Diffusion of novel treatments in healthcare

Flavio Chierichetti, Alessandro Panconesi
Sapienza University

Frank Davenport
UCSB

Luca Foschini
Evidation Health

Research Objectives and Aims

The goal of this project is to better understand how new treatments make their appearance in the health care practice and how they spread to an increasing subset of the population over time. By modeling the diffusion process of drug prescriptions and procedures we will identify common patterns of diffusion and the factors that drive these patterns (e.g., geographic separation, heterogeneous demand at different location). This will help drug developers and health care providers to predict and optimize the uptake and spread of new treatments.

Scope/Proposed Approach

Identifying "Rising Stars" First, we will consider how utilization of new drugs spread across the population. Then we will apply the same framework to procedures and diagnoses. Our initial focus will be on "rising stars": prescription drugs that have shown a prolonged and constant increase in popularity across the overall population. Finally we will define and test metrics that measure the increase of popularity and also model the distribution of such metrics for all drugs present in the dataset.

Model rising stars' spatial diffusion process over time We will begin by testing whether the change over time in prevalence of a drug in a region (fraction of prescription for D over all prescriptions in that area) is dependent on the prevalence of drugs in surrounding regions in earlier time periods. Specifically, for a given drug and year we will examine the distribution (X) of the average absolute difference in prevalence for two different locations: S and S' $X = |P_S - P'_S|$. By comparing the distributions cases in which S and S' are two neighboring geographic locations vs. when they are not, we can begin to identify the extent to which drug diffusion is a spatially dependent process.

We expect the distributions of drug prevalence in neighboring vs. non-neighboring locations to be different because the diffusion of information tends to have strong spatial correlation. We will then empirically model the diffusion of drug prescriptions using a space-time Poisson process, wherein the outcome variable is the count of prescriptions for a given drug at time t in location s. We will model the diffusion process for "Rising Star" drugs for both Metropolitan Statistical Areas (MSA) and U.S. states. By comparing the diffusion models from two different spatial scales (MSA and states) we can also investigate the extent to which the observed diffusion process is a function of either the underlying sample population or the spatial resolution of the data.

Identify similarity in diffusion processes After modeling drug diffusion process through a space-time Poisson process, the fitted parameters (regression coefficients) for each drug can be used to define a similarity metric on drug diffusion processes.

Drugs spreading processes can be then clustered based on the similarity of their fitted parameters, with the goal of identifying groups of drugs that have similar spreading patterns over time. (e.g., "the coast-to-coast spreading pattern" vs. "the LA-centric spreading pattern"). We will compare the clusters found with external grouping of the drug set (e.g., by brand, by treated condition) with the goal of identifying the causes underpinning the pattern variability. Such information will be key to formulate hypotheses on the drivers of the different diffusion mechanisms: from population center to population center (word of mouth), or from health center to health center (word of Doctor's mouths, or successful marketing.

Additional and alternative strategies We will also investigate the possibility of extending Granger Causality's framework to the spatial context. In the case the geographic resolution proves to be high enough, Granger Causality can be approximated with space-time data by comparing correlations between time lagged datasets $(s, t- > s, t+1)$ with space time lagged data sets $(j, t- > s, t+1$, where j is a neighbor to $s)$. These can be visualized/analyzed in a manner similar to time-series data by using spatial-temporal auto and cross-correlograms.

We will also investigate dependency between spreading patterns of different drugs. For instance, we expect that the diffusion of drug A would affect the pattern of diffusion of a drug B treating the same condition. Factoring in dependency relationship among different drugs can be addressed by extending the used spatial panel models to model multiple diffusion processing simultaneously. This can be accomplished by adding an additional layer to the model that assumes the space-time diffusion parameters for each drug as drawn from an underlying population of parameters, similar to well-known mixed-effects models.

Impact

The immediate benefits of a better understanding of the patterns of diffusion of new treatments (drugs and procedures) are twofold. First, it will allow us to identify areas and population segments for which the penetration of new treatment options is slower than average. That will allow deploying interventions targeted to the under-served population in that area, with the aim of strategically injecting the novel information that the natural diffusion process failed to relay. Second, being able to quantify the relative importance of catalyzers (e.g., word of mouth vs. targeting campaign) and inhibitors (e.g. geographic obstacles) of the diffusion process will allow a better allocation of resources to foster future campaigns.

Experience

Prof. Chierichetti obtained his Ph.D. in Computer Science from Sapienza University of Rome, where he currently is an Associate Professor. He is the PI of an ERC Starting Grant, a Google Focused Award, a Google Faculty Award, and of a SIR Grant. Prof. Chierichetti's main research interests lie in mathematical modelling and algorithmics — he has published numerous papers in the areas of parallel and approximation algorithms, data mining, network analysis and probabilistic inference. Prof. Chierichetti also has a history of collaboration with the industry — he has strong ties with several groups in Google (Mountain View and New York), and he visits them quite often.

Dr. Davenport received a PhD in Geography from the University of California Santa Barbara (UCSB). He currently a researcher at the Climate Hazards Group located within the same department. His research analyzes food systems by trying to unpack the relationships among weather, crop production, grain prices, health, and food security. His recent papers analyze the relationship between weather and low birth weights in Africa, price behavior in Mexican maize markets, and methods for estimating standard errors when working with spatial panel data. Prior to his research career Dr. Davenport spent 5 years as a Geographic Information Systems (GIS) analyst and strategic planner, implementing enterprise GIS and decision support systems for environmental agencies around the world.

Dr. Foschini is a co-founder and Head of Data Science at Evidation Health. His research focuses on linking behavioral phenotypes with health outcomes. He has 4 years of experience in working with medical claims, pharma benefits, and eligibility data. Dr. Foschini received a PhD in Computer Science from the University of California, Santa Barbara. He has published numerous papers in the broader area of computer science and is co-author of several patents in information clustering and behavior phenotyping. He has served as PI, co-PI, and senior personnel on DARPA, NSF, and NIH grants proposals.

Prof. Panconesi is a full professor of Computer Science at Sapienza, University of Rome. He obtained a PhD in Computer Science from Cornell University. In recent times he has been PI of several research projects, including faculty awards of IBM, Google and Yahoo!. He has received the ACM Danny Lewin Award for his work on distributed algorithms. He has published numerous papers in various areas of algorithms, network analysis and modelling, and has co-authored a monograph on the use of probabilistic concentration inequalities in computer science published by Cambridge University Press.

Timeline

February 1-February 15 Data management

February 15-March 31 Rising Stars identification

April 1-June 30 Modeling spatial diffusion process

July 1-August 31 Similarity identification in diffusion processes

September 1 - September 30 Consider procedure codes in addition to drugs.

October 1 - December 31 Iteratively test alternative methods and manuscript preparation.