



Data Science Community Hour

Thursday April 8th, 4:30–5:30 pm via [Zoom \(click to join\)](#)

Interview with Dr. Rakesh Chalasani @Spotify



Rakesh Chalasani, Ph.D., is a manager for machine learning engineers at Spotify's Boston office and gives leadership to teams that work on music and Podcast personalization for better user experience and creator value. Previously he was Vice President of Technology at nDimensional Inc., where he developed deep learning models for predictive maintenance systems. He earned the Ph.D. degree from the University of Florida in 2013 and the Bachelor of Technology from the National Institute of Technology, Nagpur, India in 2008.

“Identifying Abstracts Relevant to Drug-Drug Interactions (DDIs) and Experimental Evidence within Biomedical Text” — Gongbo Zhang, CIS, UD.

DDIs can cause adverse drug effects, which result in morbidity and mortality. Predicting, anticipating, and avoiding them is thus an area of much concern. Studies of DDIs are typically published in the biomedical literature; however, gathering DDI information remains a challenging task given the rapidly increasing publications volume. In this talk, we present a supervised learning approach to automatically collect evidence for DDIs from the biomedical literature. First, we identify the publications relevant to DDIs. The number of DDI-relevant publications is relatively small compared to the vast amount of irrelevant publications. Models that attempt to identify publications relevant to DDIs without handling this imbalance are generally biased toward the irrelevant class, as the classifiers are trained on a dataset where most examples are irrelevant to DDIs. We develop a two-stage method to tackle this challenge. Potential DDIs can be verified with higher confidence if supporting evidence is provided. Such evidence is typically described in statements conveying information about experimental methods and results. Analyzing such sentences can form the basis for effective retrieval of DDI knowledge. To identify the sentences containing evidential information, we employ learning-based text classification methods, including statistical learning models and neural networks.



Gongbo Zhang is a Ph.D. candidate in the Computer and Information Sciences Department at University of Delaware. He is a member of Computational Biomedicine and Machine Learning Lab, under the advisement of Prof. Hagit Shatkay. Gongbo's research focuses on applying text mining technologies on biomedical literature to obtain evidence for Drug-Drug Interactions (DDIs).

“PDFigCapX and FigSplit - Tools for Extracting Figures, SubFigures and Captions from Biomedical Publications” — Pengyuan Li, CIS, UD.

Figures and captions convey essential information in biomedical documents. As such, there is a growing interest within bio-databases to store and to display biomedical images as evidence for biomedical processes and for experimental results. Notably, the first fundamental step, namely extracting figures and captions from biomedical documents is neither well-studied nor yet well-addressed. Moreover, as the vast majority of published figures are compound images consisting of multiple panels, where each individual panel potentially conveys a different type of information, segmenting such images into constituent panels is another necessary step toward displaying and utilizing published images. In this talk, we introduce an effective pipeline comprising two systems: PDFigCapX for identifying and extracting figures and captions from biomedical documents, and FigSplit for splitting the extracted compound figures into their constituent subfigures.



Pengyuan Li currently is a PhD candidate in the department of Computer and Information Sciences at the University of Delaware. He is a member of the Computational Biomedicine and Machine Learning Lab, under the advisement of Prof. Hagit Shatkay. Pengyuan's research interests lie in Bio- and Medical- Informatics, Data Mining, and Machine Learning.