Machine learning approaches to Reality Mining: The Badge dataset

1 Description

One of the most challenging areas for the application of machine learning algorithms is Reality Mining [1], in particular the understanding of social dynamics from the analysis of data describing social interactions (e.g., data about mobile communications, coworking interactions, information exchange, computer use, etc.). Several recent projects have proposed experiments in which this type of data is collected [2, 3, 1, 4, 5] and a number of approaches that analyze this data have been also published [6, 7, 8, 9, 10, 11]. However, the question of exploring new and creative applications of machine learning in this domain is relevant.

In this project the students will use the *Badge Dataset* `http://realitycommons.media.mit.edu/badgedataset.html` [4] which contains the performance, behavior, and interpersonal interactions of participating employees at a Chicago-area data server configuration firm for one month. This was the first data set to contain the performance and dynamics of a real-world organization with a temporal resolution of a few seconds.

2 Objectives

The objective of this project is to apply one supervised or unsupervised machine learning algorithm (except those based on NNs for which a different project is proposed) to the *Badge* dataset that allow to produce insights about the social interactions captured by the data. The student should: 1) Select the approach to apply to the data and the Python implementation to use. 2) Preprocess the data as required for the approach chosen. 3) Apply the algorithm, describe the results, and explain why these results are useful, interesting, or reveal any insight about the process.

As in other projects, a report should describe the characteristics of the design, implementation, and results. A Jupyter notebook should include calls to the implemented function that illustrate the way it works.

3 Suggestions

- Read the description of the dataset in [4] and other papers that analyze this data.
- Think about what are relevant questions for the scenario that experiments coverts. Determine whether the available ML approaches are useful to answer these questions.
- Select any of the available ML Python libraries and implement the idea.

References

[1] Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[2] Wen Dong, Bruno Lepri, and Alex Sandy Pentland. Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, pages 134–143. ACM, 2011.

[3] Wen Dong, Katherine Heller, and Alex Sandy Pentland. Modeling infection with multi-agent dynamics. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 172–179. Springer, 2012.

[4] Daniel Olguín Olguín, Benjamin N Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):43–55, 2009.

[5] Michele Starnini, Bruno Lepri, Andrea Baronchelli, Alain Barrat, Ciro Cattuto, and Romualdo Pastor-Satorras. Robust modeling of human contact networks across different scales and proximity-sensing techniques. *CoRR*, abs/1707.06632, 2017.

[6] Tanzeem Khalid Choudhury. *Sensing and modeling human networks*. PhD thesis, Massachusetts Institute of Technology, 2003.

[7] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[8] Anmol Madan and Alex Pentland. Modeling social diffusion phenomena using reality mining. In *AAAI Spring Symposium: Human Behavior Modeling*, pages 43–48, 2009.

[9] Anmol Madan, Manuel Cebrian, David Lazer, and Alex Pentland. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 291–300. ACM, 2010.

[10] Rahman O Oloritun, Anmol Madan, Alex Pentland, and Inas Khayal. Identifying close friendships in a sensed social network. *Procedia-Social and Behavioral Sciences*, 79:18–26, 2013.

[11] R. Santana. Multi-objective optimization approach to detecting extremal patterns in social networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction (SBP-2013)*, pages 1–6, Washington, US, 2013. Submitted for publication and Winner of the second place in the SBP-2013 Challenge.