CLUSTERING IMAGES BY UNMASKING – A NEW BASELINE

Mariana-Iuliana Georgescu, Radu Tudor Ionescu

Department of Computer Science, University of Bucharest, Romania

Highlights

- We propose a novel agglomerative clustering method based on *unmasking*, a technique that was previously used for authorship verification of text documents [5] and for abnormal event detection in video [4].
- In order to join two clusters, we alternate between:
  1. training a binary classifier to distinguish between the samples from one cluster and the samples from the other cluster;
  2. removing at each step the most discriminant features.
- We conduct experiments on three benchmark data sets in order to compare our clustering method with the k-means algorithm as well as the clustering algorithm presented in [3].
- The empirical results indicate that our approach can significantly outperform the considered baselines.

Method

- Input data and parameters:
  - m training samples;
  - the number of clusters k;
  - the number of unmasking iterations n;
  - the number of features s to be removed at each iteration.
- Our algorithm starts with K clusters ($K \gg k$) and executes the following steps:
  1. For each pair of clusters i and j, we assume that the samples in cluster i belong to a different class than the samples in cluster j and compute a score that indicates the likelihood of the statement *"clusters i and j should be joined"* to be true.
  2. Randomly split the samples in each cluster into a training set and a testing set of equal size.
  3. Train a linear Support Vector Machines (SVM) classifier on the training set (until convergence) and evaluate it on the test set, retaining the accuracy rate.
  4. Sort the weights of the SVM by their absolute values in descending order, take the first s indexes of the sorted list, then remove the corresponding features from all the samples in the training and test sets.
  5. Repeat steps **3** and **4** for n iterations, retaining the accuracy rate at each iteration.
  6. Merge each cluster i with the cluster j (using a Greedy approach), such that the score of joining clusters i and j is maximum, for all $j \in \{1, 2, ..., K\}$, with $j \neq i$.
  7. If the number of clusters k is reached at any point during the merging process, halt the execution and return the current cluster assignments. Otherwise, continue by computing the merging scores for the newly-formed clusters.

Datasets

- **MNIST**. The MNIST database contains **60, 000** train samples and **10, 000** test samples of digits from **0** to **9** (**10** classes).
- **UIUCTex**. The UIUCTex data set contains **1000** texture images of **640** $\times$ **480** pixels representing different types of textures such as bark, wood, floor, water, and more (**25** classes).
- **Oxford Flowers**. The Oxford Flowers data set contains **1360** images, with a number of **80** images per category (**17** classes).

Features

- Deep supervised features provided by the first fully-connected layer (known as *fc6*) of the VGG-f model [2], which is pre-trained on ImageNet (**4096** features).
- Handcrafted features from a standard bag-of-visual-words based on dense SIFT descriptors (**4000** features).
- Deep unsupervised features that are extracted from the pre-trained AlexNet architecture provided by Caron et al. [1]. We extract features from the *conv3* layer of their unsupervised neural network (**3456** features).

Evaluation Metrics

- We report the unsupervised clustering accuracy (ACC) and the Normalized Mutual Information (NMI) score on the test set.

Results on UIUCTex and Oxford Flowers

Features	Method	UIUCTex		Oxford Flowers	
		ACC	NMI	ACC	NMI
-	Random chance	4.00%	-	6.67$	-
VGG-f	SVM	97.20%	-	95.50%	-
	K-means	48.20%	70.15%	60.35%	69.55%
	Unmasking (n=1)	19.80%	54.81%	45.50%	62.98%
	Unmasking	61.40%	74.94%	67.50%	75.82%
BOVW	SVM	94.60%	-	80.83%	-
	K-means	25.40%	46.83%	22.10%	22.83%
	Unmasking (n=1)	35.20%	55.30%	12.83%	14.64%
	Unmasking	44.60%	58.81%	25.00%	25.37%
AlexNet	SVM	96.20%	-	81.00%	-
	K-means	36.80%	58.52%	26.89%	30.43%
	Unmasking (n=1)	34.20%	62.07%	9.80%	18.19%
	Unmasking	48.60%	69.78%	33.33%	38.00%

Table: Clustering performance of various baselines versus clustering by unmasking on the UIUCTex and the Oxford Flowers test sets. Higher ACC or NMI values are better.

Results on MNIST

Method	ACC	NMI
Random chance	10.00%	-
SVM	94.40%	-
K-means	55.82%	52.18%
IDEC [3]	71.45%	69.40%
Unmasking (n=1)	72.58%	64.99%
Unmasking	81.40%	69.76%

Table: Clustering performance of various baselines versus clustering by unmasking on the MNIST test set. Higher ACC or NMI values are better.

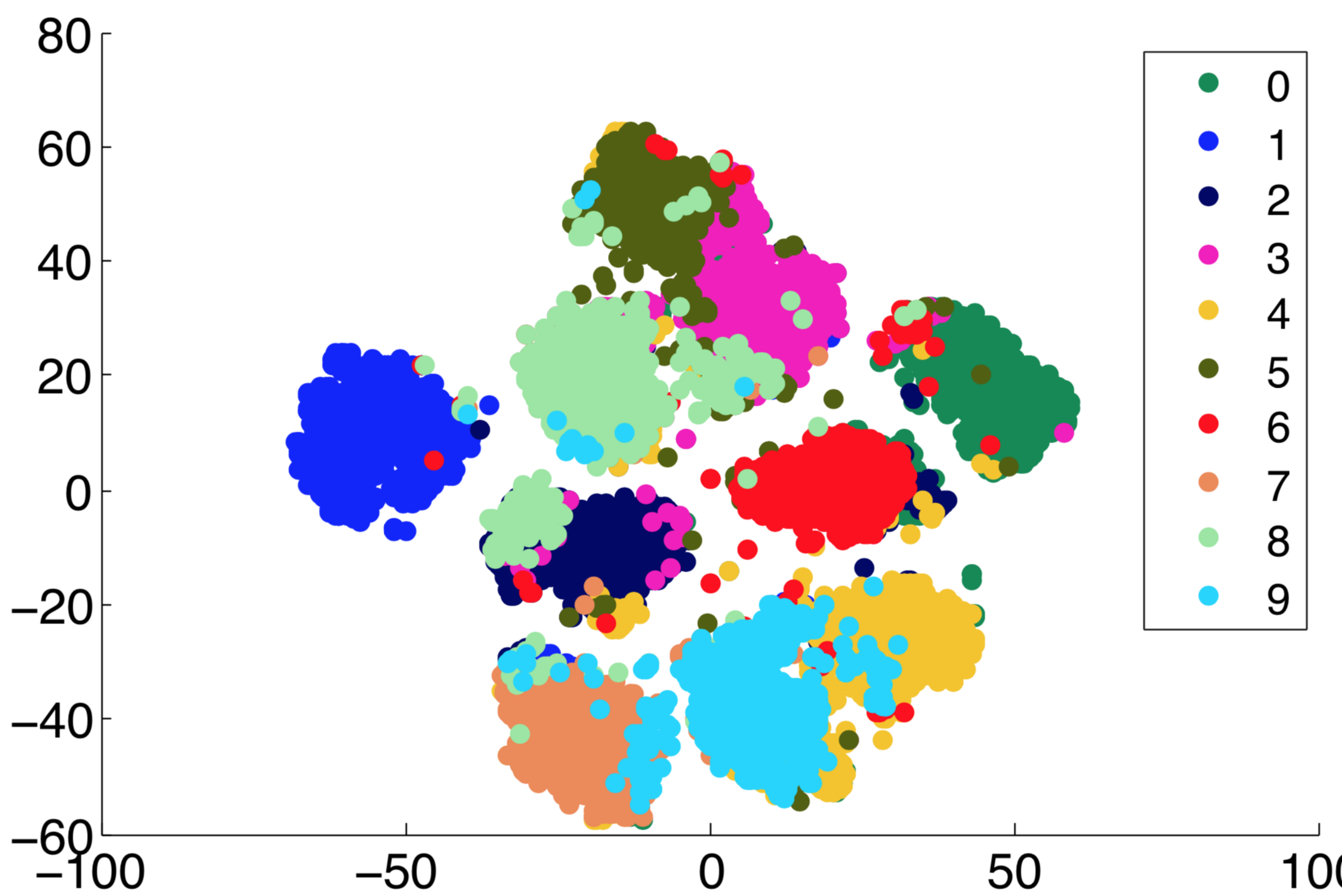

Figure: Visualization (provided by t-SNE) of unmasking-based clustering results on the MNIST test set.

References

[1] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep Clustering for Unsupervised Learning of Visual Features. In *Proceedings of ECCV*, volume 11218, pages 139–156, 2018.

[2] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the Devil in the Details: Delving Deep into Convolutional Nets. In *Proceedings of BMVC*, 2014.

[3] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved Deep Embedded Clustering with Local Structure Preservation. In *Proceedings of IJCAI*, pages 1753–1759, 2017.

[4] Radu Tudor Ionescu, Sorina Smeureanu, Bogdan Alexe, and Marius Popescu. Unmasking the abnormal events in video. In *Proceedings of ICCV*, pages 2895–2903, 2017.

[5] Moshe Koppel, Jonathan Schler, and Elisheva Bonchek-Dokow. Measuring Differentiability: Unmasking Pseudonymous Authors. *Journal of Machine Learning Research*, 8:1261–1276, December 2007.