

# CVPR2019 | 人脸聚类——Linkage Based Face Clustering via GCN

Eudora 极市平台 4月5日

加入极市专业CV交流群，与**6000+**来自**腾讯，华为，百度，北大，清华，中科院**等名企名校视觉开发者互动交流！更有机会与**李开复老师**等大牛群内互动！

同时提供每月大咖直播分享、真实项目需求对接、干货资讯汇总，行业技术交流。点击文末“[阅读原文](#)”立刻申请入群~

CVPR2019 accepted list ID已经放出，极市已将目前收集到的公开论文总结到github上（目前已收集343篇），后续会不断更新，欢迎关注，也欢迎大家提交自己的论文：

<https://github.com/extreme-assistant/cvpr2019>

今天分享一篇人脸聚类的CVPR论文~

作者 | Eudora

来源 | <https://zhuanlan.zhihu.com/p/60796909>

做过聚类的小伙伴通常都会有一个头大的问题，就是“**如何确定距离**”。这个距离是用于确定两个节点是否有连接的，它为什么那么让人头大呢？因为距离阈值设置小了，很多linkage就被断开了，导致较低的recall；而设大了，就有大量错误的linkage，precision就没法保证了。

本次介绍的工作主要就是为了解决这个问题，这是CVPR 2019的一个文章“Linkage Based Face Clustering via Graph Convolution Network”，就是通过**利用GCN**来识别graph的linkage关系的！链接如下：

<https://arxiv.org/pdf/1903.11306.pdf>

下面主要从以下几个方面来介绍这篇文章：

1. 简介
2. 方法介绍

a. Framework overview

b. Instance Pivot Subgraph (IPS) & GCN

3. 结果分析

4. Pros & Cons

## 1. 简介——人脸聚类存在的问题，以及文章的解决方法

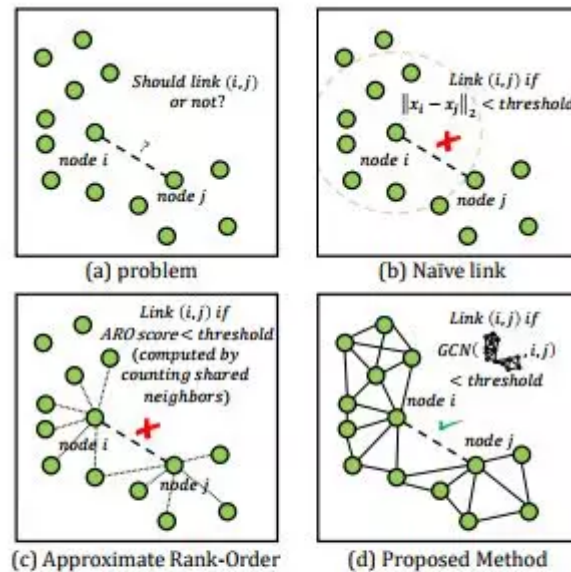


Figure 1. High-level idea of our method. (a): This paper aims to estimate whether two nodes should be linked. (b-d): Comparison of three linkage estimation methods. (b): Directly thresholding the  $l_2$  distance, *without considering context*. (c): Using a heuristic method for linkage estimation *based on context*. (d): Our method, *i.e.*, learning the linkage likelihood with a parametric model, which is *context-based*.

人脸聚类中一个主要的难题——link的判断，在figure 1给了很清晰的阐述：

- (a) 文中focus的问题是要判断两个节点i和j是否有link。
- (b) 朴素的想法是判断i、j之间的距离是否小于一定的阈值，如果不是的话，则两个节点之间不会有link。
- (c) ARO方法通过人为设定的规则来确定是否应该有link，这种手工设计的方式还是不够准确。
- (d) 文中提出的方法是通过GCN获取i、j 的neighbor信息，并学习两者之间是否有link。

这种方法不需要依赖于一个全局的阈值如 (a) 或者人为设定的局部阈值如 (c)，可以让算法自己来观察当前的context，从而确定结果，这会使得linkage更加robust。

## 2. 方法介绍

### (1) Framework Overview

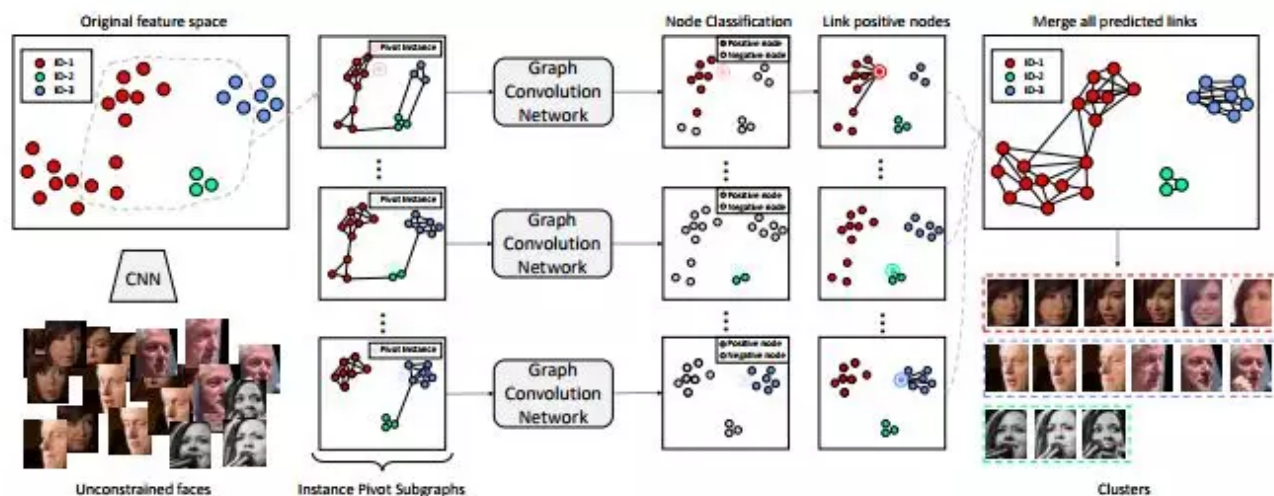


Figure 3. Framework of our method. We use every instance as a pivot, and construct an Instance Pivot Subgraph (IPS) for it. Each node in IPS models the linkage (similarity) between the pivot and the corresponding neighbor. We apply graph convolutions on IPS and classify nodes in IPS into either positive or negative. If a node is classified as positive, the corresponding neighbor should be linked to the pivot. After determining all the linkages, we transitively merge linked instances to obtain the final clusters.

该文章提出的方法主要包括以下几个步骤：

(a) 首先，是通过CNN来提取feature，这时候我们可以看到，feature虽然是有分开大致的group，但是对于一些比较困难的类来说，feature会分布得很散（这个对应于实际情况中的遮挡、模糊等情况）；另外常见的困难情况，图中没有画出来——有些类别之间会非常的相似，有一些overlap的部分。

(b) IPS：对于每个样本点，构建其对应的子图，文中称为Instance Pivot Subgraph（IPS）。对于这一步我们可以看做是对每个样本找出它的n-hop neighbor，为了给后面的节点分类提供context信息。

(c) Node Classification：把每个子图送进GCN，并判断出哪个样本与子图中的pivot是属于同类的。这个是GCN常见的一个用途——分类，用得挺到位的。

(d) 连接该有的link，并找出其中的连通图，这样就得到聚类的结果了！

## (2) Instance Pivot Subgraph (IPS) & GCN Node Classification

IPS+GCN是该文章的重点模块。

IPS的建立方式有以下几个步骤，如下图：

- (a) 首先，对于pivot node找出其子图，文中设定的子图节点包括pivot的1\2\3- hop neighbors；
- (b) 接着，对节点特征做normalization，也就是每个节点的feature都减去pivot的feature，这步可以理解为以pivot为中心，观察其附近节点的变化(context)；
  - （这点非常重要，相当于定义了图/类的中心，后面相当于对节点进行二分类，使得识别的难度大大降低。）
- (c) 对子图中的每个节点，连接其u近邻。
  - （这一步也就得到了GCN需要输入的邻接矩阵，可以直接应用GCN对子图的节点进行分类了。）

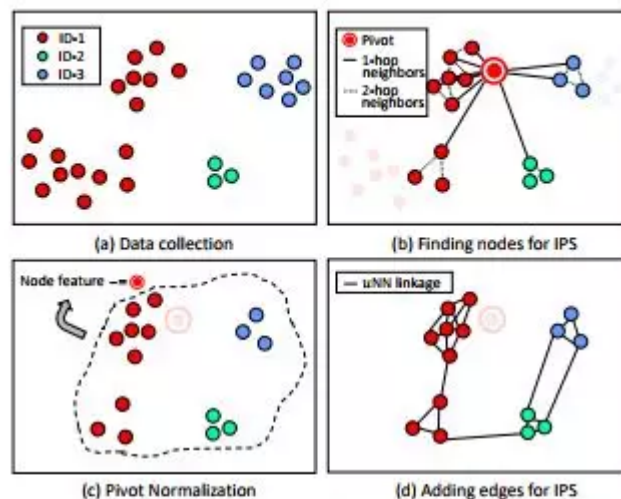


Figure 2. Construction of Instance Pivot Subgraph (IPS). (a) The collection of face representations. (b) We use each instance  $p$  as a pivot, and find its neighbors up to  $h$ -hop as the nodes of IPS. (c) These node features are normalized by subtracting the feature of the pivot. (d) For a node in IPS, we find its  $u$ NNs from the entire collection. We add an edge between a node and its  $u$ NNs if the neighbor is also a node of IPS. In this figure, we set  $h = 2$ ,  $k_1 = 10$ ,  $k_2 = 2$  and  $u = 3$ , where  $k_1$  is the number of 1-hop neighbors, and  $k_2$  is the number of 2-hop neighbors. Note that an IPS based on pivot  $p$  does not contain  $p$ . The IPS for pivot  $p$  is used to predict the linkage between  $p$  and every node in IPS.

### 3. 实验结果分析

(1) 方法比较，文章提出的方法，可见对于比heuristic的方法有比较明显的提升的。

Method	IJB-B-512		IJB-B-1024		IJB-B-1845	
	F	NMI	F	NMI	F	NMI
K-means [24]	0.612	0.858	0.603	0.865	0.600	0.868
Spectral [28]	0.517	0.784	0.508	0.792	0.516	0.785
AHC [15]	0.795	0.917	0.797	0.925	0.793	0.923
AP [11]	0.494	0.854	0.484	0.864	0.477	0.869
DBSCAN [10]	0.753	0.841	0.725	0.833	0.695	0.814
ARO [27]	0.763	0.898	0.758	0.908	0.755	0.913
PAHC* [22]	-	-	0.639	0.890	0.610	0.890
ConPaC* [29]	0.656	-	0.641	-	0.634	-
DDC [22]	0.802	0.921	0.805	0.926	0.800	0.929
GCN-M	0.821	0.920	0.819	0.928	0.810	0.927
GCN-W	0.826	0.923	0.822	0.938	0.801	0.927
GCN-A	<b>0.833</b>	<b>0.936</b>	<b>0.833</b>	<b>0.942</b>	<b>0.814</b>	<b>0.938</b>

Table 2. Comparison with baseline methods in terms of BCubed F-measure and NMI score. For all methods we tune the corresponding hyperparameters and report the best result. Suffix M, W, and A represents different aggregators. The superscript \* denotes results reported from the original papers, otherwise all methods use the same ArcFace representation.

(2) 聚类中常见的Singleton对效果的影响。

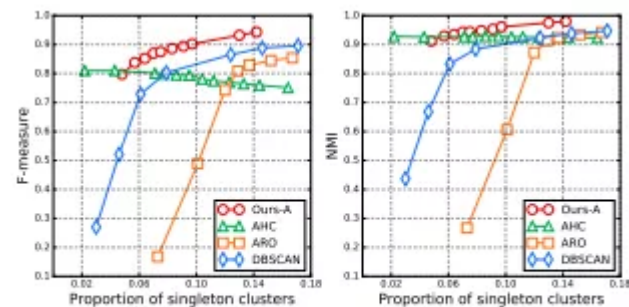


Figure 6. Performance on IJB-B-512 after removing singleton clusters v.s. proportion of removed singleton clusters.

(3) Multi-view clustering

该方法可以直接应用于Multi-view clustering上，可见也是有不错的提升的。



文章不仅验证了该方法在同质数据上的提升，还给出了异质数据上的结果，这也是不错的一点。

Method	Face		Audio		Face+Audio	
	F	NMI	F	NMI	F	NMI
K-means [24]	0.648	0.877	0.229	0.644	0.636	0.874
Spectral [28]	0.592	0.825	0.214	0.619	0.541	0.782
AHC [15]	0.755	0.913	0.358	0.704	0.833	0.934
ARO [27]	0.575	0.875	0.261	<b>0.834</b>	0.319	0.835
Ours	<b>0.801</b>	<b>0.921</b>	<b>0.395</b>	0.497	<b>0.841</b>	<b>0.940</b>

Table 5. Clustering accuracy with 512 identities.

Method	Face		Audio		Face+Audio	
	F	NMI	F	NMI	F	NMI
K-means [24]	0.589	0.871	0.152	0.650	0.582	0.871
AHC [15]	0.695	0.908	0.228	0.686	0.785	0.938
ARO [27]	0.583	0.858	0.277	<b>0.813</b>	0.370	0.873
Ours	<b>0.766</b>	<b>0.932</b>	<b>0.311</b>	0.452	<b>0.810</b>	<b>0.946</b>

Table 6. Clustering accuracy with 2,048 identities.

## 4. 总结

### Pros:

- (1) 用GCN来解决Clustering中hard core的linkage问题，方法比较新颖——GCN中利用到了数据的局部信息，能够更准确地判断节点之间的关系；
- (2) 构建IPS的想法很好，巧妙地利用了GCN做节点分类的特性，并且使用Pivot Normalization很好地定义了节点的主次关系，能更好地进行分类。

### Cons:

- (1) 在scalability和efficiency方面尚有欠缺，文中对于每个样本都会建立一个IPS，并在上面应用GCN，虽然使用的GCN计算量不大，但毕竟在大规模数据上还是吃不消的，这个方面还是有比较大的扩展空间；
- (2) 看到文章中提到Singleton的问题时，我是挺期待作者能描述一下他们是如何把这些singleton归类的，毕竟在实际问题下，这种情况实在是太太太常见了！

据我所知，近期还有不少GCN + Clustering的工作会陆续出来，坐等！

### \*延伸阅读

- CVPR2019 | 03-28日更新11篇论文及代码汇总（1篇Oral，含目标检测、人脸聚类、VQA等）
- [CVPR2019专题直播 | 王鑫龙：联合分割点云中的实例和语义](#)
- [CVPR 2019 Oral | 京东AI研究院提出 ScratchDet：随机初始化训练SSD目标检测器](#)

点击左下角“[阅读原文](#)”，即可申请加入极市[目标跟踪](#)、[目标检测](#)、[工业检测](#)、[人脸方向](#)、[视觉竞赛](#)等技术交流群，更有每月大咖直播分享、真实项目需求对接、干货资讯汇总，行业技术交流，一起来让思想之光照的更远吧~



觉得有用麻烦给个好看啦~

阅读原文