Graph-Based User Behavior Modeling: From Prediction to Fraud Detection

Alex Beutel
Carnegie Mellon University
Pittsburgh, PA, USA
abeutel@cs.cmu.edu

Leman Akoglu
Stony Brook University
Stony Brook, NY, USA
leman@cs.stonybrook.edu

Christos Faloutsos
Carnegie Mellon University
Pittsburgh, PA, USA
christos@cs.cmu.edu

ABSTRACT

How can we model users' preferences? How do anomalies, fraud, and spam effect our models of normal users? How can we modify our models to catch fraudsters? In this tutorial we will answer these questions - connecting graph analysis tools for user behavior modeling to anomaly and fraud detection. In particular, we will focus on the application of subgraph analysis, label propagation, and latent factor models to static, evolving, and attributed graphs.

For each of these techniques we will give a brief explanation of the algorithms and the intuition behind them. We will then give examples of recent research using the techniques to model, understand and predict normal behavior. With this intuition for how these methods are applied to graphs and user behavior, we will focus on state-of-the-art research showing how the outcomes of these methods are effected by fraud, and how they have been used to catch fraudsters.

Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Data mining;
H.3.5 [**Information Systems**]: Information Storage and Retrieval—*On-line Information Services*

Keywords

User Behavior Modeling; Fraud Detection; Anomalous Behavior; Outlier Detection; Recommendation Systems

1. TUTORIAL PERSPECTIVE

In this tutorial we focus on understanding anomaly and fraud detection through the lens of normal user behavior modeling. The data mining and machine learning communities have developed a plethora of models and methods for understanding user behavior. However, these methods generally assume that the behavior is that of real, honest people. On the other hand, fraud detection systems frequently use similar techniques as those used in modeling "normal" behavior, but are often framed as an independent problem. However, by focusing on the relations and intersections of the two perspectives we can gain a more complete understanding of the methods and hopefully inspire new research joining these two communities.

KDD '15, August 10-13, 2015, Sydney, NSW, Australia.
 ISBN 978-1-4503-3664-2/15/08.
DOI: http://dx.doi.org/10.1145/2783258.2789985.

2. TARGET AUDIENCE

This tutorial is aimed at anyone interested in modeling and understanding user behavior, from data mining and machine learning researchers to practitioners from industry and government. For those new to the field, the tutorial will cover the necessary background material to understand these systems and will offer a concise, intuitive overview of the state-of-the-art. Additionally, the tutorial aims to offer a new perspective that will be valuable and interesting even for researchers with more experience in these domains. For those having worked in classic user behavior modeling, we will demonstrate how fraud can effect commonly-used models that expect normal behavior, with the hope that future models will directly account for fraud. For those having worked in fraud detection systems, we hope to inspire new research directions through connecting with recent developments in modeling "normal" behavior.

3. INSTRUCTORS

Alex Beutel is a fifth year Ph.D. candidate at Carnegie Mellon University in the Computer Science Department. He previously received his B.S. from Duke University. His Ph.D. research focuses on large scale user behavior modeling, covering both recommendation systems and fraud detection systems. He has interned at Facebook on both the Site Integrity and News Feed Ranking teams, at Microsoft in the Cloud and Information Services Laboratory, and at Google Research. His research is supported by a Facebook Fellowship and the National Science Foundation Graduate Research Fellowship Program. More details can be found at `http://alexbeutel.com`.

Leman Akoglu is an Assistant Professor in the Department of Computer Science at Stony Brook University. She received her Ph.D. from the Computer Science Department at Carnegie Mellon University in 2012. She also worked at IBM T. J. Watson Research Labs and Microsoft Research at Redmond during summers. Her research interests span a wide range of data mining and machine learning topics with a focus on algorithmic problems arising in graph mining, pattern discovery, social and information networks, and especially anomaly mining; outlier, fraud, and event detection. Dr.

Akoglu's research has won 4 publication awards; Best Research Paper at SIAM SDM 2015, Best Paper at ADC 2014, Best Paper at PAKDD 2010, and Best Knowledge Discovery Paper at ECML/PKDD 2009. She also holds 3 U.S. patents filed by IBM T. J. Watson Research Labs. Dr. Akoglu is a recipient of the NSF CAREER award (2015) and Army Research Office Young Investigator award (2013). Her research is currently supported by the National Science Foundation, the US Army Research Office, DARPA, and a gift from Northrop Grumman Aerospace Systems. More details can be found at `http://www.cs.stonybrook.edu/~leman`.

Christos Faloutsos is a Professor at Carnegie Mellon University. He has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the Innovations award in KDD'10, 20 "best paper" awards, and several teaching awards. He has served as a member of the executive committee of SIGKDD; he has published over 200 refereed articles, 11 book chapters and one monograph. He holds five patents and he has given over 30 tutorials and over 10 invited distinguished lectures. His research interests include data mining for graphs and streams, fractals, database performance, and indexing for multimedia and bio-informatics data. More details can be found at `http://www.cs.cmu.edu/~christos/`.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1247489, IIS-1217559, CNS-1314632, IIS-1408924, IIS-1408287, CAREER 1452425, DGE-1252522, by the Defense Threat Reduction Agency under contract No. HDTRA1-10-1-0120, by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, by the U.S. Army Research Office (ARO) and Defense Advanced Research Projects Agency (DARPA) under Contract Number W911NF-11-C-0088, and by the ARO Young Investigator Program under Contract No. W911NF-14-1-0029.

This work is also partially supported by a Google Focused Research Award, a Facebook Fellowship, a Facebook Faculty Gift, an R&D grant from Northrop Grumman Aerospace Systems, and the Stony Brook University Office of Vice President for Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.