

ANALYSIS OF HYPERSPECTRAL DATA BY MEANS OF TRANSPORT MODELS AND MACHINE LEARNING

Wojciech Czaja, Dong Dong, Pierre-Emmanuel Jabin, Franck O. Ndjakou Njeunje

University of Maryland
Department of Mathematics and CSCAMM
4176 Campus Dr, College Park, MD, 20742

ABSTRACT

We present a new physics-inspired method for analysis of hyperspectral imagery (HSI). The method is based on the concept of transport models for graphs. The proposed approach generalizes existing dimension reduction and feature extraction algorithms, by replacing the role of diffusion processes, as a measure of estimating proximity, with dynamical systems. This approach allows us to exploit different and new relationships within the complex data structures, such as those arising in HSI. We demonstrate this by proposing a specific multi-scale algorithm in which transport models are used to translate the information about contextual similarities of material classes to enhance feature extraction and classification results. This point is illustrated with a series of computational experiments.

Index Terms— feature extraction, dimension reduction, machine learning, transport operator, advection

1. BACKGROUND

Hyperspectral images (HSI) play an important role in remote sensing [1]. The challenging problems of classifying and clustering HSI have attracted much attention in recent years [2]. The key process in many solutions is a feature extraction or dimensionality reduction technique, such as Principal Components Analysis (PCA) [3], Isomap (ISO) [4], Diffusion Maps (DIF) [5], or Laplacian Eigenmaps (LE) [6].

Inspired by LE, by the Schroedinger Eigenmaps method (SE) [7, 8], and many applications of advection operators [9, 10, 11, 12, 13], we recently introduced the concept of *Transport Eigenmap* (TE) as a new feature extraction method [14]. This is a semi-supervised learning technique with a rich set of parameters, allowing various ways to make use of partial groundtruth. The overall performance and robustness of the transport model are satisfactory. The purpose of this note is to present the transport model in a more intuitive way, to

further test it through new experiments, and to discuss new directions of applications.

2. THE TRANSPORT MODEL

We begin with a brief introduction of the transport model. For more details we refer the reader to [15] and [14]. Following the idea of LE, a collection of n data points can be associated with a weighted graph G with n nodes. The weights $w_{i,j}$ depend on the distances between pairs of nodes and mimic a diffusion process with a graph Laplacian L . Let v be a function defined on the edges of G , which can be viewed as an $n \times n$ matrix. We assume v to be anti-symmetric to model a velocity field. The associated transport operator T acting on a vector $\mathbf{y} = \{y_i\}$ is formally defined as

$$T \mathbf{y} = L \mathbf{y} - \text{div}(v \mathbf{y}).$$

A discretized version of this operator is:

$$(T \mathbf{y})_i = \sum_j (y_i - y_j) w_{ij} - \sum_j (y_i + y_j) \frac{v_{ij}}{2}. \quad (1)$$

For any positive definite matrix A , we denote the inner product $\langle \mathbf{y}, \mathbf{z} \rangle_A := \mathbf{y}^t A \mathbf{z}$.

Assume that $v_{ij} = 0$ if the nodes i and j are not connected, as $w_{ij} = 0$ in this case as well. Let $\bar{v}_{ij} := \frac{v_{ij}}{2w_{ij}}$ if i and j are connected, and $\bar{v}_{ij} = 0$ otherwise. Then $\bar{v}$ is also anti-symmetric, $\frac{v_{ij}}{2} = \bar{v}_{ij} w_{ij}$, and

$$(T \mathbf{y})_i = \sum_j [(1 - \bar{v}_{ij}) y_i - (1 + \bar{v}_{ij}) y_j] w_{ij}.$$

A desired property for the transport operator T is that it has real eigenvalues, similarly to the graph Laplacian. It was observed in [14] that if we choose $\bar{v}_{ij} = \frac{a_j - a_i}{a_j + a_i}$ for some positive a_i 's, then the operator $(T \mathbf{y})_i = \sum_j [y_i - y_j - \bar{v}_{ij} (y_i + y_j)] w_{ij}$ is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_X$, with $X = \text{diag}(ca_i)$ for some positive c .

We can easily extend this result to a more general model by introducing a symmetric matrix r . Define T_v^r to be the operator such that

$$(T_v^r \mathbf{y})_i = \sum_j [r_{ij} (y_i - y_j) - \bar{v}_{ij} (y_i + y_j)] w_{ij}. \quad (2)$$

WC was partially supported by LTS grant D00030014 and by NSF grant DMS-1738003. DD was partially supported by LTS grant DO 0052. PEJ was partially supported by NSF Grant 161453, NSF Grant RNMS (Ki-Net) 1107444 and by LTS grant D00030014.

Assume $\bar{v}_{ij} = \frac{a_j - a_i}{a_j + a_i} r_{ij}$ for some positive a_i 's. Then T_v^r is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_X$, with $X = \text{diag}(ca_i)$ for some positive c [14].

When $\bar{v}_{ij} = \frac{a_j - a_i}{a_j + a_i} r_{ij}$, the operator T_v^r can be proved to be non-negative in ℓ_X^2 , where $X = \text{diag}(ca_i)$ for some positive c .

We recall for comparison that the Laplacian operator is $(Ly)_i = \sum_j (y_i - y_j)w_{ij}$. T_v^r generalizes L in two ways. First, a_i modifies the measure/coordinate and thus makes the representation of i -th point closer to the origin if a_i is large or further away from the origin if a_i is small. Second, r_{ij} can enlarge or reduce the weight w_{ij} between two nodes i and j , serving as a weight modifier. We then choose parameters to guide the data representation based on partial knowledge about ground truth.

3. IMPLEMENTATION

Suppose there are n data points, $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ in $\mathbb{R}^d$, and we aim to find a map $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ so that $Y = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ in $\mathbb{R}^m$ satisfies $\mathbf{y}_i = \Phi(\mathbf{x}_i)$.

The implementation for Transport Eigenmap is similar to that of LE and SE. First construct the adjacency graph using the k -nearest neighbor (kNN) algorithm. Then define the weight matrix, W , on the graph, by $w_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right)$, if nodes i and j are connected.

The key step of the new algorithm is to use an appropriate transport operator/matrix in place of the classical Laplacian. Let W^r be the matrix with entries $w_{ij}^r = w_{ij} \frac{2r_{ij}}{a_i + a_j}$. Then the matrix form of T is

$$T = \text{diag}(a_i \sum_j w_{ij}^r) - W^r \text{diag}(a_i). \quad (3)$$

There are two special cases which can be used to speed up the algorithm: *transport by advection* (TA), which corresponds to the setting $a_i = 1 + \beta\mu_i$, $r_{ij} = (a_j + a_i)/2$, and *transport by gradient flows* (TG), which is the case $r_{ij} \equiv 1$. To get the matrix form of TA, we set $a_i = 1 + \beta\mu_i$ and $r_{ij} = (a_i + a_j)/2$. Then $TA = L(I + \beta\text{diag}(\mu_i))$, where $L = D - W$ is the Laplacian matrix and I is the identity. For TG, we can let $r_{ij} = 1$ in (3) to get $TG = L - (D_v + Wv)$, where $D_v = \text{diag}(\sum_j w_{ij}v_{ij})$, $Wv = (w_{ij}v_{ij})$ and $v_{ij} = (a_j - a_i)/(a_j + a_i)$.

Finally, let $\{u^0, u^1, \dots, u^{n-1}\}$ be the solution of $T\mathbf{u} = \lambda D\mathbf{u}$, in ascending order according to their eigenvalues. Define the mapping Φ_T by $\mathbf{x}_i \rightarrow \Phi_T(\mathbf{x}_i) = [u_i^1, u_i^2, \dots, u_i^m]$, which gives the final representation of the data. It is this final representation that is subject to classification tests.

In summary, the Transport Eigenmap algorithm provides us with a way to use the advection transport model and to propagate the class-specific information from a single class of choice to improve the clustering on the remaining portion of the data.

4. EXPERIMENTS ON HSI

Unlike the previous previous works [15, 14] which focused on the special forms TA and TG, in this paper we analyze TE in its most general form (3). The main parameters are the matrix r_{ij} and the vector a_i . The default values of these parameters are 1. We set $r_{ij} = 0.9$ if we believe that the points i and j are in different classes and $r_{ij} = 10^4$ if i and j should be in the same class. In general, the ratio of these numbers reflects the measure of confidence in class similarity. We set $a_i = 0$ by default, and $a_i = 20$ for a point i in a pre-identified class.

We choose the Indian Pines, the Salinas, and Pavia University scene [16] in our experiments. For testing, we pick a smaller sub-scene of the Salinas called Salinas-B. The classification task is done by the 1-nearest neighbor algorithm (after various feature extraction techniques). The following metrics are reported: the adjusted Rand index (ARI), overall accuracy (OA), average accuracy (AA), average F-score (FS) and Cohen's kappa coefficient (κ).

We first test TE on the the Indian Pines dataset, assuming the class "soybean" (class 11) is already identified. Similar to the previous testing of TG in [14], TE performs better than other methods in most cases. See Table 1 below.

IP	PCA	LE	DIF	ISO	SE	TE
ARI	0.4426	0.3745	0.4210	0.3930	0.6955	0.7106
OA	0.6761	0.6133	0.6557	0.6309	0.7354	0.7440
AA	0.6403	0.5782	0.6219	0.5979	0.6249	0.6266
FS	0.6471	0.5784	0.6212	0.5996	0.6255	0.6278
κ	0.6301	0.5592	0.6065	0.5785	0.6982	0.7081

Table 1. Classification results for Indian Pines (IP): assume soybean (class 11) is known.

If similar classes (three corn classes, three grass classes and three soybean classes) are grouped into a same class, TE gains more improvement on the new version of Indian Pines, as indicated in the following table.

IPG	PCA	LE	DIF	ISO	SE	TE
ARI	0.5330	0.4785	0.5102	0.4902	0.8929	0.9324
OA	0.7744	0.7307	0.7575	0.7418	0.9088	0.9195
AA	0.6987	0.6462	0.6883	0.6671	0.7111	0.7136
FS	0.7111	0.6479	0.6905	0.6739	0.7157	0.7154
κ	0.6996	0.6423	0.6770	0.6563	0.8788	0.8931

Table 2. Classification results for Indian Pines-G (IPG): assume soybean (class 10) is known

On the Salinas-B dataset, TE outperforms other methods in all metrics if we assume the class "corn" to be identified.

SB	PCA	LE	DIF	ISO	SE	TE
ARI	0.9429	0.9346	0.9164	0.9440	0.9678	0.9723
OA	0.9729	0.9685	0.9603	0.9733	0.9814	0.9825
AA	0.9690	0.9643	0.9564	0.9700	0.9743	0.9753
FS	0.9693	0.9638	0.9557	0.9696	0.9754	0.9757
κ	0.9682	0.9630	0.9534	0.9687	0.9782	0.9795

Table 3. Classification results for Salinas-B (SB): assume corn (class 10) is known

Then we test TE on a grouped version of Salinas-B (denoted by SBG). The four lettuce classes are regarded as a single class, i.e., classes 11-14 are all labelled as class 11. There are five classes remaining in SBG (be aware that the two broccoli classes are NOT grouped). TE is still the best method for SBG.

SBG	PCA	LE	DIF	ISO	SE	TE
ARI	0.9460	0.9421	0.9154	0.9480	0.9780	0.9819
OA	0.9791	0.9767	0.9677	0.9795	0.9900	0.9917
AA	0.9769	0.9750	0.9669	0.9784	0.9877	0.9894
FS	0.9797	0.9763	0.9697	0.9797	0.9879	0.9896
κ	0.9725	0.9694	0.9576	0.9731	0.9868	0.9891

Table 4. Classification results for Salinas-B-G (SBG): assume corn (class 10) is known

The Pavia University dataset is large, and we do not include memory consuming methods of DIF and ISO. For faster implementation, we use the special form TA here, which is good enough as indicated in the following two tables. The class Asphalt and the class Bitumen are regarded as the same in the grouped Pavia University dataset.

PU	PCA	LE	SE	TE
ARI	0.7487	0.6061	0.9267	0.9280
OA	0.8614	0.7719	0.9058	0.9059
AA	0.8404	0.7534	0.8447	0.8439
FS	0.8451	0.7556	0.8436	0.8426
κ	0.8144	0.6962	0.8753	0.8754

Table 5. Classification results for Pavia University scene (PU): assume meadow (class 2) is known

PUG	PCA	LE	SE	TE
ARI	0.7640	0.6274	0.9380	0.9400
OA	0.8762	0.7921	0.9256	0.9264
AA	0.8535	0.7729	0.8762	0.8770
FS	0.8624	0.7766	0.8761	0.8768
κ	0.8319	0.7195	0.9001	0.9013

Table 6. Classification results for Pavia University scene-G (PUG): assume meadow (class 2) is known

The improvement of accuracy comes from not only the known class, but also other classes. The following table contains the accuracy in each class of the Salinas-B. Among all

the 8 classes, TE ranks 1st in four classes and the 3rd in one class.

Class	PCA	LE	DIF	ISO	SE	TE
1	0.9752	0.9776	0.9679	0.9849	0.9791	0.9790
2	0.9984	0.9892	0.9977	0.9935	0.9903	0.9895
8	0.9672	0.9659	0.9634	0.9664	0.9763	0.9831
10	0.9583	0.9573	0.9292	0.9602	0.9990	1.0000
11	0.9576	0.9640	0.9330	0.9569	0.9761	0.9786
12	0.9984	0.9917	0.9876	0.9964	0.9979	0.9994
13	0.9731	0.9558	0.9666	0.9690	0.9547	0.9544
14	0.9235	0.9128	0.9059	0.9327	0.9208	0.9186

Table 7. Classification accuracy for each class in Salinas-B (SB): assume corn (class 10) is known

Similarly, in the grouped dataset, TE is the best in three classes and 3rd in one class among all the five classes.

Class	PCA	LE	DIF	ISO	SE	TE
1	0.9752	0.9776	0.9679	0.9849	0.9791	0.9790
2	0.9984	0.9892	0.9977	0.9934	0.9903	0.9893
8	0.9666	0.9657	0.9641	0.9665	0.9780	0.9843
10	0.9576	0.9566	0.9297	0.9602	0.9990	1.0000
11	0.9869	0.9857	0.9749	0.9867	0.9920	0.9944

Table 8. Classification accuracy for each class in Salinas-B-G (SBG): assume corn (class 10) is known

It should be noted from Table 3 and Table 4 that the performance of the transport model increases significantly after grouping similar classes together. This not only shows that the grouping is correct, but also indicates that our method is compatible with a correct grouping in a dataset with multi-tiered structure like Salinas-B (see Figure 1).

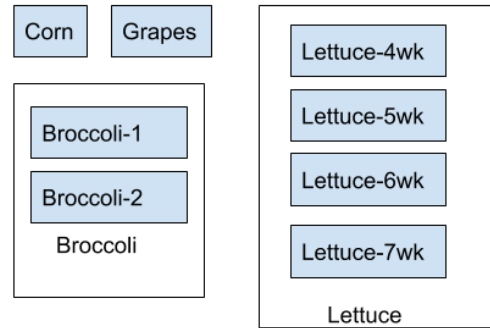


Fig. 1. Sample structure of material classes' multi-scale relations in Salinas-B.

It is natural to investigate the possibility of using the transport model to detect class similarities and multi-tiered data structure, and improve classification accuracy on HSI. For example, we could do classification on grouped classes first, and then classify the subclasses within each class. The rich

parameters in our model will allow us to make effective use of the known information, which cannot be implemented in Laplacian Eigenmaps.

5. CONCLUSIONS

In this paper, we propose and test a new approach to analyze large, complex, and noisy datasets. The proposed method takes advantage of the concept of transport models for discrete data graphs, in a manner that extends and generalizes many previous data organization and dimension reduction algorithms. The main novelty of the technique we propose is the contextual and tiered way in which it treats different material classes. In other words, this methodology allows us to build a multi-scale approach to feature extraction, by means of taking advantage of additional information about groundtruth classes of materials and how they are related to each other. This results in a more effective processing of the complex data structures, such as those present in hyperspectral remote sensing modalities. In presented experiments, we have demonstrated the superiority of this approach, as compared with standard techniques based on LE and diffusion distance-based representations.

The future work on this subject will be dedicated to optimization of the multi-scale transport model algorithm. In this regard, we shall study the suitability of several proposed advection-based transport models TE , TA , and TG , as well as the way in which they can be most effectively combined in the processing pipeline.

6. REFERENCES

- [1] M.T. Eismann, "Hyperspectral remote sensing," *SPIE Bellingham*, 2012.
- [2] J. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1829–1845, 2019.
- [3] K. Pearson, "On lines and planes of closest fit to systems of point in space," *Philosophical Magazine*, vol. 2, no. 11, pp. 559–572, 1901.
- [4] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [5] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [6] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems*, 2002, pp. 585–591.
- [7] W. Czaja and M. Ehler, "Schroedinger eigenmaps for the analysis of biomedical data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1274–1280, 2013.
- [8] N. D. Cahill, W. Czaja, and D. W. Messinger, "Schroedinger eigenmaps with nondiagonal potentials for spatial-spectral clustering of hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX*. International Society for Optics and Photonics, 2014, vol. 9088, p. 908804.
- [9] J. Bartsch, K. Brander, M. Heath, P. Munk, K. Richardson, and E. Svendsen, "Modelling the advection of herring larvae in the north sea," *Nature*, vol. 340, no. 6235, pp. 632, 1989.
- [10] W. Brutsaert and H. Stricker, "An advection-aridity approach to estimate actual regional evapotranspiration," *Water Resources Research*, vol. 15, no. 2, pp. 443–450, 1979.
- [11] H. Guo, J. Zhang, R. Liu, L. Liu, X. Yuan, J. Huang, X. Meng, and J. Pan, "Advection-based sparse data management for visualizing unsteady flow," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2555–2564, Dec 2014.
- [12] J. R. Sibert, J. Hampton, D. A. Fournier, and P. J. Bills, "An advection–diffusion–reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*)," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 56, no. 6, pp. 925–938, 1999.
- [13] H. Song, M. R. Bringer, J. D. Tice, C. J. Gerdt, and R. F. Ismagilov, "Experimental test of scaling of mixing by chaotic advection in droplets moving through microfluidic channels," *Applied Physics Letters*, vol. 83, no. 22, pp. 4664–4666, 2003.
- [14] W. Czaja, D. Dong, P-E Jabin, and F. Njeunje, "Transport model for feature extraction," *arXiv preprint arXiv:1910.14543*, 2019.
- [15] F. Njeunje, *Computational methods in machine learning: transport model, Haar wavelet, DNA classification, and MRI*, Ph.D. thesis, University of Maryland, College Park, 2018.
- [16] "Hyperspectral remote sensing scenes," http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes, Accessed: 2018-04-04.