A REGULARIZATION FRAMEWORK FOR MOBILE SOCIAL NETWORK ANALYSIS

Xiaowen Dong †‡, *Pascal Frossard* †, *Pierre Vandergheynst* ‡ *and Nikolai Nefedov* *

Signal Processing Laboratory († LTS4 / ‡ LTS2)
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
* Nokia Research Center (NRC), Lausanne, Switzerland
{xiaowen.dong, pascal.frossard, pierre.vandergheynst}@epfl.ch, nikolai.nefedov@nokia.com

ABSTRACT

Mobile phone data provides rich dynamic information on human activities in social network analysis. In this paper, we represent data from two different modalities as a graph and functions defined on the vertex set of the graph. We propose a regularization framework for the joint utilization of these two modalities of data, which enables us to model evolution of social network information and efficiently classify relationships among mobile phone users. Simulations based on real world data demonstrate the potential application of our model in dynamic scenarios, and present competitive results to baseline methods for combining multimodal data in the learning and clustering communities.

Index Terms— Multimodal data, regularization on graphs, classification and clustering

1. INTRODUCTION

Data collected from mobile phones has recently attracted an increasing interest in the scientific research community. Compared to the traditional way of conducting social surveys, mobile phone data is expected to provide much richer behavioral information with less personal bias from the subjects. As one of the pioneering works, the Reality Mining project conducted at MIT Media Lab [1] during 2005 has provided inspirational research on how to use bluetooth-enabled mobile phones as wearable sensors to measure information in complex networks. More recently, Nokia Research Center (NRC) Lausanne has conducted a new mobile phone data collection campaign since September 2009 [2], which provides us with an excellent opportunity to study behaviors of mobile phone users and apply it to social network analysis.

Mobile phone data usually comes with a multimodal nature reflecting different aspects of human activities, such as geographical locations, bluetooth scans and phone calls. With such rich information, we are particularly interested in finding the "social affinity" in the mobile social network, which represents the strength of links between mobile users in both service centric and user centric applications. Intuitively, in order to achieve this goal, we need to efficiently "merge" different modalities of information together. In this work, we represent mobile phone data from two different modalities as a graph and functions defined on the vertices of the graph. We then propose a regularization framework on the graph, in which the functions living at the vertices of the graph are regularized by a "smoothness" constraint that is defined by the graph connectivity. This method helps us to merge two different modalities of information together for classification and clustering problems in mobile social network analysis; more interestingly, by taking into account the temporal changes of graph structures, we show that it enables us to capture the evolution of information in the network, which could lead to realtime recommendation systems.

The regularization theory on graphs has been previously studied by many researchers [3][4]. However, in most of the works the graph structure remains fixed. In contrast, we consider the temporal changes of the graph, which results in the evolution of functions defined on the graph. There are also a few works in literature that aim at efficient classification or clustering by combining multimodal data [5][6], in which different modalities are represented by multiple graphs that share the same set of vertices but different edges. The research efforts are then devoted to working directly with the multiple graph structures. However, we show that representing data as functions on the graph leads to alternative and efficient solutions to this general problem.

2. REGULARIZATION FRAMEWORK

In this section we explain how to merge two different modalities of information together through a regularization framework on graphs. Let us consider a weighted and undirected graph $\mathcal{G} = \{E, V, \omega\}$ which consists of a set of vertices V, a set of edges E with associated edge weights w, and define a real scalar function $f : V \rightarrow \mathbb{R}$ on the vertices of $\mathcal{G}$. Notice that we assume that f is a scalar function for the sake of simplicity, that is, $f(i)$ denotes the function value on vertex i, but the framework can be easily extended to vector-valued function. The regularization problem on graphs is outlined as follows. We start with a prior distribution $f^{[t_0]}$ of some specific information, which is viewed as a function on the vertices of a graph at time instant t_0. When the graph evolves over time (with the same set of vertices), the function f changes accordingly. The goal is then to estimate $f^{[t_n]}$ given the graph at time instant t_n, which reflects the dynamic evolution of the information that f represents.

To study the network dynamic behavior, we adopt two criteria, namely "smoothness" (or roughness) and "closeness", of the function $f^{[t]}$ defined on the graph. More specifically, we assume that the values of $f^{[t]}$ should vary slowly between closely related vertices in the graph, while not being too different from the prior distribution $f^{[t_0]}$. These criteria thus give us a regularization term and a fidelity term in the following optimization problem:

$$\underset{f^{[t]} \in \mathcal{H}(V)}{\arg\min} \{\frac{1}{2}||f^{[t]} - f^{[t_0]}||_2^2 + \lambda \cdot \Phi_f^{[t]}\} \quad (1)$$

where $\mathcal{H}(V)$ denotes the Hilbert Space of real functions on the vertex set of the graph, $\Phi_f^{[t]}$ is the smoothness term determined by the graph at time t, and λ is a parameter specifying the trade-off between two competing terms. In order to derive such a smoothness term, we use the graph differential operators that are defined in [3]. For two vertices i and j, the graph gradient operator ∇ and the Laplacian operator Δ applied on function f are defined as:

$$(\nabla f)(i,j) = \sqrt{\frac{w(i,j)}{d(j)}} f(j) - \sqrt{\frac{w(i,j)}{d(i)}} f(i) \tag{2}$$

$$(\Delta f)(j) = f(j) - \sum_{i \sim j} \frac{w(i,j)}{\sqrt{d(i)d(j)}} f(i) \tag{3}$$

where $w(i,j)$ is the weight on the edge between i and j, $d(i)$ and $d(j)$ are the degrees of i and j respectively, and $\sim$ denotes that i and j are neighbours in the graph. The smoothness of f over the graph can then be defined as the sum of the local variations of f at each vertex:

$$\Phi_f = \frac{1}{2} \sum_{j \in V} ||\nabla_j f||^2 = \frac{1}{2} \sum_{j \in V} \sum_{i \sim j} (\nabla f)^2(i,j) \tag{4}$$

Plugging (4) into (1) results in the following problem formulation:

$$\arg\min_{f^{[t]} \in \mathcal{H}(V)} \{ \frac{1}{2} ||f^{[t]} - f^{[t_0]}||_2^2 + \lambda \cdot \frac{1}{2} \sum_{j \in V} ||\nabla_j f^{[t]}||^2 \} \tag{5}$$

Directly differentiating the objective function in (5) gives:

$$(f^{[t]} - f^{[t_0]}) + \lambda \cdot \Delta f^{[t]} = 0 \tag{6}$$

Therefore, problem (5) can be efficiently solved by iterations:

$$f^{[t]}_{(n+1)} = f^{[t]}_{(n)} - \tau \cdot [f^{[t]}_{(n)} - f^{[t_0]} + \lambda \cdot \Delta f^{[t]}_{(n)}] \tag{7}$$

where the index (n) represents the iteration number, and τ is a suitable step size for this gradient descent process. Notice that the choice of the trade-off parameter λ in (5) depends on the relative importance of the competing terms in the application at hand.

So far we have explained how to get $f^{[t]}$ from the graph structure at time t, which enables us to investigate the temporal evolution of the distribution f. Clearly, this regularization framework also helps us to merge two modalities of information together, one being the dynamic graph structure and the other being the distribution f. In the following section, we will explain two applications based on these two properties of our framework.

3. APPLICATION TO EVOLUTION MODELING

3.1. Setup

The first application of our approach is to model the evolution of network information. We evaluate our method using the mobile phone data that is currently being collected by NRC Lausanne [2]. Specifically, we take two modalities of information, namely the GPS coordinates and the bluetooth scanning records, of 68 mobile phone users in the data collection campaign, and the goal is to merge these two modalities together. On the one hand, according to the regularization approach, we build the graph from the GPS information, that is, we take GPS coordinates of all the users at one time instant, and assign an edge between each pair with edge weight being one over the physical distance between them. Then we generate a 8-nearest-neighbor version of this complete graph as a representation of location relationships among all these users. Obviously, this "GPS graph" is constantly evolving as users are moving around; On the other hand, in order to build functions on the graph, we utilize the bluetooth scanning records of each user during the nine-month data campaign period. In particular, we represent each user as a vector of all the bluetooth devices that have ever been scanned, and calculate the cosine similarity between each pair of users. This gives us an symmetric adjacency matrix B where entry b_{ij} describes the similarity of the bluetooth scanning activities of user i and user j. From this information, we define for each user i a "bluetooth affinity" function f_i, where the affinity between i and j is calculated as:

$$f_i(j) = \frac{b_{ij}}{(\sum_{v \in V_{\overline{ij}}} (b_{iv} - b_{jv})^2)^{\frac{1}{2}}} \tag{8}$$

where $V_{\overline{ij}}$ denotes the user set without user i and j. Following this definition, we see that the affinity between two users is large when they have similar scanning activities not only directly to each other, but also in terms of their relationships to the rest of the users. Since bluetooth scans can reflect physical proximity of mobile users, these functions capture the relationships among the group of users that are chosen. Clearly, each f_i is a function defined on the vertices of the "GPS graph", representing the bluetooth affinities between user i and other users. We consider these functions f_i as the prior distributions $f^{[t_0]}$ mentioned in our regularization framework, and the goal is to estimate $f_i^{[t]}$ which are the exact distributions at time t. Obviously, this could be achieved by solving (5). Moreover, by choosing different time instants, we will be able to investigate how these distributions f_i, namely bluetooth affinities between mobile users, evolve with time.

3.2. Numerical results

Now we show the simulation results. We consider five different time instants t_1 to t_5, with an eight-hour delay between each two consecutive ones. As examples, Fig. 1 shows the GPS graphs at t_1 and t_2. We compute $f_i^{[t]}$ using the regularization framework, with a fixed λ set to 10 in order to emphasize the temporal change of the graph structures. Notice that the choice of λ is quite flexible here, as long as it gives a reasonably high weight to the regularization term. In practice, choosing λ in a wide range from 5 to 15 leads to similar results. Clearly, there are 68 functions f_i in this setting and here we just plot the results for two, namely f_3 and f_7, as typical examples which are shown in Fig. 2. Specifically, row 1 are the prior distribution f_3 and the exact distributions $f_3^{[t]}$ at five different time instants, and so as for row 2 for f_7. From these results we are able to view $f_3^{[t]}$ and $f_7^{[t]}$ as functions of time and see how they evolve as time goes by, which are clearly resulted from the relevant locations of user 3 and 7 to the other users at different time instants. Hence, this gives us an interesting model in estimating the temporal evolution of social affinities between mobile phone users. In this case such affinity is defined based on bluetooth scanning activities, but the approach can also be extended to other applications where different kind of information is interested.

3.3. Analysis

In the following, we are going to validate the evolution of bluetooth affinities by examining the bluetooth scanning records in the Nokia database. In particular, we define that user i and user j have a bluetooth interaction if they have scanned the same bluetooth device within a five-minute time window. Clearly, such interactions will affect the bluetooth affinity value between user i and j. Now

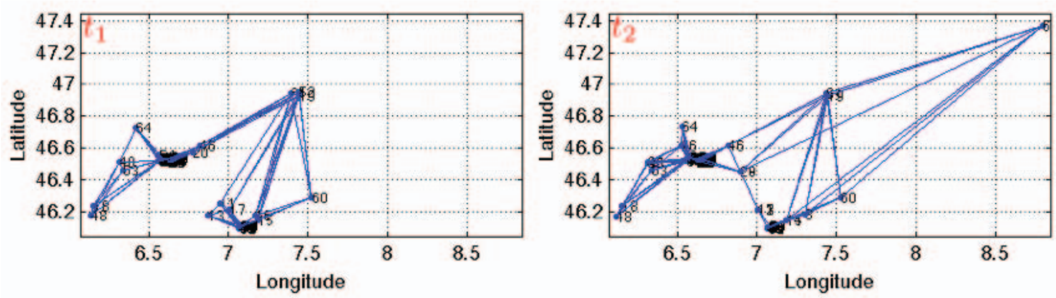


Fig. 1. GPS graphs at five different time instants

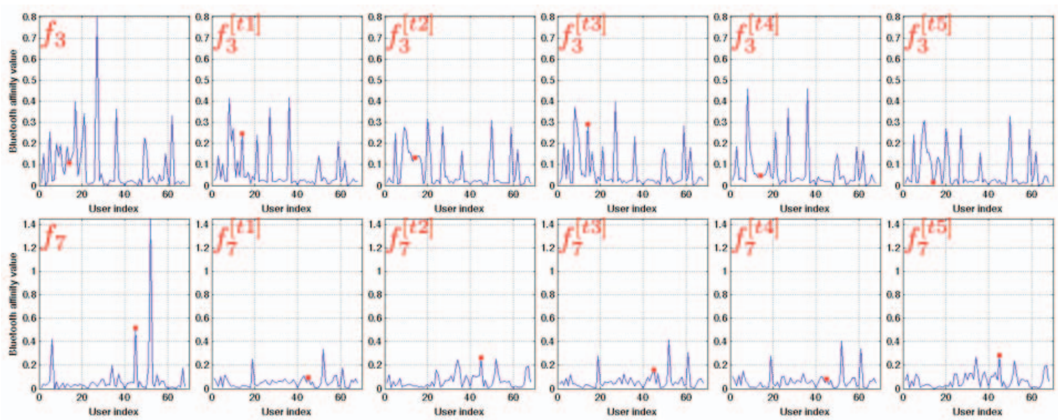


Fig. 2. Evolutions of $f_3^{[t]}$ and $f_7^{[t]}$

let us go back to the results shown in Fig. 2. In the first row the red point represents the affinity value between user 14 and 3, while in the second row it represents such value between user 45 and 7. We take these two dyadic user pairs as examples to show that the evolutions of $f_3^{[t]}$ and $f_7^{[t]}$ are supported by the bluetooth scanning records in the database. For f_3, for example, we can see that the value of $f_3(14)$ is not very large, meaning that the averaged bluetooth affinity between these two users is not strong. However, this value becomes quite large at time instants t_1 and t_3, and such increases are proven as user 3 and 14 have several interactions right around these two time instants, which makes the similarity of their scanning activities higher by the construction of f_i. Another example is shown in the second row of Fig. 2 for user 7. The averaged bluetooth affinity between user 7 and 45 is quite large in f_7, but at the five chosen time instants the values $f_7(45)$ drop significantly. This is supported by the fact in the database that, although user 7 and 45 indeed have a large number of interactions during the nine-month period, they have no interaction during the time period containing the chosen time instants. These two simple examples show that, the regularization framework enables us to capture the temporal evolution of the bluetooth affinity between each pair of the mobile phone users.

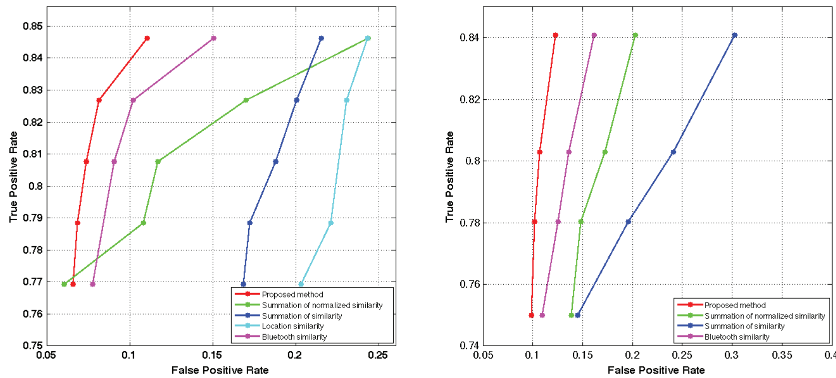
4. APPLICATION TO CLASSIFICATION & CLUSTERING

4.1. Setup

The second type of applications of our approach are classification and clustering problems. As a new way of combining multimodal data, we show that the regularization approach provides competitive results in the task of classification and clustering of objects compared to several baseline methods used for combining data represented by multiple graphs. In this task, we take averaged information on physical locations and bluetooth scanning records of 68 mobile users during nine months. More specifically, we measure how many times two users have been sufficiently close to each other, and how many times two have scanned the same bluetooth device, within a five-minute time window. Aggregating results from such windows throughout the nine-month period gives us two similarity matrices. Clearly, both of them reflect the proximity between these mobile users, and the goal is to combine them together. For our approach, we consider the similarity matrix from locations as the graph, and represent the information from bluetooth scans as functions on the graph in the same way as (8). We then apply the regularization method and combine the resulting functions f_i to get a new similarity matrix of these users. Notice that we choose λ to be around 0.005 in this scenario to give more weight to information from the bluetooth scans. Empirically, we have seen that a small range of λ from 0.002 to 0.01 provides effective performance. We compare our approach with the following three baseline methods that are described in [5]: 1) using sum of the two similarity matrices; 2) using sum of the normalized similarity matrices; 3) using sum of the spectral kernels from each similarity matrix.

4.2. Numerical results

Since we have access to the survey data that reveals the social relationships among the participants of the data collection campaign, we



(a) Nokia dataset (b) MIT Reality Mining dataset

Fig. 3. ROC curves for different methods

consider this information as the "ground truth" of proximity between users in our dataset. Each combining method can then be viewed as a binary classifier to classify these relationships. We compare the performance of different methods by plotting the Receiver Operating Characteristic (ROC) curves for each of them, which is shown in Fig. 3 (a). It can be seen that our proposed method performs better than these baseline methods in terms of the ratio of true positive rate ($i.e.$ number of true positives divided by number of positives) to false positive rate ($i.e.$ number of false positives divided by number of negatives).

In order to further verify our method, we apply the same analysis to the MIT Reality Mining dataset [7]. This dataset contains mobile phone data of 87 subjects participated in the Reality Mining project, with the only difference from the Nokia data being that it did not record the GPS coordinates of mobile users but just the IDs of the serving cell towers at each time instant. Nevertheless, we can still generate a similarity matrix representing roughly the location relationships among users by assigning the weight of edge between any two users as how many times they have been under the service of the same cell tower during a five-minute time window. We then construct functions on this location graph from bluetooth scanning records and apply our regularization approach. Similarly, we compare our approach to the baseline methods using the survey data in the dataset as the "ground truth". The result is shown in Fig. 3 (b) (the curve for using only location similarity is omitted due to its inferior performance), which again demonstrates that our proposed method outperforms the baseline methods. Moreover, we evaluate our method in the task of clustering mobile phone users using the MIT dataset. More specifically, we consider the self-reported affiliations of subjects as the "ground truth" of clusters among mobile users. We then apply spectral clustering [8] on the resulting adjacency matrices from each method (for baseline method 3 we use kernel K-means), and adopt Purity and Normalized Mutual Information (NMI) as two criteria to evaluate the clustering performance. From the result shown in Fig. 4, we see that our method shows better performance in terms of both measures.

5. DISCUSSIONS

The main contribution of this paper is two-fold: Firstly, we have proposed a regularization framework on graphs to model the temporal evolution of information that we are interested in. This is particularly interesting in social network analysis, since it enables us to capture the dynamic characteristics of the network, as shown in the bluetooth affinity example. Practical application of our approach could be powerful recommendation systems that provide mobile phone users with realtime affinity evolution. Secondly, as a new way of combining multimodal data, our method provides competitive results to the commonly used baseline methods in the task of classification and clustering. In terms of future work, we still need to find a constructive way of choosing the graph and the functions from more than two modalities of data to fit the regularization framework, which will certainly lead to more applications in social network analysis.

	Purity	NMI
Proposed method	0.7161	0.5974
Summation of normalized similarity	0.6628	0.4331
Summation of similarity	0.6368	0.3590
Summation of spectral kernels	0.6405	0.4567

Fig. 4. Evaluation of clustering for MIT Reality Mining dataset

6. REFERENCES

[1] N. Eagle and A. Pentland, "Reality Mining: Sensing Complex Social Systems," *Personal and Ubiquitous Computing*, vol. 10, no. 4, pp. 255–268, 2006.

[2] N. Kiukkonen, J. Blom, O. Dousse, D. Gatica-Perez, and J. Laurila, "Towards Rich Mobile Phone Datasets: Lausanne Data Collection Campaign," in *International Conference on Pervasive Services*, Berlin, Germany, Jul 2010.

[3] D. Zhou and B. Schölkopf, "A Regularization Framework for Learning from Graph Data," in *ICML Workshop on Statistical Relational Learning and Its Connections to Other Fields*, 2004, pp. 132–137.

[4] A. Smola and R. Kondor, "Kernels and Regularization on Graphs," in *16th Annual Conference on Computational Learning Theory*, Washington, DC, USA, Aug 2003.

[5] W. Tang, Z. Lu, and I. Dhillon, "Clustering with Multiple Graphs," in *International Conference on Data Mining*, Miami, Florida, USA, Dec 2009.

[6] A. Argyriou, M. Herbster, and M. Pontil, "Combining Graph Laplacians for Semi-Supervised Learning," in *NIPS*, 2005.

[7] N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure Using Mobile Phone Data," in *PNAS*, 2009, vol. 106, pp. 15274–15278.

[8] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug 2000.