On the Exploitation of Heterophily in Graph-Based Multimodal Remote Sensing Data Analysis

Catherine Taelman[1,2], Saloua Chlaily[2], Eduard Khachatrian[2], Fons van der Sommen[1] and Andrea Marinoni[2,3]

[1]Dept. of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands
[2]Dept. of Physics and Technology, UiT the Arctic University of Norway, P.O. box 6050 Langnes, NO-9037, Tromsø, Norway
[3]Dept. of Engineering, University of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK

Abstract

The field of Earth observation is dealing with increasingly large, multimodal data sets. An important processing step consists of providing these data sets with labels. However, standard label propagation algorithms cannot be applied to multimodal remote sensing data for two reasons. First, multimodal data is heterogeneous while classic label propagation algorithms assume a homogeneous network. Second, real-world data can show both homophily ('birds of a feather flock together') and heterophily ('opposites attract') during propagation, while standard algorithms only consider homophily. Both shortcomings are addressed in this work and the result is a graph-based label propagation algorithm for multimodal data that includes homophily and/or heterophily. Furthermore, the method is also able to transfer information between uni- and multimodal data. Experiments on the remote sensing data set of Houston, which contains a LiDAR and a hyperspectral image, show that our approach ties state-of-the-art methods for classification with an OA of 91.4%, while being more flexible and not constrained to a specific data set or a specific combination of modalities.

Keywords

label propagation, heterophily, multimodal data, remote sensing

1. Introduction

Nowadays, our planet is monitored by a wide variety of sensors that provide different information about Earth's surface [1] [2]. A multimodal data set is obtained by combining the information of different sensors. When this complementary information is put together, more complex areas on Earth can be analysed correctly. For example, LiDAR data contains height information, whereas hyperspectral data provides information on the physical-chemical composition of the elements and materials on Earth's surface. LiDAR data alone cannot differentiate objects that are at the same height, but when adding hyperspectral data it becomes possible to distinguish two objects at the same height but composed of a different material [3] [4].

An important processing step consists of providing these multimodal data sets with labels [5]. Manual labelling demands for expert knowledge and is very expensive and time-consuming [6]. This brings up the need for label propagation (LP) methods for multimodal data [5]. Graph-based methods are preferred, since they can easily handle the heterogeneity of multimodal data [7]. In graph-based LP methods, the data points form the nodes of a graph. The edges constitute the pairwise distances between pairs of data points, with a missing edge corresponding to infinite distance. A quadratic cost criterion is subsequently derived from the graph structure and optimized in order to infer the missing labels [8].

In remote sensing, the images are typically converted into a graph by taking each pixel as a node and edges are assigned based on the similarity between pixels [9]. A commonly made assumption is that the graph shows the property of homophily, meaning nodes sharing the same label tend to be linked together or have a higher edge weight and therefore influence each other's beliefs during propagation [10]. However, in many real-life data sets nodes with different labels also interact with each other. This property is known as heterophily. Although both data interaction types are frequently encountered in real life, heterophily is rarely included in LP algorithms [11]. The classic LP methods additionally assume that all nodes in the graph are of the same type, thus creating a homogeneous graph. However, many real-life data sets are heterogeneous networks with multiple node and/or edge types [10]. In the multimodal remote sensing context, each different modality or combinations thereof results in a different node type. Although LP on homogeneous graphs has been extensively studied, the extension to heterogeneous graphs is still at its early stages [12]. LP on a heterogeneous graph that includes homophily and/or heterophily during propagation has only been solved the-

CDCEO 2021: 1st Workshop on Complex Data Challenges in Earth Observation, November 1, 2021, Virtual Event, QLD, Australia.
cta014@uit.no (C. Taelman); saloua.chlaily@uit.no (S. Chlaily); eduard.khachatrian@uit.no (E. Khachatrian); fvdsommen@tue.nl (F. v. d. Sommen); andrea.marinoni@uit.no (A. Marinoni)
0000-0002-3593-2356 (F. v. d. Sommen); 0000-0001-6789-0915 (A. Marinoni)

oretically in [11] [13] [14]. None of these studies has performed any extensive tests on large, heterogeneous real-world data sets.

In Earth observation it can be challenging to find images that overlap both in space and time in order to construct multimodal data sets. In addition, certain imaging modalities are not available on large scale, due to limitations of the imaging sensors [15]. This often results in multimodal data sets that overlap only partially, which raises the need for methods that are able to transfer discriminative information across modalities in order to improve the classification of regions on Earth imaged by a single sensor [15].

In this work we aim to develop an LP algorithm for multimodal remote sensing data. We focus on two questions:
1) How can the algorithm handle unimodal and multimodal data simultaneously and consequently also perform cross-modal information transfer?
2) How can homophily and heterophily interactions be estimated and included during LP?

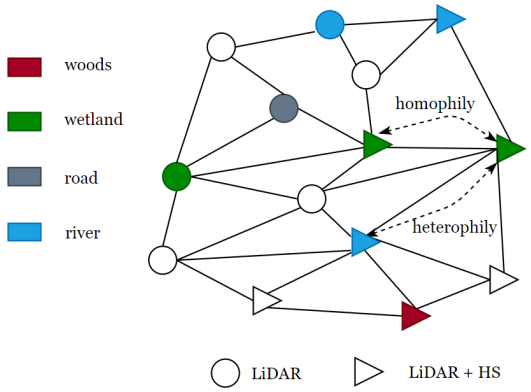

Figure 1: Example of the LP problem on multimodal data (hyperspectral (HS) + LiDAR). Solid lines indicate edges in the graph, while dashed lines show the concept of homophily and heterophily.

Illustrative example Figure 1 sketches a simple example of the LP problem on multimodal remote sensing data with two different modalities: hyperspectral (HS) and LiDAR. The HS and LiDAR images are partly overlapping, which results in a heterogeneous graph with two node types: type 1 for the pixels with only HS data ('HS'), type 2 for the pixels in the overlapping region, with both HS and LiDAR data. In Figure 1 each node type is represented by a different shape, while colours are used to indicate the different classes a node can belong to. Before LP, part of the nodes are unlabeled (white). The goal is to assign the unlabeled nodes to one of the classes.

This paper is organized as follows. Section 2 provides a brief overview of the related works. Section 3 introduces the proposed method and Section 4 reports experimental results. Section 5 discusses those results and Section 6 compares to state-of-the-art methods. Conclusions are drawn in Section 7.

Throughout this paper, vectors ($\mathbf{x}$) and matrices ($\mathbf{X}$) are denoted by lowercase and uppercase bold letters, respectively. X_{ij} refers to the element in the i-th column and j-th row of matrix $\mathbf{X}$.

2. Related Works

In recent years, multiple attempts have been made to generalize standard LP algorithms to heterogeneous networks and/or to the heterophily property. Most research has been focussing on allowing heterophily during propagation. LinBP [14], CAMLP [16], and Relational Networks (RN) [17] include heterophily during LP, but cannot handle heterogeneous graphs. Heterophily is included using a propagation matrix $\mathbf{H}$ that defines how strongly different classes influence each other during propagation. The higher a value H(i,j), the larger the chance that a node of class i will be linked to a node of class j. A diagonal $\mathbf{H}$ reflects homophily, while a non-diagonal $\mathbf{H}$ reflects heterophily. Only three algorithms [13], [18], and [11] have attempted to generalize LP to heterogeneous graphs while simultaneously including propagation via heterophily. All three methods use multiple propagation matrices $\mathbf{H}$ to handle heterogeneous networks. ZooBP [11] appears to be the more general than [13] and [18], since it works on a fully heterogeneous graph (multiple node and edge types) and allows different node types to have a different amount of classes. The main drawback of ZooBP is that the propagation matrices $\mathbf{H}$ are not estimated during optimization but are required as input.

3. Proposed Method

The following section describes our method for LP on multimodal remote sensing data. We first describe how the graph is constructed from the multimodal images, taking into account both the spectral and spatial information. Next, the propagation rules of ZooBP [11] are introduced, since they are used as a backbone in the new method. Last, we show how our method automatically estimates the propagation matrix $\mathbf{H}$ from any sparsely-labeled graph, be it homogeneous or heterogeneous. This is a major improvement compared to previous graph-based LP methods such as ZooBP that require $\mathbf{H}$ as input.

3.1. Graph Generation

Let us define two remotely-sensed images $\mathbf{X}$ and $\mathbf{Y}$, each of a different imaging modality. $\mathbf{X}$ has dimension $N_x \times D_x$ and $\mathbf{Y}$ has dimension $N_y \times D_y$, where N_x and N_y are the number of pixels and D_x and D_y are the spectral dimension of modality X and Y, respectively.

When representing data as a graph, it is crucial to select an appropriate similarity measure that can grasp the underlying data structure [5]. Remote sensing data, and especially hyperspectral images (HSI), contain non-linearities [19]. Furthermore, multi- or hyperspectral remote sensing data contain useful complementary information in both the spatial and spectral domain. Shi and Malik [20] introduce a graph construction technique that handles nonlinearity and simultaneously takes into account the spectral-spatial similarity between pixels (nodes). For each pair of pixels i and j, the similarity is calculated as follows:

$$w_{ij} = e^{\frac{-\|F(i)-F(j)\|_2^2}{\sigma_{spe}}} * \begin{cases} e^{\frac{-\|C(i)-C(j)\|_2^2}{\sigma_{spa}}} & \text{if } \|C(i)-C(j)\|_2 < r \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The left-hand side is the spectral proximity term, where $F(i)$ is the feature vector of the i-th pixel and σ_{spe} the width of the spectral kernel. The right-hand side is the spatial proximity, with $C(i)$ the spatial coordinates of the i-th pixel and σ_{spa} the width of the spatial kernel. Note that two nodes i and j will only be connected if they are less than r pixels apart. The spatial term prevents the formation of large disjoint communities in the graph, regardless of any spectral similarity [19]. We define two cases when building the graph.

Case 1 - Homogeneous Graph In this case we assume that the images $\mathbf{X}$ and $\mathbf{Y}$ fully overlap. For each pixel the different modalities are stacked together ($D_{new} = D_x + D_y$), resulting in a homogeneous graph with N nodes.

Case 2 - Heterogeneous Graph In reality, however, multimodal remote sensing images often do not cover exactly the same region on Earth. When the images $\mathbf{X}$ and $\mathbf{Y}$ are only partly overlapping, some pixels carry information from both modalities, while other pixels carry information from only one modality. The result is a heterogeneous graph with a different node type for each modality or for combinations thereof.

3.2. LP via ZooBP update rule

The new LP method builds on the propagation rules of ZooBP [11], since they are flexible in terms of heterogeneity and heterophily. ZooBP [11] defines the following update equation to calculate all nodes' beliefs in a single step:

$$\mathbf{b} \leftarrow \mathbf{e} + (\mathbf{P} - \mathbf{Q})\mathbf{b} \tag{2}$$

with $\mathbf{e}$ the prior belief vector and $\mathbf{b}$ the final belief vector, obtained as follows: $\mathbf{e} = \begin{bmatrix} \text{vec}(\mathbf{E}_1)^T & \dots & \text{vec}(\mathbf{E}_S)^T \end{bmatrix}^T$ and $\mathbf{b} = \begin{bmatrix} \text{vec}(\mathbf{B}_1)^T & \dots & \text{vec}(\mathbf{B}_S)^T \end{bmatrix}^T$. S denotes the set of node types occuring in the graph and s a single node type. $\mathbf{B}_S$ and $\mathbf{E}_S$ are the prior and final (resp.) belief matrix for type-s nodes.

The persona-influence matrix $\mathbf{P}$ is constructed as follows: $\mathbf{P} = \{P_{ij}\}_{(i,j)\in\{1,\dots,S\}^2}$; $\mathbf{P}_{ss'} = \sum_{t\in T_{ss'}} \frac{\epsilon_t}{k_s} (\mathbf{H}_t \otimes \mathbf{W}_t)$ with $T_{ss'}$ the set of edge types between nodes of type s and s', k_s the number of classes a node of type s can belong to, $\mathbf{H}_t$ the propagation matrix for edge type t, $\mathbf{W}_t$ the weighted similarity matrix for edge type t, and $\otimes$ stands for the Kronecker product. The interaction strength parameter ϵ_t allows to give more weight to particular edge types.

The echo-cancellation matrix $\mathbf{Q}$ is defined as: $\mathbf{Q} = \bigoplus_{i=1}^{S} \mathbf{Q}_s$, where $\mathbf{Q}_s = \sum_{s'\in S} \sum_{t\in T_{ss'}} \frac{\epsilon_t^2}{k_s k_{s'}} \left(\mathbf{H}_t \mathbf{H}_t^T \otimes \mathbf{D}_{st}\right)$ and $\bigoplus$ stands for the direct sum. $\mathbf{D}_{st}$ is the type-t degree matrix of s-type nodes [11].

3.3. Estimation of propagation matrix H

Most previous works define $\mathbf{H}$ as an identity matrix, which reflects strong homophily [21]. The works that include heterophily either make use of domain experts to assign the values of $\mathbf{H}$, as is the case for ZooBP [11], or use rarely-justified heuristics [21]. To overcome this issue, Kumar et al. [21] developed a method that estimates the propagation matrix $\mathbf{H}$ from a sparsely-labeled graph. However, it is only applicable to a homogeneous graph. In this section, we show how their method can be extended to the more general case of a heterogeneous graph.

Kumar et al. derive the energy function that belongs to the update equation of the label propagation method LinBP [14]. The derived energy function for LinBP is the following: $E(\mathbf{B}) = \|\mathbf{B} - \mathbf{E} - \mathbf{WBH}\|_F^2$, where $\mathbf{B}$, $\mathbf{E}$, $\mathbf{W}$ and $\mathbf{H}$ represent the same variables as described in Section 2 but used here for only a single node type. Since all variables are matrices, the Frobenius norm $\|.\|_F$ is used. Next, the authors argue that for an unlabeled node i, the algorithm infers the final label by taking the average of its neighbours: $\mathbf{B_i} = (\mathbf{WBH})_\mathbf{i}$, since there is no prior belief for this node. Based on the idea that the beliefs of a node are the average of its neighbours, the prior beliefs can be used to estimate $\mathbf{H}$ by minimizing the following energy function: $E(\mathbf{H}) = \|\mathbf{E} - \mathbf{WEH}\|_F^2$.

Since ZooBP and LinBP use a similar update scheme, $\mathbf{H}$ for ZooBP can be estimated in an analogous way [14]. The main challenge is to adapt the energy function to a heterogeneous graph, where a different matrix $\mathbf{H}$ needs to be estimated for each edge type t. When dropping the echo-cancellation term $\mathbf{Q}$, as is done in [21], the update

equation of ZooBP for a node of type s becomes:

$$\mathbf{B}_s = \mathbf{E}_s + \sum_{s' \in S} \sum_{t \in T_{ss'}} \frac{\epsilon_t}{k_s} \mathbf{W}_t \mathbf{B}_{s'} \mathbf{H}_t^T \tag{3}$$

Following the same strategy as in [21], while also taking into account that the ZooBP theory requires $\mathbf{H}$ to be residual, results in the following constrained optimization problem:

$$\begin{aligned} \underset{H}{\text{minimize}} \quad & E(\mathbf{H_{tot}}) = \|\mathbf{E_{tot}} - \mathbf{W_{tot}}\mathbf{E_{tot}}\mathbf{H_{tot}}\|_F^2 \\ \text{subject to} \quad & : \textstyle\sum_i \mathbf{H_t}(i,j) = \sum_j \mathbf{H_t}(i,j) = 0 \quad \forall t,i,j \end{aligned} \tag{4}$$

Where $\mathbf{H_{tot}}$, $\mathbf{E_{tot}}$, and $\mathbf{W_{tot}}$ are global matrices defined as: $\mathbf{E_{tot}} = diag(\mathbf{E_1}, ..., \mathbf{E_s})$, $\mathbf{H_{tot}} = (\mathbf{H_t})_{t \in T_{ss'}} = (\mathbf{H_{ij}})_{1 \le i,j \le S}$ and $\mathbf{W_{tot}} = (\mathbf{W_t})_{t \in T_{ss'}} = (\mathbf{W_{ij}})_{1 \le i,j \le S}$. The constraint enforces $\mathbf{H}$ to be residual, which means that all rows and columns sum to 0 [11].

4. Experimental Results

In this section we assess the general performance of the new LP algorithm. The classification is quantitatively evaluated using the overall accuracy (OA, in %) and the Kappa coefficient (κ, in range [-1,1]) [22]. Every time before starting LP, we randomly seed 20% of the nodes to their ground truth label. Due to computational limitations, during the experiments the test set is split into 6 folds in a stratified fashion. Each fold is classified using a separate graph and classification of the full test set is obtained by combining the results of the individual folds. The label propagation process is executed 5 times for each fold and we report the mean value for both evaluation metrics. The proposed method has four hyperparameters: ϵ, σ_{spe}, σ_{spa}, and r. Optimal values are set via a random search using 3-fold cross-validation on the following sets: ϵ a real number in the range [0.001, 0.6], σ_{spe} in the set {0.001, 0.01, 0.1, 1, 10}, σ_{spa} in the set {0.001, 0.01, 0.1, 1, 10} and r an integer in the range [0, 20].

The method is evaluated on the publicly available multimodal remote sensing data set of Houston [23] to verify the effectiveness of the proposed approach. The data set contains a HSI and a LiDAR-derived DSM image of an urban area in Houston, the USA. The images have a spatial size of 349 x 1905 pixels with a resolution of 2.5 m. The HSI consists of 144 spectral bands, ranging from 380 to 1050 nm. The data set contains 15 classes of interest and is visualised in Figure 2. The training and test set contain 2832 and 12197 samples, respectively.

The optimal values for the hyperparameters are found to be: ϵ = 0.317, σ_{spe} = 1, σ_{spa} = 10 and r = 18. The proposed method produces a smooth classification map

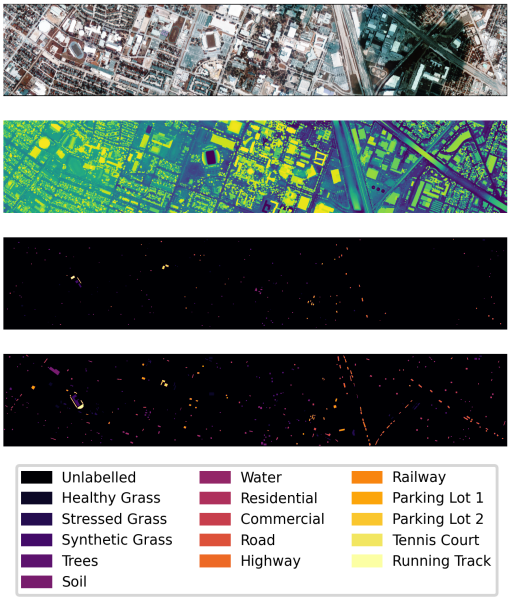

Figure 2: Houston data set [23] - From top to bottom: false-colour composite HSI (bands 10, 20, 40), LiDAR-derived DSM image, training set and test set with corresponding labels.

with low salt-and-pepper noise and achieves an OA of 91.4%. Misclassification happens at the borders of the labeled regions, more specifically where spatially adjacent pixels belong to different classes. E.g. the classes 'running track' and 'synthetic grass' or 'stressed grass' and 'healthy grass' often get confused. These are pairs of classes with a similar spectral signature. The combination of a similar spectral signature and close spatial adjacency challenges the algorithm in rightly classifying some border pixels. A crop of the classification map is provided in Figure 3 and illustrates the misclassification of 'synthetic grass' (purple) and 'running track'(white).

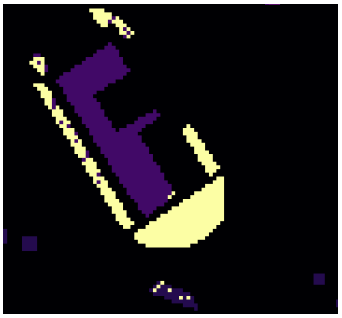

Figure 3: Crop of the classification map of Houston.

5. Discussion

The results demonstrate that the proposed method achieves good classification performance. Misclassification mainly happens for spatially near pixels that have a similar spectral signature. In that case the hyperspectral data is not able to differentiate between two classes and that is when the LiDAR data should help to make the right decision. However, since the two modalities are fused on data-level, the LiDAR data is nearly insignificant in the feature vector that is mainly composed of hyperspectral features. An adaptive feature selection method should be considered that decides for each pixel or data set which modalities should weigh more than others in the decision-making process.

The large spatial size of remote sensing images causes scalability issues in our method, since the similarity matrix **W** is computed by comparing each pair of data points. Note, however, that this scalability issue is only related to size of the input images, and not to any intrinsic methodological problem. The number of iterations until reaching convergence is constant for increasing graph sizes, which proves that the propagation rules scale well for large graphs.

6. Comparison to state-of-the-art methods

The methods listed in Related Works are not able to solve the general problem (full heterogeneity and heterophily) and are therefore left out of comparison. Instead, we focus on the application side and provide a comparison of the proposed method with five state-of-the-art classification methods: a support vector machine (SVM) with radial basis function (RBF) kernel [24], a Random Forest (RF) [25], '3-stream CNN' [26], 'patch-to-patch CNN' (PtoP CNN) [27], and 'sparse and low-rank component analysis' (SLRCA) [4]. A seed size of 30% is used here in our method to provide fair comparison. Also note that 3-stream CNN, PToP CNN and SLRCA classify the whole data set at once, which is advantageous for the classification performance. Our method largely outperforms the SVM and RF and (practically) ties the other three classification methods. Moreover, the proposed method can easily accommodate other types of multimodal data, as well as supporting a large variety of complex graph structures, which are two major advantages compared to [26], [27], and [4] which are tailored to specific remote sensing modalities and/or data sets.

7. Conclusion

We have designed and evaluated a new label propagation method for multimodal remote sensing data that contributes to:

- *Generality:* The method works on a fully heterogeneous graph (multiple node and edge types) and includes homophily and/or heterophily interactions via the propagation matrices **H**.
- *Estimation of propagation strengths:* The propagation matrices **H** are automatically estimated from any homogeneous or heterogeneous sparsely-labeled graph. No need anymore for domain experts to assign these values.
- *Cross-Modal Information Transfer:* The proposed method can propagate information between unimodal and multimodal data points.

The method achieves similar performance as state-of-the-art approaches, while being more general. Future work therefore includes testing the algorithm on multimodal data sets originating from other domains to check the robustness of the new approach. Furthermore, attention will be paid to solving the scalability issue that arises when using graphs with thousands of nodes.

	SVM [24]	RF [25]	3-stream CNN [26]	PtoP CNN [27]	SLRCA [4]	Our method
OA (%)	78.6	76.9	**92.6**	92.5	91.3	91.4
κ	0.77	0.75	**0.92**	**0.92**	0.91	0.91

Table 1
Classification accuracies on Houston obtained by state-of-the-art approaches. Reported: overall accuracy (OA, in %) and Kappa coefficient (no unit).

8. Acknowledgments

This work is funded in part by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Research Council of Norway (RCN Grant no. 237906), and the Automatic Multisensor remote sensing for Sea Ice Characterization (AMUSIC) Framsenteret "Polhavet" flagship project 2020.

References

[1] E. Kairu, An introduction to remote sensing, GeoJournal 6 (1982) 251–260. doi:10.1007/BF00210657.

[2] D. A. Tretyachenko, Y. S. Greenfieldt, O. A. Klimanova, N. N. Alekseeva, M. A. Arshinova, E. Y. Kolbowsky, Use of GIS Technologies in Order to Assess the Degree of Transformation of the Land Cover Structure, IOP Conference Series:

Earth and Environmental Science 107 (2018) 0–4. doi:10.1088/1755-1315/107/1/012107.

[3] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. Van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, F. Pacifici, Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (2014) 2405–2418. doi:10.1109/JSTARS.2014.2305441.

[4] B. Rasti, P. Ghamisi, J. Plaza, A. Plaza, Fusion of Hyperspectral and LiDAR Data Using Sparse and Low-Rank Component Analysis, IEEE Transactions on Geoscience and Remote Sensing 55 (2017) 6354–6365. doi:10.1109/TGRS.2017.2726901.

[5] O. Zoidi, E. Fotiadou, N. Nikolaidis, I. Pitas, Graph-based label propagation in digital media: A review, ACM Computing Surveys 47 (2015). doi:10.1145/2700381.

[6] S. Jia, X. Deng, M. Xu, J. Zhou, X. Jia, Superpixel-level weighted label propagation for hyperspectral image classification, IEEE Transactions on Geoscience and Remote Sensing 58 (2020) 5077–5091. doi:10.1109/TGRS.2020.2972294.

[7] M. Wang, K. Tan, X. Jia, X. Wang, Y. Chen, A deep Siamese network with hybrid convolutional feature extraction module for change detection based on multi-sensor remote sensing images, Remote Sensing 12 (2020). doi:10.3390/rs12020205.

[8] O. Chapelle, B. Schlkopf, A. Zien, Semi-Supervised Learning, The MIT Press, 2006.

[9] L. Wang, S. Hao, Q. Wang, Y. Wang, Semi-supervised classification for hyperspectral imagery based on spatial-spectral Label Propagation, ISPRS Journal of Photogrammetry and Remote Sensing 97 (2014) 123–137. URL: http://dx.doi.org/10.1016/j.isprsjprs.2014.08.016. doi:10.1016/j.isprsjprs.2014.08.016.

[10] W. Gatterbauer, Semi-Supervised Learning with Heterophily, CoRR abs/1412.3 (2014). URL: http://arxiv.org/abs/1412.3100.

[11] D. Eswaran, S. Gunnemann, C. Faloutsos, D. Makhija, M. Kumar, ZooBP: Belief propagation for heterogeneous networks, in: Proceedings of the VLDB Endowment, volume 10, 2016, pp. 625–636. URL: http://www.cs.cmu.edu/. doi:10.14778/3055540.3055554.

[12] E. F. Maleki, N. Ghadiri, M. L. Shahreza, Z. Maleki, Scalable Label Propagation Algorithms for Heterogeneous Networks, bioRxiv (2018) 477463. URL: https://doi.org/10.1101/477463.

[13] D. Deng, F. Bai, Y. Tang, S. Zhou, C. Shahabi, L. zhu, Label Propagation on K-partite Graphs with Heterophily, IEEE Transactions on Knowledge and Data Engineering (2015) 1–1. doi:10.1109/tkde.2019.2937493.

[14] W. Gatterbauer, S. Günnemann, D. Koutra, C. Faloutsos, Linearized and Single-Pass Belief Propagation, Proceedings of the VLDB Endowment 8 (2015) 581–592. doi:10.14778/2735479.2735490.

[15] D. Hong, N. Yokoya, G. S. Xia, J. Chanussot, X. X. Zhu, X-ModalNet: A semi-supervised deep cross-modal network for classification of remote sensing data, ISPRS Journal of Photogrammetry and Remote Sensing 167 (2020) 12–23. doi:10.1016/j.isprsjprs.2020.06.014.

[16] Y. Yamaguchi, C. Faloutsos, H. Kitagawa, CAMLP: Confidence-aware modulated label propagation, Proceedings of the 2016 SIAM International Conference on Data Mining (2016) 513–521. doi:10.1137/1.9781611974348.58.

[17] L. Peel, Graph-based semi-supervised learning for relational networks, Proceedings of the 2017 SIAM International Conference on Data Mining (2017) 435–443. doi:10.1137/1.9781611974973.49.

[18] W. Gatterbauer, The Linearization of Belief Propagation on Pairwise Markov Networks, arXiv preprint arXiv:1502.04956 (2015). URL: http://arxiv.org/abs/1502.04956.

[19] J. R. Stevens, R. G. Resmini, D. W. Messinger, Spectral-Density-Based Graph Construction Techniques for Hyperspectral Image Analysis, IEEE Transactions on Geoscience and Remote Sensing 55 (2017) 5966–5983. doi:10.1109/TGRS.2017.2718547.

[20] J. Shi, J. Malik, Normalized Cuts and Image Segmentation, IEEE transactions on pattern analysis and machine intelligence 22 (2000) 888–905. doi:10.1093/biomet/ass019.

[21] K. P. Kumar, P. Langton, W. Gatterbauer, Factorized Graph Representations for Semi-Supervised Learning from Sparse Data, Proceedings of the ACM SIGMOD International Conference on Management of Data (2020) 1383–1398. doi:10.1145/3318464.3380577.

[22] P. S. Bharatkar, R. Patel, Approach to Accuracy Assessment tor RS Image Classification Techniques, International Journal of Scientific & Engineering Research 4 (2013) 79–86.

[23] 2013 IEEE GRSS Data Fusion Contest (2013).

[24] J. Gualtieri, S. Chettri, Support vector machines for classification of hyperspectral remote-sensing images, International Geoscience and Remote Sensing Symposium (IGARSS) 1 (2000) 506–508. doi:10.1109/igarss.2002.1025088.

[25] M. Pal, Random forest classifier for remote sensing classification, International Journal of Remote Sensing 26 (2005) 217–222. doi:10.1080/01431160412331269698.

[26] H. Li, P. Ghamisi, U. Soergel, X. X. Zhu, Hyperspectral and LiDAR fusion using deep three-stream

convolutional neural networks, Remote Sensing 10 (2018) 1649. doi:10.3390/rs10101649.

[27] M. Zhang, W. Li, Q. Du, L. Gao, B. Zhang, Feature Extraction for Classification of Hyperspectral and LiDAR Data Using Patch-to-Patch CNN, IEEE Transactions on Cybernetics 50 (2020) 100–111. doi:10.1109/TCYB.2018.2864670.