Adaptive Graph Filtering: Multiresolution Classification on Graphs

Siheng Chen

Department of ECE
Center for Bioimage Informatics
Carnegie Mellon University
Pittsburgh, PA, USA
sihengc@andrew.cmu.edu

Aliaksei Sandryhaila

Department of ECE
Carnegie Mellon University
Pittsburgh, PA, USA
asandryh@andrew.cmu.edu

José M. F. Moura

Department of ECE
Carnegie Mellon University
Pittsburgh, PA, USA
moura@ece.cmu.edu

Jelena Kovačević

Department of BME & ECE
Center for Bioimage Informatics
Carnegie Mellon University
Pittsburgh, PA, USA
jelenak@cmu.edu

Abstract—**We present an adaptive graph filtering approach to semi-supervised classification. Adaptive graph filters combine decisions from multiple graph filters using a weighting function that is optimized in a semi-supervised manner. We also demonstrate the multiresolution property of adaptive graph filters by connecting them to the diffusion wavelets. In our experiments, we apply the adaptive graph filters to the classification of online blogs and damage identification in indirect bridge structural health monitoring.**

I. INTRODUCTION

With the growing demand in analysis of signal generated by various networks, signals with complex structures arise that require novel processing techniques. Recently, signal processing on graphs has emerged as a new approach to analyze signals with irregular structure that reside on graphs [1]–[4]. Among many applications of this theory, we consider signal classification, which models each signal as nodes of a representation graph, and their labels as a graph signal [4]. Given a subset of initial labels, unknown labels are estimated by propagating known ones along the edges of the graph using graph filters.

In this paper, we propose an adaptive graph filter that extends the application of graph filters to signal classification. Our technique uses an adaptive weighting algorithm that combines results from different graph filters into a global decision. We show that the proposed technique possesses multiresolution properties similar to diffusion wavelets [5]. We also demonstrate applications of adaptive graph filters to the classification of online blogs and damage indentification in indirect bridge structural health monitoring.

II. BACKGROUND AND PROBLEM FORMULATION

In this section, we introduce the background material necessary for the rest of the paper. We formulate the problem of signal classification and introduce the discrete signal processing on graphs theory that provides the foundation for the proposed adaptive graph filtering technique.

Classification. The purpose of classification is to assign each signal to one of the given classes [6]. Let $\mathcal{X} = \{x^{(i)} \in \mathbb{R}^d\}_{i=1}^N$ be the given dataset with $N = L+U$ elements. The first L elements form the *labeled dataset* $\mathcal{L} = \{x^{(i)} \in \mathcal{X}\}_{i=1}^L$, for which the *ground-truth labels* $\mathcal{Y} = \{y^{(i)} \in \{1, 2, \ldots, C\}\}_{i=1}^L$

Notations		
$\mathcal{X} = \{x^{(i)}\}$	input dataset	$i = 1, \ldots, N$
$\mathcal{L} = \{x^{(i)}\}$	labeled dataset	$i = 1, \ldots, L$
$\mathcal{U} = \{x^{(i)}\}$	unlabeled dataset	$i = L+1, \ldots, N$
$\mathcal{Y} = \{y^{(i)}\}$	ground-truth labels for $\mathcal{L}$	$i = 1, \ldots, L$
$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for $\mathcal{U}$	$i = L+1, \ldots, N$
$q^{(i)}$	ground-truth vector	$i = 1, \ldots, L$
Q	ground-truth matrix	$L \times C$
$\hat{q}^{(i)}$	confidence vector	$i = L+1, \ldots, N$

TABLE I: Notation used in this paper

are given. The other U elements form the *unlabeled dataset* $\mathcal{U} = \{x^{(i)} \in \mathcal{X}\}_{i=L+1}^N$. Hence, $\mathcal{X} = \mathcal{L} \cup \mathcal{U}$.

A classifier system accepts the entire dataset $\mathcal{X}$ and the *ground-truth labels* $\mathcal{Y}$ as inputs, and produces the *estimated labels* $\hat{\mathcal{Y}} = \{\hat{y}^{(i)} \in \{1, 2, \ldots, C\}\}_{i=L+1}^N$ for the unlabeled dataset $\mathcal{U}$ (see Table I).

We formulate the classification problem as a mapping of the input signal to each class with a certain probability. We regard the label as a posterior probability vector $\hat{q} \in \mathbb{R}^C$, where the cth component, $\hat{q}_c$, is the probability that a signal belongs to the cth class. Since $\hat{q}$ expresses a confidence of an assigned label, we name it a *confidence vector*. The confidence vector for a labeled signal $i = 1, 2, \ldots, L$, is the *ground-truth vector*, $q \in \mathbb{R}^C$, with 0s everywhere except 1 in position c indicating membership in class c. The *ground-truth matrix Q* of size $L \times C$ collects all L ground-truth vectors as its rows, that is,

$$Q = \begin{bmatrix} q^{(1)} & q^{(2)} & \ldots & q^{(L)} \end{bmatrix}^T.$$

Discrete Signal Processing on Graphs. Discrete signal processing on graphs is an extension of the traditional signal processing theory to signals with complex structure residing on irregular domains [4]. The dataset is represented by a *graph* $\mathcal{G} = (\mathcal{V}, P)$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the set of *nodes* that represents signals and $P \in \mathbb{C}^{N \times N}$ is an adjacency matrix of the graph called a *graph shift*. We extend the definition of graph signals from [4] from one-dimensional to D-dimensional signal as follows: a *graph signal s* is a map on the set of nodes

$$s : \mathcal{V} \to \mathbb{C}^{N \times D},$$

where D is the dimension of the graph signal at each node.

A linear shift-invariant system, or, a *graph filter*, is defined as

$$H = h(P) = \sum_{k=0}^{K} h_k P^k, \qquad (1)$$

with $h_k \in \mathbb{C}$, $k = 0, 1, \ldots, K$. A graph filter $H \in \mathbb{C}^{N \times N}$ applied to a graph signal $s \in \mathbb{C}^{N \times D}$ produces an output, which is again a graph signal,

$$Hs = h(P)s.$$

Discrete signal processing on graphs then defines other fundamental signal processing concepts including the graph Fourier transform, frequency, spectrum, spectral decomposition, and impulse and frequency responses.

III. ADAPTIVE GRAPH FILTERING

In this section, we introduce adaptive graph filters. We also show their multiresolution properties by establishing a connection to diffusion wavelets.

Graph Filtering as Semi-Supervised Classification. Let $\mathcal{G} = (\mathcal{X}, P)$ be a graph with $\mathcal{X} = \{x^{(i)}\}_{i=1}^{N}$ a given datasest and let $P \in \mathbb{R}^{N \times N}$ be a graph shift defined as

$$P_{i,j} = \frac{\exp\left(\frac{-\rho(x^{(i)}, x^{(j)})}{\sigma}\right)}{\sum_{i=1}^{N} \exp\left(\frac{-\rho(x^{(i)}, x^{(j)})}{\sigma}\right)}, \qquad (2)$$

where ρ is a local distance measurement, such as the ℓ^2 norm or the cosine distance, and σ is a scaling coefficient which scales the local distance. The graph shift defined here is the Hermitian transpose of the *transition matrix* of the graph. P thus has a probabilistic interpretation: constructed as in (2), $P_{i,j}$ gives the probability of transition from the jth node to the ith one in one step [7].

Now we define an *adaptive graph filter*

$$H = h(P) = \sum_{k=1}^{K} h_k P^k. \qquad (3)$$

This adaptive graph filter is defined similarly to (1), except that its coefficients, h_k, depend on the data and are selected adaptively. We assign $h_0 = 0$, since P^0 does not contribute to the propagation of the labels in the classification problem as will be shown later.

Let the graph signal be the confidence matrix of *all* the signals on the graph, called *prior confidence matrix*, that is, the following map: $s : \mathcal{X} \to \mathbb{R}^{N \times C}$, defined as

$$\hat{Q}_{\mathrm{pr}} = \begin{bmatrix} Q \\ \mathbf{0}_{u \times C} \end{bmatrix}.$$

In other words, the first L rows of $\hat{Q}_{\mathrm{pr}}$ are the confidence matrix Q representing the labeled dataset, while the other U rows are all zeros representing the unlabeled dataset. The prior confidence matrix thus starts with the knowledge on the labeled dataset (the ground truth) and without any knowledge on the unlabeled dataset.

By applying an adaptive graph filter (3) to the prior confidence matrix $\hat{Q}_{\mathrm{pr}}$, the *posterior confidence matrix* is

$$\hat{Q}_{\mathrm{ps}} = H\hat{Q}_{\mathrm{pr}}, \qquad (4)$$

where the ith row is the confidence vector $\hat{q}^{(i)}$ for the ith node.

The coefficients h_k of the adaptive filter (3) are then found as follows. Rewrite (4) as

$$\hat{Q}_{\mathrm{ps}} \overset{(a)}{=} \sum_{k=1}^{K} h_k P^k \hat{Q}_{\mathrm{pr}} = \sum_{k=1}^{K} h_k \hat{Q}_k, \qquad (5)$$

where (a) follows from (3) and $\hat{Q}_k$ is the $N \times C$ confidence matrix for the kth graph shift defined as $\hat{Q}_k = P^k \hat{Q}_{\mathrm{pr}}$. For each i, the desired confidence vector $\hat{q}^{(i)}$ (ith row of $\hat{Q}_{\mathrm{ps}}$) is thus a weighted linear combination of corresponding rows from each graph shift $\hat{Q}_k$. The optimal filter coefficients are designed to fit the confidence vectors of the labeled signals from all the graph shifts to the ground truth and to minimize the labeling uncertainty of the unlabeled signals. We introduce the labeling uncertainty measure [8], [9] as

$$\mathrm{M}(\hat{q}) = \mathrm{H}(\hat{q})\left(\chi_{d>T} + \lambda(d)\,\chi_{d \le T}\right), \qquad (6)$$

where χ_I is the indicator function of an interval I, $d = |\hat{q}_{(1)} - \hat{q}_{(2)}|$ with $\hat{q}_{(1)}, \hat{q}_{(2)}$ the first and second largest element in $\hat{q}$, respectively, T is the threshold, and $\lambda(d) = 1 + 5(d/T - 1)^2$ is a penalty function that is large when the first and second largest elements are close. The optimal filter coefficients are then given by the solution to the minimization problem

$$\begin{aligned} \hat{h} = \quad & \arg\min_{h} \{\alpha \sum_{i=1}^{L} \|q^{(i)} - \hat{Q}^{(i)}h\| \\ & + \quad (1-\alpha)\|h - \gamma\|\}, \end{aligned} \qquad (7)$$

where $h = \begin{bmatrix} h_1 & \ldots & h_K \end{bmatrix}^T$ with the constraint $\|h\|_1 = 1$; $\alpha = L/(L + U)$ is the labeling ratio; matrix $\hat{Q}^{(i)} = \begin{bmatrix} \hat{q}_1^{(i)} & \hat{q}_2^{(i)} & \ldots & \hat{q}_K^{(i)} \end{bmatrix}$ is an $C \times K$ confidence matrix of the ith signal that collects the graph shifts' individual confidence vectors; and $\gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \ldots & \gamma_K \end{bmatrix}$ collects the discriminative powers of each graph shift,

$$\gamma_k = \frac{e^{-(\beta/U)\sum_{i=l+1}^{N} \mathrm{M}(\hat{q}_k^{(i)})}}{\sum_{j=1}^{K} e^{-(\beta/U)\sum_{i=l+1}^{N} \mathrm{M}(\hat{q}_j^{(i)})}}.$$

Here, β is the decay coefficient that controls the distribution of the discriminant power from all the graph shifts. The first term in (7) represents the contribution from all labeled signals. The second term in (7) represents the contribution from all unlabeled signals; to obtain it, we fit weights to the confidences from all the graph shifts . We use the labeling ratio to balance these two terms. Since this is a convex optimization problem, it is numerically efficient to solve. After weighting, we compute the global decision as $\hat{y}^{(i)} = \arg\max_c \hat{q}_c^{(i)}$, where $\hat{q}^{(i)} = \hat{Q}^{(i)}\hat{h}$ (see Algorithm 1). For more details, see [9].

Relation to Diffusion Functions. Diffusion functions are the graph-based semi-supervised classifiers [10]. The classifier operates by propagating known labels using the transition matrix a finite number of times. Given a transition matrix W and a prior confidence matrix $\hat{Q}_{\mathrm{pr}}$, the posterior confidence matrix $\hat{Q}_{\mathrm{ps}}$ is obtained by $\hat{Q}_{\mathrm{ps}}^* = \hat{Q}_{\mathrm{pr}}^* W^t$, where $*$ denotes the Hermitian transpose and t is the transition time. If the filter coefficient h is a Kronecker delta impulse at position t, then

Algorithm 1 Adaptive graph filter

Input	$\mathcal{X}$	input dataset
	$\hat{Q}_{\mathrm{pr}}$	prior confidence matrix
Output	$\hat{Q}_{\mathrm{ps}}$	posterior confidence matrix
Parameters	$P_{i,j}$	graph shift
	h_k	filter coefficients
Function	$\mathbf{AGF}(\mathcal{X}, \hat{Q}_{\mathrm{pr}})$	
	$P_{i,j} = (2)$	graph shift construction
	$h = (7)$	filter coefficient optimization
	for $k = 1 : K$ **do**	filtering
	$\quad \hat{Q}_k = P^k \hat{Q}_{\mathrm{pr}}$	
	$\hat{Q}_{\mathrm{ps}} = \sum_{k=1}^K h_k \hat{Q}_k$	weighting
	return $\hat{Q}_{\mathrm{ps}}$	

$h = \delta_t$, the adaptive graph filter coincides with the diffusion function. Hence, diffusion functions form a subset of the adaptive graph filters that have only one non-zero coefficient. In practice, it can be hard to determine the required transition time; as a consequence, the performances of diffusion functions can be susceptible to initial conditions. In contrast, the adaptive graph filter learns the contribution from each transition time by using (7).

Relation to Diffusion Wavelets. Diffusion wavelets are a multiscale framework for the analysis of signals with complex structure [5]. They can be seen as an extension of the classical wavelet theory, where, the diffusion wavelet basis is learned from the geometry of the signal structure in a data-adaptive way. A diffusion wavelet basis is constructed by dilation using the dyadic powers of the transition matrix, the idea being that they propagate local relationships to global relationships throughout the graph.

Given a graph with the transition matrix T, the transition matrix at the jth resolution level is T^{2^j}, $j = 1, 2, \ldots$. Since the second singular value of the transition matrix is less than 1 to keep the graph connected, only the first singular value is 1. Consequently, if the transition matrix is raised to a high power, all the singular values disappear except for the first one,

$$\lambda_i^{2^j} \to 0, \qquad i = 2, 3, \ldots, N \quad \text{as} \quad j \to \infty.$$

When $j = 1$, the transition matrix T measures local pairwise similarities; increasing the power j gradually decreases the rank of the transition matrix T^{2^j} and causes local information of the graph to be missed since the resolution on the graph changes from finest to the coarsest. Thus, by changing j, we can both perform a multiresolution analysis as well as doing it in a computationally efficient manner.

The adaptive graph filter in (3) is formed as a linear combination of graph shifts P raised to power k. When k is large, P^k become a low-rank matrix describing the global information of the graph. Since the graph filter P is defined as the Hermitian transpose of the transition matrix T in (2), P^k works for the adaptive graph filter just as T^{2^j} does for diffusion wavelets. Each k corresponds to a different resolution on the graph, and thus, adaptive graph filter actually performs multiresolution classification on the graph. It weighs the classification results from each resolution to produce the global

result. The filter coefficients represent the discriminative power of each resolution.

IV. EXPERIMENTAL RESULTS

In this section, we apply adaptive graph filtering to classification of online blogs and damage identification in indirect bridge structural health monitoring.

A. Classification of Online Blogs

Dataset. We consider 1224 online political blogs that we wish to classify as conservative or liberal based on their context [11]. The graphs corresponding to each blog are obtained by tracing hyperlink references between blogs.

Experimental Setup. We perform a binary clasification, and vary the labeling ratio as 0.5%, 1%, 2%, 5% and 10%; in other words, we label 6, 12, 24, 60, 120 blogs, respectively. We considered two methods for which nodes to label initially: random selection, and selection of blogs with most hyperlinks. We ensure we have the same number of signals from both classes in the labeling dataset and compare the performances of the diffusion functions (DF) and adaptive graph filters (AGF). We choose the transition time $t = 2$, filter length $K = 10$ and the decay coefficient $\beta = 1$, and perform a 30-fold cross-validation.

Results. Fig. 1 shows the dependence of classification accuracy on the labeling ratio for the two selection mechanisms. Both show similar trends: adaptive graph filter works consistently better than diffusion functions. Another interesting observation is that when labeling ratio is low, initial labeling with most hyperlinks provides better accuracy, but when labeling ratio is sufficiently high, initial labeling with most hyperlinks has no advantage over initial labeling with random selection. The reason is that when labeling ratio is low, nodes with higher degree propagate labels to more nodes, but when labeling ratio is high, a large number of labeled nodes has the same effect.
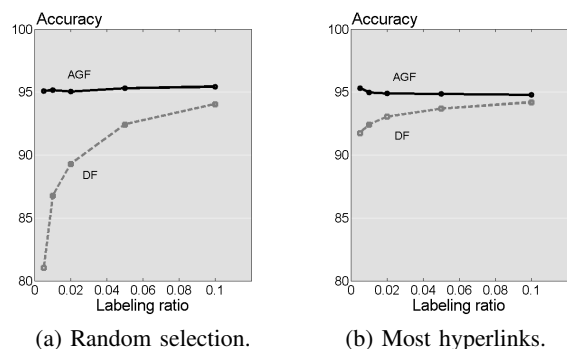

(a) Random selection. (b) Most hyperlinks.

Fig. 1: Accuracy as a function of the labeling ratio.

B. Damage Detection in Bridge Structure Monitoring

Dataset. We built a lab-scale bridge-vehicle dynamic system and put a sensor on a vehicle, and let it move across the bridge. We capture vibration characteristics of the bridge from the vibration of the traversing vehicle through the acceleration signal. We collected 30 acceleration signals for each of 13

V	S	AGF	DF	SVM
1	1	**87.98**	83.70	57.97
	2	**89.55**	85.78	70.42
	3	**89.91**	86.25	74.29
	4	**93.87**	88.08	74.82
	5	**77.64**	74.87	70.68
	6	**85.24**	82.43	67.13
	7	**66.23**	65.31	59.48
	8	**67.11**	66.52	56.53
2	1	**80.99**	78.17	49.75
	2	**70.32**	67.30	53.64
	3	**86.09**	82.50	67.96
	4	**82.86**	79.57	61.52
	5	**81.22**	77.92	62.75
	6	**81.73**	80.17	66.89
	7	**85.48**	82.87	65.09
	8	**82.57**	80.32	48.57

TABLE II: Accuracy comparison of Vehicles (V) 1 and 2, with Speeds (S) 1, 2, …, 8, and labeling ratio of 10%.

(a) Vehicle 1.   (b) Vehicle 2.

Fig. 2: Accuracy as a function of the labeling ratio.

different bridge conditions, 8 different speeds and 2 different vehicles, for a total of 6240 acceleration signals [8], [12].

Experimental Setup. Given a specific vehicle driven at a specific speed, we want to classify 13 bridge conditions, in particular with a low labeling ratio. We have 16 vehicle-speed cases for each of which there are 30 acceleration signals per 13 bridge conditions, and vary the labeling ratio as 10%, 30%, 50%, 70% and 90%; the final accuracy is the average over the 13 scenarios.

We ensure we have make the same number of signals from both classes in the labeled dataset. To reduce dimensionality of the raw acceleration signals, we conduct principal component analysis on all the signals and preverve 95% energy [6]. We then compare the performances of kernel support vector machine (SVM), which is a standard classifier that works through maximizing the margins of different classes [6], the diffusion functions and the adaptive graph filter. We construct a $k = 4$ regular graph (each vertex connects to 4 neighbors). We choose the local measurement ρ to be the cosine distance, scaling coefficient $\sigma = (1/N^2) \sum_{i,j} \rho(x^{(i)}, (x^{(j)}))$, transition time $t = 8$, filter length $K = 30$, the decay coefficient $\beta = 1$, and the penalty threshold $T = 0.02$, and perform a 30-fold cross-validation.

Results. Table II compares the performance of different classifiers with the low labeling ratio of 10%. We use V for vehicle and S for speed. In our experiments, kernel support vector machine performs poorly, diffusion function produce better accuracy, while the adaptive graph filters yields the highest accuracy.

Fig. 2 shows the dependence of classification accuracy on the labeling ratio for 2 vehicles averaged across 8 speeds. Both figures show similar trends; as the labeling ratio decreases, accuracy of SVM drops sharply; performances of diffusion function and adaptive graph filter stay relatively flat even at very low labeling ratios; and adaptive graph filter works consistently better than diffusion functions.
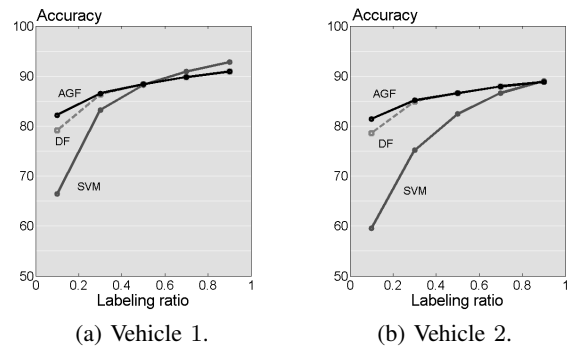
V. CONCLUSIONS

We presented an adaptive graph filtering technique for data classification based on a semi-supervised optimization algorithm that uses both labeled and unlabeled signals to learn the filter coefficients. Adaptive graph filters combine decisions from multiple graph filters to form a global decision. We also discussed the connection between adaptive graph filters and diffusion wavelets, and demonstrated their application to data classification and indirect bridge structural health monitoring.

REFERENCES

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, pp. 83–98, 2013.

[2] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, pp. 2786–2799, 2012.

[3] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, pp. 129–150, 2011.

[4] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.

[5] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, pp. 53–94, July 2006.

[6] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, Englewood Cliffs, NJ, 2001.

[7] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*, Am. Math. Soc., 1996.

[8] S. Chen, F. Cerda, J. Guo, J. B. Harley, Q. Shi, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, "Multiresolution classification with semi-supervised learning for indirect bridge structure health monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, Canada, May 2013.

[9] S. Chen, F. Cerda, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, "Semi-supervised multiresolution classification using adaptive graph filtering with application to indirect bridge structural health monitoring," *IEEE Trans. Signal Process.*, June 2013, Submitted.

[10] A. D. Szlam, R. R. Coifman, and M. Maggioni, "A general framework for adaptive regularization based on diffusion processes," *J. Mach. Learn. Res.*, , no. 9, pp. 1711–1739, Aug. 2008.

[11] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 u.s. election: Divided they blog," pp. 36–43, 2005.

[12] F. Cerda, S. Chen, J. Bielak, J. H. Garrett, P. Rizzo, and J. Kovačević, "Indirect structural health monitoring of a simplified laboratory-scale bridge model," *Int. J. Smart Struct. Syst.*, 2012, Submitted.