Amplifying Domain Expertise to Combat Antimicrobial Resistance

Protiva Rahman, B.S., Arnab Nandi, Ph.D., Erinn M. Hade, Ph.D., Emily S. Patterson, Ph.D., Courtney Hebert, M.D., M.S.
The Ohio State University, Columbus, OH

Introduction

Digitization of health records has provided an avenue for applying data-driven algorithms to healthcare challenges. One such challenge is antimicrobial resistance, which the World Health Organization (WHO) classifies as a major threat alongside cancer and dementia. Antimicrobial resistant organisms or "superbugs" kill around 700,000 people annually [1]. One way to address this is by developing tools to improve antimicrobial stewardship. This involves empowering the antimicrobial stewardship program (ASP), which consists of infectious disease (ID) domain experts (e.g., microbiologists, physicians, pharmacists), to make data-driven decisions. Domain experts have different skills and requirements from other end-users (such as crowd workers and data scientists), which must be considered when designing systems. In this work, we present a framework for developing systems to amplify domain expertise and demonstrate its use in the antimicrobial research pipeline.

Methods

Amplification of expertise refers to the process of automating redundant or inferable tasks, so that experts can focus on tasks that require domain knowledge. This is a synergy between the expert and the system which involves *summarization* of data and decisions, *guidance* towards insights, *interaction* by the expert, and *acceleration* of input.

Summarization: An amplification system should summarize large and complex datasets so that experts can meaningfully consume them. This is relevant for identifying inconsistencies as well as for open-ended exploration during analysis. It can be overwhelming for an expert to go through large datasets. Amplification systems should, therefore, statistically or visually summarize complex data. In addition to data, amplification systems need to summarize algorithmic and human decisions as well. To support algorithm transparency, amplification systems can show visual activation of features that led to the recommendation, or similar cases in the data that serve as evidence for the current recommendation. Summarizing human decisions can involve expressing data transformations as natural language rules [2] or as visual node-link diagrams [3].

Guidance: While summaries provide a global view of the data, exploratory analysis aims to find insights or data quality issues, which might require looking at a more detailed view. Amplification systems should guide the expert's navigation to meaningful regions. When guiding with visual summaries, it is important to pick optimal visual encodings for revealing the relevant insight or outlier. In addition to navigating datasets, amplification systems can also guide experts by suggesting data transformations during the cleaning stage. But care must be taken to avoid bias.

Interaction: Along with making system internals explainable, allowing experts to interact and modify data and the output of algorithms increases their trust in amplification systems. The mode of interaction also needs to cater to their background and training. For example, data transformations should be presented as natural language statements as opposed to code snippets. Further, their affinity for spreadsheet tools motivate designing systems with spreadsheet interfaces but advanced querying capabilities.

Acceleration: Time constraints of domain experts necessitate the need to accelerate their input provision. Most experts use structured interfaces such as forms or free-text notes for data entry or querying and spreadsheet interfaces for data exploration. Following user-centered interface design and adhering to interactive latency constraints is even more essential for these systems. An advantage of building systems for experts is that domain specific information can be used to accelerate their input.

Results

We demonstrate the use of our amplification framework with three systems which we developed as part of our antimicrobial stewardship pipeline. The overall goal is to help physicians in empiric antibiotic prescription for urinary tract infection (UTI). When a patient presents with a UTI, their urine sample is sent to the microbiology lab, which reports the infection-causing organism as well as a subset of antibiotics that will work against it. However, since the labs can take two days to complete testing, physicians initially prescribe antibiotics based on patient demographics and medical history. This provides an opportunity to improve prescription by modeling the probability of antibiotic coverage using patient factors.

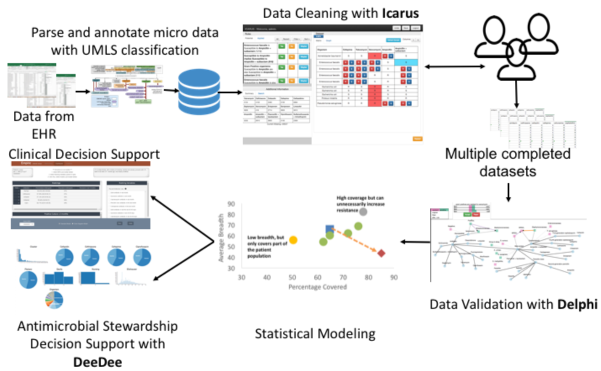

Figure 1. Amplification in the Antimicrobial Data Pipeline

At the data curation level, our domain expert has to provide the cohort definition along with variables of interest (e.g., demographics, comorbidities, allergies, etc.) to a data engineer, who pulls the relevant data from the Electronic Health Record data warehouse. The data is then annotated with organism and antibiotic classification from the UMLS metathesaurus[4]. After the data is restructured to the required format, we are faced with a data cleaning task. The microbiology lab reports susceptibilities for a subset of antibiotics based on the characteristics of the organism which grew and institutional preference. At point-of-care, physicians can infer susceptibilities on the unreported antibiotics based on the lab reports and their domain knowledge. However, when using this data for predictive modeling, we need to fill in the unreported values with domain rules. To address this, we built Icarus [2]. Icarus *guides* the expert by showing them high impact data subsets for edits. This prevents the experts from being overwhelmed by long and wide tables. Icarus allows both direct *interaction* via edits and indirect *interaction* via rules. Finally, Icarus *accelerates* task by suggesting general rules based on the expert's single edit. It also allows the expert to preview the impact of a rule by *summarizing* the cells which will be impacted.

Due to the subjective nature of this task, multiple domain experts are needed to come to consensus on the unreported values. We have three experts fill in the dataset using Icarus. To amplify the consensus process, we designed Delphi [3] which visualizes the conflicts and redundancies in expert rules. It provides an overview of the dataset by visually *summarizing* the antibiotics and related rules in a node-link diagram. The node sizes *guide* the expert to regions of high conflict by encoding the number of data points affected. It allows the experts to *interactively* edit the rule set by accepting and rejecting rules. Finally, it *accelerates* the experts' task completion by automatically removing redundant rules after each edit.

Once experts have come to consensus, the dataset is finally ready for analysis. After the model creation, the data engineer needs to build a decision support tool so that domain experts can consume the models. There are two tools: one is a point-of-care tool aimed at physicians who are prescribing antibiotics, while the second one DeeDee, is a tool aimed at the antibiotic stewardship group to validate and create guidelines. DeeDee amplifies expertise by *summarizing* population characteristics, *guiding* towards high-risk or concerning data points, allowing experts to *interactively* specify guidelines, and *accelerating* their task by suggesting guidelines.

Conclusion

Effectively engaging domain experts is crucial for the success of data-driven workflows. We provide a novel framework for developing systems which amplify domain expertise. Amplification systems should summarize data, guide experts' data navigation, allow experts to interact and update algorithms, and finally accelerate their task by learning from their interactions. Empowering stakeholders to directly interact with the data can lead to faster and more impactful insights and decision-making, which is important for democratizing data to benefit society.

References

1. Blanchard S, Mailonline. The Ganges is teeming with 'astronomical' amounts of antibiotic-resistant bacteria. Daily Mail Online 2019. https://www.dailymail.co.uk/health/article-7824481/The-Ganges-teeming-astronomical-amounts-antibiotic-resistant-bacteria.html (accessed December 24, 2019).
2. Rahman P, Hebert C, Nandi A. ICARUS: Minimizing human effort in iterative data completion. Proceedings of the VLDB Endowment. 2018 Sep 1;11(13):2263-76.
3. Rahman P, Chen J, Hebert C, et al. Exploratory Visualizations of Rules for Validation of Expert Decisions. IEEE VIS Workshop on Data Systems for Interactive Analysis. 2018.
4. Rahman P, Hebert CL, Lai AM. Parsing complex microbiology data for secondary use. AMIA 2016.

Research reported in this publication was supported by the National Institute of Allergy And Infectious Diseases of the NIH under Award Number R01AI116975.