Retrieving mobility information in Social Security Administration clinical records: evaluation of a neural relevance tagger

Denis Newman-Griffis, MS1,2, Jonathan Camacho Maldonado, MD1, Pei-Shu Ho, PhD1, Albert M Lai, PhD3, Eric Fosler-Lussier, PhD2
1National Institutes of Health Clinical Center, Bethesda, MD
2The Ohio State University, Columbus, OH; 3Washington University, St. Louis, MO

Introduction
The US Social Security Administration (SSA) manages federal disability programs in the United States. A challenge for SSA is identifying evidence in a claimant’s clinical records relevant to their alleged limitations. Limitations in mobility are some of the primary sources of disability in the US1. Thus, developing tools to identify information related to mobility status in clinical records could improve SSA’s disability determination process under heavy workloads. Mobility information is complex, and typically documented in free text, involving the interaction of an individual, a specific action, and a specific environment; for example, “The patient was able to walk 300’ in the clinic.” We investigate the use of a neural network model that predicts the relevance of each token in a document for mobility information in heterogeneous collections of clinical documents at SSA, as a first step in an automated pipeline.

Methods
We analyzed two SSA datasets. Collection 1 contains 304 consultative exam (CE) reports with a detailed evaluation of the disability claimant, annotated for the descriptions of mobility status. Collection 2 contains 1,200 documents, both CEs and third-party EHR notes provided to SSA, with binary annotation for whether the document contains mobility information or not. We applied our recent neural network model for token-level relevance tagging3, to run three sets of experiments. Experiment 1 evaluated the accuracy of our relevance tagger at the token level on Collection 1, as a strict measure of exactly recovering mobility information. Experiment 2 evaluated the efficacy of ranking the documents in Collection 1 by the predicted amount of mobility-relevant information, to measure the utility of our relevance tagger as a document triage tool. Experiment 3 expanded our ranking evaluation to the highly heterogeneous Collection 2, and measured our ability to rank the relevant documents higher than the non mobility-relevant records.

Results
Experiment 1. Our best relevance tagging model achieved 72.3% F-2 measure (favoring recall over precision) on identification of relevant tokens in the 304 CEs, measured via 5-fold cross validation. Our model achieves high recall of 82.3%, at the expense of a high false positive rate (precision 48.9%). Experiment 2. Comparing to ranking the 304 CEs by the true number of relevant segments, ranking our model’s predictions after a smoothing process to reduce noise achieved a Spearman’s rank correlation coefficient of 0.887, indicating very strong positive correlation. Experiment 3. Our model ranked the truly relevant documents with an average precision (AP) of 97.1%, demonstrating clear success in bringing documents with relevant information to the fore. Furthermore, qualitative review of the relevance tagging results suggests that documents with more mobility information were ranked higher.

Conclusion
We found that a neural network model for predicting word-level relevance to mobility status successfully indexed heterogeneous SSA records with high positive correlation to the true amount of mobility information in them. Our findings suggest that a simple relevance model can be useful for document triage in large-scale review for evidence of functional limitations in clinical records.

This research was supported by the Intramural Research Program of the National Institutes of Health, Clinical Research Center and through an Inter-Agency Agreement with the US Social Security Administration.

References