Nonlinear Ranking Loss on Riemannian Potato Embedding

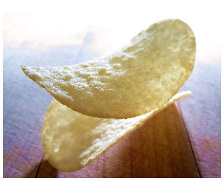
Byung Hyung Kim[1], Yoon-Je Suh[1], Honggu Lee[2], Sungho Jo[1*]

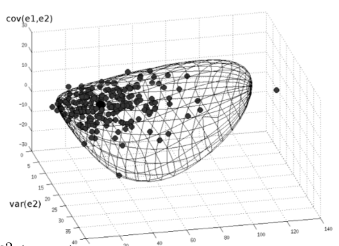
[1]KAIST, Republic of Korea
[2]Looxid Labs, Republic of Korea
https://nmail.kaist.ac.kr, [*]Corresponding author: shjo@kaist.ac.kr

Abstract / Introduction

We propose a rank-based metric learning method by leveraging a concept of the Riemannian Potato for better separating non-linear data. By exploring the geometric properties of Riemannian manifolds, the proposed loss function optimizes the measure of dispersion using the distribution of Riemannian distances between a reference sample and neighbors and builds a ranked list according to the similarities. We show the proposed function can learn a hypersphere for each class, preserving the similarity structure inside it on Riemannian manifold. As a result, compared with Euclidean distance-based metric, our method can further jointly reduce the intra-class distances and enlarge the inter-class distances for learned features, consistently outperforming state-of-the-art methods on three widely used non- linear datasets.

Motivation





$$z_t = \frac{\log(d_t/\mu_t)}{\log(\sigma_t)} \qquad \bar{\Sigma} = \arg\min_{\Sigma \in \mathcal{M}} \sum_{i=1}^{N_I} \delta_R^2(\Sigma_i, \Sigma)$$

$$\mu_t = \exp(\frac{1}{t}\sum_{i=1}^{t} \log(d_i)) \qquad \sigma_t = \exp(\sqrt{\frac{1}{t}\sum_{i=1}^{t}(\log(d_i/\mu_i))^2}$$

Experimental Results

Table 1. Comparison with the state-of-the-art methods on DEAP, AFEW, and HDM05..

Method	DEAP-4				AFEW				HDM05			
	F1	R@1	R@3	NMI	F1	R@1	R@3	NMI	F1	R@1	R@3	NMI
DSML-Triplet	38.7	35.5	37.8	36.1	29.3	31.3	34.2	31.9	53	57.3	58.5	52.4
Triplet-Random	33.5	31.4	32.5	28.7	25.8	25.1	25.4	27.4	48.3	44.5	47.5	45.6
Triplet-Semihard	35.5	30.1	31.4	27.3	27.4	24.4	30.5	28	51.3	50.4	55.3	47.5
Lifted Struct	35.3	35.2	35.8	33.4	32.5	35.4	38.4	39.4	55.5	59.3	59.2	53.4
N-pair-mc	41.5	38.4	39.5	34.8	34.4	33.5	36.4	35.1	59.8	60	61	59.4
Proxy NCA	39.8	41.3	41.4	**38.1**	34.5	34.2	36	36.3	59	63.3	64.5	62
NRA	42.2	44.4	**46.2**	37.2	35.2	**36.5**	38.6	**36.8**	59.2	64.3	65.2	64.1
RPL	**43.3**	**44.7**	**46.2**	37.5	**36.4**	**36.5**	**39.4**	36.4	**59.4**	**66.7**	**68.8**	**65.4**

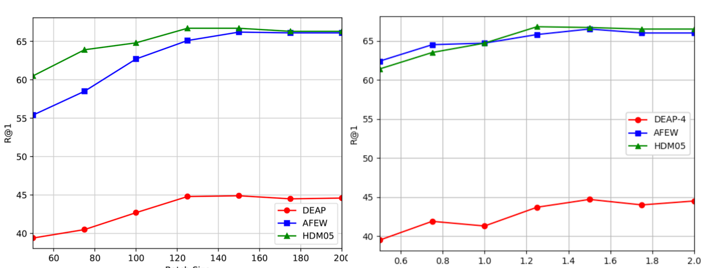


Figure 2. Effect of the proposed RPL

Methods

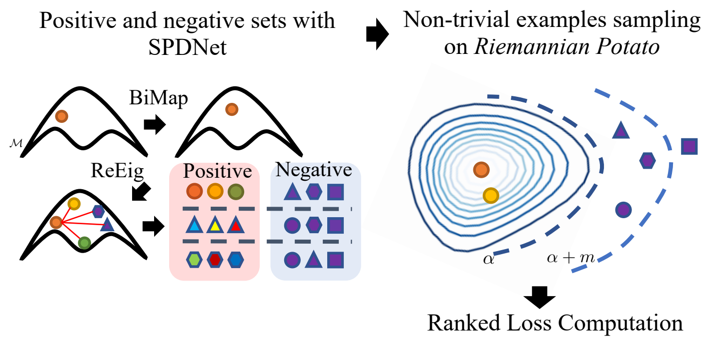


Figure 1. Overview

Riemannian Potato-based Ranking Loss (RPL)

- Our goal is to pull positive points closer than the potato-shaped region of acceptability (z-score) and push negative points out of the boundary.

Algorithm 1 Riemannian Potato-based Ranking Loss

Input: $\{\{z_i^c\}_{i=1}^{N_c}\}_{c=1}^{C} = \{(\Sigma_i, y_i)\}_{i=1}^{N}$, the embedding function $f(\cdot)$, the learning rate β_L and β_δ

Output: Updated $f(\cdot)$

1: **for** all embeddings $f(z_i^c) \in \{\{f(z_i^c)\}_{i=1}^{N_c}\}_{c=1}^{C}$ **do**
2: Sample less trivial positive points in $\mathcal{P}_{c,i}^{*}$.
3: Sample less trivial negative points in $\mathcal{N}_{c,i}^{*}$.
4: Compute the joint loss in Eq. (17).
Update geometric statistics of the RP in Eq.(19) $\sim$ (21).
5: **end for**
Compute the averaged loss in Eq. (18).
Compute the gradient $\nabla f = \partial \bar{L}_{RP}(\Sigma_i; f)/\partial f$.
6: **return** $f(\cdot) = f(\cdot) - \beta_\delta \cdot \nabla f$

$$\hat{\mathcal{P}}_i^c = \{\forall \Sigma_j | j \neq i \wedge c = y_i = y_j, z_j^c > z_{th}\}$$

$$\hat{\mathcal{N}}_i^c = \{\forall \Sigma_j | c = y_i, y_i \neq y_j, z_j^c < z_{th} + m\}$$

$$L_P(\Sigma_i, y_i; f) = \frac{1}{|\hat{\mathcal{P}}_i^c|} \sum_{\Sigma_j \in \hat{\mathcal{P}}_i^c} L(\Sigma_i, \Sigma_j, y_i; f)$$

$$L_N(\Sigma_i, y_i; f) = \frac{1}{|\hat{\mathcal{N}}_i^c|} \sum_{\Sigma_j \in \hat{\mathcal{N}}_i^c} L(\Sigma_i, \Sigma_j, y_i; f)$$

$$L_{RP}(\Sigma_i, y_i; f) = L_P(\Sigma_i, y_i; f) + \lambda L_N(\Sigma_i, y_i; f)$$

$$\bar{L}_{RP} = \frac{1}{N} \sum L_{RP}(\Sigma_i, y_i; f),$$

Conclusion

We proposed a rank-based metric learning method for learning discriminative embeddings and showed the efficacy on classifying non-linear data, reducing the intra-class distances and enlarging the inter-class distances for learned features. Our next work will study the non-stationary nature of brain activity as revealed by EEG, which has been subject to noises from various artifacts, low signal-to-noise ratio (SNR) of sensors, and inter- and intra-subject variability.

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021
KAIST School of Computing