Local Selection of Features for Image Search and Annotation

Jichao Sun
New Jersey Institute of Technology
Newark, New Jersey
js87@njit.edu

ABSTRACT

In image applications, direct representations of images typically involve hundreds or thousands of features and not all the features are relevant for any given object. Errors introduced into similarity measurements by irrelevant or noisy features are detrimental to the semantic performance of content-based image retrieval. Feature selection techniques can be used to identify indiscriminative features from the entire image database. However, such global approach neglects the possibility that the feature importance may vary across different images or classes of images. We propose several techniques for the local selection of features for image databases. By checking the local neighborhood of each image, our methods determine the feature importance with respect to the image and select different feature sets for individual images. We also design methods based on the proposed local selection schemes for K-NN graph construction, image search, and graph-based image annotation. We provide experimental results on two image datasets to demonstrate the effectiveness of our methods.

Categories and Subject Descriptors

H.2.4 [**Database Management**]: Systems—*Multimedia Databases*

Keywords

feature selection, local selection of features, semantic quality, subjective feature space, K-nearest neighbor graph, content-based image retrieval, graph-based image annotation

1. INTRODUCTION

Nearest neighbor search is a fundamental component of many established methods for content-based image retrieval (CBIR) and automated image annotation. The semantic quality of the search results, which can be measured as the proportion of images in the query result sharing the same class label as their query images, relies heavily on the representation of images. The existence of irrelevant features in the image representation may overwhelm the contributions of the relevant ones in similarity search. Image feature selection which constructs a reduced feature set from the original feature space is a well-known approach for boosting the effectiveness of semantic image retrieval.

Traditional feature selection approaches are global in the sense that they compute a single set of features across the entire database. In image retrieval, supervised feature selection is widely used when image labels are available. The quality of a candidate feature set can be evaluated using a learning algorithm such as the Bayesian classifier adopted in [8]. Distribution-based approaches are popular in scoring the importance of individual features. For example, the mutual information which measures the degree of mutual dependence between two variables was adopted in [3] to guide feature selection.

Supervised feature selection has its limitations on image databases when the semantic labels are few or missing. An unsupervised feature selection method was proposed in [2] for medical image retrieval. However, the computational cost is high due to the use of a clustering wrapper for feature evaluation. Laplacian Score (LS), an unsupervised method for feature selection of generic data, ranks individual features according to their locality-preserving abilities [4]. Spectral feature selection ($SPEC$) [10] presents a unified framework based on spectral graph theory for both supervised and unsupervised feature selection. The unsupervised discriminative feature selection ($UDFS$) algorithm incorporates discriminative analysis and $L_{2,1}$-norm minimization into a joint framework [9]. Better results were achieved by LS, $SPEC$ and $UDFS$ in clustering and classification tasks, although no evidence was provided to indicate that their direct use in CBIR improves the performance of retrieval tasks.

The methods mentioned above discard noisy features from all data points to reduce the dimensionality of the database. Such global approaches, however neglect the possibility that a feature that is relevant for one image (or one image class) may be irrelevant for another. It is reasonable to find a feature set local to each image in an effort to reduce the negative impact of locally irrelevant features. To this end, we study the following two problems in this work: (1) designing methods for the detection of locally noisy features with respect to individual images; and (2) utilizing the computed (different) feature sets in applications to image search and annotation.

MM'14, November 3–7, 2014, Orlando, FL, USA.

http://dx.doi.org/10.1145/2647868.2654863.

In Section 2, we present three supervised/unsupervised techniques for the local selection of image features. These techniques are then incorporated into different algorithms for K-NN graph construction, image search and graph-based image annotation. We provide experimental results on two image datasets to demonstrate the effectiveness of the proposed methods for different image applications in Section 3. A summary of the proposed methods and a discussion on the ongoing work and other possible research directions are given in Section 4.

2. LOCAL SELECTION OF IMAGE FEATURES AND ITS APPLICATIONS

We propose techniques for the local selection of image features, based on which several methods are developed for the applications to CBIR and image annotation.

2.1 Local Laplacian Score and Feature Sparsif cation for K-NN Graph Construction

K-NN graphs are an essential component of many established methods for CBIR and image annotation. The performance of these methods relies heavily on the semantic quality of the K-NN graphs, which can be evaluated by: *graph correctness* $= (\#\textit{correct neighbors})/(\#\textit{images}\times K)$, where a correct neighbor is one whose class label coincides with that of the query image. For example, in image label propagation, each graph edge connecting two unrelated image nodes is a source of error, in that it suggests that these two images should share the same label.

To enhance the semantic quality of the K-NN graph, in [5] we propose the Local Laplacian Score for the detection of locally noisy features, and embed it into an approximate K-NN graph construction framework *NN-Descent* [1].

Given a dataset D consisting of n data points represented by m-dimensional feature vectors, we denote the r-th feature by $\mathbf{f}_r = (f_{r1}, f_{r2}, \ldots, f_{rn})^T$, where $r = 1, \ldots, m$, and f_{ri} ($i = 1, \ldots, n$) is the feature value of $\mathbf{f}_r$ taken from data point $x_i \in D$. Given a nearest neighbor graph G of dataset D, the Laplacian Score [4] of the r-th feature can be computed as:

$$LS(r) = \frac{\sum_{ij}(f_{ri} - f_{rj})^2 S_{ij}}{var(\mathbf{f}_r)}, \tag{1}$$

where $var(\mathbf{f}_r)$ is the estimated variance of the values of feature $\mathbf{f}_r$, and S_{ij} is the RBF kernel on feature vectors x_i and x_j representing the i-th and j-th data points, respectively.

For the identification of noisy features relative to each data point, we define as follows the Local Laplacian Score (LLS) of the r-th feature for item x_i which represents the contribution to the LS in Eq. 1 that can be attributed to x_i:

$$LLS_i(r) = \frac{\sum_{j}(f_{ri} - f_{rj})^2 S_{ij}}{var(\mathbf{f}_r)}. \tag{2}$$

It can be found that, if a feature is noisy for a given class, many data points from this class would tend to agree on its identification as such. A consensus, however, does not in general occur among data points drawn from different classes. We adopt a straightforward method for the detection of noisy features local to node i using LLS, in which the m features are sorted in descending order of $LLS_i(r)$, and returns the first z ($z > 0$) features. We refer to these z features as the *locally noisy features* of x_i, and to the remaining $(m - z)$ features as the *subjective features* of x_i.

Traditional feature selection methods cannot be applied to the reduction of the locally noisy features identified by LLS, as the importance of a feature varies from one data point to another. Instead of discarding a feature from the entire dataset, we modify the noisy feature values for individual data points in an effort to reduce intra-class distances. As a heuristic solution, we propose to change the values of the locally noisy features to 0, which is the the global mean for standardized features. We refer to this process as *sparsification*. The LLS feature ranking and sparsification is then integrated into an iterative K-NN graph construction method *NN-Descent*. The combined method, named *NNF-Descent*, first computes an approximate K-NN graph based on the standardized feature vectors. In each iteration of *NNF-Descent*, a small number of locally noisy features detected by LLS are sparsified for each image. One *NN-Descent*-based iteration is then performed, which essentially involves the re-computation of the distance values from each image to its current neighbors, and the K-NN updating through checking whether two neighbors of a same image could serve as better neighbors in each other's K-NN list. During the sparsification, the centers of the classes can change. However, the data points in each class converge towards their new centers as the sparsification rate increases.

The iterative feature ranking and K-NN updating are mutually beneficial: an updated K-NN graph improves the accuracy of feature ranking, and the sparsification of noisy features improves the semantic quality of the K-NN graph in return. We refer the reader to [5] for a detailed analysis of the complete algorithm.

2.2 Generalized Laplacian Score and Image Ranking in Subjective Feature Spaces

In this section, we propose a method that takes advantage of the different feature sets computed for individual images without modifying the original feature values. First, we define the Generalized Laplacian Score (GLS) of the r-th feature for x_i as a linear combination of $LLS_i(r)$ with the average contribution to $LS(r)$:

$$GLS_i(r) = (1 - \beta) \cdot \frac{LS(r)}{n} + \beta \cdot LLS_i(r), \tag{3}$$

where β is a weighting factor in the range of $[0, 1]$, controlling the contributions of LS and LLS.

The selection of a subset of features for data point i can be accomplished in the same way as in Section 2.1. We can represent the *subjective feature set* of x_i by a mask vector: $F_i = (b_1, b_2, \ldots, b_m) \in \{0,1\}^m$, where $F_i[r] = b_r$ ($r = 1, \ldots, m$) is a boolean value equal to 1 if and only if the r-th feature is a subjective feature of x_i. F_i also straightforwardly defines an $(m-z)$-dimensional feature space for x_i. We refer to this space as a *subjective feature space* of item x_i, and use F_i to denote both the subjective feature set and the corresponding subjective feature space for x_i.

Let $d(\cdot,\cdot)$ be a distance function over the items of D with respect to the full set of features. Given a subjective feature space F_i for x_i, we denote the distance between x_i and x_j in F_i by $d_{F_i}(x_i, x_j) = d(F_i(x_i), F_i(x_j))$, where $F_i(\cdot)$ is the projection of a feature vector from the full feature space to the subspace F_i.

To utilize the different sets of features computed by GLS in an image search application, we propose a novel querying strategy which ranks the query image in the subjective feature spaces of the candidate images and selects the candidate images which correspond to the subjective feature spaces wherein the query image is ranked highly.

First, a subjective feature space F_i is computed for each item $x_i \in D$. Given a query q, the distance values $d_{F_i}(x_i, q)$ are computed for all $x_i \in D$. A direct comparison of these distance values would have no intuitive meaning, as they are computed from different feature spaces. To make proper use of these distances, we first compute $\overline{d_{F_i}}$, the *mean distance* from x_i to the other items of D with respect space F_i, then normalize $d_{F_i}(x_i, q)$ by $\overline{d_{F_i}}$. The ratio — referred to as the *ranking score* (RS) of q in F_i with respect to x_i — is used as the rank of q in the subjective feature space F_i.

Based on GLS and RS, we outline an image search prototype (GLS+RS) as follows: (1) Compute an approximate K-NN graph for the image database in the full feature space for GLS; (2) Compute a subjective feature space for each database image, and the ranking scores of the query image in the computed subjective feature spaces; (3) Return the images corresponding to the subjective features spaces wherein the query image is ranked highly.

2.3 Local Selection of Features for Labeled Images

In typical graph-based image annotation algorithms, a nearest neighbor graph (for example, a K-NN graph) is built based on the similarities between image feature vectors. The image labels are propagated from the labeled images to the unlabeled images along the graph edges according to certain weighting and combination rules. The quality of the edges leading from labeled images to unlabeled images is critical for the success of these algorithms.

To improve the quality of the edges from labeled images, we propose a supervised method for computing reduced feature sets for individual labeled images. The idea is that each feature of a labeled image is used in isolation to rank other labeled images; the features that assign high ranks to related neighboring images are treated as more important. By deleting the least important features, a different feature set is computed for each labeled image, and is used in the ranking of unlabeled images.

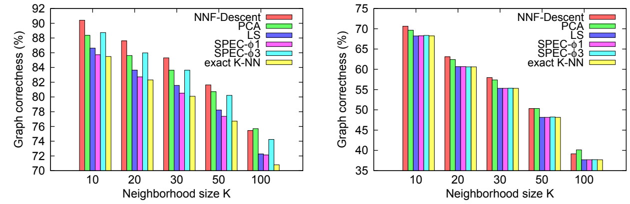
This idea is adopted as a preprocessing step in a graph-based image annotation method *SW-KProp+* [6]. When the proportion of correct edges leading from labeled images increases, better annotation performance can be achieved.

3. EXPERIMENTS

3.1 Datasets

We provide experimental results on two image datasets MNIST and Google-23, to show the effectiveness of our local selection schemes on image search and annotation. MNIST has $70,000$ images of handwritten digits [7]. The 784 pixel values of each image were treated as its image features. Except for the experiments in Section 2.2, we constructed a reduced subset of MNIST containing 10,000 images, by randomly selecting 1000 images for each digit. Google-23 is described in [6], which consists of 6686 faces extracted from web images of 23 celebrities. The total dimension of each feature vector is 1937.

We also conducted experiments on some other datasets including web images and non-image data; the experimental results are omitted in this abstract due to the space limit.



(a) MNIST (b) Google-23

Figure 1: *NNF-Descent* vs. competing methods.

3.2 Similarity Graph Construction

We compared *NNF-Descent* with principle component analysis (PCA), LS, $SPEC$-$\phi1$ and $SPEC$-$\phi3$ with respect to the correctness of produced K-NN graphs.

The neighborhood size for the target graph was set at $10, 20, 30, 50$ and 100. The number of features sparsified in each iteration was 5. For each experimental run of *NNF-Descent*, we computed the best graph correctness score over 50 iterations. The average computed from 5 runs was reported for each dataset. For the other methods, for each data point, 5 least important features were discarded per iteration, and an exact K-NN graph was computed from the resulting set of reduced feature vectors. Over all K-NN graphs produced the best correctness value achieved was reported. The performance of the exact K-NN graph computed from the original feature vectors is plotted as a baseline.

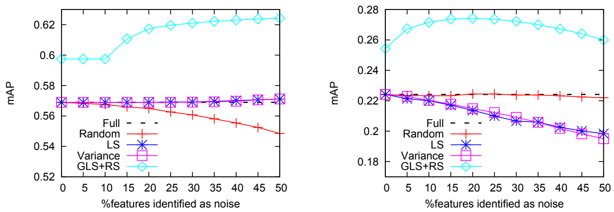
It can be seen from Fig. 1 that for all choices considered for the value of K, *NNF-Descent* achieved graph correctness scores better than those of the exact K-NN graphs. In almost all cases, our method clearly outperformed its competitors. When K is large, PCA outperformed the other methods. This outcome can be explained by the semantic quality of the K-NN graphs upon which *NNF-Descent*, LS and $SPEC$ rank features — when the semantic quality of the K-NN graph degrades, the detection of noisy features becomes less reliable.

3.3 Image Similarity Search

We compared GLS+RS with approaches based on LS, data variance and random feature selection. In each experimental run, MNIST and Google-23 were firstly split into a query set containing 100 random images and a candidate set containing the rest images. We also collected 100K faces from Wikipedia Commons as noise (referred to as Wiki-Faces) and appended them to the candidate set of Google-23. Except for GLS+RS, the other methods selected features globally — that is, if a feature was identified as a noise feature in the candidate set, it was also discarded from the query image, and the original distance function was used directly on the lower-dimensional feature vectors.

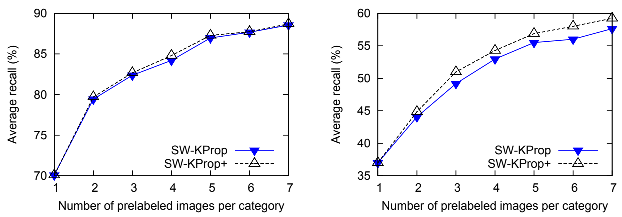
The results reported in Fig. 2 are the averages over 5 test runs. The proportion of features identified as noise was varied from 0 to 50%. As a baseline for comparison, the performance of sequential search in the full feature space was plotted as a dashed line in each figure (labeled as '*Full*').

It is clear from Fig. 2 that GLS+RS achieved consistently better results over its competitors which used the original distance function. This improvement can be attributed to two aspects: the image ranking strategy (when $z = 0$), and the GLS feature selection (when $z > 0$). On the other hand, traditional methods for filter-based unsupervised feature se-



(a) MNIST (b) Google-23+WikiFaces

Figure 2: *GLS+RS* vs. competing methods.



(a) MNIST (b) Google-23

Figure 3: *SW-KProp* vs. *SW-KProp+*.

lection yield little or no improvement in the semantic quality of CBIR results.

3.4 Graph-based Image Annotation

We compared *SW-KProp* with *SW-KProp+* in an image labeling task. Both methods propagate label scores from labeled images to unlabeled images in a similarity graph; *SW-KProp+* extends *SW-KProp* by utilizing the technique described in Section 3.3 in the computation of the similarity graph [6]. The number of prelabeled images ranged from 1 to 7. The overall annotation performance is evaluated, that is: $recall = (\#correctly\ labeled\ test\ items)/(\#test\ items)$.

The results can be found in Fig. 3. On Google-23, *SW-KProp+* outperformed *SW-KProp*, when the number of prelabeled images per category was larger than 1. This implies that with only a few images of the same category, *SW-KProp+* can effectively select a subset of features with better discriminative ability for each prelabeled image, and enhance the quality of the similarity graph by recomputing the neighborhood of prelabeled images.

However on MNIST, the use of the feature selection technique for prelabeled images did not lead to a significant improvement for this dataset. One possible reason is that MNIST is a relatively easy dataset whose original image descriptors are already of sufficient quality. The improvement on the edges from labeled images had little effect on the overall performance of label propagation.

4. DISCUSSION AND FUTURE WORK

In this thesis work, we summarize the methods we proposed for the local selection of image features. We proposed Local Laplacian Score (*LLS*) and Generalized Laplacian Score (*GLS*) as two unsupervised methods to construct a reduced feature set for each database image. *LLS* is embedded into *NN-Descent* for efficient and effective K-NN graph computation; a new ranking scheme is applied to the feature sets produced by *GLS* in an image search prototype. For graph-based image annotation, a supervised method is proposed for the computation of a discriminative feature set for each labeled image, by which, semantically related labeled-unlabeled image pairs are more likely to be connected in the similarity graph. Note that we are proposing methods for the identification of important feature dimensions for individual images rather than a new type of image feature.

Our ongoing work is focused on the extension of *GLS+RS*, which suffers from large computational cost as the query image is ranked with respect to all database images. One possible solution is to utilize an index structure to filter similar images of the query image. The query image is then ranked in the subjective feature spaces of the filtered images for query refinement. Other possible research directions include: (1) the applications of the proposed methods to other types of data; (2) a combination of query expansion and *GLS+RS* in CBIR approaches; (3) a study of the local selection methods on sparse feature vectors, such as bag-of-visual-words; and (4) the use of the MapReduce framework in the parallelization of the graph construction.

5. REFERENCES

[1] W. Dong, M. Charikar, and K. Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the 20th International Conference on World Wide Web*, pages 577–586, 2011.

[2] J. G. Dy, C. E. Brodley, A. C. Kak, L. S. Broderick, and A. M. Aisen. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(3):373–378, 2003.

[3] E. Guldogan and M. Gabbouj. Feature selection for content-based image retrieval. *Signal, Image and Video Processing*, 2(3):241–250, 2008.

[4] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In *Advances in Neural Information Processing Systems 18*, pages 507–514. MIT Press, 2006.

[5] M. E. Houle, X. Ma, V. Oria, and J. Sun. Improving the quality of K-NN graphs for image databases through vector sparsification. In *Proceedings of the International Conference on Multimedia Retrieval*, pages 89–96, 2014.

[6] M. E. Houle, V. Oria, S. Satoh, and J. Sun. Annotation propagation in image databases using similarity graphs. *TOMCCAP*, 10(1):7, 2013.

[7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

[8] Y. Sun and B. Bhanu. Image retrieval with feature selection and relevance feedback. In *Proceedings of the 17th IEEE International Conference on Image Processing*, pages 3209–3212, Sept 2010.

[9] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou. $l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning. In *Proceedings of the International Joint Conferences on Artificial Intelligence*, pages 1589–1594, 2011.

[10] Z. Zhao and H. Liu. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1151–1157, 2007.