Effective and Generalizable Graph-Based Clustering for Faces in the Wild

Nisim Hurst

Friday 3 May 2019

Effective and Generalizable Graph-Based Clustering for Faces in the Wild

The article was written by (Chang, Perez-Suarez, and Gonzalez-Mendoza 2017). The task performed was face clustering, i.e. identifying unique face identities. They used Pairwise Fmeasure and BCubed Fmeasure metrics to measure the external quality of the clusters.

Hypothesis

A 128-dimensional face representation obtained by a pretrained ResNet combined with an improved Chinese Whispers (CW) graph clustering algorithm produce results comparable to commercial solutions.

Evidence and Results

Evidence is restricted to the main contribution of the paper, i.e. the idea that cluster number can be reduced by using a higher level of abstraction and reapplying CW twice. Evidence is presented in terms of cluster quality and duration. Three approaches besides the one proposed are used for comparison, namely k-means, GLC and Approximate Rank-order. The main comparison algorithm is the Approximate Rank-order, the rest are used as common baseline to previous works. The results of ConPac were also copied directly from the paper given that no public implementation is available.

Dataset

The authors tested four datasets. The four datasets cover a wide range of variations on expression, illumination, pose, resolution, background, occlusions and resolution. The main characteristics of these datasets are shown in Table 1.

Table 1: Main characteristics of the four datasets that were used to test the improved CW.

	# Instances	Resolution	Scenery	Author
LFW	13233 images of 5749. Only 1680 subjects have two or more photos.	??, variable head angle	Color, different Poses and Backgrounds.	(**huang2008**)
YTF	3425 videos of 1595 subjects.	100x100, variable enclosing area	Color, different Poses and Backgrounds.	(**wolf2011**)
EYaleB	1710 images of 38 subjects.	168x192	Black and White, Frontal.	(**sanchez2006**)
AR	3200 of 126 subjects.	120x165	Color, Frontal.	(**mart1998**)

Results

First, the authors provide clusters quality evaluation using *Pairwise FMeasure* and *BCubed FMeasure* for

each approach. These results are contained in a table per each dataset. They also provide a third column with the number of clusters found by each approach.

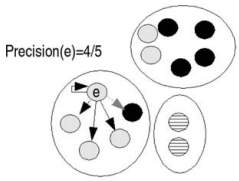
The *BCubed FMeasure* (**amigo2009**) is given by the following formula:

$$\text{BCubed-FMeasure} = 2 \cdot \frac{\text{BCubed-Precision} \cdot \text{BCubed-Recall}}{\text{BCubed-Precision} + \text{BCubed-Recall}}$$

Let $L(e)$ and $C(e)$ denote the category (true cluster) and the cluster (predicted cluster) of an item e. Then, correctness is calculated using:

$$\text{Correctness(e,e')} = \begin{cases} 1 & \text{if } L(e) = L(e') \iff C(e) = C(e') \\ 0 & \text{otherwise} \end{cases}$$

$$\text{BCubed-Precision} = Avg_e \left[\frac{\sum_{C(e)=C(e')} \text{Correctness(e,e')}}{\sum_{C(e)=C(e')} 1} \right]$$



$$\text{BCubed-Recall} = Avg_e \left[\frac{\sum_{L(e)=L(e')} \text{Correctness(e,e')}}{\sum_{L(e)=L(e')} 1} \right]$$

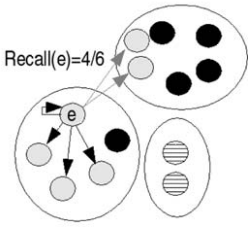


Figure 1: BCubed clustering example.

Equation 1 shows the BCubed FMeasure for clusters shown in Figure 1.

$$
\begin{aligned}
\text{BCubed-Precision} &= \frac{4 * \frac{4}{4} + 1 * \frac{1}{3} + 2 * \frac{2}{3} + 1 * \frac{1}{7} + 1 * \frac{1}{7} + 1 * \frac{1}{7} + 4 * \frac{4}{7}}{14} \\
&= 0.5986394557823128 \\
\text{BCubed-Recall} &= \frac{4 * \frac{4}{5} + 1 * \frac{1}{5} + 2 * \frac{2}{6} + 1 * \frac{1}{1} + 1 * \frac{1}{1} + 1 * \frac{1}{1} + 4 * \frac{4}{6}}{14} \\
&= 0.6952380952380952 \\
\text{BCubed-FMeasure} &= 2 \cdot \frac{0.5986394557823128 \cdot 0.6952380952380952}{0.5986394557823128 + 0.6952380952380952} \\
&= 0.6433328326072805
\end{aligned}
\tag{1}
$$

Second, the threshold is evaluated independently using a range of combinations that cover values around the one that produced the most performance. The authors present the best results from values between 0.25 and 0.55 in steps of 0.05. A single value of 0.40 was found to be the best among the four datasets.

Finally, computation time is shown for each of the approaches and datasets with the best threshold for each. The authors fixed the face descriptor to ResNet-29 to isolate the performance of the clustering algorithm from confounders. Significant improvements were found in the YTF dataset in contrast to the Approximate Rank-order approach.

Contribution

The main contribution of this paper is that it proposes to add 2 post processing steps to the CW graph clustering approach to reduce the number of clusters found. The paper is mainly presented as an extension of the results obtained by (**otto2018**) using Approximate Rank-order.

A second contribution is the introduction of the BCubed metric in the evaluation protocol for face clustering. The BCubed metric meets the four constraints proposed in (**amigo2009**) and weights clusters linearly based on their size advantage. See Figure 2 for a depiction of these constraints.

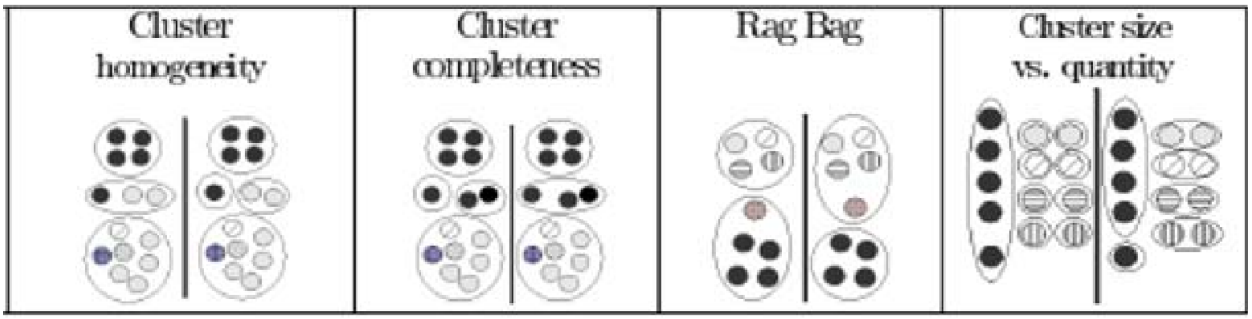


Figure 2: The four constraints proposed by (**amigo2009**).

A third contribution of the paper is that the approach depends on finetuning only two parameters: (1) number of iterations for the CW algorithm and (2) the threshold for building a higher level of cluster abstraction and reapply CW. Of the two parameters, higher values for the former lead the algorithm to converge, so just selecting a high value here will do. For the latter, the authors present the best results from values between 0.25 and 0.55 in steps of 0.05.

Finally, the authors found a single threshold value that performs best for all the datasets under the current face representation. This is a great advantage over previous approaches because in real world scenarios there is no training data to finetune any parameter. By contrast, the paper by (**otto2018**) depends on finetuning its several parameters for the particular scenario.

Weaknesses

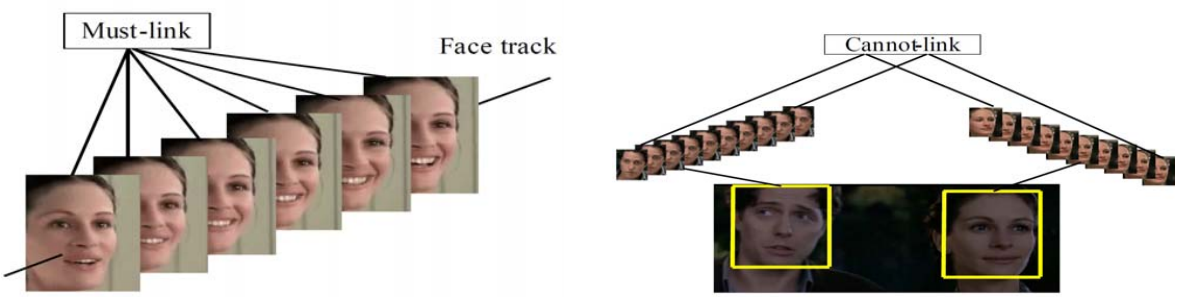
Face embeddings are extracted using a 29-ResNet. Samples for training this primal network are assumed to be independent and identically distributed as the samples from the other 4 datasets. The independence assumption might hold but the primal sample distribution might be in fact biased towards one of the following 4 testing datasets.

Also, the authors hint that some datasets have a great number of outliers, i.e. subjects that have less than two instances. Conclusions derived from the results using such imbalanced datasets must take this fact into account or opt to filter those cases out.

Finally, the authors affirm that the single manually found threshold is scalable to other datasets. However, this threshold is bound to the specific 128D vector representation and probably even to the ResNet architecture. There is no evidence that the aforementioned representation generalize equally well to all datasets in the real world. Thus, this assertion is more general than it should.

Future Work

Chang points out the necessity to enforce pairwise constraints, i.e. must-link and cannot-link relations to improve face clustering accuracy in scenes containing multiple faces, as proposed by (**otto2018**).



(a) Must-link constraint (b) Cannot-link constraint

Figure 3: Pairwise contraints from (**wu2013**)

References

Chang, L., A. Perez-Suarez, and M. Gonzalez-Mendoza (2017). In: 4. DOI: 10.1109/ACCESS.2017.DOI.