An Efficient Algorithm for Learning Distances that Obey the Triangle Inequality

Arijit Biswas http://www.xrci.xerox.com/profile-main/67 David Jacobs http://www.cs.umd.edu/~djacobs/

Semi-supervised clustering of images has been an interesting problem for machine learning and computer vision researchers for decades. Pairwise constrained clustering is a popular paradigm for semi supervision that uses knowledge about whether two images belong to the same category (must-link constraint) or not (can't-link constraint). Performance of constrained clustering algorithms can be improved if the supervision on some image pairs is used to modify the pairwise distances of other image pairs, for which no supervision is available. There are several excellent metric learning approaches when the image distances can be represented as Euclidean distances in vector spaces [6]. However, in many cases distances are computed robustly [3], on manifolds [4] or as the output of algorithms [1] or classifiers. These do not lead to a natural embedding in a metric space, and may not even obey the triangle inequality. We are particularly interested in semi-supervised clustering of fine-grained categories. If we look at the top 10 distance measures in two important domains (LFW faces and leaf shapes) related to fine-grained classification, we find that 80% of the methods use non-vector space distances. These distances can be used for clustering images, but are not suitable for existing metric learning algorithms.

In order to propagate constraints from supervised to unsupervised pairs, some structure must be assumed on the set of possible distances. Otherwise, the distance between supervised pairs could be altered without affecting the distance between unsupervised pairs. Perhaps the weakest assumption that we can make about a distance is that it obeys the triangle inequality. Enforcing the triangle inequality allows us to propagate constraints; if a constraint alters one distance, other distances must also change to maintain triangle inequalities. For many interesting distances, the triangle inequality is not guaranteed to hold. However, we empirically find that the triangle inequality almost always holds for distances computed for fine-grained classification even when not explicitly enforced. This strongly motivates us to enforce the triangle inequalities when we alter distances to incorporate the pairwise constraints. We then find empirically that by enforcing the triangle inequality we can improve performance on several real world datasets.

Our main contribution is to formulate distance learning with pairwise constraints as a metric nearness problem¹ [2] and then provide an efficient algorithm to solve metric nearness for clustering. First, we formulate a quadratic optimization problem, where the pairwise distances between images are modified such that pairwise constraints and triangle inequality constraints are satisfied as much as possible. Since enforcing $O(N^3)$ triangle inequalities is computationally expensive, we propose a graph based approach, where only O(n(M + C)) triangle inequalities are sampled for use in the QP (*N* is the total number of images, *n* is the number of nearest neighbors in the *n*-nearest neighbor graph, *M* and *C* are the number of must-link and can't-link constraints respectively). We empirically show that this sampling approach works well in practice. We use the distances obtained by our approach along with a constrained clustering algorithm [5] to achieve state-of-the-art clustering results.

We theoretically analyze a simplified case in which only one pairwise constraint is present, to gain insights into our fast approach. Our sampling approach is based on the intuition that clustering is predominately affected by small distances, and is not sensitive to the exact value of larger distances. We prove that our sampling approach produces the same set of small distances that would be obtained by enforcing all constraints. We perform experiments on leaf and face image/video datasets and show that distances obtained by our method achieve state-of-the-art clustering results.

Formulation: We begin with a set of *N* unlabeled images \mathcal{U} ($x \in \mathcal{U}$) from *K* classes. We are also provided with initial distances between all

Xerox Research Centre India Bangalore, India Computer Science Department University of Maryland College Park, USA

the image pairs, i.e., $d_I(\mathbf{x}_i, \mathbf{x}_j)$ is given $\forall i, j$ (note that d_I denotes initial distance). However we do not have any vector representation of the images. We have access to a set of must-link and can't-link constraints. Let \mathcal{M} denote the set of must-link constraints such that any pair $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}$ implies that \mathbf{x}_i and \mathbf{x}_j belong to the same class. Similarly \mathcal{C} denotes the set of can't-link constraints such that any pair $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{H}$ implies that \mathbf{x}_i and \mathbf{x}_j belong to the same class. Similarly \mathcal{C} denotes the set of can't-link constraints such that any pair $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}$ implies that \mathbf{x}_i and \mathbf{x}_j belong to different classes.

We learn a new set of distances for all the image pairs given the pairwise constraints \mathcal{M} and \mathcal{C} . A quadratic optimization problem is formulated to find these distances. Let us assume that the set of final distances are given by $d_F(\mathbf{x}_i, \mathbf{x}_j)$, which we can obtain by solving the following quadratic optimization problem:

$$\begin{array}{ll}
\begin{array}{ll} \underset{d_{F}}{\text{minimize}} & \sum_{(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \notin \mathcal{M} \cup \mathcal{C}} (d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) - d_{I}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}))^{2} \\
\text{subject to} & (i) \quad d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \leq U, \quad (\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \in \mathcal{M} \\
& (ii) \quad d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \geq L, \quad (\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \in \mathcal{C} \\
& (iii) \quad d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) + d_{F}(\boldsymbol{x}_{j}, \boldsymbol{x}_{k}) \geq d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{k}), \forall i, j, k \\
& (iv) \quad d_{F}(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) \geq 0, \forall i, j
\end{array}$$
(1)

In the formulation in Eq. 1, distances between all image pairs are modified such that the constraints (i) to (iv) are satisfied. The objective function minimizes the total sum of changes in pairwise distances. Constraint (i) causes distances corresponding to must-link constraints to be reduced so they are upper-bounded by U, a user defined constant. Similarly, constraints in (ii) move can't-link image pairs as far as possible. The triangle inequality constraints are added in (iii). These triangle inequality constraints propagate information about must-link and can't-link constraints to other image pairs. The final set of constraints in (iv) ensure that all the quadratic optimization variables, i.e., the pairwise distances, remain non-negative.

In this QP formulation, there are large number of QP variables $(O(N^2))$ and triangle inequality constraints $(O(N^3))$. Also with all the pairwise and triangle inequality constraints there may not exist a feasible solution. To avoid these issues we reduce the size of the QP significantly by determining which triangle inequalities are crucial for clustering using a novel graph-based formulation (please see the paper for further details).

- Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, editors, *NIPS*, pages 831–837. MIT Press, 2000.
- [2] J Brickell, I. S. Dhillon, Suvrit Sra, and J Tropp. The metric nearness problem. SIAM J. Matrix Analysis and Applications, April 23 2008. URL http://eprints. pascal-network.org/archive/00004443/;http: //link.aip.org/link/?SML/30/375.
- [3] David W. Jacobs, Daphna Weinshall, and Yoram Gdalyahu. Classification with nonmetric distances: Image retrieval and class representation. *IEEE Trans. Pattern Anal. Mach. Intell*, 2000. URL http:// doi.ieeecomputersociety.org/10.1109/34.862197.
- [4] F. R. Schmidt, M. Clausen, and D. Cremers. Shape matching by variational computation of geodesics on a manifold. In *DAGM*, pages 142–151, 2006. URL http://dx.doi.org/10.1007/ 11861898_15.
- [5] Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schroedl. Constrained K-means clustering with background knowledge. In *ICML*, 2001.
- [6] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart J. Russell. Distance metric learning with application to clustering with side-information. In *NIPS*, 2002. URL http://books.nips.cc/papers/files/nips15/AA03.pdf.

¹Given a dissimilarity matrix, find the "nearest" matrix of distances that satisfy the triangle inequalities.