

Contribution of Natural Language Processing in Predicting Rehospitalization Risk

To the Editor:

Greenwald et al¹ propose using free text in patient records to estimate hospital readmission risk. They use expert knowledge to identify 35 groups of phrases indicative of 30-day rehospitalization, and use 16 of these in logistic regression. We believe the use of natural language processing (NLP) for predicting rehospitalization is an interesting approach, and provide suggestions to improve the model.

NLP METHODS

The proposed terms are all n -grams ($n \leq 4$) and therefore a subset of simpler bag-of-words,² which can be extracted with lighter expert workload. Grouping terms to create variables can be done automatically using topic modeling.³ Taking context into account and normalizing abbreviations and word variants, as discussed by the authors, can be done using common-off-the-shelf software such as cTAKES.⁴ Graph modeling is another document representation for classification that has been shown to have good interpretability by experts.⁵

COLLINEARITY

The distortion of the coefficients in table 3 and the modest improvements over the baseline suggest that the variables may share the same information. The Pearson correlation coefficients of all variables would help determine whether this is the case.

MODEL EVALUATION

Another concern is that the proposed method is only compared with a baseline of prior hospitalizations. To evaluate the added value of the proposed variables, a stronger baseline could use all available structured data in the patient records that have been shown to have predictive value, that is, age, sex, comorbidity index.⁶ This also contributes to measuring the true effect of the proposed variables when adjusting for potential confounders.

CONCLUSIONS

The study of rehospitalization risks presents an excellent opportunity to assess the contribution of NLP to predicting important clinical outcomes. With this letter we want to encourage a more thorough evaluation of NLP methods toward this goal.

Christopher Norman, MSc*†
Thu Van Nguyen, PharmD, MPH‡§
Aurélie Névéol, PhD*

*LIMS, CNRS, Université Paris Saclay,
 Orsay, France

†Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
 ‡Equipe METHODS, Sorbonne Paris Cité
 Epidemiology and Statistics Research
 Center, University Paris Descartes,
 Paris, France
 §University of Liverpool, Liverpool, UK

On behalf of the Methods in Research on Research (MiRoR) project.

Supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 676207.

The authors declare no conflict of interest.

REFERENCES

- Greenwald JL, Cronin PR, Carballo V, et al. A novel model for predicting rehospitalization risk incorporating physical function, cognitive status, and psychosocial support using natural language processing. *Med Care*. 2017;55:261–266.
- Jurafsky D, Martin JH. *Speech and Language Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall Inc.; 2009.
- Rumshisky A, Ghassemi M, Naumann T, et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Transl Psychiatry*. 2016;6:e921.
- Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010;17:507–513.
- Luo Y, Xin Y, Hochberg E, et al. Subgraph augmented non-negative tensor factorization (SANTF) for modeling clinical narrative text. *J Am Med Inform Assoc*. 2015;22:1009–1019.
- Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA*. 2011;306:1688–1698.

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.
 ISSN: 0025-7079/17/5508-0781