Unsupervised Data Labeling on Graphs by Self-Assignment Flows

Matthias Zisler\textsuperscript{1,*}, Artjom Zern\textsuperscript{1}, Bastian Boll\textsuperscript{1}, Stefania Petra\textsuperscript{2}, and Christoph Schnörr\textsuperscript{1}

\textsuperscript{1} Image & Pattern Analysis Group, Universität Heidelberg, 69120 Heidelberg, Germany
\textsuperscript{2} Mathematical Imaging Group, Universität Heidelberg, 69120 Heidelberg, Germany

This paper extends the recently introduced assignment flow approach for supervised image labeling to unsupervised scenarios where no labels are given. The resulting self-assignment flow takes a pairwise data affinity matrix as input data and maximizes the correlation with a low-rank matrix that is parametrized by the variables of the assignment flow, which entails an assignment of the data to themselves through the formation of latent labels (feature prototypes). A single user parameter, the neighborhood size for the geometric regularization of assignments, drives the entire process. By smooth geodesic interpolation between different normalizations of self-assignment matrices on the positive definite matrix manifold, a one-parameter family of self-assignment flows is defined. Accordingly, our approach can be characterized from different viewpoints, e.g. as performing spatially regularized, rank-constrained discrete optimal transport, or as computing spatially regularized normalized spectral cuts. Regarding combinatorial optimization, our approach successfully determines completely positive factorizations of self-assignments in large-scale scenarios, subject to spatial regularization. Various experiments including the unsupervised learning of patch dictionaries using a locally invariant distance function, illustrate the properties of the approach.

1 Motivation and Overview

Assignment flows, as introduced in [1], constitute smooth dynamical systems for data labeling on graphs. To this end, the assignment of labels to data is encoded as a point on an elementary statistical manifold. A flow on this manifold evolves the unique unbiased initial assignment towards desired pixel-label decisions. See [2] for additional background, details and related work. In the supervised scenario, data are assumed to lie in a known metric space in which the unique unbiased initial assignment towards desired pixel-label decisions. See [2] for additional background, details and related work. In the supervised scenario, data are assumed to lie in a known metric space in which each associated with a class label in $\{1, \ldots, c\}$, are given. However, this constitutes strong prior knowledge which may not be available in applications. Therefore, an adaptation of the assignment flow framework to unsupervised scenarios was introduced in [3] where the feature space has a known Riemannian manifold structure and prototypes are dynamically adapted over the course of the labeling process.

In the present paper, we discuss a different and more general unsupervised assignment flow variant which does not require knowledge of feature space structure other than a metric. Additionally, prototypes are not sought explicitly but emerge from data self-assignment. The latter constitutes a shift in perspective away from assigning labels to data and towards grouping data into classes by deciding for each pair of data if they belong to the same class. The resulting procedure thus performs clustering of the data. Simultaneously, regularization is induced by underlying graph structure without introducing bias through feature augmentation. Many additional aspects are discussed in [4].

2 Clustering and Self-Assignment Matrices

For a graph with $n$ nodes, matrices $W \in \mathcal{W} = \{W \in \mathbb{R}^{n \times c} : W \geq 0, W1_c = 1_n\}$ encode soft assignment of labels to nodes. Specifically, each row of $W$ can be seen as a probability distribution over the set $[c]$ of labels. In the context of clustering, the matrix $C(W) := \text{Diag}(W^T 1_n) = W^T W$ carries the cardinalities of $c$ clusters on its diagonal if $W$ is restricted to the subset $\mathcal{W}_c$ of integer assignment matrices with full rank. Let $K_F \in \mathbb{R}^{n \times n}$ be given with entries $(K_F)_{i,k}$ denoting similarities between vectors $f_i, f_k$ in a feature space $\mathcal{F}$. The combinatorially hard task of partitioning $n$ data points $f_i \in \mathcal{F}$ into $c$ clusters such that within-cluster variance is minimized and between-cluster variance is maximized can be written as

$$\max_W \langle K_F, WC(W)^{-1} W^T \rangle \quad \text{subject to } W \in \mathcal{W}_c. \quad (1)$$

If the integrality constraint in (1) is relaxed, the equality $\text{Diag}(W^T 1_n) = W^T W$ no longer holds, leading to the consideration of a family of self-assignment matrices

$$A_s(W) = W \gamma_s(W)^{-1} W^T, \quad s \in [0, 1], \quad W \in \begin{cases} \mathcal{W}_c & \text{if } s = 0 \\ \mathcal{W}_c & \text{if } s > 0 \end{cases} \quad (2)$$

where the normalizing matrix $\gamma_s(W)$ is chosen on the geodesic between $C(W)$ and $W^T W$ in the cone $\mathcal{P}$ of symmetric positive definite matrices, i.e. $A_0(W) = WC(W)^{-1} W^T$ and $A_1(W) = W(W^T W)^{-1} W^T$.

* Corresponding author: e-mail zisler@math.uni-heidelberg.de
3 Self-Assignment Flows

In [1], data is incorporated into the assignment flow vector field by lifting the gradient of an objective to be maximized onto $W$. Analogously, a family of self-assignment flows is defined by lifting the gradient of the relaxed clustering objective $E_s(W) = \langle K, A_s(W) \rangle$ and otherwise proceeding without further modification to the supervised scenario. In particular, existing schemes for numerical integration of resulting flows [5] still apply.

Looking at $A_s$ in the relaxation of (1) clarifies the term self-assignment. By viewing the entries of a matrix $W \in W$ as posterior probabilities $W_{i,j} = P(j|i)$ of assigning label $j$ to node $i$, we obtain the label distribution $P(j) = \sum_{i \in \mathcal{V}} P(j|i) P(i) = \frac{1}{n} C(W)_{j,j}$ by marginalizing with uniform prior. By invoking Bayes’ rule, we find

$$P(k|j) = \frac{P(j|k)P(k)}{P(j)} = \frac{W_{k,j}}{\sum_{i \in [n]} W_{i,j}} = (C(W)^{-1}W^\top)_{j,k}, \quad k \in [n]$$

such that the probability of nodes $i$ and $k$ having the same label can be seen as the self-assignment ($i \leftrightarrow k$) probability

$$P(k|i) = \sum_{j \in [c]} P(k|j)P(j|i) = \sum_{j \in [c]} W_{i,j}(C(W)^{-1}W^\top)_{j,k} = (WC(W)^{-1}W^\top)_{i,k} = (A_0)_{i,k}.$$

This makes clear that the objective $E_s(W)$ is to maximize agreement between similarity in $F$ and (self-)assignment to the same label for each pair of nodes.

4 Numerical Experiments and Conclusion

Figure 1 illustrates application of the proposed self-assignment flow for $s = 0$. $F$ is chosen as the space of $7 \times 7$ image patches with rotation invariant distance. Graph connectivity is induced by neighborhoods of size $3 \times 3$ in the image grid. To break symmetry, the unbiased barycenter initialization was slightly perturbed as informed by preliminary $k$-center clustering. Once the flow has converged to a low-entropy assignment state, latent prototypes can in this case be recovered by computing means in $F$, yielding a learned patch dictionary of prototypes. Reconstruction of the image by assigned patches illustrates how the majority of image structure is clearly captured in the learned patch dictionary. In large-scale scenarios, the matrix $K_F$ may require a large amount of memory such that computational efficiency is greatly reduced by using a low-rank factorization. The analysis in [4] shows that this can lead to very similar results while using less than 1% of otherwise required memory.

The whole process is driven by a single parameter – the neighborhood size inducing graph connectivity. In particular, choosing large neighborhoods, i.e. strong assignment regularity, leads to prototypes dying out, highlighting the plug-in-and-play nature of self-assignment flows. We conclude that self-assignment flows provide a flexible mathematical framework for clustering under additional regularization induced by a graph structure underlying the data. This is achieved in a geometrically natural way, without the need for many user parameters and without feature augmentation. A number of additional aspects is discussed in [4].

Acknowledgements This work is supported by Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster).

Open access funding enabled and organized by Projekt DEAL.

References

[1] F. Aström, S. Petra, B. Schmitzer, and C. Schnörr, J. Math. Imag. Vision 58(2), 211–238 (2017).
[2] C. Schnörr, Assignment Flows, in: Handbook of Variational Methods for Nonlinear Geometric Data, edited by P. Grohs, M. Holler, and A. Weinmann (Springer, 2020), p. 235–260.
[3] A. Zern, M. Zisler, S. Petra, and C. Schnörr, Journal of Mathematical Imaging and Vision (2020).
[4] M. Zisler, A. Zern, S. Petra, and C. Schnörr, SIAM Journal on Imaging Sciences (2020).
[5] A. Zeilmann, F. Savarino, S. Petra, and C. Schnörr, Inverse Problems 36(3), 034004 (33pp) (2020).