

## Morning Tutorial

# Big Data Analytics for Healthcare

Jimeng Sun  
IBM TJ Watson Research Center

Chandan K. Reddy  
Wayne State University

### Abstract

Large amounts of heterogeneous medical data have become available in various healthcare organizations (payers, providers, pharmaceuticals). Those data could be an enabling resource for deriving insights for improving care delivery and reducing waste. The enormity and complexity of these datasets present great challenges in analyses and subsequent applications to a practical clinical environment. In this tutorial, we introduce the characteristics and related mining challenges on dealing with big medical data. Many of those insights come from medical informatics community, which is highly related to data mining but focuses on biomedical specifics. We survey various related papers from data mining venues as well as medical informatics venues to share with the audiences key problems and trends in healthcare analytics research, with different applications ranging from clinical text mining, predictive modeling, survival analysis, patient similarity, genetic data analysis, and public health. The tutorial will include several case studies dealing with some of the important healthcare applications.

### Instructors

Jimeng Sun is a research staff member at IBM TJ Watson Research Center. Dr. Sun graduated with PhD in Computer

Science in Carnegie Mellon University in the fall 2007. His advisor was Prof. Christos Faloutsos. He studied in Computer science department at Carnegie Mellon University from 2003 to 2007. His research focus is on healthcare analytics and informatics, large-scale data mining, graph mining, high dimensional data mining such as time series, matrices, and tensors (data cubes) and visual analytics. Dr. Sun has received ICDM best research paper in 2007 and KDD Dissertation runner-up award in 2008 and SDM best research paper in 2007. For more details, one can refer to his personal homepage at <http://www.dasfa.net/jimeng>.

Chandan K. Reddy is an Assistant Professor in the Department of Computer Science at Wayne State University. He received his PhD from Cornell University and MS from Michigan State University. His primary research interests are in the areas of data mining and machine learning with applications to healthcare, 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 45 peer-reviewed articles in leading conferences and journals. He received the Best Application Paper Award at the ACM SIGKDD conference in 2010 and was a finalist of the INFORMS Franz Edelman Award Competition in 2011. He is a member of IEEE, ACM, and SIAM.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
Copyright is held by the author/owner(s).

*KDD '13*, August 11–14, 2013, Chicago, Illinois, USA.  
ACM 978-1-4503-2174-7/13/08.