Exploiting Socially-Generated Side Information in Dimensionality Reduction

Alejandro Marcos Alvarez*
University of Liège
amarcos@ulg.ac.be

Makoto Yamada
Yahoo! Labs
makotoy@yahoo-inc.com

Akisato Kimura
NTT CS Labs
akisato@ieee.org

ABSTRACT

In this paper, we show how side information extracted from socially-curated data can be used within a dimensionality reduction method and to what extent this side information is beneficial to several tasks such as image classification, data visualization and image retrieval. The key idea is to incorporate side information of an image into a dimensionality reduction method. More specifically, we propose a dimensionality reduction method that can find an embedding transformation so that images with similar side information are close in the embedding space. We introduce three types of side information derived from user behavior. Through experiments on images from Pinterest, we show that incorporating socially-generated side information in a dimensionality reduction method benefits several image-related tasks such as image classification, data visualization and image retrieval.

Categories and Subject Descriptors

K.4 [**Computing Methodologies**]: Machine Learning; J.3 [**Human-Centered Computing**]: Collaborative And Social Computing

Keywords

Dimensionality reduction; side information; social media

1. INTRODUCTION

In multimedia applications, traditional approaches to solve image-related problems, such as image classification [9, 5] or image retrieval, usually consist in using low-level image features, such as SIFT [4] or GIST features [6], as input of a dedicated algorithm. However, since the appearance of the objects varies as a function of a number of factors including pose, shape, and lighting, solving those problems with only low-level information is rather hard.

*This work was completed when the author was intern at NTT CS Labs.

SAM'13, October 21, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2509916.2509923.

Nowadays, various kinds of side information, such as textual description, geotags, "likes", or link structure of a web graph, are often available in addition to multimedia content. These various kinds of side information have recently drawn much attention from the research community that tries to take advantage of it, as an additional source of knowledge about the data. For example, side information obtained from users' interactions has been used through *graph regularization* to improve the sentiment analysis of messages shared through microblogging services [3]. As another example, Zhang et al. [12] proposed to use side information to construct *clean* training datasets, that are then used to train better video classifiers. In the context of images, it has recently been demonstrated that side information is useful for image classification when used either directly as input of the classifier [10], or via *graph regularization* [5]. Although all those approaches have been shown to be useful in their domains, most of them use side information within a classification method and are not applicable to other tasks.

In this paper, we investigate to what extent *socially generated side information* is useful for several image-related tasks. We propose a general approach that fully leverages the side information of an image through dimensionality reduction, thus making the method applicable to any kind of problem. The key idea behind the proposed approach is to find an embedding transformation so that images with similar side information are close in the embedding space. Based on this idea, we propose a simple approach that incorporates socially-generated side information into local Fisher discriminant analysis (LFDA) [8]. Note that the side information is only required for the training dataset and that the dimensionality of new objects can be reduced as well, even if their side information is not available. In this paper, we consider types of side information derived from human actions such as 'who are the owners of an image' or 'how does a user organize his images'. Through experiments on an image dataset obtained from Pinterest, we show that integrating socially-generated side information into dimensionality reduction actually improves the results obtained on tasks such as image classification, data visualization and image retrieval.

The contributions of this paper are threefold. First, we show that socially-generated side information can be advantageously introduced into dimensionality reduction in order to yield improved results on several tasks. Secondly, we propose a simple, yet effective, way to take advantage of side information for dimensionality reduction. Finally, we introduce three different types of side information generated from socially-curated data.

2. PROPOSED APPROACH

In this section, we formulate a dimensionality reduction problem using side information. Then, based on LFDA [8], we propose a linear dimensionality reduction method using side information. Although the method proposed in this paper is a supervised method, the proposed approach can be extended without effort to unsupervised dimensionality reduction methods such as LPP [2].

2.1 Problem Formulation

Let $\mathcal{X} \subset \mathbb{R}^d$ be the domain of input features $\boldsymbol{x}$ describing images, $\mathcal{V} \subset \mathbb{R}^p$ be the domain of side information $\boldsymbol{v}$, and $\mathcal{Y} \subset \mathbb{N}$ be the domain of labels $y \in \{1, 2, \ldots, C\}$. Suppose we are given n independent and identically distributed (i.i.d.) triplets,

$$D = \{(\boldsymbol{x}_i, \boldsymbol{v}_i, y_i) \mid \boldsymbol{x}_i \in \mathcal{X},\ \boldsymbol{v}_i \in \mathcal{V},\ y_i \in \mathcal{Y}, i = 1, \ldots, n\},$$

drawn from a joint distribution. Section 3.1 describes the kind of side information that is used in more detail.

The goal of dimensionality reduction is to find a low-dimensional representation $\boldsymbol{z} \in \mathbb{R}^m$ $(m \leq d)$ of input $\boldsymbol{x}$. In this paper, we focus on linear dimensionality reduction methods:

$$\boldsymbol{z} = \boldsymbol{T}^\top \boldsymbol{x},$$

where $\boldsymbol{T} \in \mathbb{R}^{d \times m}$ is a transformation matrix and $^\top$ denotes the matrix transpose.

Traditional dimensionality reduction methods only use input vector $\boldsymbol{x}$ and its corresponding label y to learn a transformation matrix $\boldsymbol{T}$. In this paper, we propose the additional use of side information $\boldsymbol{v}$ to generate $\boldsymbol{T}$. Note that side information is only needed to compute the transformation matrix $\boldsymbol{T}$. Once $\boldsymbol{T}$ is computed from the training set D, the dimensionality of new unseen images $\boldsymbol{x}'$ can be reduced without the need of having the corresponding side information for the new images, since the computed $\boldsymbol{T}$ can be re-used.

2.2 Local Fisher Discriminant Analysis with Side Information

In this section, we propose a simple, yet effective, dimensionality reduction approach based on local Fisher discriminant analysis (LFDA) [8]. The novelty of our approach resides in the computation of the affinity matrix from the available side information.

LFDA is a linear dimensionality reduction method that takes account of the local structure of the data through an affinity matrix. The key idea behind LFDA is to maximize the *local* between-class scatter $\boldsymbol{S}_\mathrm{b}$ while minimizing the *local* within-class scatter $\boldsymbol{S}_\mathrm{w}$. This amounts to finding a transformation matrix $\boldsymbol{T}^*$ that yields an embedding space in which the separability between classes is maximized and the variance in each class is minimized, while preserving the local structure.

The local between-class scatter $\boldsymbol{S}_\mathrm{b}$ and local within-class scatter $\boldsymbol{S}_\mathrm{w}$ are given as follows:

$$\boldsymbol{S}_\mathrm{b} = \frac{1}{2} \sum_{i,j=1}^{n} D^\mathrm{b}_{i,j} \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^\top,$$

$$\boldsymbol{S}_\mathrm{w} = \frac{1}{2} \sum_{i,j=1}^{n} D^\mathrm{w}_{i,j} \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^\top,$$

where

$$D^\mathrm{b}_{i,j} = \begin{cases} A_{i,j}(1/n - 1/n_c) & \text{if } y_i = y_j = c, \\ 1/n & \text{if } y_i \neq y_j, \end{cases}$$

$$D^\mathrm{w}_{i,j} = \begin{cases} A_{i,j}/n_c & \text{if } y_i = y_j = c, \\ 0 & \text{if } y_i \neq y_j. \end{cases}$$

Here, n_c is the number of samples in class c and $\boldsymbol{A}$ is an affinity matrix whose element $A_{i,j}$ quantifies the *affinity* between the i-th and j-th objects.

The transformation matrix $\boldsymbol{T}^*$ is then given by

$$\boldsymbol{T}^* = \underset{\boldsymbol{T} \in \mathbb{R}^{d\times m}}{\operatorname{argmax}} \operatorname{tr}\left[\left(\boldsymbol{T}^\top(\boldsymbol{S}_\mathrm{w} + \eta \boldsymbol{I}_\mathrm{d})\boldsymbol{T}\right)^{-1} \boldsymbol{T}^\top \boldsymbol{S}_\mathrm{b} \boldsymbol{T}\right], \quad (1)$$

where $\boldsymbol{I}_\mathrm{d} \in \mathbb{R}^{d\times d}$ is the identity matrix, $\eta > 0$ is a regularization parameter used to avoid overfitting and $\operatorname{tr}[\cdot]$ is the matrix trace operator. The solution of Eq. (1) is given by

$$\boldsymbol{T}^* = \left(\boldsymbol{\varphi}_1 | \boldsymbol{\varphi}_2 | \ldots | \boldsymbol{\varphi}_m\right),$$

where $\{\boldsymbol{\varphi}_i\}_{i=1}^d$ represent the generalized eigenvectors associated with the generalized eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ of the following generalized eigenvalue problem:

$$\boldsymbol{S}_\mathrm{b}\boldsymbol{\varphi} = \lambda(\boldsymbol{S}_\mathrm{w} + \eta \boldsymbol{I}_\mathrm{d})\boldsymbol{\varphi}.$$

Affinity matrix with side information: In the original LFDA paper [8], the affinity matrix $\boldsymbol{A}$ is determined from input $\boldsymbol{x}$ using the *local scaling method* [11]. However, if the input features $\boldsymbol{x}$ are very noisy, the affinity matrix computed by the local scaling method can provide poor local information.

In this paper, exploiting the fact that side information of an image is useful for image classification [5], we propose using the side information $\boldsymbol{v}$ to compute the affinity matrix as

$$A_{i,j} = \exp\left(-\frac{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|^2}{2\sigma^2}\right),$$

where σ is a parameter. Side information is then naturally included in dimensionality reduction via the affinity matrix, and LFDA can find a transformation matrix such that images that have similar side information are close in the embedding space.

3. EXPERIMENTS

In this section, we experimentally investigate the interest of using socially-generated side information within dimensionality reduction. The proposed dimensionality reduction method is compared with existing methods on an image dataset gathered from Pinterest.

3.1 Pinterest Dataset and Side Information

In this paper, we use images obtained from Pinterest, which is an emerging social curation service focused on image sharing. A Pinterest user organizes his images through one or more *boards* to which he can add either new images (a.k.a. *pin*) or images pinned by other Pinterest users before (a.k.a. *repin*). A typical example of the structure of Pinterest is shown in Figure 1.

To explain how Pinterest can provide socially-generated side information, we report two observations made from Pinterest data. First, we notice that, although the users are free to organize their pins as they want, most of the boards tend

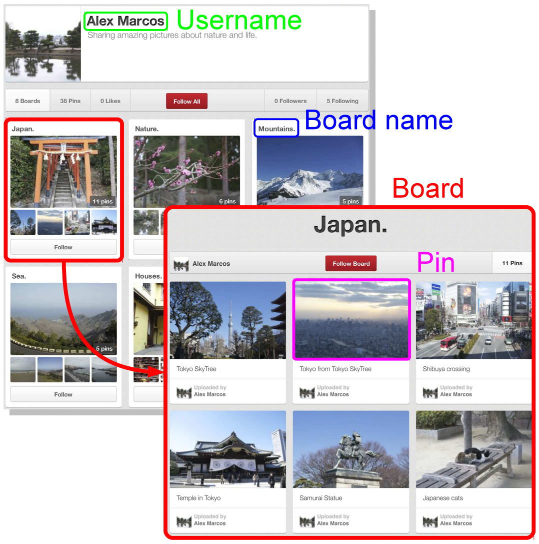


Figure 1: An example of Pinterest user. Source: pinterest.com.

to focus on a particular topic (e.g. trips, cooking, movies, fashion). Thus, one can assume that two images found in the same board are very likely related to the same topic. Second, we find that one user tends to show an interest in a relatively small number of different topics and thus tends to pin images related to one of those topics. The above two observations show intuitively why Pinterest user behavior naturally creates useful side information that can be used within our method.

We now describe the three types of socially-generated side information that we propose to extract from Pinterest data, namely *users information*, *boards information* and *clusters information*. This side information is derived from the behavior of the user in the social network.

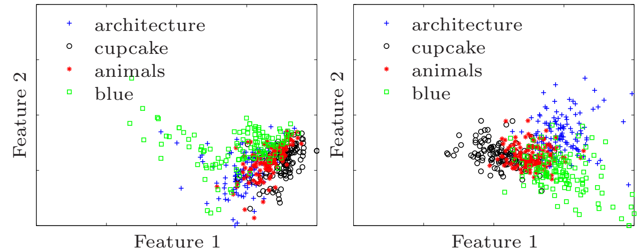
Users information: Let U be the set of Pinterest users, the vector $\boldsymbol{v}_i^{\text{user}} \in \{0,1\}^{|U|}$ associated with each image i contains the users information. If user j pinned or repinned image i, then the j-th element of $\boldsymbol{v}_i^{\text{user}}$ is set to 1 (0 otherwise).

Boards information: Let B be the set of all Pinterest boards. We define a vector $\boldsymbol{v}_i^{\text{board}} \in \{0,1\}^{|B|}$. If image i has been pinned to board l, then the l-th element of $\boldsymbol{v}_i^{\text{board}}$ is set to 1 (0 otherwise).

Clusters information: We create clusters of boards by adaptive hierarchical clustering with the Louvain method [1] based on a board diffusion graph. This graph contains one vertex for each board and weights an edge between two vertices based on the number of images that are shared by those two boards. Note that an image belongs to one or several boards and that each board belongs in turn to one cluster. Let K denote the set of all clusters. We define a vector $\boldsymbol{v}_i^{\text{cluster}} \in [0,1]^{|K|}$ for each image i. The k-th element of $\boldsymbol{v}_i^{\text{cluster}}$ is given by $n_{i,k}/n_i$ where n_i is the number of times the image i has been pinned and $n_{i,k}$ is the number of times the image i is pinned to a board that is part of cluster k.

Note that the proposed side informations could change over time. For example, new users can sign in to the service and new boards can be added. This could be an applicability issue in some cases as it would imply to recompute the side informations for all images.

Images are collected together with the side information $\boldsymbol{v}$ from Pinterest. We create a database composed of 12,500 images and their side information that are uniformly distributed among 10 classes whose labels are 'architecture', 'fashion', 'cupcake', 'animals', 'chocolate', 'flowers', 'blue', 'sea', 'Christmas' and 'green'. The images are resized to 256×256 and split into 4×4 grid sub-regions on which GIST features [6] are computed. Finally, each image is described by a 960-dimensional feature vector $\boldsymbol{x}$. The numbers of users $|U|$, boards $|B|$ and clusters $|K|$ are 171, 634 and 389, respectively.



(a) LFDA (b) Proposed (clusters)

Figure 2: 2-D data visualization example for 4 classes. The two most relevant embedded features obtained by LFDA and the proposed method with clusters affinity matrix are selected for plotting.

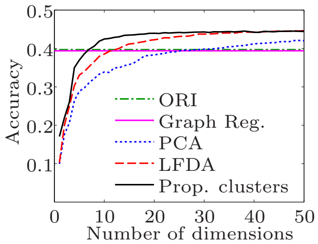
		ORI	Graph Reg. $\gamma = 0.002$	PCA	LFDA	Prop.
	m	960	960	10	10	10
Linear reg.		39.7	39.4	33.6	38.7	**42.5**
NN	k=15	37.9	–	35.7	42.5	**45.1**
	k=20	38.1	–	36.2	42.7	**45.4**
	k=25	38.2	–	36.4	42.8	**45.6**

Table 1: Classification accuracy (in %) averaged over 100 runs. 'ORI' denotes the original data without dimension reduction. 'Prop.' denotes the features embedded by the proposed method with clusters information. For each learning method, a bold font indicates the best dimensionality reduction method that is significantly better than the others according to a one-tailed Student's t-test ($p = 0.01$).

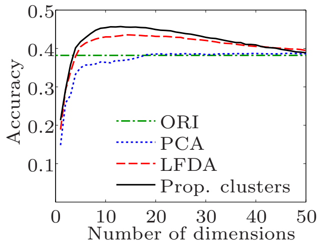
3.2 Results

In this section, we illustrate the behavior of the proposed dimensionality reduction method. Our results clearly show the interest of using socially-generated side information for dimensionality reduction.

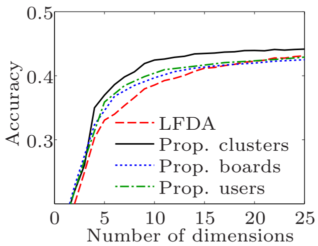
Setup: We compare the proposed method with principal component analysis (PCA), original LFDA without side information, and a graph regularized method ('Graph reg.') [5] with squared loss. We experimentally set the regularization parameter for Graph reg. to $\gamma = 0.002$. Classification is performed with linear regression and k-nearest neighbors for dimensionality reduction methods. Then, the classification performance is assessed based on the classification accuracy averaged over 100 runs. Each classifi-



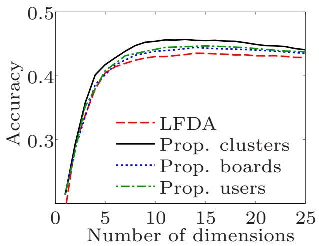
(a) Linear regression



(b) 25-Nearest neighbors



(c) Linear regression



(d) 25-Nearest neighbors

Figure 3: (a)-(b) show the evolution of classification accuracy vs. the number of dimensions for different dimensionality reduction methods. The 'Proposed' method uses an affinity matrix computed from clusters information. (c)-(d) compare the classification accuracy of the traditional LFDA to three types of affinity matrices used with the proposed method.

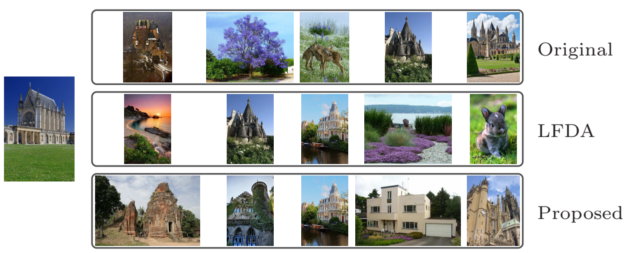


Figure 4: Image retrieval example for a query of architecture class. The leftmost image represents the query. The retrieved images are shown in the boxes. Source: pinterest.com.

cation run randomly selects $5,000$ images for both learning and testing. The regularization parameter in LFDA is experimentally set to $\eta = 0.1$. The kernel width is set to $\sigma = 2^{-0.5}\text{median}(\{\|\boldsymbol{v}_i - \boldsymbol{v}_j\|\}_{i,j=1}^{n})$ following a common heuristic [7].

Data visualization: Figure 2 shows embedding results realized by LFDA and the proposed method for the architecture, cupcake, animals and blue classes. The proposed method tends to separate samples of different classes from each other, while LFDA tends to mix the samples together.

Classification performance: Figure 3-(a)(b) show the classification accuracy as a function of the dimensionality m of the low-dimensional space. Table 1 shows the performance with $m = 10$, where a bold font indicates the best dimensionality reduction method, which is significantly better than the others according to a one-tailed Student's t-test ($p = 0.01$). The graphs and table clearly show that the proposed side-information-based dimensionality reduction method compares favorably with existing methods. For example, in Table 1 the features embedded by the proposed method lead to 19% higher accuracy than the full GIST features (i.e., ORI) and to 6% higher accuracy than the features embedded by LFDA without side information. Figure 3-(c)(d) compare results for the proposed method with different types of side information. We observe that the clusters information tends to give the best performance.

Image retrieval: Figure 4 shows an image retrieval result for a query of architecture class. The images are retrieved by a 5-NN search. The original and LFDA approaches recover 3 and 2 architecture images, respectively, while our method retrieves 5 of them.

4. CONCLUSION

In this paper, we showed that *socially-generated side information* can be used to improve dimensionality reduction in the context of several image-related tasks. More specifically, we proposed a dimensionality reduction method based on LFDA that leverages socially-generated side information. The key idea behind the proposed approach is to seek an embedding transformation so that images that have similar side information are close in the embedding space. Through experiments on an image dataset obtained from Pinterest, we showed that incorporating socially-generated side information within a dimensionality reduction method is useful for image classification, data visualization and image retrieval.

References

[1] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008.

[2] X. He and P. Niyogi. Locality preserving projections. In *NIPS*, 2004.

[3] X. Hu, L. Tang, J. Tang, and H. Liu. Exploiting social relations for sentiment analysis in microblogging. In *ACM workshop WSDM*, 2013.

[4] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[5] D. K. Mahajan and M. Slaney. Image classification using the web graph. In *ACM MM*, 2010.

[6] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 2001.

[7] B. Schölkopf and A. J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press, 2001.

[8] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 2007.

[9] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, 1991.

[10] X. Wu, W.-L. Zhao, and C.-W. Ngo. Towards google challenge: combining contextual and social information for web video categorization. In *ACM MM*, 2009.

[11] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *NIPS*, 2004.

[12] J. R. Zhang, Y. Song, and T. Leung. Improving video classification via youtube video co-watch data. In *ACM workshop SBNMA*, 2011.