Enabling Effective Data Interaction for Domain Experts

Protiva Rahman
Computer Science and Engineering
The Ohio State University
Columbus, Ohio
rahmanp@cse.ohio-state.edu

Courtney Hebert
Biomedical Informatics
The Ohio State University
Columbus, Ohio
courtney.hebert@osumc.edu

Arnab Nandi
Computer Science and Engineering
The Ohio State University
Columbus, Ohio
arnab@cse.ohio-state.edu

The growth of electronic health records (EHRs) and digital devices has lead to an increase in availability of data which require interaction from domain experts i.e., healthcare practitioners such as nurses and physicians. This introduces two key challenges: interface usability and expert interaction with data.

With increase in use of EHRs, all patient information has to be recorded on electronic interfaces using forms and notes. This increase has been accompanied with the growth of touch interfaces such as smartphones, smartwatches and tablets. Electronic versions of forms are often directly copied from their paper designs with multiple text fields. While writing text information on paper is quick, entering text on tablets is difficult due to the small screen size and lack of physical keyboard. This can lead to increased errors and higher input time. Instead, selection via tapping and larger user interface widgets are preferable on tablets. Keeping these in mind, we developed *Transformer* [2], a system that uses a data-driven approach to calculate the cost of human input for different widgets, e.g., radio buttons, drop downs, range sliders, text input, based on the screen size of the given device to find the optimal form layout. Our experimental evaluation show that forms redesigned with *Transformer* had a 50% improvement in user completion time.

While *Transformer* addresses the user input challenges, there is still work to be done in information loss when going from paper to digital. On paper forms, physicians are able to add annotations next to form fields to note down other relevant information [6]. While it is still possible to do this for electronic forms, this information is not propagated to the backend database. The advantage of digital forms over paper ones is that they can be stored in structured databases which can later be queried for information. Annotations that do not fit into form fields are either lost or stored in a catch-all unstructured *"other"* field and can then be overlooked at query time. For example, if a physician notes a patient's allergies next to their culture results in the absence of an allergy field, allergies are stored in the *"other"* field. This information will not be available to the next doctor who searches for all allergies. There is thus a need to automatically structure such annotations based on prior data. Further, these signals can be used to update the form design.

On the flip side, EHR data is now readily accessible for analysis, which often requires expert interpretation. While data scientists can extract outliers and trends, domain experts are needed to meaningfully interpret them. Simply showing results to the expert is not enough, as they might want to drill down and explore information iteratively on their own. Since domain experts have limited access and knowledge of computation, they require tools that allow them to meaningfully interact and communicate with the data. Further, an expert's time is limited, hence there is an emphasis in reducing their effort.

A specific example of this includes filling in unreported microbiology lab results. The unreported results are known by practitioners, hence not required at point of care, but may have to be filled in when using the data for analysis (e.g., creating syndromic antibiograms [1], modeling antibiotic coverage [5]). The missing values cannot be predicted through imputation methods since there is no evidence in the data for those values. Multiple experts are then required to fill in thousands of unreported values (10,797 in our case), which is infeasible manually. Experts can specify domain dependent rules to fill in these values, but this is a time-consuming process if they are unable to interact with the data meaningfully. To address these challenges, we developed *Icarus* [3], a system that shows the user informative subsets and uses underlying annotations [4] and database structure to generalize the users single edit to domain dependent rules. These rules are suggested to the user and not applied until the user accepts them. Experimental evaluations show that users could, on average fill 56,000 cells in just 148 edits.

Once multiple experts fill in unreported data, we want to identify areas of agreements and disagreements across experts, so that they can come to consensus. This requires them to visualize their conflicts, make edits to resolve them and then validate those decisions. While it is easy enough to compute conflicts and outliers, comprehensibly presenting them to a domain expert is not straightforward. What might be evident from looking at numbers is not necessarily apparent on visualizations, but raw data is hard to digest. Thus, finding the optimal visualization for a given dataset and then providing cues to the expert on possible insights remains an open problem we hope to address.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation under awards IIS-1422977, IIS-1527779, CAREER IIS-1453582 and by the National Institute Of Allergy And Infectious Diseases of the NIH under Award Number R01AI116975.

REFERENCES

[1] Hebert C et al. *Demonstration of the weighted-incidence syndromic combination antibiogram: an empiric prescribing decision aid.* Infection Control & Hospital Epidemiology. 2012 Apr;33(4):381.

[2] P Rahman and A Nandi. *Transformer: Rewriting Forms for Constrained Interaction.* In Submission.

[3] P Rahman, C Hebert and A Nandi. *Icarus: Minimizing Human Effort in Iterative Data Completion.* In Submission.

[4] P Rahman, CL Hebet and AM Lai. *Parsing Complex Microbioloogy Data for Secondary Purposes.* In AMIA 2016.

[5] C Hebert, E Hade, P Rahman, M Lustberg, K Stevenson, P Pancholi. *Modeling Likelihood of Coverage for Narrow Spectrum Antibiotics in Patients Hospitalized with Urinary Tract Infections.* InOpen forum infectious diseases 2017 (Vol. 4, No. Suppl 1, p. S281).

[6] A Sarcevic, Z Zhang, I Marsic, RS Burd. *Checklist as a memory externalization tool during a critical care process.* In AMIA Annual Symposium Proceedings 2016 (Vol. 2016, p. 1080).