Consensus-Driven Propagation in Massive Unlabeled Data for Face Recognition

2020.5.20

CUHK - SenseTime Joint Lab   ECCV2018

Content

motivation

Face recognition has witnessed great progress in recent years, mainly attributed to the high-capacity model designed and the abundant labeled data collected. However, it becomes more and more prohibitive to scale up the current million-level identity annotations

The following problem is that how to make the unlabeled data effective as labeled data ?

challenges

Intuition : According the idea of semi-supervised, Cluster the unlabeled data

However, the setting of face recognition is difference from semi-supervised in tow aspects

First, the unlabeled data are collected from unconstrained environments, where pose, illumination, occlusion variations are extremely large. It is non-trivial to reliably compute the similarity between different unlabeled samples.

Second, there is usually no identity overlapping between the collected unlabeled data and the existing labeled data. Thus, the popular label propagation is no longer feasible here.

Corresponding solution

One key insight here is that although unlabeled data do not provides us with the straightforward semantic classes, its inner structure, which can be represented by a graph, actually reflects the distribution of high-dimensional face representations.

With the graph, we can sample instances and their relations to establish an auxiliary loss for training our model.

Our Approach

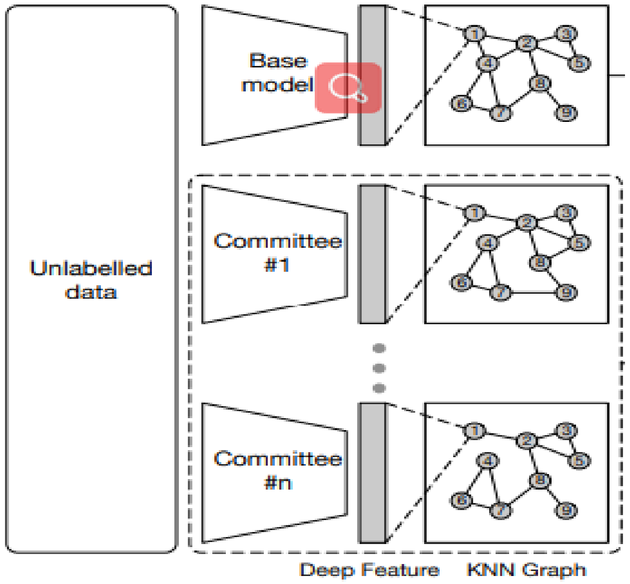
Author propose a novel Consensus-Driven Propagation (CDP) approach .It consists of three modules: Base model, committee model, mediator model.

i. Train base model and committee model

Use labeled data to train base model and committee model, the committee model consists N different classification models.

ii. Building K-NN Graphs.

we feed the base model and committee model with unlabeled data and extract deep features With the features build N+1 versions of k-NN graphs.



Our Approach

iii. Collecting opinions from Committee

For every candidate (a pair connected together) from KNN graph of base model, N Committee will give three factors

1) The relationship

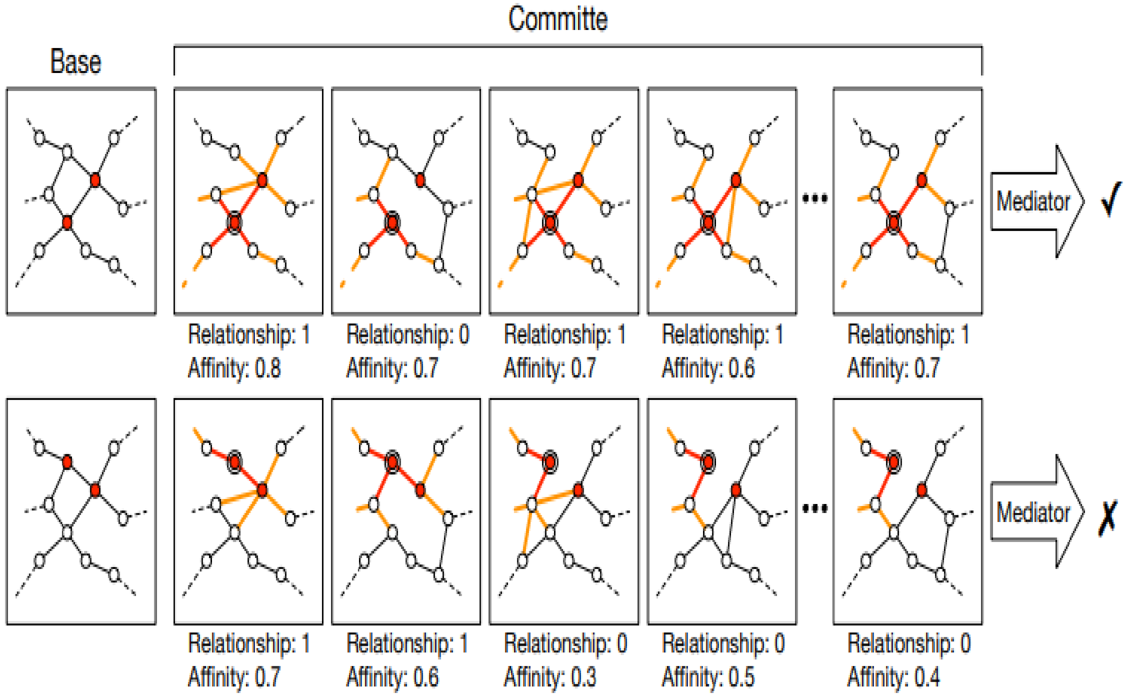
$$R_{C_i}^{(n_0,n_1)} = \begin{cases} 1 & \text{if } (n_0, n_1) \in \mathcal{E}(\mathcal{G}_{c_i}) \\ 0 & \text{otherwise.} \end{cases}, \quad i = 1, 2, \ldots, N, \qquad (1)$$

2) The affinity

$$A_{C_i}^{(n_0,n_1)} = \cos\left(\langle \mathcal{F}_{C_i}(n_0), \mathcal{F}_{C_i}(n_1) \rangle\right), \quad i = 1, 2, \ldots, N. \qquad (2)$$

3) The local structures

$$D_{C_i}^{x} = \{\cos\left(\langle \mathcal{F}_{C_i}(x), \mathcal{F}_{C_i}(x_k) \rangle\right), k = 1, 2, \ldots, K\}, \quad i = 1, 2, \ldots, N. \qquad (3)$$



Our Approach

iV. Aggregate Opinions via Mediator.

The role of a mediator is to aggregate committee members' opinions for pair selection. The author formulate the mediator as a Multi-Layer Perceptron (MLP) classifier

The input to the mediator for each pair (n0, n1) is a concatenated vector containing three parts

1) "relationship vector" $I_R \in \mathbb{R}^N$: $I_R = \left(\dots R_{C_i}^{(n_0,n_1)} \dots\right), i = 1, 2, \dots, N$, from the committee.

2) "affinity vector" $I_A \in \mathbb{R}^{N+1}$: $I_A = \left(\dots A_{C_i}^{(n_0,n_1)} \dots\right), i = 0, 1, 2, \dots, N$, from both the base model and the committee.

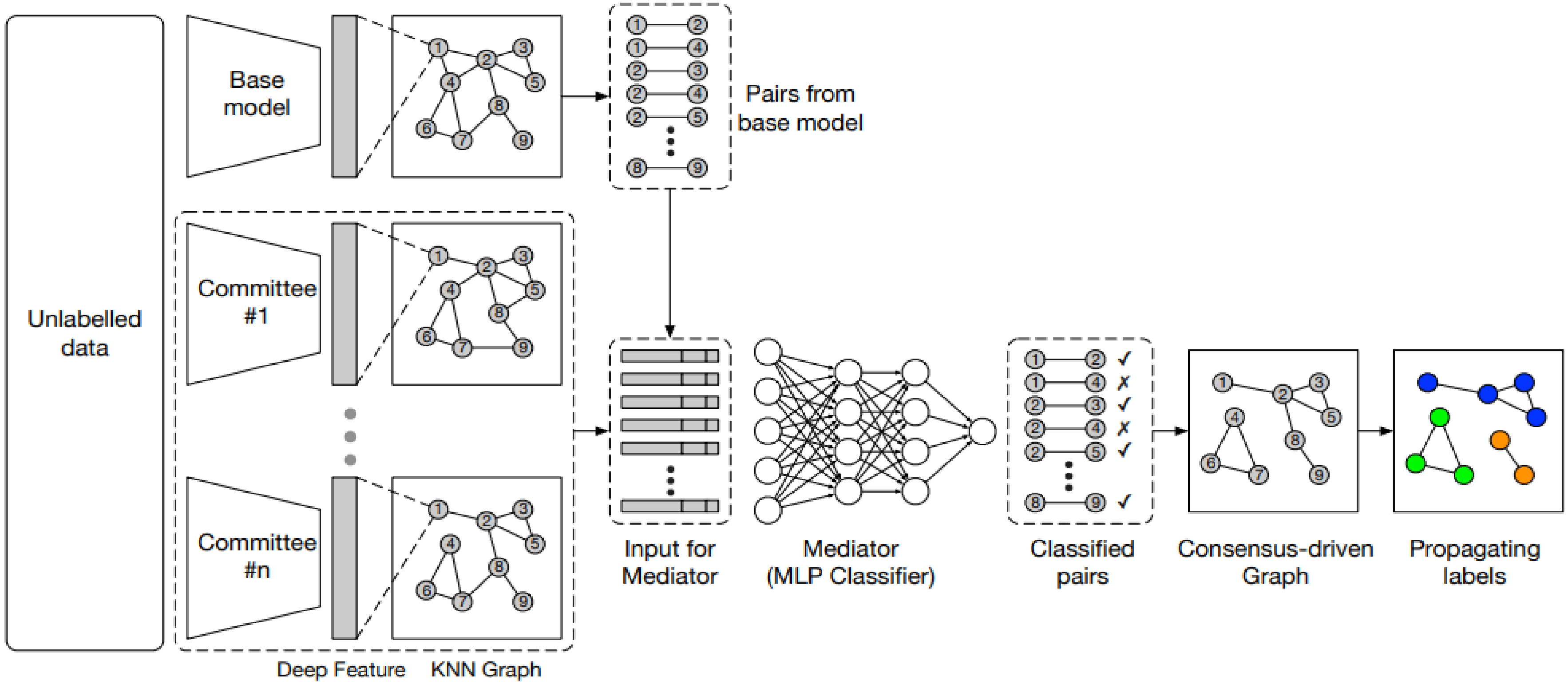
3) "neighbors distribution vector" including "mean vector" $I_{D_{mean}} \in \mathbb{R}^{2(N+1)}$ and "variance vector" $I_{D_{var}} \in \mathbb{R}^{2(N+1)}$:

$$\begin{aligned} I_{D_{mean}} &= \left(\dots E\left(D_{C_i}^{n_0}\right) \dots, \dots E\left(D_{C_i}^{n_1}\right) \dots\right), i = 0, 1, 2, \dots, N, \\ I_{D_{var}} &= \left(\dots \sigma\left(D_{C_i}^{n_0}\right) \dots, \dots \sigma\left(D_{C_i}^{n_1}\right) \dots\right), i = 0, 1, 2, \dots, N, \end{aligned} \tag{4}$$

The mediator is trained on labeled data , and the objective is to minimize the corresponding Cross-Entropy loss function

Our Approach

V. Pseudo Label Propagation

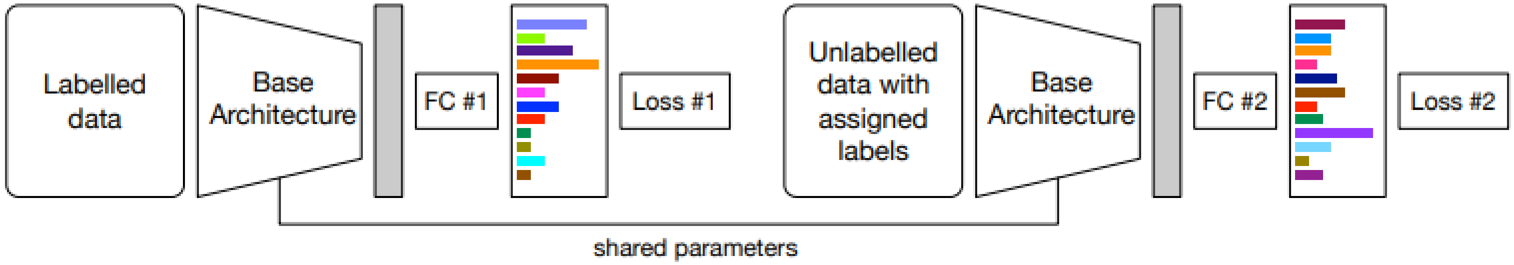


Finally, we propagate labels in the graph, and the propagation for each category ends by recursively eliminating low-confidence edges.

Our Approach

VI Joint Training using Labeled and Unlabeled

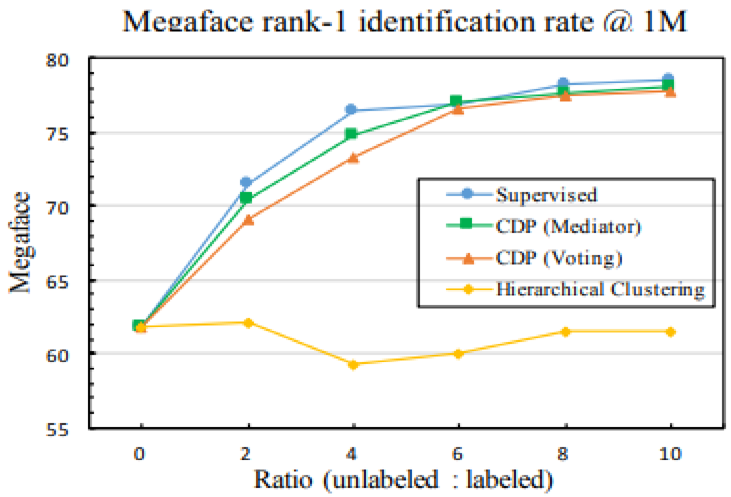
After building pseudo labels, we can use pseudo labeled data and labeled data to train the base model. Since the identity intersection of two data sets is unknown, the author formulate the learning in a multi-task training.

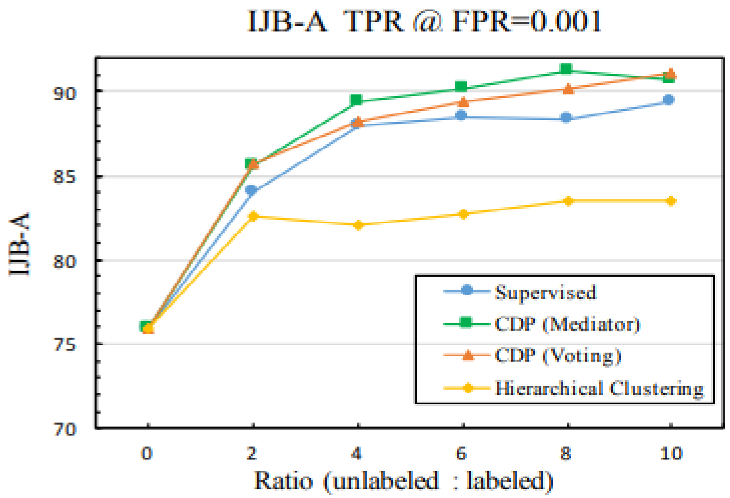


The loss function is same as me as the one for training the base model and committee model

Experiment

The author splits the dataset into 11 balanced parts, one part is labeled data others is unlabeled





1. CDP obtains significant and steady improvements given different quantities of unlabeled data

2. CDP surpasses the baseline “Hierarchical Clustering” by a large margin, obtaining competitive or even better results over the fully-supervised counterpart.

3.CDP by the “mediator” performs better than by naive voting.

4.In the IJB-A face verification task, both settings of CDP surpass the fully supervised counterpart

Experiment

The method is robust in pinpointing wrongly annotated faces (group 1), extremely low-quality faces (e.g., heavily blurred face, cartoon in group 2), which do not help training.

Conclusion

1. CDP can scale up data set at low cost and it may be applied in other classification task

2. Maybe CDP could be a way to clean data, according the result on IJB-A data set

Thanks