Use of Natural Language Processing to Extract Information from Clinical Text
Summary of the Workshop

August 4, 2017

A public workshop organized by the U.S. Food and Drug Administration (FDA), the UCSF-Stanford Center of Excellence in Regulatory Science and Innovation (UCSF-Stanford CERSI), and San Francisco State University was held at the FDA White Oak Campus on June 15, 2017. The objective of the workshop was to identify current and emerging natural language processing (NLP) efforts being applied to unstructured text such as clinical notes or narratives in electronic health records (EHRs).

In this report, we briefly summarize the main discussion points and ideas presented at the workshop. This summary complements the speaker slides and recordings posted on the workshop website at www.ucsfstanfordcersi.org/nlp-workshop.

Our summary is structured around several key areas identified at the workshop. For each, we summarize the challenges that were identified, followed by some ideas presented for moving forward. These areas are:

1. NLP and other analysis algorithms (e.g., machine learning, deep learning, convolutional neural networks (CNN), text mining, rule-based systems)
2. Data
3. Output/usage issues
4. Tools and resources
5. Organization, collaboration and other non-technical issues needed to enable progress

1. NLP and other analysis algorithms

Several speakers provided their thoughts on the current state of NLP and related methods:

- While this workshop focused on NLP, it was noted by several speakers that it is only one part of the solution pipeline (sequence of software tools) – it is most often preceded by complex data acquisition and pre-processing, and followed up with some combination of machine learning or rule-based systems to produce desired decisions or interpretations. In many cases, NLP is in fact performed as a set of rules, or has been replaced with text mining and statistical methods with good success in specific areas especially when a large number of records are available.

- Classical NLP technology, designed for regular text, encounters significant challenges in processing information from clinical text in medical records such as EHRs. The challenges come from two key areas: a) NLP was not originally designed to process data in this format and with the "noise" inherent in EHRs, and b) most NLP algorithms require large gold standard databases for training ground truth data, which are hard to come by. It was also noted that for NLP to be successful in these areas, it will need to adequately handle negative statements (negations) and medical context.

- In spite of the above challenges, a literature review presented by one of the speakers showed that NLP and related methods were successful for some specific applications (e.g., radiology diagnosis, filling in certain EHRs, extraction of specific medical terms, etc.). Since each of these applications were developed separately, the issue of generalization remains.

To move forward, several possible approaches and directions were identified by one or more speakers/panelists:

- Given that NLP is only part of the analysis pipeline that may include (deep) machine learning, it is important to optimize the *whole pipeline*, from data capture to final data/decision representation.

- Leveraging new analytics methods that reduce dependence on large training databases (e.g., deep learning, CNN, text mining etc.) would be beneficial.

- Developing general solutions using a single generalizable NLP/analytics method is difficult, so it may be worthwhile to work on solving specific problems first, then analyzing commonalities among successful solutions, leading to possible generalizations.

2. Data

Some speakers identified data as being critical for the success of NLP and other analytics approaches as a source of the information to make desired decisions and actionable intelligence, as well as for training and tuning the analytics algorithms themselves. There are several critical obstacles that were noted in relation to this by one or more speakers:

- Electronic Health Records (EHRs) and other structured clinical documents were originally designed for very different purposes (patient care, billing, reporting). The clinical notes or text contained in these documents may contain additional information and context about the medical encounter but the clinical information it contains is non-standardized, has errors, typos, omissions and often miss key information necessary for envisioned FDA applications (e.g., confounders, prescriber and patient intent or behavior). This poses significant challenges for classical NLP tools designed to work on regular text. Further complicating the situation, the applications of NLP addressed at the workshop are many, and often require information and context not originally coded or missing in EHRs and related documents (e.g., confounders, temporal components, state of instrumentations).

- The issue of providing adequate training databases with ground truth (often called "gold standard") data came up as one of the key obstacles in the future advances of NLP algorithms that require such data for training. First, the establishment of these databases (including evaluation of existing data) is very costly and challenging (e.g., requiring human annotation to label data into desired classes, often inconsistent or unavailable ground truth information, and noise in the original data). A crowdsourcing approach, which has been very successful in creating large training databases (e.g., in computer vision) is not applicable here due to privacy and legal issues. Second, even if available, such databases are very difficult to share due to legal and privacy issues. As such, many stay within organizations and hence are of limited use to the wider community. FDA speakers also noted that it would be difficult to require vendors to provide such data for sharing.

- It was noted that having data from a variety of sources to complement EHRs would greatly help. The time component of data and data trends were also mentioned as important additional information.

Several possible approaches and directions were identified by one or more speakers to address and mitigate the issues outlined above:

- Increased use of new algorithms complementing NLP such as deep learning, CNN, text mining, etc. which use a large volume of data and a variety of data sources that do not necessarily require perfect labeling appears to be very promising

- Novel methods for data annotation called "distant supervision" using small "seed" data, followed by iterations to get more "golden data," also appear to be promising

- Leveraging a variety of complementary data sources (including other patient records, observations during the exam etc.) in addition to EHRs would help to provide missing information, redundancy, as well as context, all of which are important for making the correct decisions.

- Data input and collection at the source, using modern technology and better usability engineering is needed. Additionally, incorporation of smart user interfaces can improve the quality of captured data at the source, thus alleviating challenges in subsequent analysis.

- Efforts to share training databases would be beneficial (e.g., promoting sharing of data among researchers, providing the right environment for industry and vendors to share data (safe areas)).

- During panel discussion, some speakers suggested that the use of mobile devices, media and sensor technology needs to be leveraged more in medical practice, despite their limitations, to augment EHRs and to provide a variety of additional information seamlessly at a low cost (e.g., video, GPS, sensor data, voice, images).

3. Output/usage issues

Some speakers described the concerns and relevance of NLP for end-users:

- Too little attention has been devoted to understanding the types of outputs needed from NLP systems and how they can be designed to provide optimal utility for end users. Specifically, what kind of actionable intelligence the system has to provide, what kind of accuracy for which applications are needed, and what kind of representation and visualization can be provided to help the end user.

- The output of such systems would be needed to help FDA decision makers assess important actionable information.

To help in this area, several possible approaches and directions were identified by one or more speakers/ panelists:

- Issues related to the actual usage and output requirements of NLP pipeline needs to be specified in more detail in the future, following the best practices of software engineering (e.g., use cases for actual usage, requirements for functionality, accuracy, data visualization and user interfaces, etc.).

- End-users (e.g. clinicians, risk evaluators etc.) need to be kept in the loop throughout the process. This would not only make the task of NLP/machine learning easier, but could also increase the confidence of users in such systems and help when considering for example actionable regulatory decisions. This could be achieved by various means (e.g. with proper user-oriented design of system output; better user interface and visualizations; more explanation of system decisions; allowing end-users to control tradeoffs between accuracy and explainability, costs vs. benefits etc.)

4. **Tools and resources**

Some speakers articulated their thoughts on current approaches used with NLP and machine learning :

- There does not appear to be a single (or very few) best approaches so far in NLP and machine learning for clinical text

- The best chance of success is to start by solving specific problems, which in turn requires experts using available tools and quick prototyping methods.

The following ideas were suggested by one or more speakers/panelists as possibilities to help develop tools to satisfy the above needs:

- Using and leveraging best practices of open source software.

- Developing software environments in the form of interactive workbench where one can easily create and test analytic pipelines (aggregate of software tools used) by integrating available tools, as has been done for other areas.

- Provide and disseminate open source NLP and machine learning tools with adequate distribution, licensing and documentation.

- Availability of accessible, easy to share gold standard databases remains critical. The positive experience of other areas on machine learning where such databases have helped to make significant advances needs to be leveraged. This remains, however, a technical/cost issue as well as a policy and legal issue due to data privacy considerations.

5. **Organization, collaboration and other non-technical issues needed to enable progress**

Several speakers/panelists identified some non-technical issues and solutions related to use of NLP:

- Policies and rules governing data usage and sharing remain an obstacle, and ways to mitigate these issues would be beneficial.

- Forming a community of researchers, government organizations (e.g., FDA), industry and healthcare professionals with a common place to share ideas, information, tools, data and resources would be very beneficial.