guibas, L. J. 2012. Supervised earth movers distance learning and its computer vision applications. *In European Conference on Computer Vision*, 442–455. Springer.

Earth Mover's Distance

Ground Distance Metric A

$$d_{ij} = \sqrt{\left(fea_{Pi} - fea_{Qj}\right)^T \mathbf{A} \left(fea_{Pi} - fea_{Qj}\right)}$$

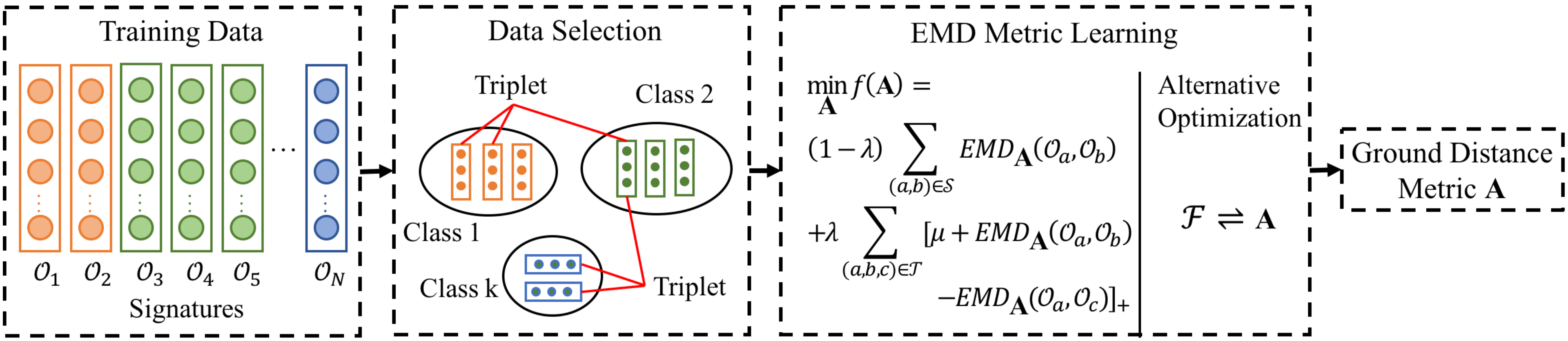
Ground Distance Matrix D



$$EMD(P,Q) = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} \boxed{d_{ij}} f_{ij}^{*}}{\sum_{i=0}^{m} \sum_{j=0}^{n} f_{ij}^{*}}$$

Contents

The Framework of EMD Metric Learning

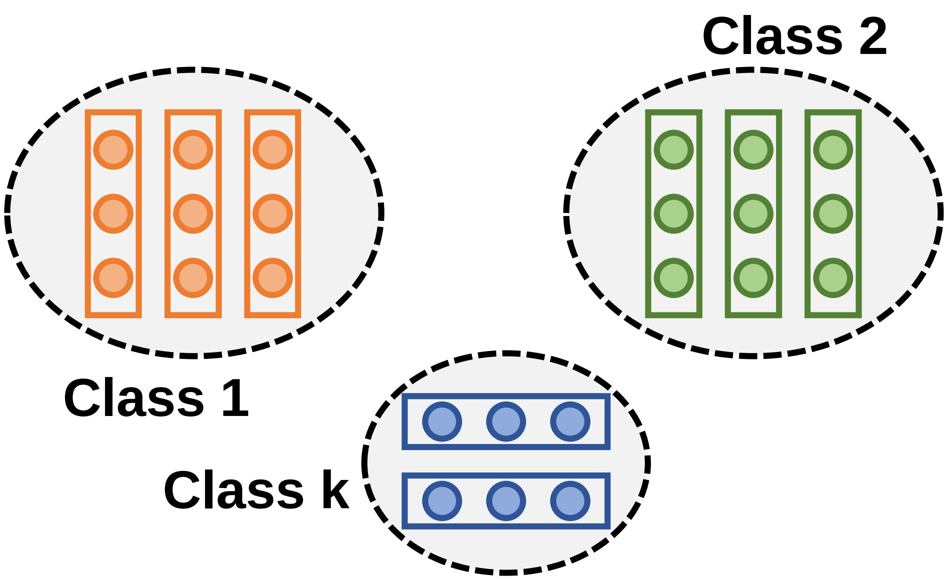


The Formulation of EMD Metric Learning

The objective is to preserve the topological structure of the data and satisfy triplet constraints simultaneously.

Cost function:

$$\min_{\mathbf{A}} f(\mathbf{A}) = (1-\lambda) \sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_b) + \lambda \sum_{(a,b,c)\in\mathcal{T}} [\mu + EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_c)]_+$$



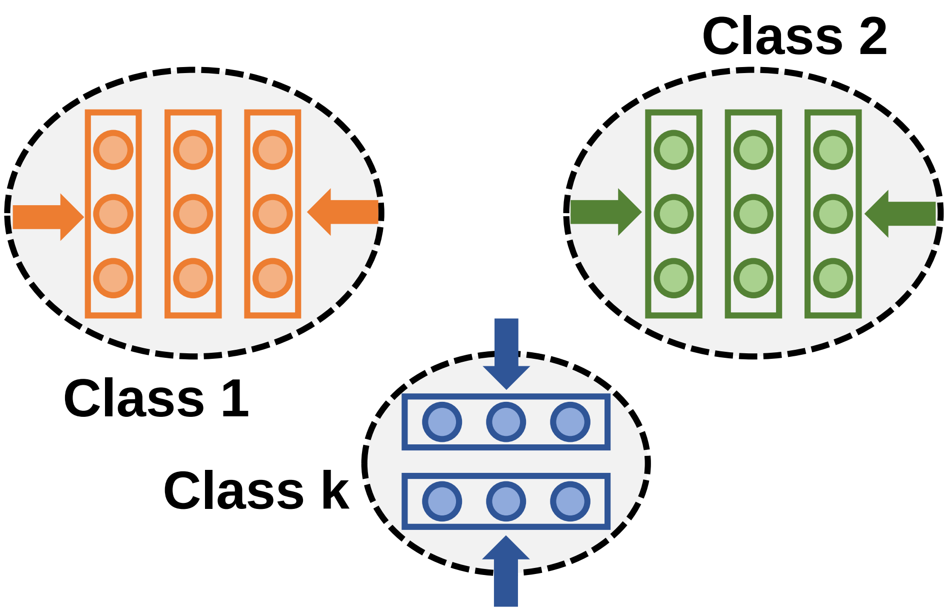
The Formulation of EMD Metric Learning

The objective is to preserve the topological structure of the data and satisfy triplet constraints simultaneously.

Cost function:

$$\min_{\mathbf{A}} f(\mathbf{A}) = \boxed{(1-\lambda) \sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_b)} + \lambda \sum_{(a,b,c)\in\mathcal{T}} [\mu + EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a, \mathcal{O}_c)]_+$$

Preserve the topological structure of the data



The Formulation of EMD Metric Learning

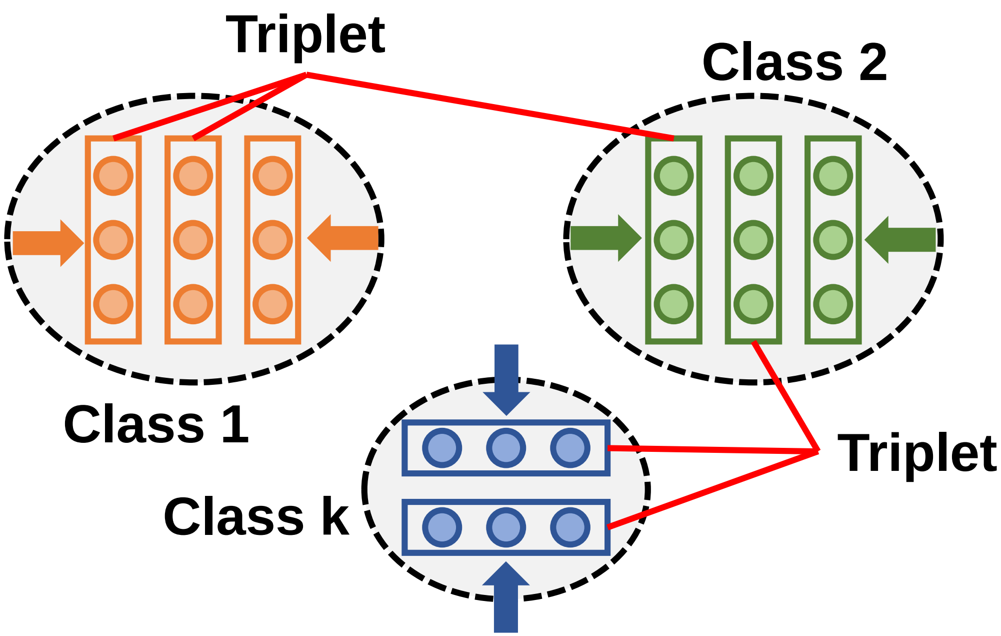
The objective is to preserve the topological structure of the data and satisfy triplet constraints simultaneously.

Cost function:

$$\min_{\mathbf{A}} f(\mathbf{A}) = \boxed{(1-\lambda)\sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b)}$$
$$\boxed{+\lambda \sum_{(a,b,c)\in\mathcal{T}} [\mu + EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_c)]_+}$$

Preserve the topological structure of the data

The triplet constraints

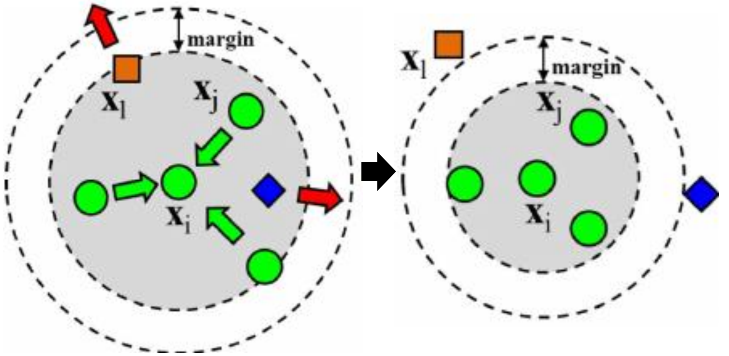


Training Data Selection

- How to construct triplet constraints?

  Use ~~all possible triplet constrains~~
  nearest neighbors to construct triplets

  The number of triplets: $N^3 \rightarrow k_i k_g N$

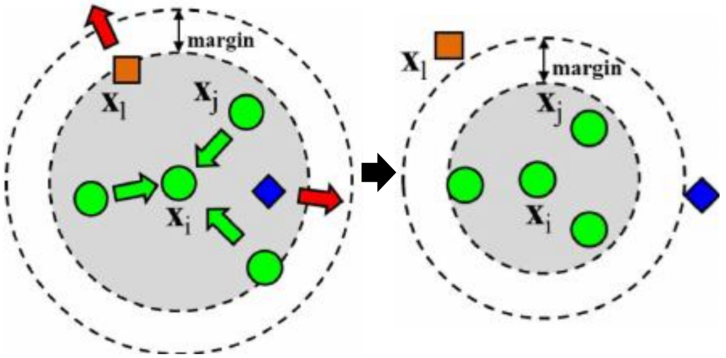


Training Data Selection

- How to construct triplet constraints?

  Use ~~all possible triplet constrains~~
  nearest neighbors to construct triplets

  The number of triplets: $N^3 \rightarrow k_i k_g N$



- How to search nearest neighbors?

  Calculate the ~~exact~~ EMD between all pairs of signatures.
  relaxed

$$REMD(\mathcal{P}, \mathcal{Q}) = \begin{cases} \dfrac{\sum_{i=1}^{n} d_{ij^*} w_{P_i}}{\sum_{i=1}^{n} w_{P_i}} & \sum\limits_{i=1}^{n} w_{P_i} \leq \sum\limits_{j=1}^{m} w_{Q_j} \\ \dfrac{\sum_{j=1}^{m} d_{i^*j} w_{Q_j}}{\sum_{j=1}^{m} w_{Q_j}} & \sum\limits_{i=1}^{n} w_{P_i} > \sum\limits_{j=1}^{m} w_{Q_j} \end{cases}$$

  Time complexity: $O(n^3 \log n) \rightarrow O(n^2)$

Optimization

Cost function:

$$
\begin{aligned}
&\min_{\mathbf{A},\mathcal{F}} f(\mathbf{A},\mathcal{F}) \\
&= (1-\lambda)\sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) \\
&+\lambda \sum_{(a,b,c)\in\mathcal{T}} \left[\mu + EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_c)\right]_+ \\
&= (1-\lambda)\sum_{(a,b)\in S} \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d(x_a^i, x_b^j)\right) \\
&+\lambda \sum_{(a,b,c)\in T} \left[\mu + \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d\left(x_a^i, x_b^j\right)\right)\right. \\
&\left. - \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_c} F_{a,c}(i,j)\, d\left(x_a^i, x_c^j\right)\right)\right]_+
\end{aligned}
$$

Fix $\mathcal{F}$

The problem of Mahalanobis distance metric learning

Fix A

$|\mathcal{S}|+2|\mathcal{T}|$ independent traditional EMD sub-problems

Optimization

Cost function:

$$
\begin{aligned}
&\min_{\mathbf{A},\mathcal{F}} f(\mathbf{A},\mathcal{F}) \\
&= (1-\lambda) \sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) \\
&+\lambda \sum_{(a,b,c)\in\mathcal{T}} [\mu + EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_c)]_+ \\
&= (1-\lambda) \sum_{(a,b)\in S} \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d(x_a^i, x_b^j)\right) \\
&+\lambda \sum_{(a,b,c)\in T} \Bigg[\mu + \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d\left(x_a^i, x_b^j\right)\right) \\
&- \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_c} F_{a,c}(i,j)\, d\left(x_a^i, x_c^j\right)\right)\Bigg]_+
\end{aligned}
$$

Fix $\mathcal{F}$ → **The problem of Mahalanobis distance metric learning**

Fix A → $|\mathcal{S}|+2|\mathcal{T}|$ independent traditional EMD sub-problems

Optimization

Cost function:

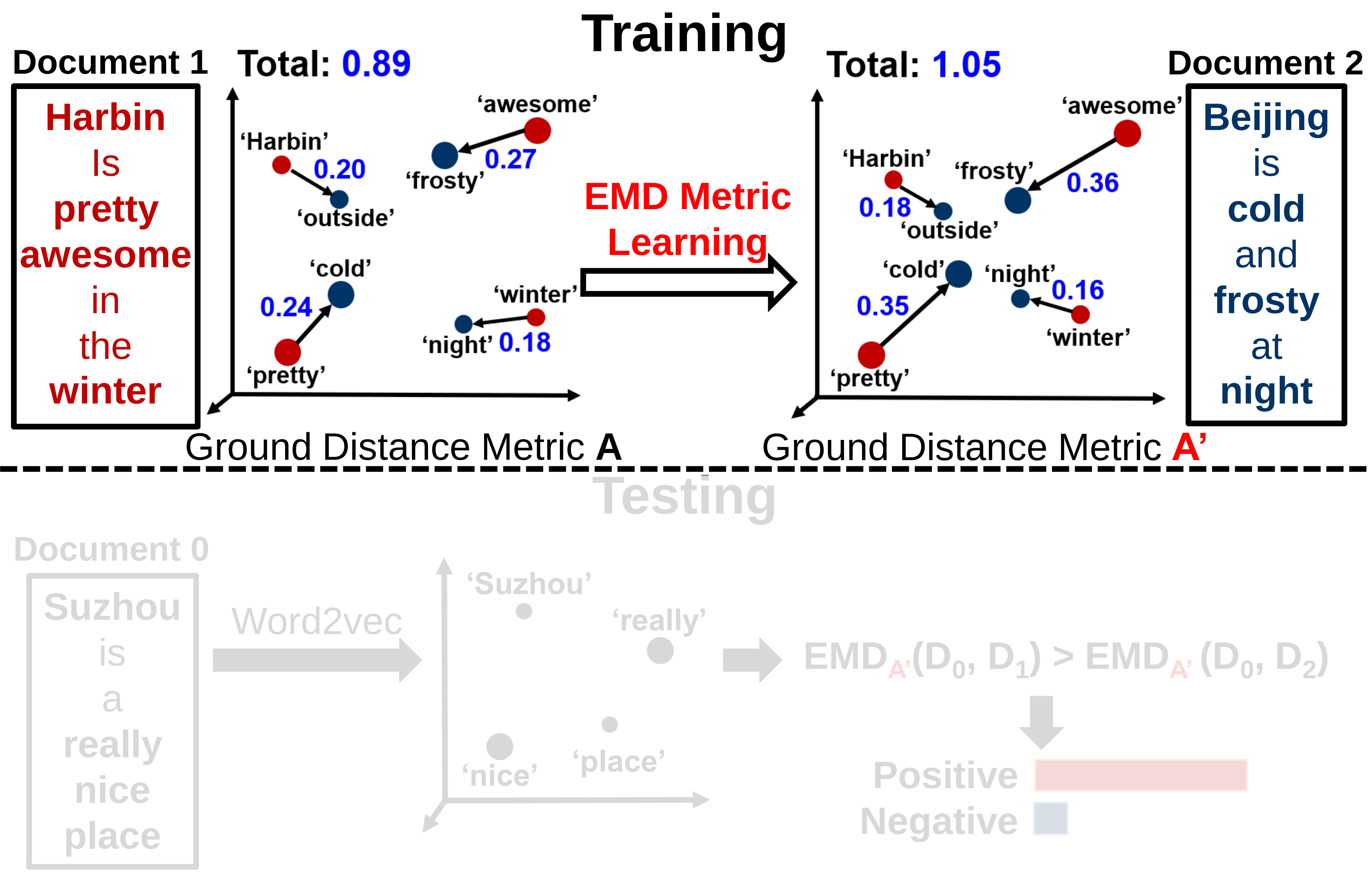
$$\min_{\mathbf{A},\mathcal{F}} f(\mathbf{A},\mathcal{F})$$
$$= (1-\lambda)\sum_{(a,b)\in\mathcal{S}} EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b)$$
$$+\lambda \sum_{(a,b,c)\in\mathcal{T}} \left[\mu + EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_b) - EMD_{\mathbf{A}}(\mathcal{O}_a,\mathcal{O}_c)\right]_+$$
$$= (1-\lambda)\sum_{(a,b)\in S}\left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d(x_a^i,x_b^j)\right)$$
$$+\lambda \sum_{(a,b,c)\in T}\left[\mu + \left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_b} F_{a,b}(i,j)\, d\left(x_a^i,x_b^j\right)\right)\right.$$
$$\left.-\left(\sum_{i=1}^{n_a}\sum_{j=1}^{n_c} F_{a,c}(i,j)\, d\left(x_a^i,x_c^j\right)\right)\right]_+$$

Fix $\mathcal{F}$ → **The problem of Mahalanobis distance metric learning**

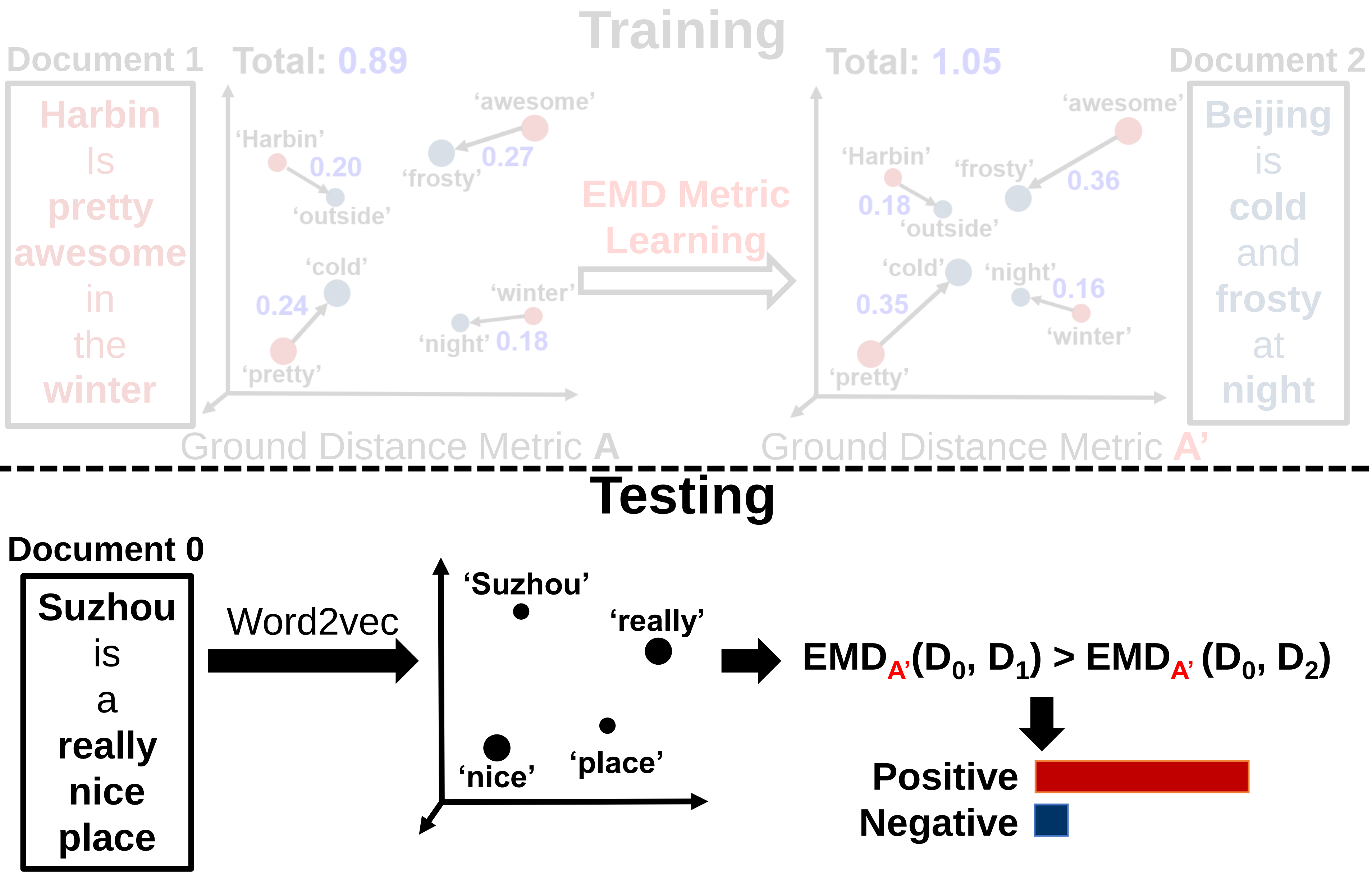
Fix A → **$|\mathcal{S}|+2|\mathcal{T}|$ independent traditional EMD sub-problems**

Contents

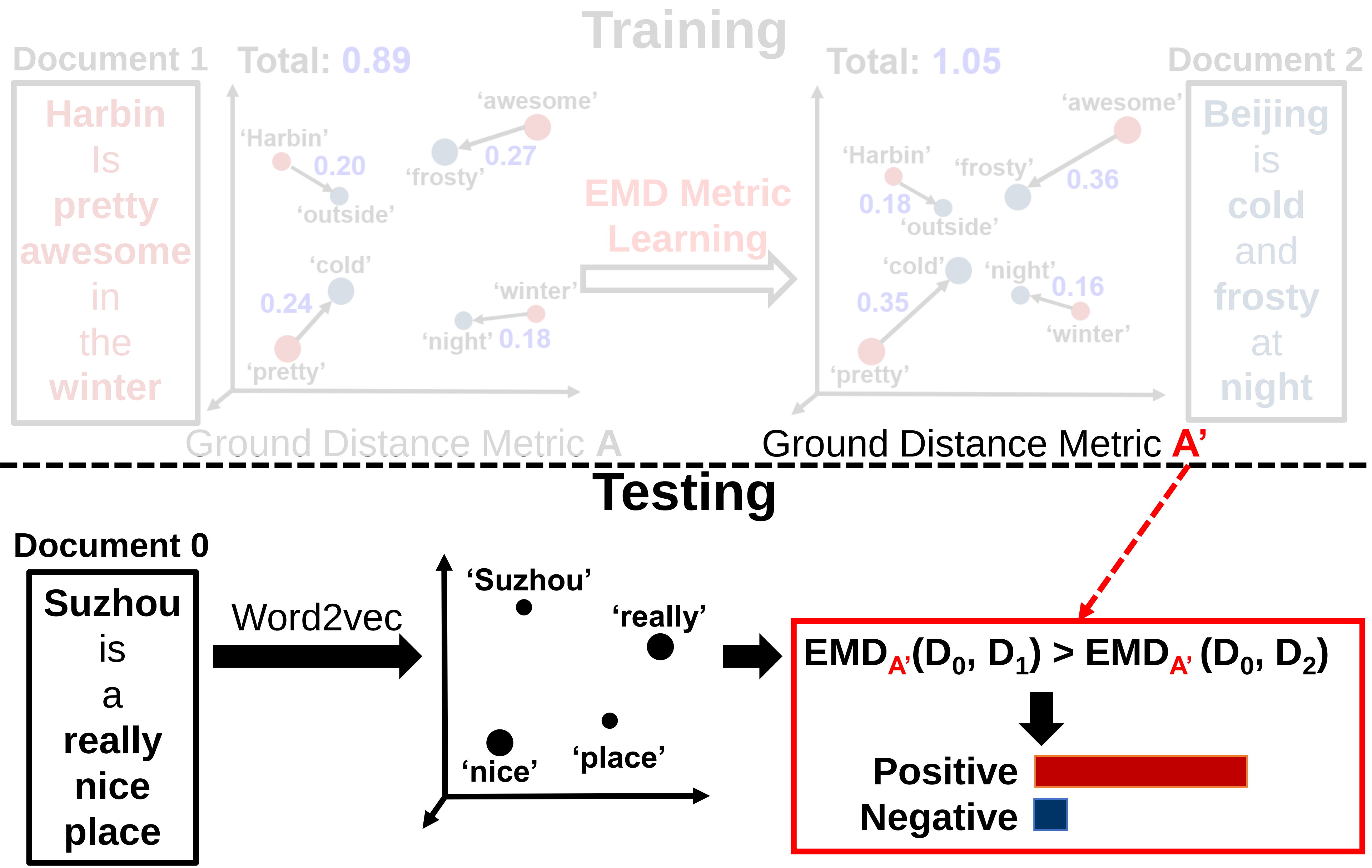
Document Classification



Document Classification



Document Classification



Document Classification

The TWITTER dataset

Contains 2176 objects labeled with ‘positive’, ‘negative’, and ‘neutral’.

Each nonstop word is represented by a 300-d feature.

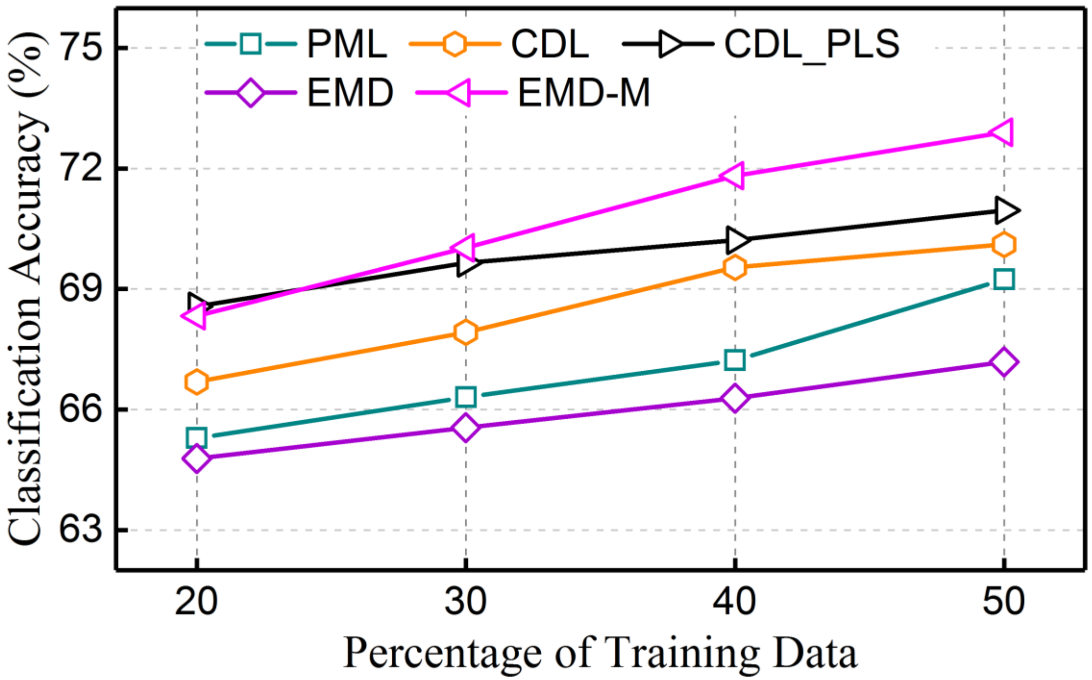
The average number of unique words per document is 9.9

State-of-the-art methods

Covariance Discriminative Learning

Covariance Discriminative Learning with PLS *CVPR’12*

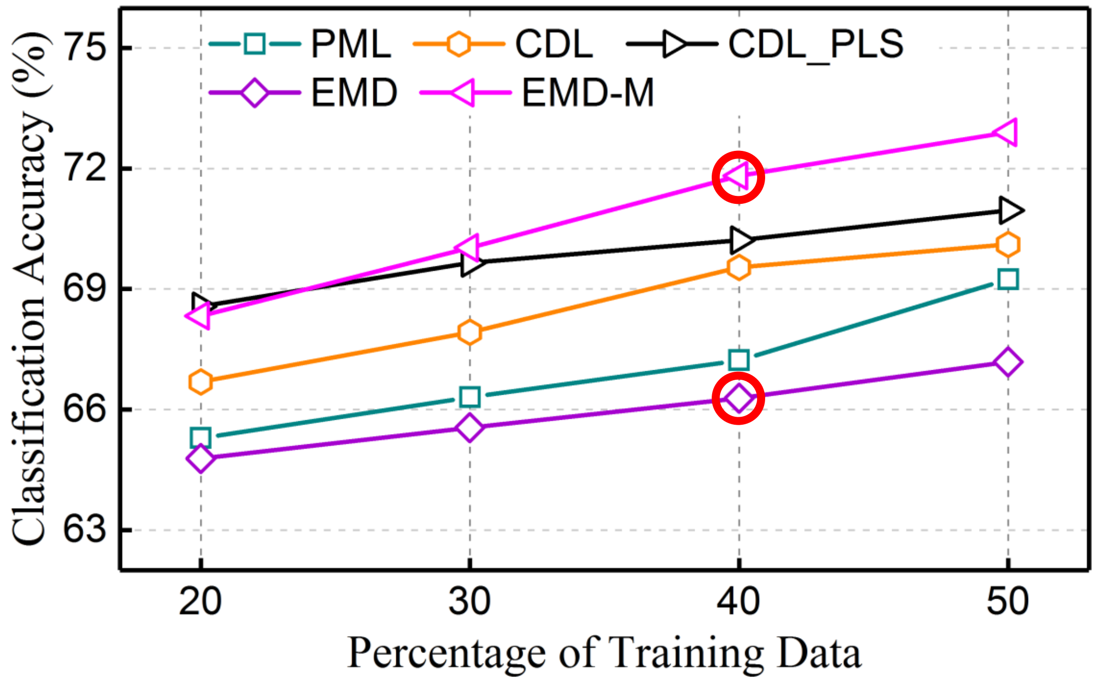
Projection Metric Learning *CVPR’15*



Document Classification

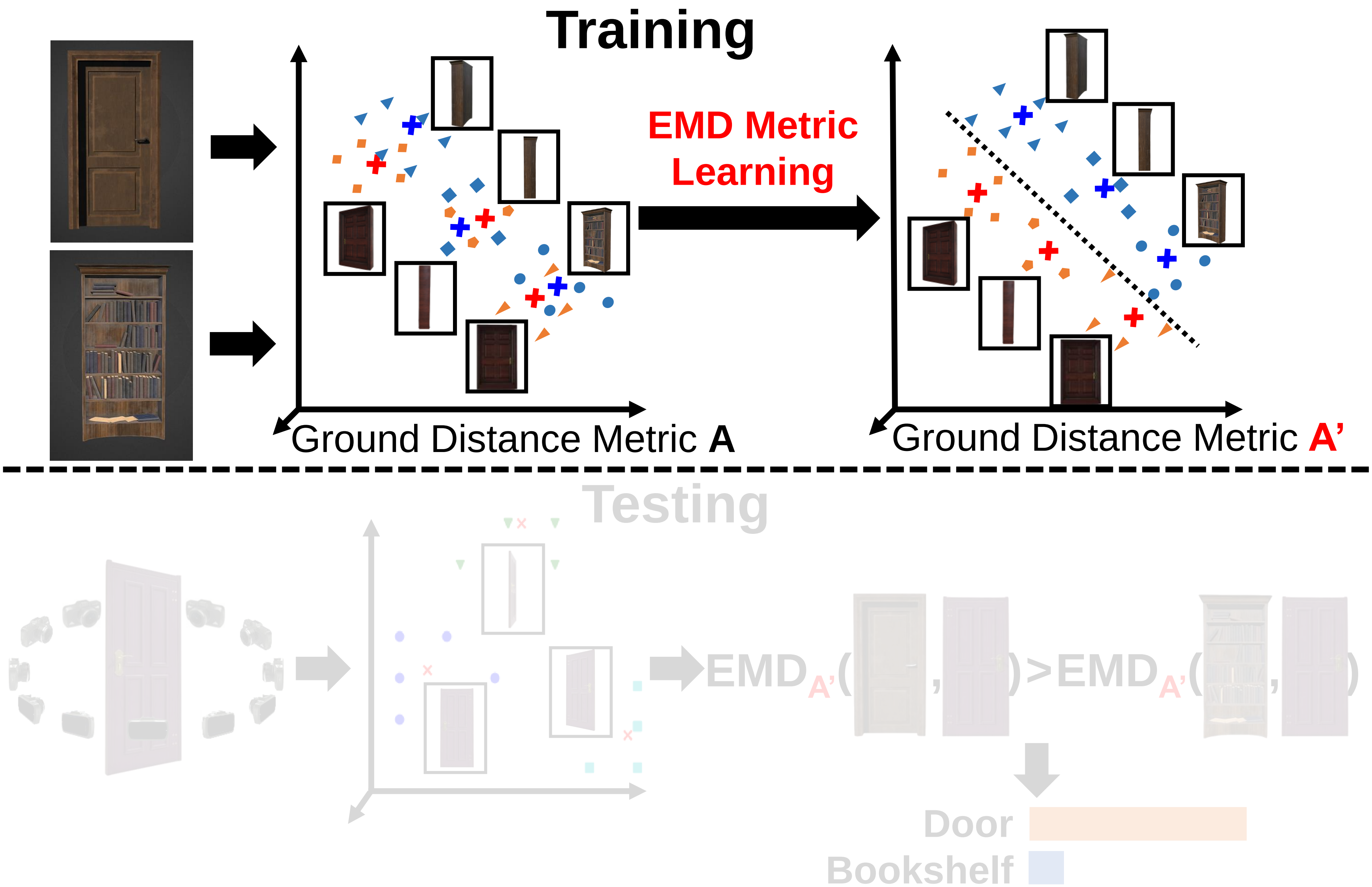
The TWITTER dataset
Contains 2176 objects labeled with 'positive', 'negative', and 'neutral'.
Each nonstop word is represented by a 300-d feature.
The average number of unique words per document is 9.9

State-of-the-art methods
Covariance Discriminative Learning
Covariance Discriminative Learning with PLS *CVPR'12*
Projection Metric Learning *CVPR'15*

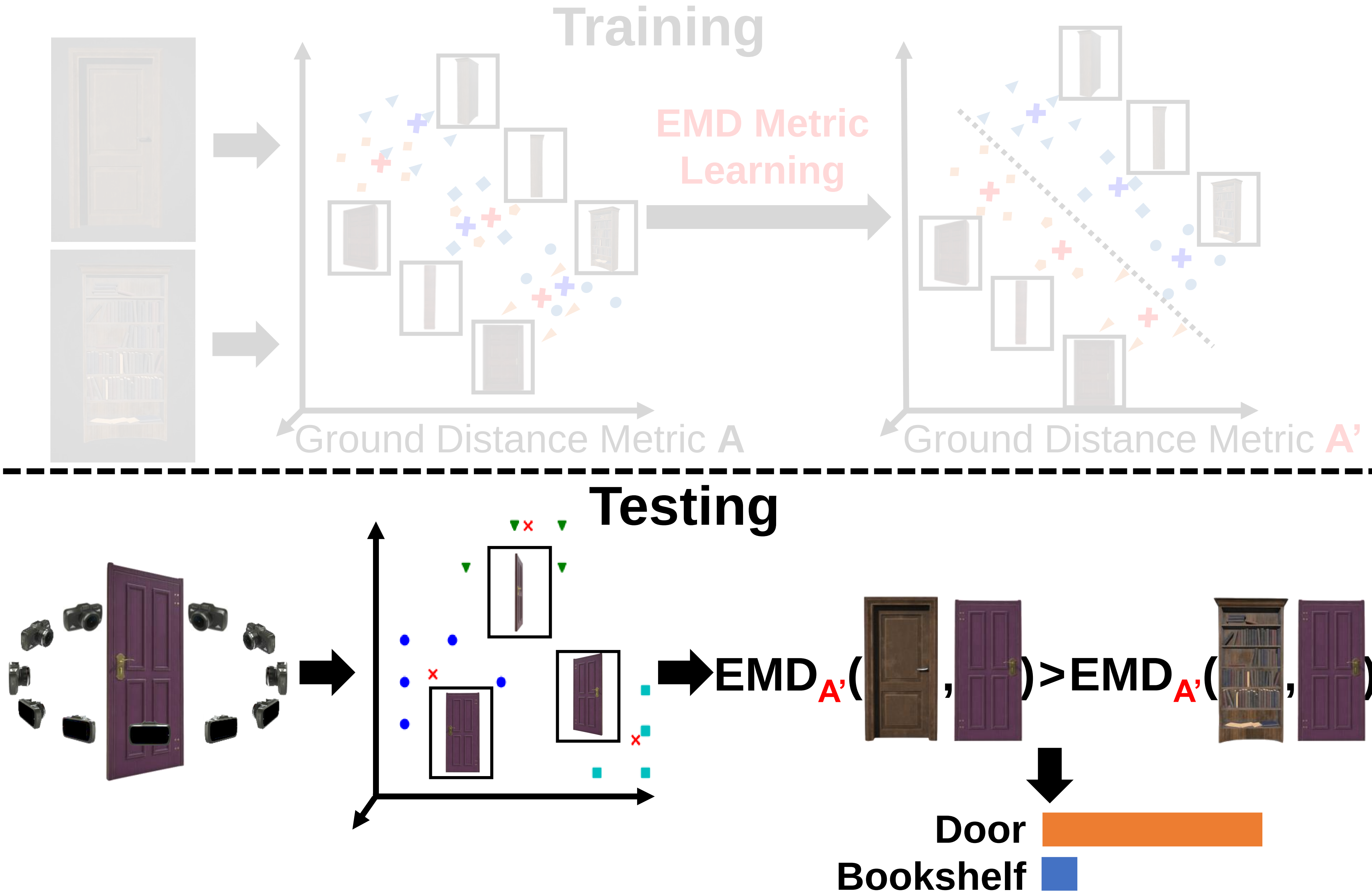


Contents

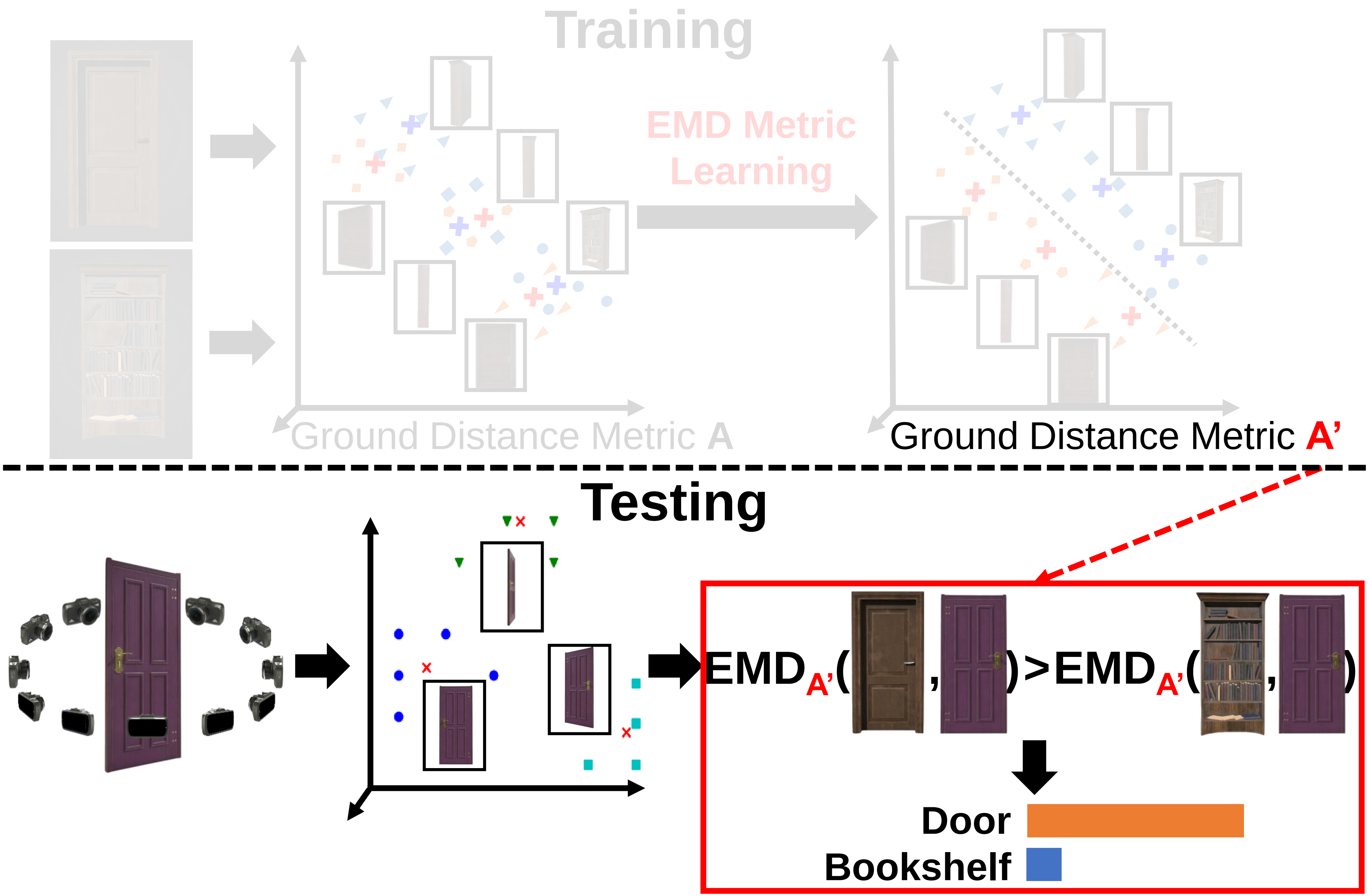
Multi-View Object Classification



Multi-View Object Classification



Multi-View Object Classification



Multi-View Object Classification

The NTU dataset

Contains 401 objects from 16 classes

Each object contains 60 views.

Extract the 4096-d CNN feature for each view

State-of-the-art methods

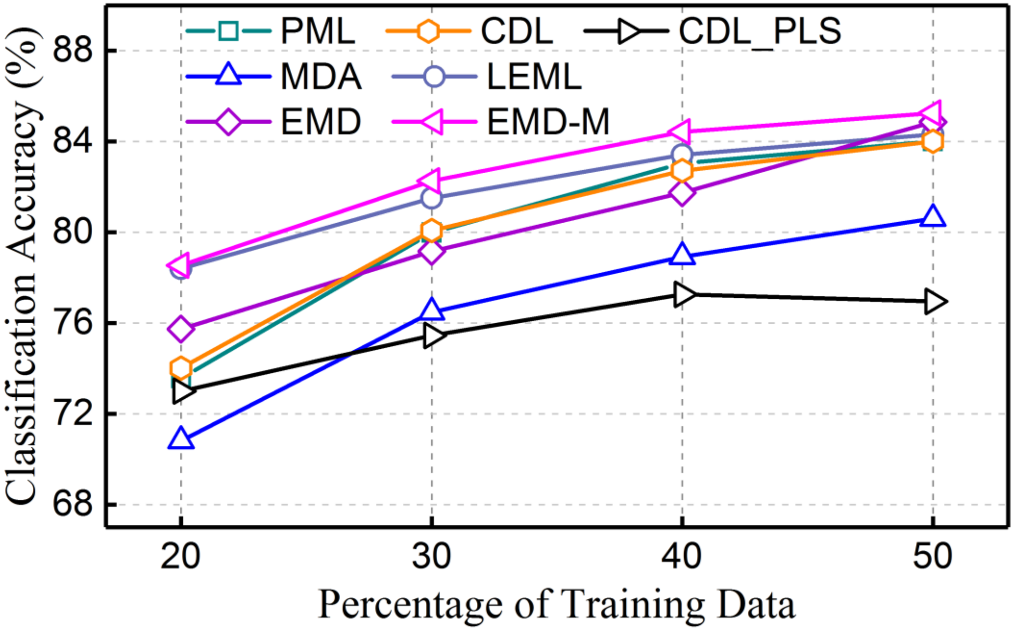
Manifold Discriminant Analysis *CVPR'09*

Covariance Discriminative Learning

Covariance Discriminative Learning with PLS *CVPR'12*

Log-Euclidean Metric Learning *ICML'15*

Projection Metric Learning *CVPR'15*



Multi-View Object Classification

The NTU dataset

Contains 401 objects from 16 classes

Each object contains 60 views.

Extract the 4096-d CNN feature for each view

State-of-the-art methods

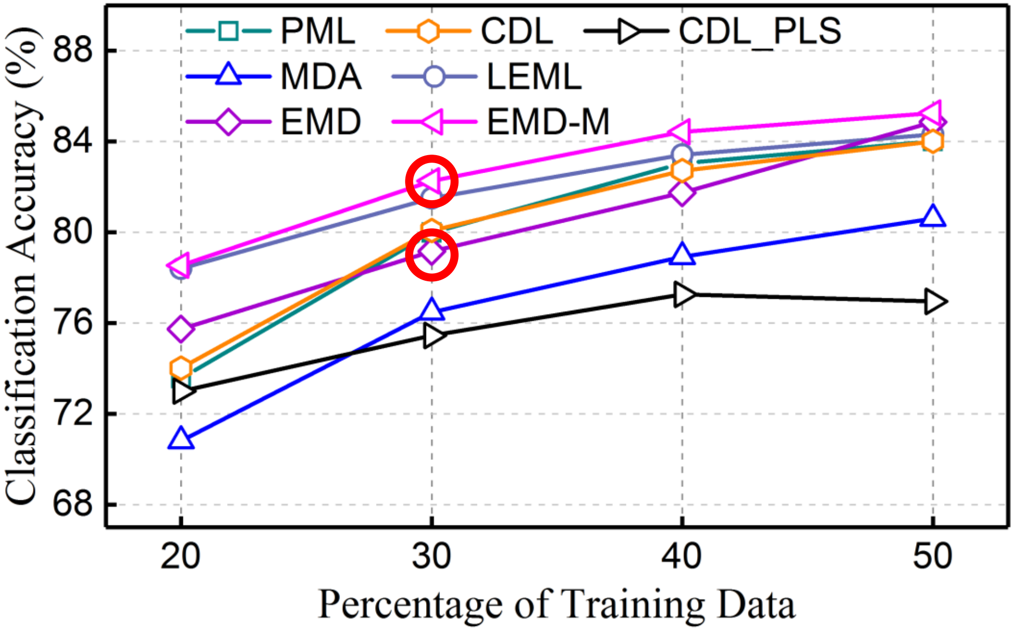
Manifold Discriminant Analysis *CVPR'09*

Covariance Discriminative Learning

Covariance Discriminative Learning with PLS *CVPR'12*

Log-Euclidean Metric Learning *ICML'15*

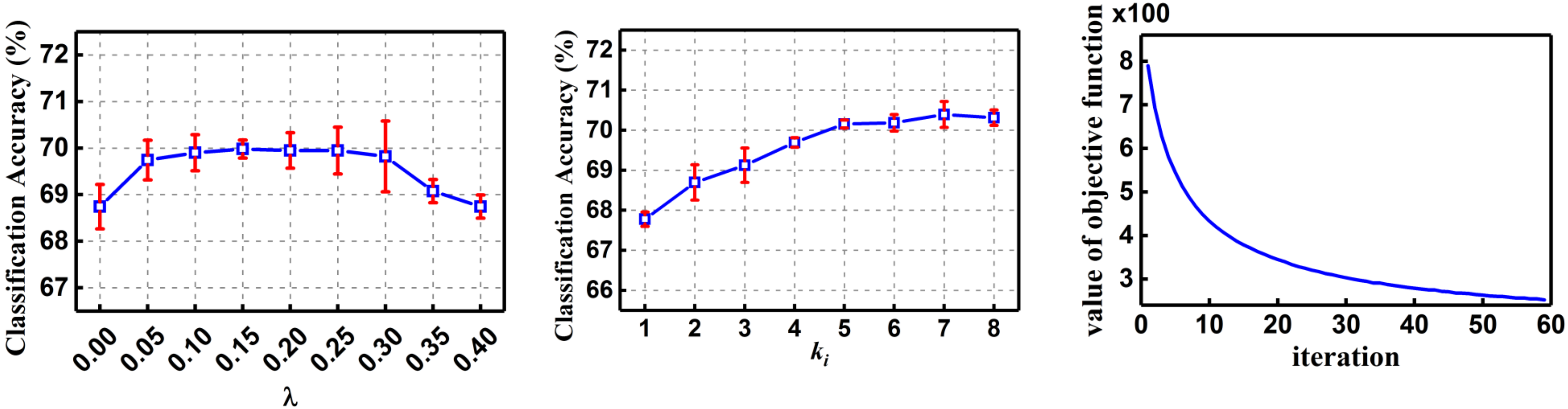
Projection Metric Learning *CVPR'15*



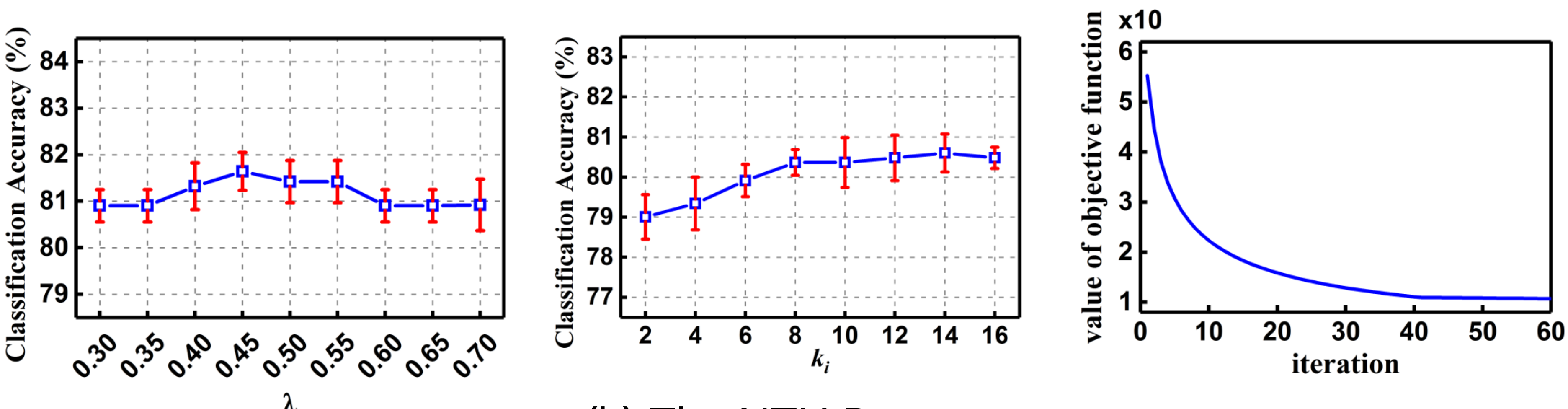
On parameter sensitivity and convergence

Parameter sensitivity

Convergence curve



(a) The TWITTER Dataset



(b) The NTU Dataset

Conclusion

- We propose an **EMD metric learning algorithm** targeting on a more general setting.
- We apply EMD metric learning on the tasks of **multi-view object classification** and **text classification**.
- EMD metric learning can achieve about **5% improvement** compared with the traditional EMD.

References

[1] Goldberger, J.; Gordon, S.; and Greenspan, H. 2003. An efficient image similarity measure based on approximations of kl-divergence between two gaussian mixtures. *In Proceedings of IEEE International Conference on Computer Vision*, 487. IEEE.

[2] Huang, Z.; Wang, R.; Shan, S.; and Chen, X. 2015a. Projection metric learning on grassmann manifold with application to video based face recognition. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 140–149.

[3] Huang, Z.; Wang, R.; Shan, S.; Li, X.; and Chen, X. 2015b. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. *In Proceedings of International Conference on Machine Learning*, 720–729.

[4] Kusner, M.; Sun, Y.; Kolkin, N.; and Weinberger, K. 2015. From word embeddings to document distances. *In Proceedings of International Conference on Machine Learning*, 957–966.

[5] Li, P.;Wang, Q.; and Zhang, L. 2013. A novel earth mover’s distance methodology for image matching with gaussian mixture models. *In Proceedings of IEEE International Conference on Computer Vision*, 1689–1696.

[6] Rubner, Yossi, Carlo Tomasi, and Leonidas J. Guibas. 2000. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40(2): 99-121.

[7] Wang, F., and Guibas, L. J. 2012. Supervised earth movers distance learning and its computer vision applications. *In European Conference on Computer Vision*, 442–455. Springer.

[8] Wang, R.; Guo, H.; Davis, L. S.; and Dai, Q. 2012. Covariance discriminative learning: A natural and efficient approach to image set classification. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2496–2503. IEEE.

Thanks!