Graphs Regularization for Data Sets and Images: Filtering and Semi-Supervised Classification

Vinh Thong Ta VINH-THONG.TA@ETU.INFO.UNICAEN.FR
Olivier Lezoray OLIVIER.LEZORAY@UNICAEN.FR
Abderrahim Elmoataz ELMOATAZ@CHBG.UNICAEN.FR
LUSAC EA 2607, Vision and Image Analysis, 6 Bd. Maréchal Juin, F-14050 Caen, FRANCE

Abstract

In this paper, we propose a general discrete regularization framework on weighted graphs of arbitrary topology. This framework unifies filtering and classification methods and can be applied on data sets or images. Experimental results show the abilities of the proposed algorithm on classical images and data sets problems and the similarities between both.

1. Introduction

Recent years have seen an important interest on graph based differential geometry methods in image processing or semi-supervised learning methods. They use random walks, diffusions maps or spectral methods (Coifman et al., 2005; Qiu & Hancock, 2005; Zhou et al., 2004). Inspired by continuous regularization of images, a general discrete regularization framework based on weighted graphs of arbitrary topology is proposed. This framework unifies images and data sets processing by modeling and solving problems within graph theory. In this paper, some notations and definitions for the proposed general framework are described in section 2. In section 3, application results of this regularization for data sets filtering and classification problems are shown. Moreover, relationship between image and data analysis and the benefits they can gain one to each other are also described.

2. Weighted Graph Regularization Framework

Let $G_w = (V, E)$ be a weighted graph described by a set V of *vertices* (representing for instance image pixels, data features, etc.), a subset $E \subseteq V \times V$ of *edges*, and a weight function w reflecting the dissimilarity between two vertices of the graph. The graphs are considered as simple, always connected, undirected, with no self loops and no multiple edges. Let $\mathcal{H}(V)$ be the Hilbert space defined on the vertices of a graph G_w. The regularization of a given function $f^0 \in \mathcal{H}(V)$ consists in seeking a function $f^* \in \mathcal{H}(V)$ which is sufficiently smooth to respect the graph structure and also close enough to f^0. This discrete minimization problem can be formalized by:

$$f^* = \min_{f\in\mathcal{H}(V)} \left\{ \sum_{v\in V} \|\nabla f(v)\|^p + \lambda \|f - f^0\|^2_{\mathcal{H}(V)} \right\} \quad (1)$$

where $p \in [1, +\infty]$ measures the smoothness of the function over the graph and $\lambda \geq 0$ is the fidelity parameter specifying the tradeoff between the two competing terms. The weighted *gradient* operator of a function f at vertex v is defined as $||\nabla_v f|| = \sqrt{\sum_{u\sim v} w_{uv}(f_v - f_u)^2}$. The weighted *p-Laplace* operator of a function f at the vertex v can be expressed as : $(\Delta_p f)_v = \sum_{u\sim v} w_{uv}(||\nabla_v f||^{p-2} + ||\nabla_u f||^{p-2})(f_v - f_u)$. Here $u \sim v$ means that an edge exists between vertices u and v. For the most frequently used $p = 2$ case, the solution of (1) is also the solution of the system of equations defined as: $\Delta f^* + 2\lambda(f^* - f^0) = 0$ which can be solved by using a Gauss-Jacobi iterative resolution. The linearized algorithm is defined by, for all $v \in V$:

$$\begin{cases} f^{(0)} = f^0 \\ f_v^{(t+1)} = (\lambda + \sum_{u\sim v} w_{uv})^{-1}(\lambda f_v^0 + \sum_{u\sim v} w_{uv} f_v^{(t)}) \end{cases} \quad (2)$$

For more details, please refer to Bougleux et al. (2007) and Lezoray et al. (2007).

3. Applications

In this section, application of the proposed algorithm (2) to filtering and semi-supervised clustering on both data sets and images fields are exposed.

3.1. Images and Data Sets Filtering

In image segmentation context, image filtering is a well know important pre-processing step to improve segmentation results. Figure 1 clearly shows the benefits

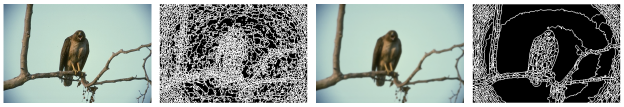

Figure 1. From left to right: original image, its segmentation, filtered image, its segmentation.

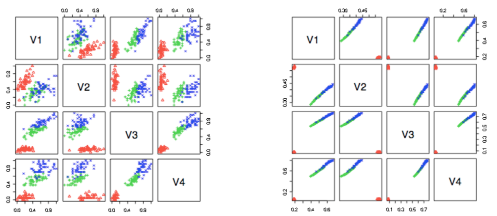

Figure 2. Left: original Iris data set, right: filtered Iris data set (all features pairs projections where V_i denotes a feature).

of image filtering (accomplished by proposed regularization framework) on segmentation results. Here a classical *watershed* segmentation algorithm is used to segment either the original and filtered images. Therefore, filtering is of interest to simplify further analysis. We propose to improve data sets classification by the same means. Figure 2 shows considered filtering method on the popular multivariate *Iris* data set. Table 1 presents classification accuracies with the well know *k-nearest neighbors* classifier on both original and filtered database. Results clearly show that filtering efficiently enhances the clustering.

Table 1. Classification accuracies with *k-nn* classifier ($k = 3$) using sepal length ($V1$) and width ($V2$) features in original and filtered iris data sets.

Original iris data set	66.7 ± 0.1 %
Filtered iris data set	93.3 ± 0.1 %

3.2. Data Sets and Images Semi-Supervised Classification

Recent years have seen a surge of efficient graph based label propagation methods. Our approach is used here for image semi-supervised classification. Starting from a user labeled image, labeled vertices are propagated over the graph and unlabeled ones are found by *transductive learning* (Zhou et al., 2004). Figure 3 shows a semi-supervised segmentation result by applying proposed diffusion framework.

4. Conclusion

We have considered a general framework unifying data sets and images analysis methods. The relationships between data sets and image fields have been shown by

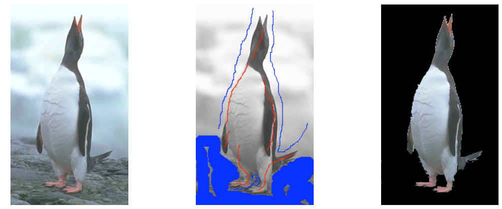

Figure 3. From left to right: original image, user labeled image (blue for the background and red for the foreground), segmentation result (black background).

applying a method from one context to the other to improve results. For instance, data sets filtering as a preprocessing step can enhance data clustering results and in the same way data sets classification methods can be used for image segmentation. As future works, we plan to extend this regularization framework to other fields and to combine it to others techniques.

Acknowledgments

This work was partially supported under a research grant of the ANR Foundation (ANR-06-MDCA-008-01/FOGRIMMI), and a doctoral grant of the Conseil Régional de Basse-Normandie and of the Cœur et Cancer association in collaboration with the Department of Anatomical and Cytological Pathology from Cotentin Hospital Center.

References

Bougleux, S., Elmoataz, A., & Melkemi, M. (2007). Discrete regularization on weighted graphs for image and mesh filtering. *SSVM'07: Proc. of the 1st International Conference on Scale Space and Variational Methods in Computer Vision, to appear.*

Coifman, R. R., Lafon, S., Lee, A., Maggioni, M., Nadler, B., Warner, F., & Zucker, S. (2005). Geometric diffusions as a tool for harmonic analysis and structure definition of data. *Proc. of National Academy of Sciences*, 7426–7431.

Lezoray, O., Bougleux, S., & Elmoataz, A. (2007). Graph regularization for color image processing. *Computer Vision and Image Understading*, in press.

Qiu, H., & Hancock, E. R. (2005). Commute times for graph spectral clustering. *CAIP* (pp. 128–136).

Zhou, D., Bousquet, O., Lal, T., Weston, J., & Schlkopf, B. (2004). Learning with local and global consistency. *Advances in Neural Information Processing Systems* (pp. 321–328). Cambridge, MA, USA: MIT Press.