

Non-metric affinity propagation for unsupervised image categorization

Delbert Dueck and Brendan J. Frey
Probabilistic and Statistical Inference Group
University of Toronto

10 King's College Road, Toronto, Ontario, Canada M5S 3G4

<http://www.psi.toronto.edu>

Abstract

Unsupervised categorization of images or image parts is often needed for image and video summarization or as a preprocessing step in supervised methods for classification, tracking and segmentation. While many metric-based techniques have been applied to this problem in the vision community, often, the most natural measures of similarity (e.g., number of matching SIFT features) between pairs of images or image parts is non-metric. Unsupervised categorization by identifying a subset of representative exemplars can be efficiently performed with the recently-proposed 'affinity propagation' algorithm. In contrast to k-centers clustering, which iteratively refines an initial randomly-chosen set of exemplars, affinity propagation simultaneously considers all data points as potential exemplars and iteratively exchanges messages between data points until a good solution emerges. When applied to the Olivetti face data set using a translation-invariant non-metric similarity, affinity propagation achieves a much lower reconstruction error and nearly halves the classification error rate, compared to state-of-the-art techniques. For the more challenging problem of unsupervised categorization of images from the Caltech101 data set, we derived non-metric similarities between pairs of images by matching SIFT features. Affinity propagation successfully identifies meaningful categories, which provide a natural summarization of the training images and can be used to classify new input images.

1. Introduction

Many vision tasks either produce as output a categorization of input features or require unsupervised categorization of input features as a preprocessing step for subsequent analysis. The features to be categorized may take a wide variety of forms, ranging from raw data such as vectors of pixel intensities [1] to collections of SIFT features representing scale-invariant keypoints [2].

A powerful approach to representing image categories is to identify a relatively small number of images or image fragments, called 'exemplars'. Exemplars have been used with success in a variety of vision tasks, including image synthesis [3, 4], super-resolution [5, 6], image and video completion [7, 8] and combined tracking and object detection [9, 10].

The use of exemplars is attractive for several reasons. A relatively small number of representative exemplars can capture high-order statistics, since each exemplar can simultaneously express dependencies between a large number of image features. In contrast to general statistical methods for which many configurations of the parameters do not correspond to realistic image data, each exemplar is an image or an image fragment so it naturally corresponds to realistic image data. For this reason, exemplars can be used to make realistic predictions for missing image data and avoid the blurring that often occurs when parametric methods are applied. Exemplars are represented efficiently as pointers into the training data (e.g., a subset of image features), so the number of bits of information needing to be specified during exemplar learning is quite small [11].

If N training cases to be categorized are labeled $1, \dots, N$, the exemplar learning problem consists of identifying a set of exemplars and assigning every other training case to an exemplar so as to maximize a fitness function. Denoting the index of the exemplar representing training case i by c_i and the input similarity between training case i and k by $s(i, k)$, the fitness function is $\mathcal{S}(\mathbf{c}) = \sum_{i=1}^N s(i, c_i)$. An example of a simple metric similarity is the negative Euclidean distance between input vectors: $s(i, k) = -||x_i - x_k||^2$, where x denotes an input vector. If training case i is an exemplar, we take $c_i = i$, in which case the fitness function will include a term $s(i, i)$. This value is also taken as input and represents the *a priori* preference that training case i be chosen as an exemplar. Note that $s(i, i)$ is *not* computed in the same way as $s(i, k)$, since it does not represent an assignment similarity. Learning consists of maximizing $\mathcal{S}(\mathbf{c})$ w.r.t \mathbf{c} , subject to the

constraints that for all i , if there exists a $j \neq i$ s.t. $c_j = i$ then $c_i = i$ (i.e., if training case j is assigned to i , then i must be an exemplar). This problem is NP-hard [12], but approximate methods can be used to search for good solutions [13, 14, 15].

As observed in [10], one of the primary advantages of using exemplar-based learning to categorize images or images parts is that the input similarities need not be metric (i.e., need not be symmetric or satisfy the triangle inequality). Non-metric similarity functions that have been used in vision include Chamfer distance [9], Hausdorff distance [16] and shuffle distance [17].

Recently, the ‘affinity propagation’ algorithm was proposed as a new technique for exemplar learning and was applied to machine learning tasks, including clustering face images using Euclidean distance, finding genes using microarray data, document summarization and airline routing [15]. While affinity propagation was shown to achieve larger values of the fitness function than other methods, several questions of specific interest to vision researchers were left open, including 1) Do improvements in the fitness function translate into significant improvements in image classification rates? 2) Affinity propagation was applied to cluster face images using a metric similarity (Euclidean distance) – can it successfully be applied to vision tasks using non-metric measures of similarity? 3) The authors of [15] previously described a different version of affinity propagation [18] — which version obtains better results on vision-related tasks? Here, we seek to answer these questions. In particular, the major contributions of this paper are to provide the first ever comparison between the two versions of affinity propagation; to demonstrate for the first time the use of affinity propagation for unsupervised image categorization with results reported in terms of classification error; to develop useful non-metric similarity measures for unsupervised image categorization — one based on translation-invariance and the other based on matching SIFT features.

2. Background

Most techniques for identifying exemplars (e.g. k -centers clustering) keep track of a fixed set of candidate exemplars while searching for good solutions. The affinity propagation algorithms introduced in [18] and [15], however, simultaneously consider *all* data points as candidate exemplars. As mentioned earlier, given a data set of N training cases, the clustering problem can be concisely described by introducing N state variables, c_1, \dots, c_N , where the index of the exemplar representing training case i is c_i .

The original affinity propagation algorithm [18] was derived as an instance of the loopy belief propagation algorithm on a factor graph reflecting the constraints that no cluster can be without an exemplar and that clusters must contain at least one member in addition to said exemplar

(i.e. no singleton clusters); see [18] for details. The message sent from each c_i variable node to the k^{th} constraint function is referred to as the *responsibility* of cluster k for data point i , denoted by $r(i, k)$. The message sent from each f_k function to each c_i variable is referred to as the *availability* of x_k as a candidate exemplar for x_i . These would typically be N -ary probability vectors, but there are many repeated quantities which, if exploited, make it possible to pass scalar quantities between data points themselves. This leads to the following update equations:

Affinity propagation (NIPS, 2006)

Initialization:

$$r(i, k) = 0, \quad a(k, i) = 0 \text{ for all } i, k$$

Responsibility updates:

$$r(i, k) \leftarrow s(i, k) - \max_{j:j \neq k} (a(j, i) + s(i, j))$$

Availability updates:

$$a(k, k) \leftarrow \max_{j:j \neq k} \min(0, r(j, k)) + \sum_{j:j \neq k} \max(0, r(j, k))$$

$$a(k, i) \leftarrow \min \left(\begin{array}{l} r(k, k) + \sum_{j:j \notin \{k, i\}} \max(0, r(j, k)), \\ -\max_{j:j \notin \{k, i\}} \min(0, r(j, k)) \end{array} \right)$$

Making assignments:

$$c_i^* \leftarrow \underset{k}{\operatorname{argmax}} \quad r(i, k) + a(k, i)$$

For J input similarities, the above update rules take $O(J)$ scalar binary operations per iteration.

In [15], affinity propagation is re-derived using a simpler constraint that drops the restriction on singleton clusters. The use of this new constraint dramatically simplifies the update equations and increases the speed and stability of the algorithm. Most importantly, as shown for the first time in our experimental results, the new algorithm achieves significantly higher fitness function values. The new algorithm is as follows:

Affinity propagation (Science, 2007)

Initialization:

$$r(i, k) = 0, \quad a(k, i) = 0 \text{ for all } i, k$$

Responsibility updates:

$$r(i, k) \leftarrow s(i, k) - \max_{j:j \neq k} (a(j, i) + s(i, j))$$

Availability updates:

$$a(k, k) \leftarrow \sum_{j:j \neq k} \max\{0, r(j, k)\}$$

$$a(k, i) \leftarrow \min \left(0, r(k, k) + \sum_{j:j \notin \{k, i\}} \max\{0, r(j, k)\} \right)$$

Making assignments:

$$c_i^* \leftarrow \underset{k}{\operatorname{argmax}} \quad r(i, k) + a(k, i)$$

The responsibility update rule is unchanged: corresponding similarities, $s(i, k)$ are normalized by the *best* alternative, modulated by its availability. The update rules for availabilities, however, are now simpler than before: $a(k, i)$ are set to the responsibility $r(k, k)$ plus the responsibilities from other training cases, thresholded to be positive so there is no penalty from distant training cases with low responsibility. The entire quantity is thresholded to be non-positive, because it is the log-ratio of $P(x_i | x_i \text{ in cluster } x_k)$ to $P(x_i | x_i \text{ in cluster } x_{j \neq k})$ given all incoming messages to function node k , where the latter log-probability is the former rectified at zero.

3. Comparison of the *NIPS* (2006) and *Science* (2007) Algorithms

Previous publications have never contained a comparison between the two published affinity propagation algorithms. We find that the original constraint in [18] disallowing singleton clusters unnecessarily prevents the algorithm from moving through regions of the search space on the way to better (higher fitness) solutions.

The newer affinity propagation algorithm of [15] is also less sensitive to the order in which messages are updated. The original algorithm passed messages with a sequential schedule (requiring an ordering of nodes), but the new algorithm also works well using a less-arbitrary parallel message-passing schedule; oscillations can be effectively managed with moderate damping (see [15] for details).

In [18], affinity propagation was applied to several distinct problems in machine learning, including exon/gene detection, character recognition and image segmentation. To compare the proposed new affinity propagation algorithm to the original version, we examined the task of clustering patches taken from an image, which was described in the original paper.

Briefly (see [18] for details) a tiling of 24×24 non-overlapping patches was extracted from the image and translation-invariant similarities between patches were computed by comparing smaller 16×16 windows within each patch. The lowest squared error between windows (over all possible translations) was chosen as a similarity measure. Patches were clustered using affinity propagation and the resulting likelihood was compared with 1000 restarts of k -centers clustering (we use 100,000 restarts), garnering results ranking 3188th of 100,000 for $K = 6$ clusters. We downloaded the similarities from the web site reported in [18] and ran the original version of affinity propagation and reproduced these results.

We ran the later version of affinity propagation from [15] and compared it to 100,000 restarts of k -centers clustering (requiring roughly 1000 times the computation as both affinity propagation algorithms), achieving better re-

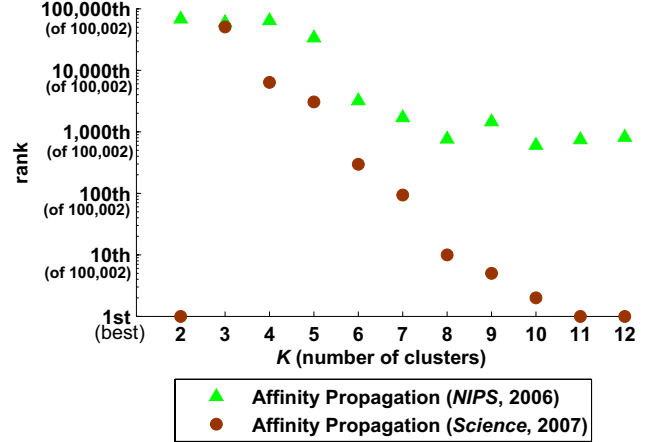


Figure 1. Performance (measured by ranking within 100,000 runs of k -centers clustering plus one run of each affinity propagation algorithm) for various numbers of clusters on a patch clustering task. A common value for the exemplar preferences was tuned to produce $K = \{2, \dots, 12\}$ clusters and the result was compared to the that obtained from 100,000 random initializations of k -centers. In most cases, the new version of affinity propagation finds a configuration at or near the best of the many k -centers initializations, while the older version performs comparatively worse. For $K = 3$, the clustering task is trivial (search space of size $\binom{64}{3} = 41664$, so k -centers clustering works well.

sults that rank in the top 300 of k -centers runs for the same $K = 6$ setting. We believe the results would have been even more favorable on a more difficult problem with a larger search space, as the task of partitioning the 64 training cases into 6 clusters only has $\binom{64}{6} \approx 2.8$ billion possibilities. For instance, modifying the exemplar preference to output 9 clusters yields results ranked in the top 5 of 100,000 (the old version of affinity propagation gives rank 1447). Rankings for varying K , the number of clusters, between 2 and 12 are shown in Figure 1.

4. Unsupervised Categorization of Olivetti Face Images

In the next section, we develop a framework for categorizing general images using scale-invariant keypoints (SIFT features). Here, we study the more straight-forward problem of unsupervised categorization of face images extracted from the Olivetti database. In [15], these images were analyzed using Euclidean distance and comparisons were made with k -centers clustering in terms of squared error. Here, we compare affinity propagation with several other methods and additionally study a non-metric similarity derived from considering relative translations of the two images being compared. Importantly, instead of reporting squared error, we present results on unsupervised classification error,

which is more relevant to many vision applications.

The Olivetti face database consists of ten 64×64 grey-scale images of each of 40 individuals, where each individual appears with a range of in- and out-of-plane pose variations. From each 64×64 image, we extracted a centered 50×50 region, so as to avoid pixel intensities in the corners of the images, which often contain background. To remove overall image brightness or darkness as a cue that would make unsupervised categorization easy, we normalized the pixel intensities in each image so that the mean was zero and the variance was 0.1. To examine the effect of a wider range in image variation for each individual, we constructed a second data set by extracting the images of 10 individuals and for each of the resulting 100 images, applying 3 in-plane rotations and 3 scalings, producing a data set of 900 images.

When applying affinity propagation, we set the exemplar preferences $s(i, i)$ for all training cases equal to the same common value and then applied the new affinity propagation algorithm. This value was varied to obtain different numbers of categories and we report results for all values that we tried. In general, the preferences were chosen to span the range of input similarities, so that the number of detected clusters ranged from small to large.

4.1. Performance on squared error

Using the 900 images including rotations and scales, we set the similarity between image i and image k to the negative of the sum of squared pixel differences (squared error). In addition to applying affinity propagation (which took 5 sec for each number of clusters), we applied 10,000 runs of k -centers clustering with different random initializations (which took 116 sec for each number of clusters). For each number of clusters, we then defined the baseline error to be the 1st percentile of error found by the 10,000 runs of k -centers clustering. Fig. 2 shows the error relative to the baseline achieved by affinity propagation versus the number of clusters, as well as the range of error values achieved by the many runs of k -centers clustering. While the error achieved by the best of 10,000 runs of k -centers clustering is comparable to that of affinity propagation for small numbers of categories, for most solutions affinity propagation achieves significantly lower error.

In Fig. 2, for limited numbers of clusters, we also show the error achieved by 1) the best of one million runs of k -centers clustering (4 hours), 2) k -centers clustering initialized by placing centers uniformly along the first principal component of the data, 3) the best quantized output of 10 runs of the EM algorithm applied to isotropic mixtures of Gaussians (240 sec), 4) hierarchical agglomerative clustering using the similarities to pick the best new exemplar at each agglomeration step (5 hours). Affinity propagation uniformly achieves lower errors in much less time.

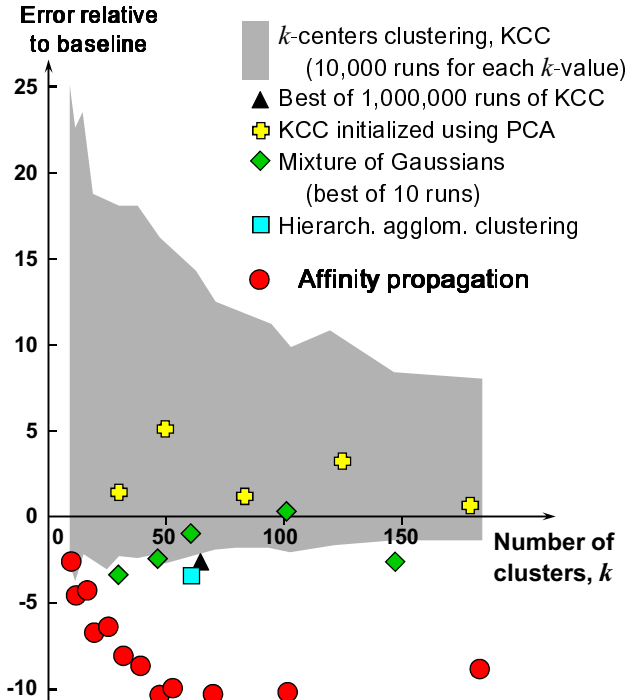


Figure 2. A comparison of the errors achieved relative to the baseline, for a variety of unsupervised image categorization methods applied to images extracted from the Olivetti face database. Affinity propagation achieves significantly lower error than other methods when the number of clusters is reasonably large.

4.2. Performance on unsupervised image classification

In several vision tasks, such as image or video summarization, labels are unavailable and the goal is to detect meaningful image categories in an unsupervised fashion. Even in supervised tasks, it can be helpful to first perform unsupervised categorization of images or image parts so as to reduce the dimensionality of the input and simplify supervised learning. Here, we explore for the first time the performance of affinity propagation in terms of classification error.

In this paper, we take two approaches to measuring the unsupervised classification error of the learned categories based on the true categories. In the first approach, each learned category is associated with the true category that accounts for the largest number of training cases in the learned category. In this case, the classification rate will approach 100% as the number of learned categories approaches the number of training cases. So, we report classification rate as a function of the number of learned categories. The second approach is to determine, for each solution, the ‘rate of true association’, which is the fraction of pairs of images from the same true category that were correctly placed in the same learned category, as well as the ‘rate of false asso-

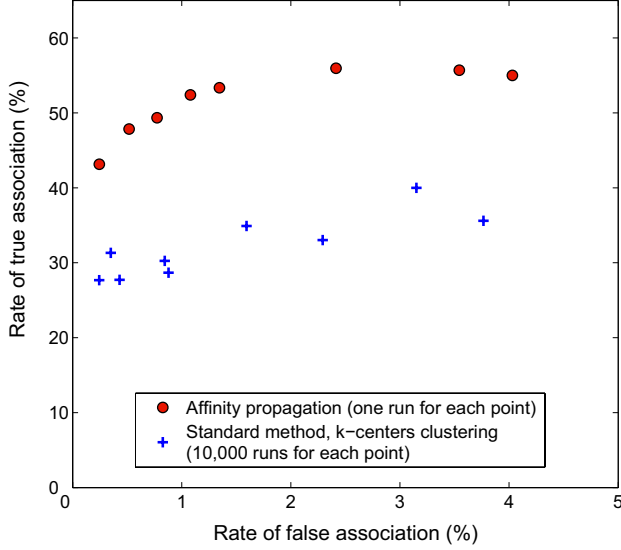


Figure 3. Performance on unsupervised classification error for the Olivetti face data. The true association rate is plotted against the false association rate for affinity propagation and the best of 10,000 runs of k -centers clustering. Affinity propagation achieves significantly higher true association rates.

ciation’, which is the fraction of pairs of images from different true categories that were erroneously placed in the same learned category. In this section, we use the latter approach.

Fig. 3 plots the true association rate against the false association rate for affinity propagation (circles) and the best of 10,000 runs of k -centers clustering (pluses), on the data set containing 400 images. Affinity propagation achieves a true association rate that is roughly 70% higher than that of k -centers clustering, for the same false association rate.

4.3. Performance using non-metric similarities

When comparing two face images, Euclidean distance ignores the fact that certain facial features may appear in different positions in the two images. In the next section, we describe a general-purpose non-metric similarity function based on SIFT features, but here, we study a non-metric similarity function that is tailored toward matching face images and show that by making the similarity non-metric, we can achieve higher classification rates.

Denoting the vector of pixel intensities for images i and k by \mathbf{x}_i and \mathbf{x}_k , the previous two subsections used the following definition of similarity: $s(i, k) = -\|\mathbf{x}_i - \mathbf{x}_k\|^2$. Here, we compute the similarity of image i to image k by extracting a sub-image from the center of image i and finding its best match to all sub-images (not necessarily centered) in image k . Let \mathbf{T} denote an operator that cuts a window of a fixed size out of the image it is operating on. There will be many operators corresponding to different possible positions from which the window may be extracted. Let \mathbf{T}_0

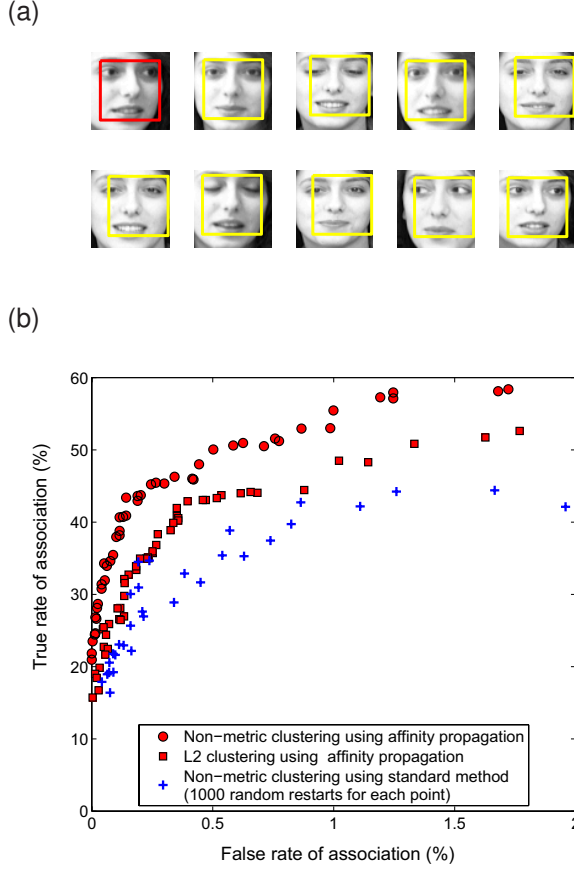


Figure 4. Unsupervised image categorization using non-metric similarities. (a) The similarity of an image (upper left) to each other images is determined by finding the best match (in Euclidean distance) between a window centered in the first image and all possible windows of the same size in the second image. (b) The true association rate is plotted against the false association rate for affinity propagation and the best of 1000 runs of k -centers clustering, using the non-metric similarity function. Also shown is the plot for affinity propagation applied using the Euclidean similarity function described previously. Affinity propagation applied using the non-metric similarity function achieves the highest classification rate.

denote the operator that cuts the window out of the center of the image. The similarity that we use in this section is given by

$$s(i, k) = -\min_{\mathbf{T}} \|\mathbf{T}_0 \mathbf{x}_i - \mathbf{T} \mathbf{x}_k\|^2.$$

We use the original Olivetti images of size 64×64 with a window size of 50×50 . Fig. 4a shows an example of an image \mathbf{x}_i (upper left) and the windows that achieve the minimum in the above expression, for the other 9 images in the same true category.

Fig. 4b plots the rate of true association against the rate of false association for affinity propagation applied to this

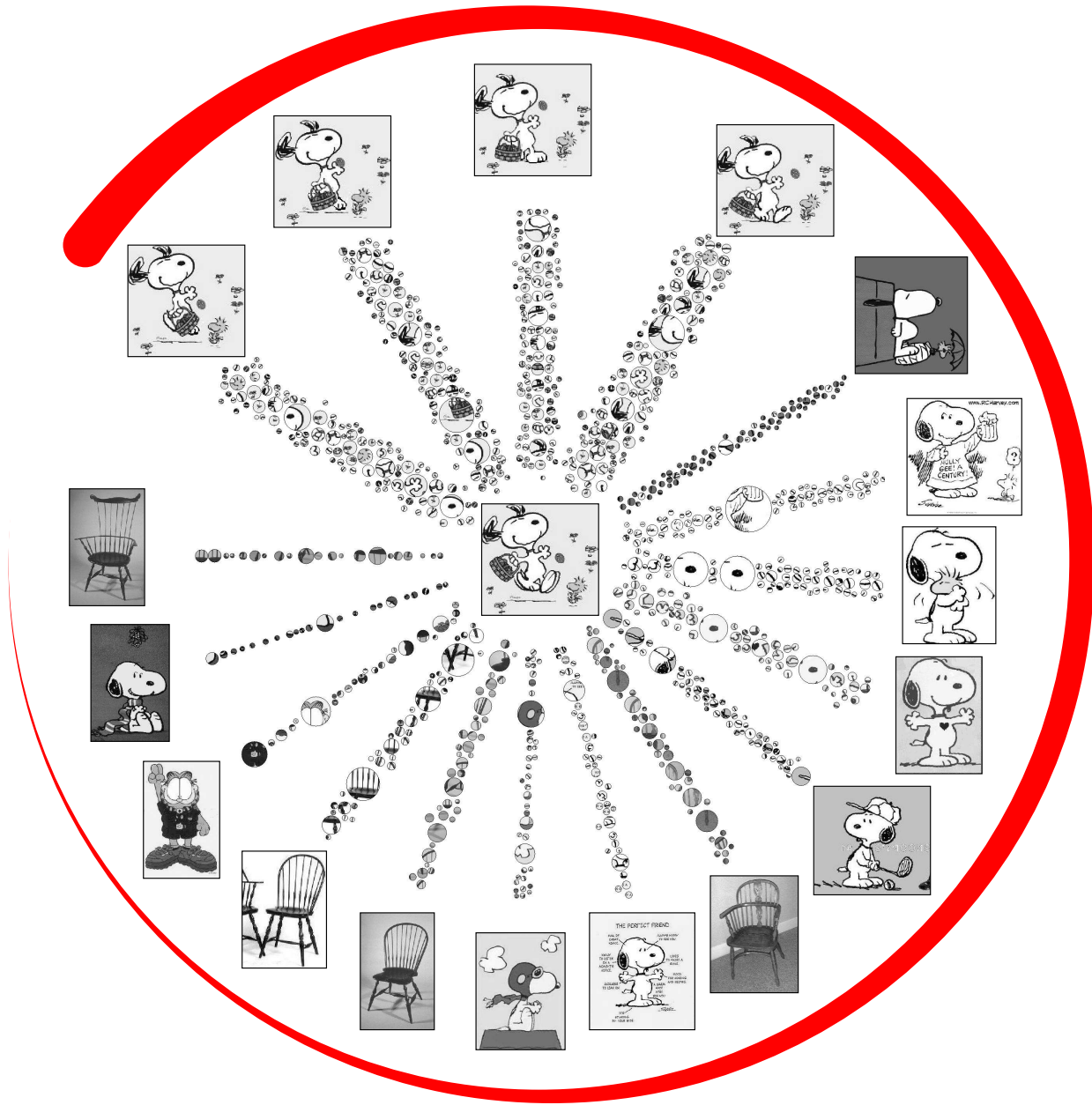


Figure 5. A sample category of images learned by affinity propagation. The central “Snoopy” image is the exemplar; the 17 other category members are shown with a stream of SIFT features matching the exemplar leading up to them. The thickness of the red “swoop” around the perimeter indicates the relevance of each image to this category as measured by the rank of its normalized similarity. Notice that the shared SIFT features for the three (weakly) misclassified chair images contain pieces that look similar to Snoopy’s basket.

non-metric similarity function (circles). Included for comparison is the plot obtained using the best of 1000 runs of k -centers clustering applied to the same non-metric similarity function (pluses). Also included is the plot obtained using the Euclidean similarities described in the previous section (boxes). When comparing with Fig. 3, note the different x -axis scale. The non-metric similarity function facilitates a significant increase in the classification rate and affinity

propagation achieves higher classification rates, compared to k -centers clustering.

5. Unsupervised Categorization of Caltech101 Images Using SIFT Features

The Caltech101 image dataset [19] contains 8677 pictures of objects, each with approximately 0.1 megapixel

resolution, belonging to 101 categories. We extract features believed to be important to human perception from a subset of images in the dataset and cluster the images based on this information, providing results for affinity propagation and another algorithm for comparison. Previous work [20], [21] utilizing SIFT features for image matching and category learning has focussed on the smaller Caltech4 image database or used supervised learning [21].

Input images are first subjected to the Scale-invariant feature transform (SIFT) [2], an algorithm that identifies local appearance features invariant to scale and rotation. Each feature is described by a 128-dimensional vector that has been shown to be quite useful for matching images. Lowe describes an algorithm used for matching the features extracted from one image with those from another: for each local feature from the first image, the nearest and second-nearest features are found in the second image (comparing descriptors by Euclidean distance). If the distance ratio between the nearest and second-nearest neighbors is greater than 0.8, the match is considered significant. Note that this comparison is asymmetric *i.e.* comparing image i to image j differs from comparing image j to image i .

We denote the number of significant feature matches found comparing image i with image k (with an upper threshold of 100) as $m(i, k)$, and define the similarity $s(i, k)$ between image i and image k to be the number of significant feature matches (subject to the threshold), normalized by subtracting means across both dimensions:

$$s(i, k) = m(i, k) - \frac{1}{n} \sum_{j=1}^n m(i, j) - \frac{1}{n} \sum_{j=1}^n m(j, k)$$

Note that this similarity function is non-metric due to $s(i, k)$ not being equal to $s(k, i)$.

We selected 20 of the 101 classes that represented a wide range of objects; they are as follows (with numbers in parenthesis representing the number of images in each class): faces (435), leopards (200), motorbikes (798), binocular (33), brain (98), camera (50), car_side (123), dollar_bill (52), ferry (67), garfield (34), hedgehog (54), pagoda (47), rhino (59), snoopy (35), stapler (45), stop_sign (64), water_lilly (37), windsor_chair (56), wrench (39), yin_yang (60). Some classes contained a very large number of images (*e.g.* faces, leopards, motorbikes, car_side), so for those classes we used only the first 100 images in each, yielding a total dataset of 1230 images.

We computed $s(i, k)$ similarity values between all possible image pairings and ran affinity propagation and k -centers clustering (best of 100 restarts) on the resulting set of similarities. Fig. 5 shows an example of a category learned by affinity propagation.

We report classification rates as a function of K , the number of clusters, in Fig. 6. (See section 4.2 for a description of how unsupervised classification rates are computed.) The number of exemplars (clusters) is varied in

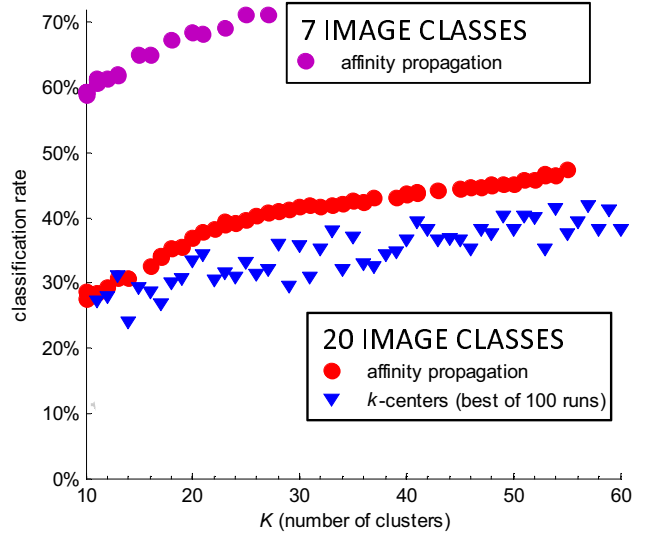


Figure 6. Performance of unsupervised classification on Caltech101 images. The correct classification rate is plotted against the number of categories learned (clusters). Affinity propagation achieves consistently better classification rates than the best of 100 k -centers runs for a dataset of 20 image classes. Affinity propagation’s classification rate is also shown for a different subset of the data containing 7 image classes.

affinity propagation by adjusting the preference *i.e.* $s(i, i)$ over a range of values. For almost all values of K , affinity propagation outperforms k -centers clustering, usually achieving correct classification rates around 7% higher than k -centers clustering. Both setups (affinity propagation and 100 k -centers runs) take roughly the same amount of computation time, 1–2 minutes depending on the number of exemplars identified.

We also ran affinity propagation on a smaller 7-category subset of the data (faces, motorbikes, dollar_bill, garfield, snoopy, stop_sign, windsor_chair) and achieved correct classification rates in the 60%-70% range, as shown.

6. Conclusions

Many vision tasks such as categorization can benefit from the identification of a set of exemplars in images or image fragments. Most methods, including the k -centers clustering technique, keep track of a fixed set of K exemplars while iteratively refining this set. The affinity propagation algorithm [18], [15] takes a conceptually novel approach by simultaneously considering all training cases as candidate exemplars and using a probabilistic message-passing procedure to gradually identify a good set of exemplars. Our objectives in this paper were to demonstrate how affinity propagation can be used to achieve high classification rates for unsupervised image categorization, to de-

velop non-metric similarity functions based on translation-invariance and SIFT features, and to compare the two published versions of the algorithm.

Using the Olivetti face database, we compared affinity propagation to k -centers clustering and found that even when the latter algorithm was re-run tens of thousands to millions of times, affinity propagation achieved higher fitness values (lower reconstruction error) and higher true positive classification rates. We also find that using a non-metric similarity function that accounts for image translation significantly increases classification rates.

In addition, we perform unsupervised categorization on images from the Caltech101 dataset and again demonstrate that, with a non-metric similarity function based on the number of matching SIFT features in two images, affinity propagation is able to identify meaningful categories and achieves competitive classification rates, which are higher than those found using k -centers clustering.

References

- [1] M. Turk, A. Pentland 1991. Eigenfaces for recognition. *J. Cognitive Neuroscience* **3**:1. 1
- [2] D. G. Lowe 2004. Distinctive image features from scale-invariant keypoints. *Intern. J. of Computer Vision* **60**:2, 91-110. 1, 7
- [3] T. Vetter, T. Poggio 1996. Image synthesis from a single example image. In *Proc. 4th European Conf. on Computer Vision*, Cambridge, UK, 652-659. 1
- [4] A. Efros, T. Leung 1999. Texture synthesis by non-parametric sampling. In *Proc. Int. Conf. on Computer Vision*, IEEE Press, 1033-1038. 1
- [5] W. Freeman, E. Pasztor 1999. Learning to estimate scenes from images. In *Advances in Neural Information Processing Systems 11*, MIT Press. 1
- [6] E. Shechtman, Y. Caspi, M. Irani 2005. Space-time super-resolution. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**:4, 531-545. 1
- [7] N. Jovic, B. J. Frey, A. Kannan 2003. Epitomic analysis of appearance and shape. In *Proc. Int. Conference on Computer Vision*, IEEE Press. 1
- [8] Y. Wexler, E. Shechtman, M. Irani 2004. Space-time video completion. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE Press. 1
- [9] D. Gavrilu, V. Philomin 1999. Real-time object detection for smart vehicles. In *Proc. Int. Conf. on Computer Vision*, 87-93. 1, 2
- [10] K. Toyama, A. Blake 2002. Probabilistic tracking with exemplars in a metric space. *Int. J. of Computer Vision* **48**:1, 9-19. 1, 2
- [11] G. E. Hinton, M. Revow 1995. Using pairs of data-points to define splits for decision trees. In *Advances in Neural Information Processing Systems* 8, Morgan Kauffman, 507-513. 1
- [12] M. Charikar, S. Guha, A. Tardos, D. B. Shmoys 2002. *J. Computer and System Science* **65**, 129. 2
- [13] J. MacQueen 1967. Some methods for classification and analysis of multivariate observations. In *Proc. of the Fifth Berkeley Symp. on Mathematical Statistics and Probability*, Univ. of California Press, 281. 2
- [14] S. Dasgupta, L. J. Schulman 2000. A two-round variant of EM for Gaussian mixtures. In *Proc. Uncertainty in Artificial Intelligence*, 152. 2
- [15] B. J. Frey, D. Dueck 2007. Clustering by passing messages between data points. *Science* **315**, 972-976. 2, 3, 7
- [16] D. Huttenlocher, J. Noh, W. Rucklidge 1993. Tracking non-rigid objects in complex scenes. In *Proc. Int. Conf. on Computer Vision*, IEEE Press, 93-101. 2
- [17] K. Kutulakos 2000. Approximate N-view stereo. In *Proc. European Conf. Computer Vision 1*, 67-83. 2
- [18] B. J. Frey, D. Dueck 2006. Mixture modelling by affinity propagation. In *Advances in Neural Information Processing Systems 18*, MIT Press. 2, 3, 7
- [19] L. Fei-Fei, R. Fergus, and P. Perona 2004. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *IEEE CVPR 2004, Workshop on Generative-Model Based Vision*. 6
- [20] K. Grauman, T. Darrell 2006. Unsupervised Learning of Categories from Sets of Partially Matching Image Features. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE Press. 7
- [21] K. Mikolajczyk, B. Leibe, B. Schiele 2006. Multiple Object Class Detection with a Generative Model. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, IEEE Press. 7