Project Proposal

Sensing Semantic Information from Mobile Social Networks

Ting Zhang and Wen Yi
School of Industiral Engineering, School of Electrical and Computer Engineering
zhan1013@purdue.edu, yi35@purdue.edu

September 29, 2015

1 Introduction

Human society consists of extensive communications and interactions between individuals, via the use of mobile sensors, such as mobile phones, tablets and GPS. The understanding of individual relations from these sensors, can greatly facilitate and promote the interactions between individuals. For example, listing phone contacts in semantic orders according to the time and location when a person wants to make a phone call, would save the person both time and memory load, to find a specific contact from a phone book with tons of contacts based on alphabetic order. Therefore, in this project, we will focus on the inference of friend relationships with the detection of communities using data collected from mobile phones. Detailed explanation of dataset is stated in section 4.

2 Related work

2.1 Relationship Inference

Relationship inference in social networks has been studied in various fields and domains. In this context, we refer to friendship inference between pairs of individuals. Representing social networks with topology structures provides insights to predict relationships between individuals based on topology and probability distribution of the links in the topology. Liben-Nowell and Kleinberg proposed different measurements to compute the s̈imilarity b̈etween two nodes (individuals) in the graph, including the distance between two nodes, number of shared neighbors, and m̈eta-approaches ẗhat integrate different measurements. Beyond topological structures, individual attribute and context information have also been utilized to facilitate the construction of relations between individuals. In the study from Taskar et al., correlations between individuals were constructed using user attributes with relational Markov Networks. For instance, they proposed a transitivity pattern that is useful in relationship prediction, where the presence of A-B relation and B-C relation promotes the probability of A-C relation. Context information, such as locations and periods of time, has also shown potentials to predict social ties. Using location information alone may not be a sufficient predictor. In Crandall et al.'s work, only 0.1% of the relations were

predicted with a confidence of 60%. However, when network structure are analyzed together with location information, over 90% friendship were detected with confidence over 80%, illustrated in the study from Sadilek et al.. Although location information alone is not a good indicator for friendship inference, a number of researches have indicated the importance between social ties and distance. The integration of location information and other features are also proved to be of high accuracy in friendship inference.

2.2 Community Detection

In previous research such as the paper by Xie and Szymanski, from the analysis of individual data collected, community detection in social networks focus on grouping the people into overlap or distinct communities while the people from same community have stronger general relationships or share the similar characteristics. Further more, in the paper by Mislove et al., community detection were used to inferring the profile information of people based on the profile information from other members in the same community, while some of the members in community have profile information vacant.

3 Problem formulation

Describe your project as a machine learning problem, identify inputs objects, labels, possible features

4 Data and Evaluation plan

We will use a reality mining dataset from MIT media lab Eagle et al.. The dataset consists of phone logs of 94 subjects from September 2004 to June 2005. Among these 94 subjects, 68 were colleagues working in the same building (90% were graduate students, while 10% were staff). The remaining 26 subjects were incoming students from the business school. The dataset was collected from Nokia 6600 phones programmed to automatically run a log application as background process, including phone log, bluetooth and location. The format of each log is summarized as following:

* Phone log: (TIME) 20060720T211505 (DESCRIPTION) Voice call (DIRECTION) Outgoing (DURATION seconds) 23 (NUMBER) 6175559821
* Bluetooth: (TIME) 20060721T111222 devices: 000e6d2a3564 [Amys Phone]000e6d2b06ea [Jons PalmPilot]
* Location:(TIME) 20060721T111222 (CELL AREA) 24127, (CELL TOWER) 111, (SERVICE PROVIDER) AT&T Wirel (USER DEFINED LOCATION NAME) My Office

How will you evaluate your algorithm? What is a reasonable baseline?

Submission Instructions:

delete this section when submitting

You are required to use LaTeX to type your solutions to questions, and report of your programing as well. Other formats of submission will **not** be accepted. A template named "homework.tex" is also provided for your convenience.

After logging into data.cs.purdue.edu (physically go to the lab or use ssh remotely, you are all granted the accounts to CS data machines during this class), please follow these steps to submit your assignment:

1. Make a directory named *'your Name_your Surname'* and copy all of your files there.
2. While in the upper level directory (if the files are in /homes/dan/dan_goldwasser, go to/homes/dan), execute the following command:

   `turnin -c cs578 -p PROPOSAL *your_folder_name*`

   (e.g. your prof would use: `turnin -c cs578 -p PROPOSAL dan_goldwasser` to submit his work)

   Keep in mind that old submissions are overwritten with new ones whenever you execute this command.

3. You can verify the contents of your submission by executing the following command:

   `turnin -v -c cs578 -p PROPOSAL`

   Do **not** forget the -v flag here, as otherwise your submission would be replaced with an empty one.

References

D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. 107:22436–22441. ISSN 0027-8424. doi: 10.1073/pnas.1006155107. URL http://adsabs.harvard.edu/abs/2010PNAS..10722436C.

Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. 106(36):15274–15278. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0900282106. URL http://www.pnas.org/content/106/36/15274.

David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. 58(7):1019–1031. ISSN 1532-2882. doi: 10.1002/asi.v58:7. URL http://dx.doi.org/10.1002/asi.v58:7.

Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. You are who you know: Inferring user profiles in online social networks. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 251–260. ACM. ISBN 978-1-60558-889-6. doi: 10.1145/1718487.1718519. URL http://doi.acm.org/10.1145/1718487.1718519.

Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham. Finding your friends and following them to where you are. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 723–732. ACM. ISBN 978-1-4503-0747-5. doi: 10.1145/2124295.2124380. URL http://doi.acm.org/10.1145/2124295.2124380.

Ben Taskar, Ming-fai Wong, Pieter Abbeel, and Daphne Koller. Link prediction in relational data. In S. Thrun, L. K. Saul, and B. Schlkopf, editors, *Advances in Neural Information Processing Systems 16*, pages 659–666. MIT Press. URL http://papers.nips.cc/paper/2465-link-prediction-in-relational-data.pdf.

Jierui Xie and B.K. Szymanski. Community detection using a neighborhood strength driven label propagation algorithm. In *2011 IEEE Network Science Workshop (NSW)*, pages 188–195. doi: 10.1109/NSW.2011.6004645.