AHRQ National Web Conference on Applying Advanced Analytics in Clinical Care Q&A Responses

October 14, 2020

"Artificial Intelligence and Natural Language Processing of EHR Data: Identification of Patients with Low Life Expectancy and Other Applications"

Alexander Turchin, MD, MS - Harvard Medical School Director, Brigham and Women's Hospital

QUESTION: Please clarify the limitation of ICD coding on metastatic cancer and how your natural language processing (NLP) can be used to enhance such coding for both clinical and billing purposes inasmuch as many academic medical centers use NLP to improve ICD coding.

ANSWER: For most malignancies, there are no ICD codes that indicate that the cancer is metastatic. Therefore, it is not possible to distinguish metastatic from non-metastatic malignancy on the basis of ICD codes alone. On the other hand, patient's prognosis can be markedly changed by the presence of a metastasis. Natural language processing can identify characterization of a malignancy as metastatic in narrative EHR data (e.g. provider notes, imaging reports or pathology reports) and thus improve accuracy of population-based predictive modeling for patients with malignancies.

QUESTION: The comment made by Dr. Turchin that the value of machine learning was more evident when new data were obtained was interesting. Are you talking about a transfer learning scenario?

ANSWER: What we found was that a predictive model will usually get a greater accuracy boost from an additional data source than from a better analytical method applied to existing data. One can think of it by imagining a blind person with only one ear. They will learn more about the world around them by gaining vision (a new data source) than from adding another ear (improving the existing method – hearing). Similarly, in our analysis, we found a greater accuracy boost from adding natural language processing than from changing analytical methods from logistic regression to support vector machines to neural networks.

QUESTION: By looking at frequency of lifestyle counseling, do you risk selecting patients who are most motivated and come for routine care frequently? How might you tackle this problem more broadly?

ANSWER: It is true that there is no perfect way to analyze the effect of lifestyle counseling – both clinical trials and observational studies have their drawbacks. The selection bias pointed out in this question may well be present (it is even stronger in clinical trials). However, there were further indicators of a causal relationship in our data that we didn't have time to discuss during the talk. These include a) the apparent mediation of the relationship between lifestyle counseling and clinical outcomes by HbA1c (adding HbA1c during the treatment ascertainment period into the model attenuated the relationship); and b) the dose-response relationship (longer or more diverse documentation of lifestyle counseling were also associated with a decreased risk of CV events and death). The bottom line is it seems unlikely that we will ever be able to prove this beyond reasonable doubt, but advanced analytical technologies, like NLP, allow us to get pretty close to preponderance of evidence.

QUESTION: Noted clinical teams are fond of # hashtags in notes too, instead of numbering medical conditions.

ANSWER: Agree and this is something that needs to be taken into account when designing natural language processing tools.

QUESTION: The machine learning methods are comparable in most ways, hence there wasn't a marked difference in performance using the area under the curve (AUC) values.

ANSWER: Beauty is in the eye of the beholder. Most sources I am familiar with would consider logistic regression,

support vector machines and neural networks reasonably different from each other, but of course everything is relative.

QUESTION: Does your NLP algorithm incorporate negation detection? If so what tools, or methods or approach do you use to detect negated text among clinical data?

ANSWER: The NLP algorithm for lifestyle counseling did not have significant negation detection because it is not common for providers to write in their notes “I did not talk to the patient about their diet”. The NLP algorithm for the mortality predictive model could likely benefit from negation detection – it would be a reasonable next step. Many of our other studies that included NLP did incorporate negation detection – usually similar to what was described in the NegEx tool implementation.

QUESTION: Are there any commercially available programs that any of the presenters used to create their systems?

ANSWER: We did not use any commercially available programs (apart from programming languages, such as python).

QUESTION: On average, how long does it take natural language processing and machine learning to correctly identify and predict meaningful data when introduced with a new dataset, such as orthopedics specialty to neurology specialty?

ANSWER: The time it takes to adapt natural language processing tools to a new source of data can range from a few weeks to 2-3 months, depending on the complexity of the concepts that are being identified and on the differences in language used by providers. Teaching NLP tools to identify completely new concepts will take longer.

QUESTION: Where can a beginner or a data/informatics student go to learn more about these topics that you presented?

ANSWER: All of the tools described in the talk (Canary, python, cTAKES) are publicly available and have online manuals that describe how to use them.

QUESTION: Do you include the Clinical Documentation Specialist teams at your hospitals to help educate the clinical teams on documentation?

ANSWER: Our hospitals do have Clinical Documentation Specialists – their focus is typically on compliance with healthcare insurance requirements for documentation required for payments. They do not concern themselves with how narrative EHR data could be used by researchers or analysts, whether for natural language processing or for any other reason.

“Optimal Methods for Notifying Clinicians About Epilepsy Surgery Patients”
Judith Dexheimer, PhD - Cincinnati Children’s Hospital Medical Center

QUESTION: For Dr. Dexheimer, the notion that NLP-based research can help providers figure out when best to pick up a scalpel was quite illuminating. Can you say more about how providers respond to this? What kinds of comments have they made?

ANSWER: The providers have been supportive of the entire project. They helped us from the beginning by being involved in the algorithm development. We made an effort to include their opinions every step of the way. They have mentioned liking the algorithm and two of the providers have been heavily involved in working to expand and improve the algorithm.

QUESTION: What analytical methods did you use to measure bias in your models?

ANSWER: Allow me to provide a shameless link to our paper: https://onlinelibrary.wiley.com/doi/full/10.1111/epi.16320. We used cross-validation to evaluate the model, and there are many excellent papers doing this already.

QUESTION: Do providers need coaching on use of certain words, phrases in documentation to assist in natural language

algorithm locating useful data?

ANSWER: We do not coach on the use of certain words or phrases. The general structure of provider notes does help in identifying key pieces of information (like past medical history).

QUESTION: How long did it take to build the algorithm from time of conception/idea?

ANSWER: It took 2-3 years from start to finish for the algorithm development. We did a lot of note annotation (which was time consuming) with the initial build.

QUESTION: On average, how long does it take natural language processing and machine learning to correctly identify and predict meaningful data when introduced with a new dataset, such as orthopedics specialty to neurology specialty?

ANSWER: That's a great question. It depends on a lot of factors including how similar the data sets might be, how the identifiers are structured, and how clear your outcome is. You can certainly use some off the shelf machine learning algorithms and start getting results almost immediately. It helps if your outcome is very clear (e.g. did they break their arm or not) compared to vague (e.g. is the patient presenting with sepsis). If I were planning a research study, I would plan 1-2 months to start getting preliminary data and plan an additional 4 months to tinker with the algorithm (while other things happened).

QUESTION: What technologies are used to launch finalized NLP models into production for automated running?

ANSWER: We use Oracle SQL Programming and MySQL, and a combination of Perl and Python.

QUESTION: Does your NLP algorithm incorporate negation detection? If so what tools, or methods or approach do you use to detect negated text among clinical data?

ANSWER: We do not incorporate negation detention.

QUESTION: Are there any commercially available programs that any of the presenters used to create their systems?

ANSWER: Unfortunately, in this case, all of our algorithms are home grown, but most academic projects would be happy to collaborate and share some of their code or work to perform a study at another site.

QUESTION: Where can a beginner or a data/informatics student go to learn more about these topics that you presented?

ANSWER: I asked my informatics students and they recommended the free machine learning course by Andrew Ng on Coursera (https://www.coursera.org/learn/machine-learning), and the book An Introduction to Statistical Learning by James, Witten, Hastie, and Tibshirani (https://link.springer.com/book/10.1007/978-1-4614-7138-7)

QUESTION: Do you include the Clinical Documentation Specialist teams at your hospitals to help educate the clinical teams on documentation?

ANSWER: We do not provide any documentation-specific education. Individual divisions provide some education, but we do not have a specialized role for this right now in our hospital.

"Anesthesiology Control Tower: Feedback Alerts to Supplement Treatment (ACTFAST)"
Michael S. Avidan, MBBCh FCASA - Washington University School of Medicine

QUESTION: The idea of creating an anesthesiology control tower is really cool. A few years ago, providers in the critical care units at the University of Michigan Hospital created a way to monitor equipment used in ICUs and CCUs there. There are over 30 such units. They found that there was useful info created by a machine learning approach that in their case helped prevent sepsis. I'd like to learn more about the work in your anesthesiology units.

ANSWER: The best way to learn more about the work we are doing is to follow our publications in this area.
1: Fritz BA, Abdelhack M, King CR, Chen Y, Avidan MS. Update to 'Deep-learning model for predicting 30-day postoperative mortality' (Br J Anaesth 2019; 123: 688-95). Br J Anaesth. 2020 Aug;125(2):e230-e231. doi: 10.1016/j.bja.2020.04.010. Epub 2020 May 7. PMID: 32389391.

2: Cui Z, Fritz BA, King CR, Avidan MS, Chen Y. A Factored Generalized Additive Model for Clinical Decision Support in the Operating Room. AMIA Annu Symp Proc. 2020 Mar 4;2019:343-352. PMID: 32308827; PMCID: PMC7153157.

3: King CR, Abraham J, Kannampallil TG, Fritz BA, Ben Abdallah A, Chen Y, Henrichs B, Politi M, Torres BA, Mickle A, Budelier TP, McKinnon S, Gregory S, Kheterpal S, Wildes T, Avidan MS; TECTONICS Research Group. Protocol for the Effectiveness of an Anesthesiology Control Tower System in Improving Perioperative Quality Metrics and Clinical Outcomes: the TECTONICS randomized, pragmatic trial. F1000Res. 2019 Nov 29;8:2032. doi: 10.12688/f1000research.21016.1. PMID: 32201572; PMCID: PMC7076336.

4: Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, Ben Abdallah A, King CR, Avidan MS. Deep-learning model for predicting 30-day postoperative mortality. Br J Anaesth. 2019 Nov;123(5):688-695. doi: 10.1016/j.bja.2019.07.025. Epub 2019 Sep 23. PMID: 31558311; PMCID: PMC6993109.

5: Lockhart EM, Hincker A, Klumpner TT, Hofer J, Cahill AG, Palanisamy A, Boyle W, Ginosar Y. Consultation, Surveillance, Monitoring, and Intensive Care (COSMIC): A Novel 4-Tier Program to Identify and Monitor High-Risk Obstetric Patients From the Clinic to Critical Care. Anesth Analg. 2019 Jun;128(6):1354-1360. doi: 10.1213/ANE.0000000000004141. PMID: 31094812.

6: Murray-Torres T, Casarella A, Bollini M, Wallace F, Avidan MS, Politi MC. Anesthesiology Control Tower-Feasibility Assessment to Support Translation (ACTFAST): Mixed-Methods Study of a Novel Telemedicine-Based Support System for the Operating Room. JMIR Hum Factors. 2019 Apr 23;6(2):e12155. doi: 10.2196/12155. PMID: 31012859; PMCID: PMC6658281.

7: Gregory S, Murray-Torres TM, Fritz BA, Ben Abdallah A, Helsten DL, Wildes TS, Sharma A, Avidan MS; ACTFAST Study Group. Study protocol for the Anesthesiology Control Tower-Feedback Alerts to Supplement Treatments (ACTFAST-3) trial: a pilot randomized controlled trial in intraoperative telemedicine. F1000Res. 2018 May 22;7:623. doi: 10.12688/f1000research.14897.2. PMID: 30026931; PMCID: PMC6039946.

8: Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Ben Abdallah A, Kronzer A, McKinnon SL, Budelier T, Helsten DL, Wildes TS, Sharma A, Avidan MS. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. BMJ Open. 2018 Apr 10;8(4):e020124. doi: 10.1136/bmjopen-2017-020124. PMID: 29643160; PMCID: PMC5898287.

9: Murray-Torres TM, Wallace F, Bollini M, Avidan MS, Politi MC. Anesthesiology Control Tower: Feasibility Assessment to Support Translation (ACT-FAST)-a feasibility study protocol. Pilot Feasibility Study. 2018 Jan 25;4:38. doi:

10.1186/s40814-018-0233-4. PMID: 29416871; PMCID: PMC5785885.

QUESTION: Does your NLP algorithm incorporate negation detection? If so what tools, or methods or approach do you use to detect negated text among clinical data?

ANSWER: We have not used NLP.

QUESTION: Are there any commercially available programs that any of the presenters used to create their systems?

ANSWER: Yes. We have used the program AlertWatch in our Anesthesiology Control Tower. https://www.alertwatch.com

QUESTION: On average, how long does it take natural language processing and machine learning to correctly identify and predict meaningful data when introduced with a new dataset, such as orthopedics specialty to neurology specialty?
ANSWER: We do not use NLP (yet).

QUESTION: Epic now has the capability to perform key term or key word search. Have you looked into using this for your research?

ANSWER: Not yet. Thanks for the great suggestion.

QUESTION: Any use of regex (regular expressions) in your word searching or chunking (focusing on specific parts of speech like nouns, verbs, and adverbs in particular proximity in the text?

ANSWER: We do not use NLP (yet).

QUESTION: What approaches are used to calibrate risk/mortality models and what processes implemented for continuous updating (in light of new data)?

ANSWER: Our calibration approach (observed versus expected) is shown in our manuscripts. We are in the process of refining our algorithms with new data. We are not doing this continuously because we do not yet have a continuously updating output of the relevant outcomes (e. g. death).

Fritz BA, Abdelhack M, King CR, Chen Y, Avidan MS. Update to 'Deep-learning model for predicting 30-day postoperative mortality' (Br J Anaesth 2019; 123: 688-95). Br J Anaesth. 2020 Aug;125(2):e230-e231. doi: 10.1016/j.bja.2020.04.010. Epub 2020 May 7. PMID: 32389391.

Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, Ben Abdallah A, King CR, Avidan MS. Deep-learning model for predicting 30-day postoperative mortality. Br J Anaesth. 2019 Nov;123(5):688-695. doi: 10.1016/j.bja.2019.07.025. Epub 2019 Sep 23. PMID: 31558311; PMCID: PMC6993109.

QUESTION: Where can a beginner or a data/informatics student go to learn more about these topics that you presented?
ANSWER: I would recommend one of the online courses on machine learning that are offered free by institutions like Stanford. https://www.coursera.org/learn/machine-learning

QUESTION: You mentioned potential improvement in prediction by adding structural data variables, presumably on patients. I would suggest including structural variables on the consultative physician, office location, years in practice, and perhaps doctor-patient pair.

ANSWER: Thank you for these very helpful suggestions.

QUESTION: Random Forest (RF) may work better since it robustly averages over predictors of single but general outcome variable.

ANSWER: Thanks for this interesting thought.

QUESTION: It has become more apparent that F1 scores, PPV values, or precision-recall curves be reported in analytical methods. Medical data is heavily imbalanced with respect to positive and negative cases.

ANSWER: Thank you. We agree with this perspective.

QUESTION: Do you include the Clinical Documentation Specialist teams at your hospitals to help educate the clinical teams on documentation?

ANSWER: We have not yet been doing this. Thank you for the suggestion.