Improving Image Distance Metric Learning by Embedding Semantic Relations

Fang Wang[1,2], Shuqiang Jiang[1,2], Luis Herranz[1,2], and Qingming Huang[1,2,3]

[1] Key Lab of Intell. Info. Process,
Chinese Academy of Sciences, Beijing 100190, China
[2] Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100190, China
[3] Graduate University of Chinese Academy of Sciences, Beijing 100049, China
{wangfang,sqjiang,qmhuang}@jdl.ac.cn,
luis.herranz@vipl.ict.ac.cn

Abstract. Learning a proper distance metric is crucial for many computer vision and image classification applications. Neighborhood Components Analysis (NCA) is an effective distance metric learning method which maximizes the kNN leave-out-one score on the training data by considering visual similarity between images. However, only using visual similarity to learn image distances could not satisfactorily cope with the diversity and complexity of a large number of real images with many concepts. To overcome this problem, integrating concrete semantic relations of images into the distance metric learning procedure can be a useful solution. This can more accurately model the image similarities and better reflect the perception of human in the classification system. In this paper, we propose Semantic NCA (SNCA), a novel approach which integrates semantic similarity into NCA, where neighborhood relations between images in the training dataset are measured by both visual characteristics and their concept relations. We evaluated several semantic similarity measures based on the WordNet tree. Experimental results show that the proposed approach improves the performance compared to the traditional distance metric learning methods.

Keywords: Metric Learning, kNN, Image Classification, NCA, Semantic Relations.

1 Introduction

A simple method to classify a data point is by comparing it with its neighbors. The k-Nearest Neighbor (kNN) rule[1] classifies each point using the majority class of its k nearest (most similar) neighbors in the training set. Recently, there has been an increasing interest in non-parametric kNN for image classification[2], with a competitive classification performance compared to other parametric classification methods.

Distance metric learning plays an important role in computer vision, machine learning and multimedia retrieval. In particular, due to the very nature

L. Weisi et al. (Eds.): PCM 2012, LNCS 7674, pp. 424–434, 2012.

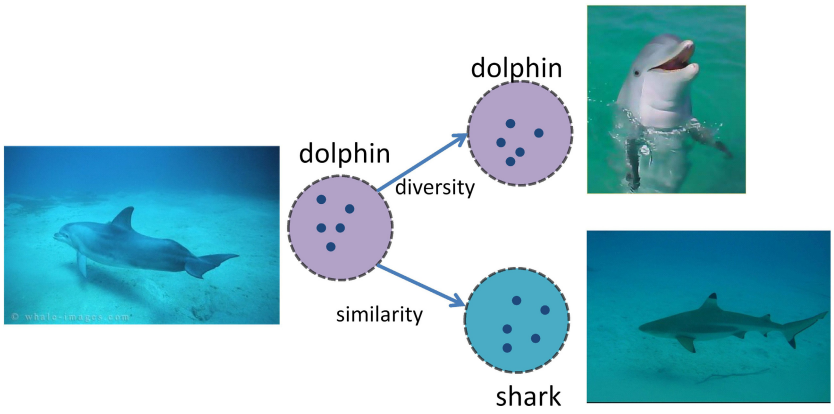


Fig. 1. Intra-class diversity and inter-class similarity

of kNN classification, an appropriate distance metric is critical to improve the kNN classification performance[3]. In image classification, most distance metric learning approaches try to employ the information contained in visual features and class labels, so that an appropriate Mahalanobis distance metrics are obtained to achieve better classification performance. The Mahalanobis distance is usually characterized by a positive semidefinite (PSD) matrix, which depends on the training data, and its estimation is the objective of distance metric learning. A variety of distance metric learning methods have been proposed in the literature[4,5,6,7,8,3,9,10,11,12,13], such as Neighborhood Components Analysis (NCA)[10], Large Margin Nearest Neighbor (LMNN)[6], Maximally Collapsing Metric Learning (MCML)[5] and Information-Theoretic Metric Learning (ITML)[8] and so on. NCA tries to maximize the probability of each sample assigned to those of same class using the visual information in each class label. The nearest neighborhood relation between two images in NCA is characterized by their visual similarity. A prevalent idea in metric learning is that points in the same class is made to be near to each other, however those points belonging to different concepts are pushed away and MCML explicitly constructs a convex optimization building on the basis of the idea. The same idea is applied in LMNN based on the large margin framework.

However, in conventional distance metric learning methods, semantic relations between concepts are not taken into account. These solutions may suffer from the following limitations:

1) First, using only low level visual features could not appropriately model the intra-class diversity and inter-class similarity. In a large scale scenario, each class contains a large number of images, which leads to very heterogeneous classes, with diverse shapes and visual characteristics. Besides, the number of concepts may be large, and discriminating between the distributions of the different concepts can be very complex, as images belonging to different classes may be visually similar. So these facts pose a tremendous challenge for measuring image similarity using only visual features. For instance, in Fig. 1 two images with

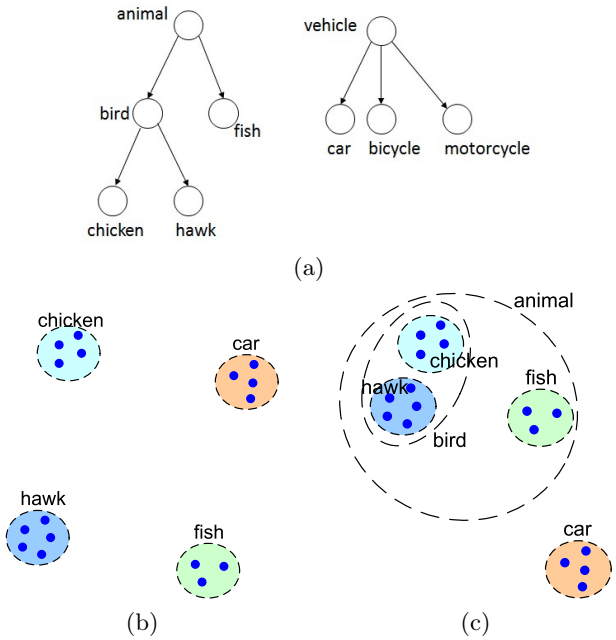


Fig. 2. Semantic information and the transformed space in kNN: a) tree with related concepts, b) distribution in the transformed space considering only visual similarity, and c) distribution in the transformed space considering visual and semantic similarity.

the same label *dolphin* have wide visual differences in shape, color and texture (i.e. intra-class visual diversity). At the same time, another image with the label *shark* is visually similar to one of the *dolphin* images (i.e. inter-class visual similarity).

2) Second, using only low level visual features could not satisfactorily reflect humans' real perception on image similarity. As the final objective of distance metric learning is to obtain a metric which better reproduces human perception, ignoring the semantic relations of concepts may not well satisfy this requirement. However, this kind of information can better reveal more meaningful high level similarities between images[14]. For instance, as shown in Fig. 2a, the concepts of *chicken* and *hawk* should be closer between them than to the concept *car*, as both are related to the concepts *animal* and *bird*. Data points are projected in such a way that they are optimized for classification according to the given training class labels, as shown in Fig. 2b. However, using semantic relations, the projection can reflect a more semantically meaningful structure (see Fig. 2c). A kNN classifier can also benefit from this projection, as related classes are closer than non-related.

Motivated by the above observations, in this paper we study the integration of semantic similarity in distance metric learning in the case of NCA and propose Semantic NCA (SNCA). In order to learn a more suitable matrix A to improve the performance of kNN, the nearest neighborhood relation between

two images is characterized by their visual similarity as well as the semantic relationship between their class labels, using an appropriate semantic measure based on the WordNet database[15]. Following the NCA formulation, the probability that a data point selects another point as its neighbor in the transformed space is measured based on both visual similarity and semantic correlation. As the performance of the proposed approach highly depends on the way in which semantic similarity is measured, we also study several semantic metrics based on the WordNet database.

The rest of this paper is organized as follows. The proposed Semantic NCA method is described in Section 2. Section 3 introduces the semantic similarity measures between concepts. Experimental evaluation is detailed in Section 4. The last section draws the conclusions.

2 Semantic Neighborhood Component Analysis

In this section, we first introduce the distance metric problem and the NCA framework for kNN multi-class image classification. Then, we integrate semantic similarity into this framework, resulting in the proposed Semantic NCA (SNCA) algorithm.

2.1 Distance Metric Learning

Suppose that we have a training dataset $C = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ with N images, where $\boldsymbol{x}_i$ represents the feature vector of the element i in the dataset. The element i belongs to a class with the label y_i. Given two elements i and j, the (squared) Mahalanobis distance between their feature vectors is calculated as:

$$d^2(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} - \boldsymbol{x_j})^T M (\boldsymbol{x_i} - \boldsymbol{x_j}) \tag{1}$$

where M is the PSD matrix ($M \geq 0$) we want to learn. As the classification procedure is based on neighbors, an optimal M should have the following property: data points belonging to the same class should have a low distance and data points belonging to different classes are separated as much as possible (see Fig. 2b).

Using $M = A^T A$, then Eq. (1) can be also rewritten as:

$$d^2(\boldsymbol{x_i}, \boldsymbol{x_j}) = (\boldsymbol{x_i} - \boldsymbol{x_j})^T A^T A (\boldsymbol{x_i} - \boldsymbol{x_j}) \tag{2}$$

Following this transformation, the distance between two points is calculated as the Euclidean distance between the projected points $A\boldsymbol{x_i}$ and $A\boldsymbol{x_j}$. Thus, the Mahalanobis distance is transformed to Euclidean distance via the matrix A.

2.2 Including Semantic Similarity in NCA

In NCA[10], the objective is to learn a metric A maximizing the classification performance for future test images in a kNN multi-class image classifier. However, the only available resource is the training dataset C. NCA applies the

leave-one-out (LOO) rule to maximize the performance to obtain A. During the learning stage, the assignment of neighbors is stochastic, which means that an element j is selected as a neighbor of another element i with certain probability $p_{ij}^{(NCA)}$. Thus, it is not certain whether j is considered as a neighbor of i or not. Such certainty would correspond to $p_{ij} = 1$ and $p_{ij} = 0$, respectively. A higher value of $p_{ij}^{(NCA)}$ shows that the two closer points are more likely to be neighbors. Therefore the similarity between i and j is calculated as[10] :

$$I^{(NCA)}(i,j) = v(\boldsymbol{x_i}, \boldsymbol{x_j}) = \exp\left(-d^2(\boldsymbol{x_i}, \boldsymbol{x_j})\right) = \exp(-\|A\boldsymbol{x}_i - A\boldsymbol{x}_j\|^2) \quad (3)$$

where $v(\boldsymbol{x_i}, \boldsymbol{x_j})$ is the visual similarity between two elements via their feature vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. An important characteristic of NCA is the description of neighborhood relations using a stochastic assignment rule in the LOO-kNN framework. When dealing with images, visual features are used to measure neighborhood relation between two points. However, visual features are not the only information available in a supervised classification system, such as kNN. As high-level textual descriptions of the content of images, class labels should not be ignored as they can help to better describe the relation between two images.

Fig. 1 illustrates both visual and semantic description of images. In the examples, each image is described by two parts, including visual feature $\boldsymbol{x_i}$ and textual label y_i. In general, y is the class label which is closely related to classification. For a given pair of images i and j, their visual similarity is measured using the feature vectors as $v(\boldsymbol{x_i}, \boldsymbol{x_j})$. In SNCA we also use the semantic similarity $s(i,j)$ between concepts, in order to obtain a classification system being more consistent with human cognition. Thus, not including the semantic component in neighborhood relations can prevent NCA from using important information which could help to improve the classification performance.

Therefore, we compute the similarity between two images i and j using both visual and semantic similarity as:

$$I^{(SNCA)}(i,j) = v(\boldsymbol{x_i}, \boldsymbol{x_j})\, s(i,j) = \exp(-\|A\boldsymbol{x}_i - A\boldsymbol{x}_j\|^2) s(i,j) \quad (4)$$

Comparing Eq. (3) and (4), this new similarity between two images also depends on semantic information. In order to be consistent with the definition of probability, we normalize the previous expression. Then the probability $p_{ij}^{(SNCA)}$ including both visual and semantic similarity can be rewritten as:

$$p_{ij}^{(SNCA)} = \begin{cases} \frac{1}{\Omega} I_{ij}^{(SNCA)}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{s(i,j)\exp(-\|A\boldsymbol{x}_i - A\boldsymbol{x}_j\|^2)}{\Sigma_{k\neq i} s(i,k)\exp(-\|A\boldsymbol{x}_i - A\boldsymbol{x}_k\|^2)}, & i \neq j \\ 0 & i = j \end{cases} \quad (5)$$

which denotes the probability that a point i selects another point j as its neighbor given their visual features and corresponding concept relation.

The objective is to try to find a projection which maximizes the probability that points with the same label are neighbors in projected space. For that

purpose, we use the objective function proposed in [10], which maximizes the LOO score of the of all data points in the training set C:

$$g^{(SNCA)}(A) = \sum_{i=1}^{N} \log p_i^{(SNCA)} = \sum_{i=1}^{N} \log \sum_{j \in C_i} p_{ij}^{(SNCA)} \tag{6}$$

where C_i is a set of all points which have the same class label to point i. In Eq. (6) the only factor involving semantic correlation is $s(i,j)$, with the matrix A remaining independent. Thus we can compute the gradient of $g^{(SNCA)}(A)$ as:

$$\frac{\partial g^{(SNCA)}(A)}{\partial A} = 2A \sum_{i=1}^{N} \left(\sum_{k=1}^{N} p_{ik}^{(SNCA)} \boldsymbol{x_{ik}} \boldsymbol{x_{ik}}^T - \frac{\sum_{j \in C_i} p_{ij}^{(SNCA)} \boldsymbol{x_{ij}} \boldsymbol{x_{ij}}^T}{\sum_{j \in C_i} p_{ij}^{(SNCA)}} \right) \tag{7}$$

where $\boldsymbol{x_{ij}}$ means $\boldsymbol{x_i} - \boldsymbol{x_j}$.

The same optimization method used in NCA can still be used to estimate A, while we also take advantage of semantic similarity. After obtaining A, the input data can be projected in the transformed space, in which conventional kNN classification can be performed.

3 Semantic Similarity between Concepts

In this paper, semantic similarity is measured based on WordNet[15]. In the experiment we tested four different measures: node count (*path*), Resnik (*res*)[16], Leacock and Chodorow (*lch*)[17], and the least common subsumer measure[18] (*LCS*). Except for *LCS*, the other three measures can be found in the JAVA WordNet Similarity (JWS) package[19] which implements several widely used semantic similarity measures between concepts in WordNet. Table 1 details how these measures are computed.

In WordNet[15], each concept is represented as a node in the tree taxonomy, with the term synonym set (synset). We denote depth(i) as the length of the path from root to node i. The most common subsumer CS(i,j) is the most specific concept which is a common ancestor of the concepts i and j. The information content IC(i) of a node i is computed as described in [16].

4 Experimental Results

4.1 Dataset and Feature Representation

We evaluate the proposed method over the Caltech256[20] and ImageNet[21] datasets. We selected a total of 4546 images from the Caltech256 dataset covering 40 subconcepts of the concept *animal* (Caltech40). For the second dataset ImageNet20 we selected 20 concepts covering subconcepts of the broad concepts *animal*, *vegetable*, *flower* and *vehicle*, represented by about 21100 images from

Table 1. Concept semantic similarity measures used in the experiments

Measure	Formulation	Description
path	$s_{path}(i,j) = \frac{1}{\min(\text{depth}(i),\text{depth}(j))}$	The reciprocal of the number of nodes along the shortest path between i and j
res	$s_{res}(i,j) = \text{IC}(\text{CS}(i,j))$	$\text{CS}(i,j)$ is the least common subsumer of node i and j, $\text{IC}(i)$ is the information content of node i
lch	$s_{lch}(i,j) = -\log(L/2D)$	L is the length of the shortest path between i and j and D is the maximum depth of the taxonomy
LCS	$s_{\mathcal{LCS}}(i,j) = \frac{\text{depth}(\text{CS}(i,j))}{\max(\text{depth}(i),\text{depth}(j))}$	The length of the least common subsumer node normalized by the longest branch

Table 2. Comparison of the classification accuracy of NCA, LMNN and SNCA with different semantic measures in Caltech40

Accuracy(%)	Caltech40			
Method	color		GIST	
	$k = 20$	$k = 40$	$k = 20$	$k = 40$
kNN	9.78	10.43	13.48	14.72
NCA	11.40	11.27	20.37	19.71
LMNN	10.26	10.92	13.83	13.70
SNCA (*path*)	**12.23**	11.75	18.56	18.16
SNCA (*res*)	11.71	**12.01**	21.56	20.28
SNCA (*lch*)	12.01	11.79	20.11	20.24
SNCA (*LCS*)	11.93	11.79	**22.18**	**20.86**

the ImageNet dataset. For each concept approximately half of the images were used for training and the remaining were considered as test images.

We used color histograms in the HSV space (16x4x4 bins) and GIST[22] to represent the images in the visual feature space.

4.2 Results and Analysis

In the first experiment, we studied the classification accuracy of SNCA, NCA, LMNN and basic kNN over both datasets, with different values of k. SNCA uses *path*, *res*, *lch*, *LCS* as semantic similarity measures. The dimensionality is reduced using PCA to 80 dimensions for both color and GIST feature, in order to accelerate the classification process.

Table 2 shows the classification accuracy in Caltech40. For both features, the best performance is obtained using SNCA with different semantic metrics, as shown in Table 2. However, the classification accuracy using color features is still very limited. Using GIST features, it improves considerably. In both cases SNCA has better performance than NCA and LMNN for most of the semantic measures. In general the classification accuracy is reasonably high, considering that the dataset

Table 3. Comparison of the classification accuracy of NCA, LMNN and SNCA with different semantic measures in ImageNet20

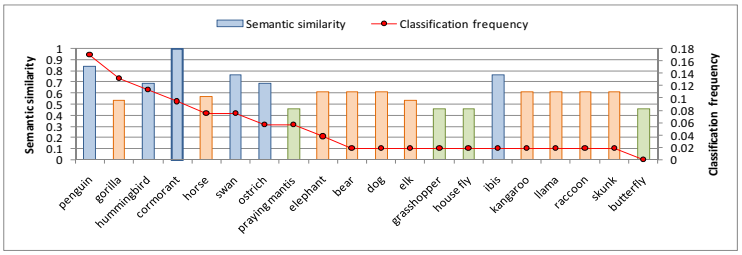
Accuracy(%)	ImageNet20			
Method	color		GIST	
	$k = 20$	$k = 40$	$k = 20$	$k = 40$
kNN	31.46	30.13	38.36	37.93
NCA	33.47	33.75	41.05	40.97
LMNN	33.75	33.63	41.72	41.22
SNCA (*path*)	32.99	34.03	41.26	41.09
SNCA (*res*)	34.59	34.84	42.16	41.20
SNCA (*lch*)	**34.63**	33.83	42.34	41.93
SNCA (*LCS*)	34.07	**34.88**	**42.69**	**42.22**

Table 4. Comparison of the five most frequently predicted concepts using instances of the *cormorant* concept as test images in Caltech40. F: frequency, SS: semantic similarity using *LCS* metric.

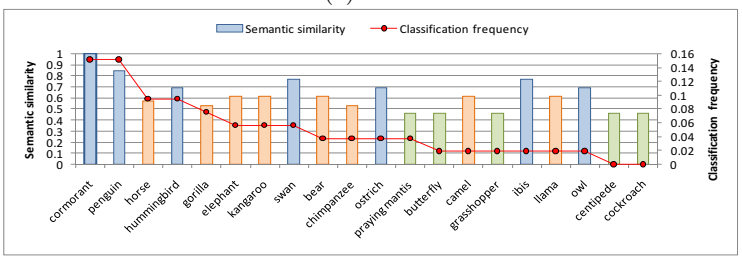
Method								
NCA			SNCA (*LCS*)			SNCA (*res*)		
Concept	F (%)	SS (%)	Concept	F (%)	SS (%)	Concept	F (%)	SS (%)
penguin	16.98	84.62	***cormorant***	15.09	100	***cormorant***	18.87	100
gorilla	13.21	53.33	*penguin*	15.09	84.62	*hummingbird*	11.32	69.23
hummingbird	11.32	69.23	*horse*	9.43	57.14	*penguin*	11.32	84.62
cormorant	9.43	100	*hummingbird*	9.43	69.23	*gorilla*	9.43	53.33
horse	7.55	57.14	*gorilla*	7.55	53.33	*horse*	9.43	57.14

has forty concepts. Particularly, SNCA improves the accuracy of conventional NCA. The classification accuracy over ImageNet20 is shown in Table 3. The performance of SNCA is better than NCA with *res*, *lch*, *LCS* measures using the color feature. In the case of the GIST feature, we can observe that NCA has a worse performance than LMNN. However, the performance of NCA is significantly improved by integrating semantic information, resulting in a better accuracy than LMNN. Thus, integrating semantic information in NCA helps to better discriminate between images, and a better performance can be achieved.

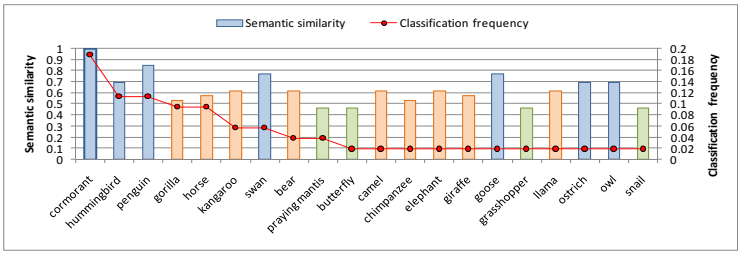
However, the improvement of SNCA is still limited. Even though the number of concepts considered in the experiment is high compared to other datasets used in the evaluation of distance metric learning methods[6], this number still falls short to fully exploit semantic relations in the WordNet hierarchy, especially in ImageNet20 where the concepts were selected randomly among all the concepts in ImageNet. Although the effect in small datasets with few concepts is still unsatisfactory, such kind of semantic relations may be significant in large datasets with high diversity. Thus, most distance learning methods using only visual similarity may have good performance when dealing with specialized datasets with relatively few and narrow concepts (e.g. faces, letters, plants)[6], but may fail when they are used in scenarios with larger datasets, as they may not be able to cope with all the variability in the dataset.



(a) NCA



(b) SNCA (LCS)



(c) SNCA (res)

Fig. 3. Most frequently predicted concepts (top 20) using instances of the *cormorant* concept as test images in Caltech40: a) NCA, b) SNCA (LCS) and, c) SNCA (res). Frequency and semantic similarity (using LCS metric) between the true and the predicted concepts are shown combined in each figure (better viewed in color).

Apart from improving the classification accuracy of the classifier, another objective of this work is to project concepts in a more semantically meaningful structure, in which semantically related concepts are projected closer. In order to illustrate how the resulting space using SNCA can be more suitable, Fig. 3 shows an example in which instances of the concept *cormorant* are classified using both NCA and SNCA (using both LCS and res metrics) in Caltech40, with the GIST feature and $k = 20$. The figure shows the 20 most frequently predicted concepts for each of the methods. The classifier did not assign any test image to any of the remaining concepts, so we did not include them in the figure. These concepts are shown in the horizontal axis sorted in descending order by frequency. In the same plot, the semantic similarity between the true and the predicted concept is also shown. In this case, to measure semantic similarity we used the LCS metric, as it is bounded between 0 and 1. For better visualization, concepts are represented

in different colors depending on the type of animal (*bird*, *mammal*, *amphibian*, *invertebrate*, *reptile* or *fish*), and the concept *cormorant* is emphasized. Similarly, Table 4 shows the numerical results for the five most predicted concepts.

A first observation is that in NCA the correct concept is only the forth most predicted result, being *penguin* the most predicted one with a much higher frequency, which is still a kind of *bird*. However, the second most predicted concept is *gorilla*, which is a kind of *mammal*, thus less semantically related to the correct concept. However, in this case SNCA can return a more reasonable predictions, in which *cormorant* is the most predicted concept, and also with higher accuracy, especially using the *res* metric. Besides, in both cases the most confusing results are also kinds of *bird*. However, in this example even using semantic relations the result is still limited, as the input image is often confused with some kind of mammals (e.g. *horse* and *gorilla* in the top five) with relatively high frequency. One reason may be that the similarities between *cormorant* and other birds and between *cormorant* and mammals are not very different, so the semantic relations may not be fully exploited.

5 Conclusion

In this paper, we explored the integration of semantic relations into an image classification system via distance metric learning. Thus, a distance metric learning method using both semantic and visual similarities can project the input data to a space with a more meaningful structure. This observation motivates the proposed SNCA method, which has been studied with different semantic similarity measures and datasets, improving the classification performance. Although the improvement is still limited in these datasets, we expect that the gain can be higher in large scale scenarios, in which the number of concepts is high and their semantic relations can be fully exploited. The proposed framework to integrate semantic similarity in metric learning is generic, so we expect that it can be extended to other learning methods.

Acknowledgement. This work was supported in part by National Basic Research Program of China (973 Program):2012CB316400, in part by National Natural Science Foundation of China: 61070108,61025011, and 61150110480, in part by Chinese Academy of Sciences Fellowships for Young International Scientists: 2011Y1GB05

References

1. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13(1), 21–27 (1967)
2. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: Proc. of IEEE Conf. Computer Vision and Pattern Recognition, pp. 1–8 (2008)

3. Shen, C., Kim, J., Wang, L.: A scalable dual approach to semidefinite metric learning. In: Proc. of IEEE Computer Vision and Pattern Recognition, pp. 2601–2608 (2011)
4. Singh-Miller, N., Collins, M., Hazen, T.J.: Dimensionality reduction for speech recognition using neighborhood components analysis. In: INTERSPEECH 2007, pp. 1158–1161 (2007)
5. Globerson, A., Roweis, S.: Metric learning by collapsing classes. In: Proc. of the Conference on Advances in Neural Information Processing Systems (2006)
6. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10, 207–244 (2009)
7. Wang, Z., Hu, Y., Chia, L.-T.: Image-to-Class Distance Metric Learning for Image Classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 706–719. Springer, Heidelberg (2010)
8. Davis, J., Kulis, B., Sra, S., Dhillon, I.: Information-theoretic metric learning. In: Proc. of the International Conference on Machine Learning, pp. 209–216. ACM, New York (2007)
9. Bronstein, M.M., Bronstein, A.M.: Data fusion through cross-modality metric learning using similarity-sensitive hashing. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 3594–3601 (2010)
10. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood components analysis. In: Proc. of the Conference on Advances in Neural Information Processing Systems (2005)
11. Sugiyama, M.: Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. Journal of Machine Learning Research 8, 1027–1061
12. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. In: Proc. of the Conference on Advances in Neural Information Processing Systems, vol. 40 (2003)
13. Hwang, S.J., Grauman, K., Sha, F.: Learning a tree of metrics with disjoint visual features. In: Proc. of the Conference on Advances in Neural Information Processing Systems (2011)
14. Li, L., Jiang, S., Huang, Q.: Learning hierarchical semantic description via mixed-norm regularization for image understanding. IEEE Transactions on Multimedia 14 (2012)
15. Fellbaum, C.: WordNet: An electronic lexical Database (1998)
16. Resnik, P.: Using information content to evaluate semantic similarity. In: Proc. of the International Joint Conference on Artificial Intelligence, pp. 448–453 (1995)
17. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: WordNet: An Electronic Lexical Database. MIT Press (1998)
18. Fergus, R., Bernal, H., Weiss, Y., Torralba, A.: Semantic Label Sharing for Learning with Many Categories. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 762–775. Springer, Heidelberg (2010)
19. Hope, D.: Java WordNet similarity (2008), `http://www.sussex.ac.uk/Users/drh21`
20. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Tech. Rep. 7694, California Institute of Technology (2007)
21. Deng, J., Dong, W., Socher, R., Jia Li, L., Li, K., Fei-Fei, L.: ImageNet: a large scale hierarchical image database. In: Proc. of IEEE Computer Vision and Pattern Recognition, pp. 248–255 (2009)
22. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175 (2001)