At the Intersection of Healthcare, Drug Discovery, and Big Data

Nirmal Keshava
Senior Principal Informatics Scientist, Research & Development Information
AstraZeneca, PLC
35 Gatehouse Drive, Waltham, MA 02451
Nirmal.Keshava@astrazeneca.com, (781) 839-4677

Public and private institutions are increasingly turning to data to make decisions that improve the cost and effectiveness of healthcare. However, the sources, quantity, and complexity of data are all increasing at exponential rates. As an integral part of improving healthcare, pharmaceutical companies are also looking to harness the collection and analysis of data to improve the yield of drug development and reverse the trend of decreasing efficiency that has accompanied increased R&D spending.

While Big Data has promised to impact all sectors of business, the challenges of extracting value vary greatly. Retail and commercial interests exploit data that is usually highly structured and are less interested in mechanistic explanations for their data as they are in how to exploit significant correlations to improve customer stratification, product recommendations, and to target advertising.

In contrast, pharmaceutical companies are being inundated with data from state-of-the-art measurement platforms as well as altogether new data sources enabled by technological breakthroughs such as mobile technologies and genomics. FDA drug approval can also hinge upon an understanding of mechanism-of-action and cannot rely simply on correlations to demonstrate the safety and efficacy of drugs.

In this talk we briefly review the history and landscape of drug discovery and the sources of data that are driving Big Data in the pharmaceutical industry before examining how one key source of data, Real-World Evidence (RWE), is being used to improve insight on a number of key problems facing drug discovery. In particular, claims data from medical insurers provide a longitudinal medical history for millions of patients that can be modeled thru the use of graphs to represent the causal and temporal nature of human health.

In the context of Type 2 diabetes, we will demonstrate how graphs can host such RWE and how analysis of patient graphs can build altogether new interpretations of how diseases evolve before diagnosis as well as modeling the relationships between members of a cohort. The graphs are extensible to include other forms of data and also can scale to significantly larger populations. Finally, we will examine the evolving trajectory of skills and expertise that are increasingly in demand to find value in Big Data in healthcare and the pharmaceutical industry.