SPARSE REPRESENTATION-BASED ARCHETYPAL GRAPHS FOR SPECTRAL CLUSTERING

Ribana Roscher, Lukas Drees, Susanne Wenzel

University of Bonn, Institute of Geodesy and Geoinformation

ABSTRACT

We propose sparse representation-based archetypal graphs as input to spectral clustering for anomaly and change detection. The graph consists of vertices defined by data samples and edges which weights are determines by sparse representation. Besides relationships between all data samples, the graph also encodes the relationship to extremal points, so-called archetypes, which leads to an easily interpretable clustering result. We compare our approach to k-means clustering performed on the original feature representation and to k-means clustering performed on the sparse representation activations. Experiments show that our approach is able to deliver accurate and interpretable results for anomaly and change detection.

Index Terms— Sparse representation, spectral clustering, sparse graphs, anomaly detection, change detection

1. INTRODUCTION

The task of finding clusters in data sets has been an active research area for a long time. Especially if the cluster distribution is complex and different clusters cannot be separated linearly in input space, there is a demand for powerful clustering techniques. Therefore, spectral clustering has become a popular clustering approach (e.g., [1, 2]), since it often outperforms conventional cluster algorithms such as k-means or EM-algorithm. For spectral clustering, singular value decomposition of a matrix, representing relationships between data samples, is used to derive the clustering result. One of the key factors, influencing the performance of spectral clustering, is the construction of the relationship-encoding graph form which the matrix is derived. Commonly used approaches build dense graphs comprising pairwise distances between all data samples. More efficient sparse graphs comprise pairwise distances between each sample and its neighbors in a small neighborhood (e.g., [3]). The authors of [4] propose an alternative based on sparse representation, which fulfill the desired graph characteristics of high discriminating power, sparsity and an adaptive neighborhood. They show that so-called L_1-sparse graphs achieve a higher accuracy for clustering and for semi-supervised classification, and its success is also underlined by [5].

Our contribution is the introduction of a sparse representation-based graph as input to spectral clustering, which includes information about relationships between all data samples, but also information about the relation to archetypes (i.e., extremal data samples). Archetypes has been proved to be suitable representatives of data samples in e.g., [6] and [7], which can be efficiently extracted using simplex volume maximization (SiVM, [8][1]). Due to the relation of all samples to archetypes, the approach is able to find a suitable number of clusters and provides a spectral information which can be easily interpreted as anomaly in single images or change in time-series data. In our experiments, we show the suitability of sparse representation-based graphs for spectral clustering by means of plant disease symptom detection in a hyperspectral image and by means of a change detection task. These tasks have already been addressed by approaches using sparse representation, e.g., [9], [10]), but none of them uses sparse representation-based graphs.

2. METHODS

2.1. Spectral Clustering

Spectral clustering is the task of finding clusters using a spectral decomposition of a matrix, interpretable as graph, derived from the data. In order to represent relationships between the samples, a weight matrix $\boldsymbol{W}$ is defined by weights $w_{nm} > 0$ being connections between two samples $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ and $w_{nm} = 0$, otherwise. We further define the normalized graph Laplacian $\boldsymbol{L}_{\text{sym}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}}$ with $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$. The degree matrix is given by $\boldsymbol{D} = \operatorname{diag}\left(\sum_m \boldsymbol{w}_m\right)$, where $\boldsymbol{w}_m$ is the m-th row of $\boldsymbol{W}$. Since spectral decomposition of the Graph Laplacian is intractable for large matrices, we use Nystrm approximation[11] to derive a smaller matrix for decomposition. We use the singular vectors to the smallest singular values as input to k-means clustering to obtain the final clustering result.

2.2. Graph Construction with Sparse Representation

In our approach, we follow the work of [4] by designing a sparse graph in which the estimated sparse coefficients, computed by a nonnegative least squares optimization, are

[1] www-ai.cs.uni-dortmund.de/weblab/code.html

used to characterize relationships between data samples $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \ldots, \boldsymbol{x}_N]$ with N being the number of samples. We define a directed graph $\mathcal{G} = \{\boldsymbol{X}, \boldsymbol{W}\}$ with the samples $\boldsymbol{X}$ being the vertices and the matrix $\boldsymbol{W}$ being the edge weights. The graph is constructed in an unsupervised way using sparse representation, where the activations are interpreted as weights in the graph.

In terms of sparse coding a sample $\boldsymbol{x}_n$ is represented by a weighted linear combination of a few elements taken from a $(M \times (N-1))$-dimensional dictionary $\boldsymbol{T}$, such that $\boldsymbol{x}_n = \boldsymbol{T}\boldsymbol{\alpha}_n + \boldsymbol{\gamma}$ with $\|\boldsymbol{\gamma}\|_2$ being the reconstruction error. The dictionary $\boldsymbol{T} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}, \boldsymbol{x}_{n+1} \ldots, \boldsymbol{x}_N]$ is embodied by all other samples except $\boldsymbol{x}_n$. The coefficient vector, comprising the activations, is given by $\boldsymbol{\alpha}_n$, from which we can derive $\boldsymbol{w}_m = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{n-1}, 0, \boldsymbol{\alpha}_{n+1}, \ldots, \boldsymbol{\alpha}_N]$, being the rows of $\boldsymbol{W}$. The optimization problem for the determination of optimal $\hat{\boldsymbol{\alpha}}_n$ is given by

$$\hat{\boldsymbol{\alpha}}_n = \operatorname{argmin} \|\boldsymbol{T}\boldsymbol{\alpha}_n - \boldsymbol{x}_n\|_2, \quad \text{subject to} \quad \boldsymbol{\alpha}_n \succeq \boldsymbol{0} \quad (1)$$

where the first term is the reconstruction error and the second term is the non-negativity constraint. Since non-negativity alone leads to a sufficient sparsity, we do not introduce a further sparsity enforcing prior.

Let us call the weight matrix comprising the relationships between all samples $\boldsymbol{W}_X$. For spectral clustering, we construct the weight matrix $\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_X & \boldsymbol{W}_A \end{bmatrix}$ by extending $\boldsymbol{W}_X$ by an additional weight matrix $\boldsymbol{W}_A$ based on sparse representation on archetypal dictionaries [6]. The weight matrix $\boldsymbol{W}_A$ includes the independently computed result of (1) of all samples with archetypes serving as dictionary. Therefore the relationship between all samples is encoded, but also the relationship to archetypes leading to an easily interpretable clustering result. More specifically, the cluster assignments of the archetypes can be related to all samples in the same cluster and thus, each cluster can be interpreted by an expert by means of its assigned archetype.

3. EXPERIMENTAL SETUP AND RESULTS

3.1. Datasets

3.1.1. Hyperspectral Sugar Beet Plant

We use a hyperspectral image of a plant, *cv.* Pauletta (KWS, Einbeck, Germany), which was cultivated for 8 weeks in a controlled environment in a greenhouse. The plant was inoculated with *Cercospora beticola*, the causal agent of *Cercospora* leaf spot. We use the hyperspectral pushbroom sensor unit VISNIR-camera ImSpector V10E (Specim, Oulu, Finland) with 1600 pixel, observing a spectral signature with 211 spectral bands in the range of $400 - 1000$ nm. For evaluation purposes, we manually annotated disease symptoms. Due to the error-prone labeling of the exact area of the symptoms, we decided to exclude this effect from the analysis by relying only on the symptom centers which are labeled more robust. This dataset is used to evaluate our proposed approach for anomaly detection, where disease symptoms are defined as anomalies.

3.1.2. The Bastrop Complex Wildfire

This dataset comprises 4 satellite images of size 1534×808 pixel of different sensors acquired over the Bastrop County, Texas (USA). In September 2011, most of the area has been destroyed by wildfire. One image acquired by Landsat 5 TM sensor depicts the area before the event and three images acquired by Landsat 5 TM, Advanced Land Imager (ALI) from the EO-1 mission, and Landsat 8 Operational Land Imager show the area after the event. The data was collected by NASA Land Processes Distributed Active Archive Center Program and ground truth is available indicating changed areas [12]. We use this dataset to evaluate our proposed approach for change detection. Here, it is common to take the differences of images as input features for clustering. Instead, for our experiments, we stack all images and use this as input to our clustering approach, because we found that spectral clustering with sparse representation-based graphs is more stable for high dimensional feature spaces. However, for indicating change only Landsat 5 TM images can be directly compared.

3.2. Experimental Setup

For our experiments, we construct an archetypal sparse graph as explained in Section 2. In order to be efficient in deriving $\boldsymbol{W}_X$, we restrict the dictionary for sparse representation to the 3000 spatially nearest neighbors. The extraction of archetypes follows the idea of [13] by using multiple initial points in order to ensure a complete final set. The final number of clusters is fixed to the number of extracted archetypes for all tested algorithms. We compare our approach, using spectral clustering on a sparse representation-based archetypal graph (`SCASR`), with k-means in original data representation (`k-meansO`), and k-means on coefficients obtained from sparse representation with archetypal dictionaries (`k-meansASR`).

For our approach, we define the archetypes as indicator for change. Please note, that an archetype represents a pixel in the stacked image at four points in time. An archetype is supposed to indicate change, if its RMS difference between the two Landsat 5 TM images, before and after the event, is higher than a threshold. The threshold is given as the 0.9 quantile value obtained from all archetypes' Landsat 5 TM RMS differences. For `k-meansO` we use the obtained cluster centers in the same way, as indicators for change.

3.3. Results

Fig. 2 shows the obtained results for the hyperspectral plant dataset. By visual inspection, it can be seen that our approach `SCASR` is able to find different clusters for healthy plant parts, specular reflections, leaf veins and disease symptoms. More

precisely, it is able to cluster two types of disease symptoms (blue and red areas in Fig. 2 (d) and (h)), which are assigned to two disease archetypes, illustrated in Fig. 1. However, also the leaf borders are assigned to disease symptom, resulting from erroneous measurements similar to the spectra of disease symptoms. In contrast to this, `k-meansO` is not able

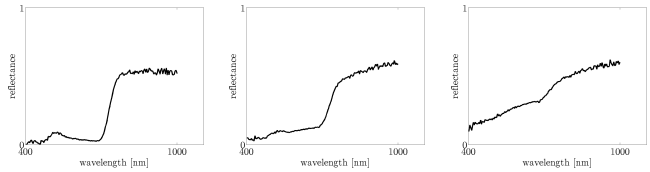

Fig. 1. Left: Spectrum of healthy plant; Middle and right: Two spectra of disease symptoms, which were identified by archetypal analysis.

to cluster disease symptoms. The approach `k-meansASR` detects disease symptoms, however it assigns only one cluster to all of them. Moreover, it gives less smooth results and noisy detections for some disease symptoms in comparison to `SCASR`. After a visual inspection of the archetypes and a manual assignment to specific archetypes to disease symptoms, a comparison with our annotation yield 163 out of 173 (94.2%) detected disease symptoms for `SCASR` and 137 out of 173 (79.2%) for `k-meansASR`.

Fig. 3 shows the result for the Bastrop fire dataset. By visual inspection, all three approaches are able to distinguish most of the areas affected and not affected by fire by assigning different cluster sets to both areas. That means, that the cluster centers detected by k-means as well as the obtained archetypes are able to describe clusters for change. However, a comparison of change gives significant differences. Although `k-meansO` delivers multiple clusters which could be assigned to change, some parts are missing and are thus, assigned to clusters defined by no change. For `SCASR`, the region affected by fire is clearly visible and less effected by noise. A qualitative comparison to the ground truth yielded a kappa coefficient of 0.61 for `k-meansO`, 0.49 for `k-meansASR` and 0.80 for `SCASR`, which underlines that our approach is suitable for change detection with limited manual input. We tested various quantile values to decide on indicators for change, but for `k-meansO` as well as `k-meansASR` a decrease of the quantile value led to an increase of falsely detected change.

4. CONCLUSION

We propose a novel design for a directed graph as input to spectral clustering. The graph is designed using sparse representation, which includes information about relationships between all data samples, but also information about the relation to archetypes. Our experiments confirm that the clustering result yield discriminative and interpretable clusters for the task of change detection and anomaly detection. Our approach can also be used for multi-sensoral datasets, since the archetypes are interpretable by experts.

Acknowledgements

The authors would like to thank Jan Behmann and Anne-Katrin Mahlein for providing the hyperspectral image data and helpful conversations. The Bastrop data is available from the U.S. Geological Survey (http://earthexplorer.usgs.gov/).

References

[1] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Adv. Neural. Inf. Process. Syst.*, vol. 2, pp. 849–856, 2002.

[2] F. Hu, G.S. Xia, X. Wang, Z.and Huang, L. Zhang, and H. Sun, "Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 8, no. 5, 2015.

[3] U. Von Luxburg, "A tutorial on spectral clustering," *Stat. and Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[4] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[5] S. Yan and H. Wang, "Semi-supervised learning by sparse representation.," in *SDM*. SIAM, 2009, pp. 792–801.

[6] R. Roscher, C. Römer, B. Waske, and L. Plümer, "Landcover classification with self-taught learning on archetypal dictionaries," in *IEEE IGARSS*, July 2015, pp. 2358–2361.

[7] C. Römer, M. Wahabzada, A. Ballvora, F. Pinto, M. Rossini, C. Panigada, J. Behmann, J. Léon, C. Thurau, and C. Bauckhage, "Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis," *Funct. Plant Biol.*, vol. 39, no. 11, pp. 878–890, 2012.

[8] C. Thurau, K. Kersting, and C. Bauckhage, "Yes we can: simplex volume maximization for descriptive web-scale matrix factorization," in *Proc. CIKM*. ACM, 2010, pp. 1785–1788.

[9] R. Roscher, J. Behmann, A.-K. Mahlein, J. Dupuis, H. Kuhlmann, and L. Plümer, "Detection of disease symptoms on hyperspectral 3D plant models," in *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, pp. 89–96.

[10] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," *J. Signal. Process. Syst.*, vol. 79, no. 2, pp. 179–188, 2015.

[11] Anna Choromanska, Tony Jebara, Hyungtae Kim, Mahesh Mohan, and Claire Monteleoni, "Fast spectral clustering via the nyström method," in *ALT*. Springer, 2013, pp. 367–381.

[12] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 50–63, 2015.

[13] R. Roscher, S. Wenzel, and B. Waske, "Discriminative archetypal self-taught learning for multispectral landcover classification," in *Proc. PRRS, Workshop at ICPR*, 2016.

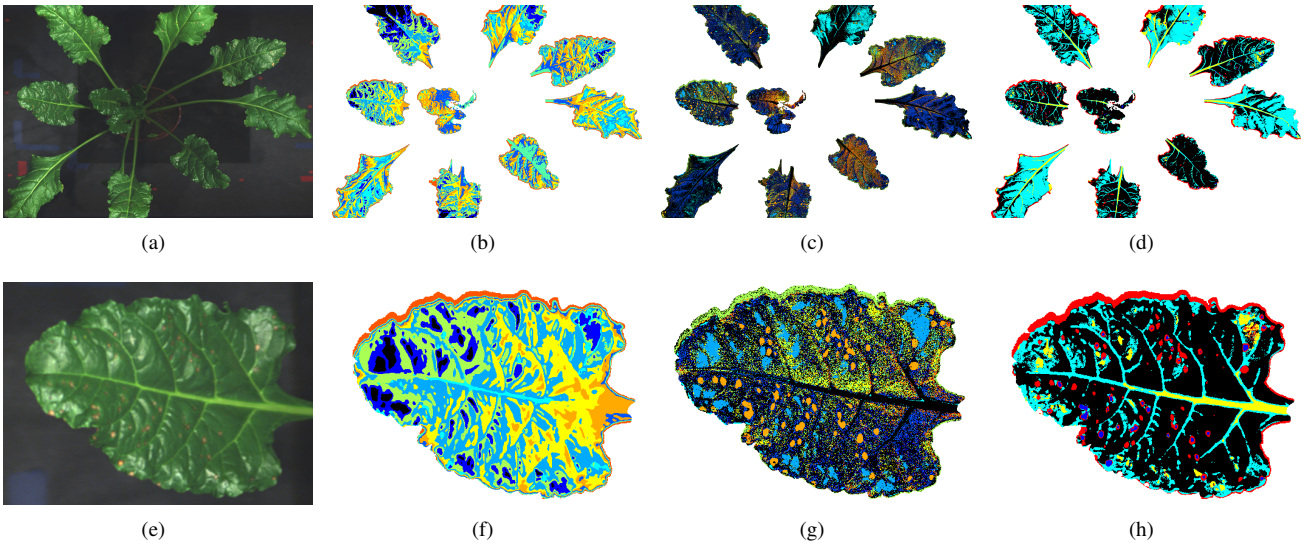

Fig. 2. Results obtained for hyperspectral plant dataset. (a) Image data, (e) Image data, detail, (b) and (f) Clustering result with k-means on original representation, (c) and (g) Clustering result with k-means on sparse representation coefficients, (d) and (h) Spectral clustering result on a sparse representation-based archetypal graph. Colors are arbitrary and indicate assignment to clusters. Therefore, they are not related between the images.

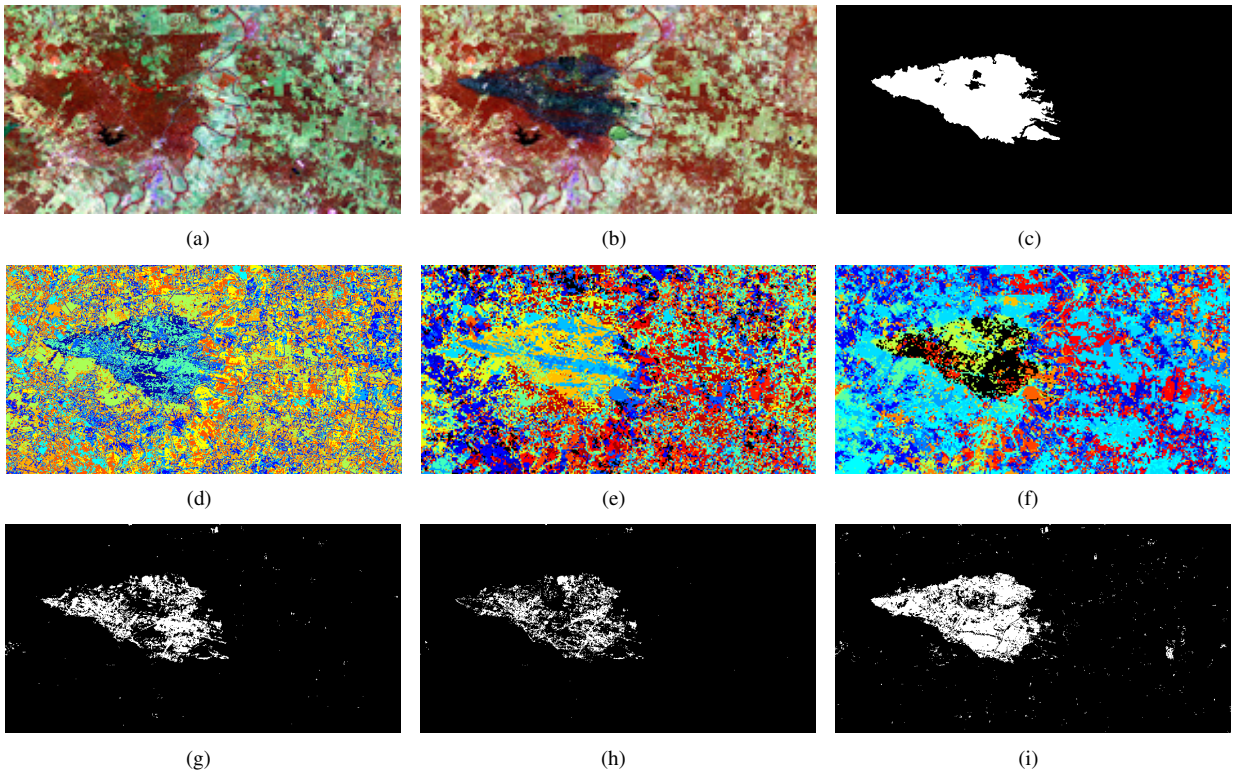

Fig. 3. Results obtained for the Bastrop fire dataset. (a) Image data pre-event, (b) Image data post-event, (c) Ground-truth information for change, (d) Clustering result with k-means on original representation, (e) Clustering result with k-means on sparse representation coefficients, (f) Spectral clustering result on a sparse representation-based archetypal graph, (g) Detected change using k-means on original representation, (h) Detected change using k-means on sparse representation coefficients, (i) Detected change using spectral clustering result.