VirginiaTech
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Computer Science Seminar Series

National Capital Region

Scalable and Interactive Data Analytics

Speaker: Prof. Chandan Reddy
Wayne State University
Friday, February 6, 2015
1:00PM - 2:00PM, NVC 325

Abstract

Big data has become ubiquitous. In recent years, data acquisition has become easier and data storage has become cheaper which has led to the accumulation of large volumes of data in a wide range of applications. Analytical methods that can provide critical insights from such voluminous datasets are yet to catch up with these rapid developments. This talk consists of two parts. In the first part, I will present mapreduce-based scalable ensemble machine learning algorithms that can efficiently handle large-scale data by facilitating simultaneous participation of multiple computing nodes. I will demonstrate the superior performance of the proposed algorithms in terms of speedup and scaleup while maintaining the generalizability of the corresponding original versions. Some of the related topics such as privacy-preserving data mining and multi-task learning in the context of big data will also be discussed. In the second part of the talk, I will present a novel interactive topic modeling approach for analyzing document collections. Most of the widely-used topic modeling methods based on probabilistic models, such as Latent Dirichlet Allocation (LDA), have drawbacks in terms of consistency from multiple runs, empirical convergence, and incorporating user feedback. To overcome these challenges, we developed a reliable and flexible visual analytics system for topic modeling called UTOPIAN (User-driven Topic modeling based on Interactive Nonnegative Matrix Factorization). I will describe the UTOPIAN system and how it enables users to interact with the topic modeling method and steer the results in a user-driven manner. I will end this talk with some of our ongoing works in healthcare and social media.

Biography

Chandan Reddy is an Associate Professor in the Department of Computer Science at Wayne State University. He received his Ph.D. from Cornell University and M.S. from Michigan State University. He is the Director of the Data Mining and Knowledge Discovery (DMKD) Laboratory and a scientific member of Karmanos Cancer Institute. His primary research interests are Data Mining and Machine Learning with applications to Healthcare Analytics, Bioinformatics and Social Network Analysis. His research is funded by the National Science Foundation, the National Institutes of Health, the Department of Transportation, and the Susan G. Komen for the Cure Foundation. He has published over 50 peer-reviewed articles in leading conferences and journals including TPAMI, TKDE, DMKD, SIGKDD, ICDM, SDM, and CIKM. He received the Best Application Paper Award in ACM SIGKDD conference in 2010, and was a finalist of the INFORMS Franz Edelman Award Competition in 2011. He is a senior member of IEEE and life member of ACM.