

Tools, Techniques and Case Studies for Text Analytics for Real World Problems

(Tutorial)

Sameep Mehta
IBM Research India
sameepmehta@in.ibm.com

Dhaval Patel
Indian Institute of Technology Roorkee
patelfec@iitr.ac.in

ABSTRACT

In this tutorial, we will present the commonly used techniques for Entity Extraction, Entity Resolution & Text Mining to handle real world problems. The tutorial will focus on few case studies based on social data. We will present the unique challenges which were encountered during client engagements and how we extended the traditional techniques to handle the nuances. We will also cover some of the state-of-the-art tools towards this purpose. Finally, we will conclude the tutorial with some open ended problems and in-progress work. The brief tentative outline is

- Introduction to Big Text Data
- Key Steps in Extracting Insights from Text Data
- Motivating examples to show nuances and finer points in Sentiment Mining
- Introduction to Entity Extraction and Integration Languages
- Case Study - Monitoring Social Channels for Elections
- Joint Analysis of Text and Links - Followers, Friends etc.
- Open/WIP Research Problems: Identifying News Broadcasters & Activity Miner

Biographies

Sameep Mehta is a Senior Researcher and Manager at IBM Research - India. He received his Ph.D. in Data Mining and Visualization from Ohio State University. His current research interests are Data Mining, Text Mining, Machine Learning, Big Data Technologies, Social Data Analytics and Knowledge Graph. He has published extensively in top conferences in Data Mining, Services and Visualization. He is a regular speaker at conferences and is PC chair for Big Data Analytics Conference 2014. He also serves as Adjunct Faculty at IIIT-Delhi in the area of Data Analytics.

Dr. Dhaval Patel received PhD from National University of Singapore in 2011. His PhD work focus on analyzing complex medical data for building efficient classifier. After

2011, Dr. Patel worked two year as post-doctoral research fellow in the area of spatio-temporal data analysis. Dr. Patel joined IIT-Roorkee in 2013. His current research interest lies in analyzing real world dataset (mostly text and location-aware) for building useful applications. He continues to publish papers in top-tier conferences (IJCAI-2015, CIKM-2015, DASFAA-2015, DAWAK-2015).

Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Articles from this volume were invited to present their results at The 21st International Conference on Management of Data. *International Conference on Management of Data, COMAD*, Copyright 2016 Computer Society of India (CSI).