SSL using Improved Unsupervised Discriminant Projection(UDA)

- Unsupervised Discriminant Projection(UDA) is a dimension reduction algorithm.
- This paper uses UDA as regularization which utilize both local and non-local data distribution for semi-supervised learning.
- · Original version: applicable for small scale dataset only.

Basic idea of UDP

• Direction of projection, w based on the ratio of local scatter to non-local scatter. Therefore, the optimal projection:

$$w^* = arg\,min J(w) = rac{J_L}{J_N}$$
 $J_L = Local\,scatter, J_N = Non-local\,scatter$

· Means we want:

$$J_L\downarrow J_N\uparrow$$

Close data closer, distant data more distant.

Mathematical representation

Some notations before diving deeper

• The weighted adjacency matrix(RBF kernel), Kij indicates closer data larger value:

$$K_{ij} = egin{cases} exp(-||x_i - x_j||^2/t) & if \ x_j \ is \ among \ K \ nearest \ neighbors \ of \ x_i \ or \ x_i \ is \ among \ K \ nearest \ neighbors \ of \ x_j \ othervise \end{cases}$$

Hij:

$$H_{ij} = egin{cases} K_{ij} & if \, x_j \, is \, among \, K \, nearest \, neighbors \, of \, x_i \ & or \, x_i \, is \, among \, K \, nearest \, neighbors \, of \, x_j \ 0 & otherwise \end{cases}$$

Original version

Note: y_i and y_j are projected samples Local scatter:

$$J_L(w) = rac{1}{MM} \sum_{i=1}^{M} \sum_{j=1}^{M} K_{ij} (y_i - y_j)^2$$

Multiplying K ij involves distance measures of neighbouring samples only

Non-local scatter:

$$J_N(w) = rac{1}{MM} \sum_{i=1}^M \sum_{j=1}^M (K_{ij} - H_{ij}) (y_i - y_j)^2$$

Multiplying (K_ij-H_ij) involves distance measures of distant samples only

Improved version

The local scatter is same as original version. Non-local scatter is defined as below:

$$J_D = rac{1}{m} \sum_{i=1}^{M} W_{ij} ||y_i - y_j||_2^2$$

where Wij is defined as below:

$$W_{ij} = egin{cases} exp(-||x_i - x_j||^2/t) & if \ x_j \ is \ among \ N \ furthest \ samples \ from \ x_i \ or \ x_i \ is \ among \ N \ furthest \ samples \ from \ x_j \ othervise \end{cases}$$

The objective function of the improved UDP:

$$J_R(W) = rac{J_L}{J_D} = \sum_{i=1}^{M} rac{\sum_{j \in U^K} H_{ij} ||y_i - y_j||_2^2}{\sum_{b \in D^N} W_{ij} ||y_i - y_b||_2^2}$$

Using UDP in semi-supervised learning

The objective function:

$$J = \sum_{i=1}^{L} l(f_i, y_i) + \lambda \sum_{i=1, \ j \in U^K, \ k \in D^N}^{L+U} J_R(g_i, g_j, g_k, H_{ij}, W_{ik})$$

I(.)----labeled loss
JR-----UDP regularization term
gi,gj,gk-----embeddings of the samples through deep network
Uk=local set, DN=distant data set

Conclusion

Similarity between two samples are measured using Euclidean distance.

- **Weighted adjacency matrix** is used to measure the degree of "nearest" of K neighbors, instead of labeling all K neighbors as identical neighbours.
- Scatter function is related to Laplacian(refer Laplacian note).

Reference

• Improved UDP Paper

Doubts

- Unlabeled data->feed into deep network->gi,gj,gk->UDP?
- What is the difference between improved distant scatter function and original distant scatter function?

Author

Author: Lee Zhicheng Date: 02/05/2020