PLOS:
Encouraging Open Science

Jack Prominski
Pragati Shah
Dr. Phil Bourne, Advisor
Peter Alonzi & Daniel Mietchen, Technical Advisors
Public Library of Open Science, Sponsor
Bourne Laboratory, Sponsor

10.11.2017

UNIVERSITY OF VIRGINIA
DATA SCIENCE
INSTITUTE

PLOS is on the forefront of the Open Science Movement

Why Open Science is Important[1]

- Improved Reproducibility of Research
- Accelerated Scientific Discovery
- Public Enrichment
- Improved Education

UNIVERSITY OF VIRGINIA
DATA SCIENCE INSTITUTE

But, Open Science has drawbacks…

Problem Statement:

PLOS receives 25,000 article submissions per year and publishes 50% of them.[2] Readers struggle sifting through the overwhelming amount of scholarly work, and PLOS' manuscript review process is slowed by the tedious manual checks they must perform on each submission.

Goals and Metrics

Goal: *Recommend more relevant content to PLOS readers*
Metrics: *Conversion rate on "Related Reading" section, F1 score*

Goal: *Improve the efficiency of PLOS manuscript review process*
Metric: *Time savings in manuscript review*

Our Technical Approach

Recommendation System

- Collaborative Filtering
- Content-based Filtering
- Recurrent Neural Network

Automated Data Checks

- Text Mining & Regular Expressions
- Semantic Role Labeling
- Discourse Analysis

Relevant Related Work

Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breitinger. “Research Paper Recommender Systems: A Literature Survey.” International Journal on Digital Libraries (2015):1–34. doi:10.1007/s00799-015-0156-0.

Di Noia T., Ostuni V.C. (2015) Recommender Systems and Linked Open Data. In: Faber W., Paschke A. (eds) Reasoning Web. Web Logic Rules. Reasoning Web 2015. Lecture Notes in Computer Science, vol 9203. Springer, Cham

Neustein, A. (Ed.), Bellika, J., Bravo-Salgado, A., et al. (2014). *Text Mining of Web-Based Medical Content*. Berlin, Boston: De Gruyter.

Ghazinour K., Sokolova M., Matwin S. (2013) Detecting Health-Related Privacy Leaks in Social Networks Using Text Mining Tools. In: Zaïane O.R., Zilles S. (eds) Advances in Artificial Intelligence. AI 2013. Lecture Notes in Computer Science, vol 7884. Springer, Berlin, Heidelberg

M. C. Stamm and K. J. R. Liu, "Forensic detection of image manipulation using statistical intrinsic fingerprints," in *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 492-506, Sept. 2010.

Expected Results

Recommendation System

- An algorithm to recommend relevant content to readers that can be implemented on PLOS' website

Automated Data Checks

- A set of algorithms and/or methodologies that PLOS can incorporate into their manuscript review process

References

[1] "WhyOpenAccess? | PLOS." https://www.plos.org/open-access/. Accessed 29 Sep. 2017.

[2] Interview with PLOS Executives. 5 Sep. 2017.