Evidence Based Library and Information Practice

Machine Learning Offers Opportunities to Advance Library Services

Wang, Y. (2022). Using machine learning and natural language processing to analyze library chat reference transcripts. Information Technology and Libraries, 41(3). https://doi.org/10.6017/ital.v41i3.14967

Samantha J. Kaplan

Volume 19, Number 2, 2024

URI: https://id.erudit.org/iderudit/1112195ar
DOI: https://doi.org/10.18438/eblip30527

See table of contents

Publisher(s)

University of Alberta Library

ISSN

1715-720X (digital)

Explore this journal

Cite this review

Kaplan, S. (2024). Review of [Machine Learning Offers Opportunities to Advance Library Services / Wang, Y. (2022). Using machine learning and natural language processing to analyze library chat reference transcripts. Information Technology and Libraries, 41(3). https://doi.org/10.6017/ital.v41i3.14967]. *Evidence Based Library and Information Practice*, *19*(2), 142–144. https://doi.org/10.18438/eblip30527

Evidence Based Library and Information Practice

Evidence Summary

Machine Learning Offers Opportunities to Advance Library Services

A Review of:
Wang, Y. (2022). Using machine learning and natural language processing to analyze library chat reference transcripts. Information Technology and Libraries, *41*(3). https://doi.org/10.6017/ital.v41i3.14967

Reviewed by:
Samantha J. Kaplan
Research & Education Librarian, Liaison to the School of Medicine
Duke University Medical Center Library & Archives
Durham, North Carolina, United States of America
Email: samantha.kaplan@duke.edu

Received: 12 Mar. 2024 **Accepted:** 26 Apr. 2024

DOI: 10.18438/eblip30527

Abstract

Objective – The study sought to develop a model to predict if library chat questions are reference or non-reference.

Design – Supervised machine learning and natural language processing.

Setting – College of New Jersey academic library.

Subjects – 8,000 Springshare LibChat transactions collected from 2014 to 2021.

Methods – The chat logs were downloaded into Excel, cleaned, and individual questions were labelled reference or non-reference by hand. Labelled data were preprocessed to remove nonmeaningful and stop words, and reformatted to lowercase. Data were then stemmed to group words with similar meaning. The feature of question length was then added and data were transformed from text to numeric for text vectorization. Data were then divided into training and testing sets. The Python

packages Natural Language Toolkit (NLTK) and scikit-learn were used for analysis, building random forest and gradient boosting models which were evaluated via confusion matrix.

Main Results – Both models performed very well in precision, recall and accuracy, with the random forest model having better overall results than the gradient boosting model, as well as a more efficient fit time, though slightly longer prediction time.

Conclusion – High volume library chat services could benefit from utilizing machine learning to develop models that inform plugins or chat enhancements to filter chat queries quickly.

Commentary

This article was appraised using the Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML) tool developed by Al-Zaiti et al. (2022). The data set seems large enough to train the model on this task. However, there are minimal details provided about who coded the data. The ROBUST-ML tool specifically asks about criteria and procedures to label the training set for supervised machine learning, unfortunately the author does not indicate if labels were provided by the sole author or a team, if there were any inter-rater reliability issues, missing data problems, duplicate data, or discarded data. For example, the number of 8,000 chat questions is likely an approximation and one would assume that within the 8,000 questions, some had to be discarded, perhaps for not actually having a question. The author provides no detail about this, or what the ultimate number of questions was. In the data preprocessing, the author does not indicate if a chat session with multiple questions was split or abbreviated, so the unit of analysis is somewhat murky. There is also no raw or trained data available for readers to review. This is not in accordance with an important component of the ROBUST-ML tool, which asks if the experiment is reproducible.

The ROBUST-ML also asks about feature omission bias. The only feature this model utilized was question length. This single feature, along with the label of reference or non-reference, were the only components of this model. Additional features or labels that could have informed the model might have been question duration length, time of day or day of week. This information is typically available in chat logs and could have an influence on type of question asked. The author does provide sample labelled questions but does not supply a clear definition of the two categories, forcing one to wonder if reference or non-reference were overly simplistic.

Additional components of the ROBUST-ML ask about best fitting algorithm selection, bias-variance tradeoff, evaluation bias, and algorithmic bias. For the first, they indicate comparing less than three to four predictive models is an important red flag; this article only compared two and it is difficult to determine if there is enough information in the bias-variance tradeoff. The results and analysis section does not mention or assess for any other kind of bias, either.

While these reporting issues are significant, the study still has value to the field. As we are in the early stages of implementing machine learning into library work, this is one of the few studies to do so and demonstrate proof of concept. The author also had to balance reporting methods while educating the audience about these methods. Admittedly, there are possible missed opportunities – such as using the machine learning tools to analyze the transcripts for trends to inform library services and policies or to identify hot and cold times for the chat service. As the work is a pilot study, it surprisingly lacks a clear step forward, calling for the model to be used to assist in building a plugin or library chat enhancement to classify questions and forward them to the appropriate staff. It is unclear if the library staff at the institution from where the chat questions originated even desire this. However, this study is important for considering the possibility of using machine learning with existing sources of library data to optimize services.

References

Al-Zaiti, S. S., Alghwiri, A. A., Hu, X., Clermont, G., Peace, A., Macfarlane, P., & Bond, R. (2022). A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). European Heart Journal. *Digital Health*, *3*(2), 125–140. https://doi.org/10.1093/ehjdh/ztac016

Wang, Y. (2022). Using machine learning and natural language processing to analyze library chat reference transcripts. *Information Technology and Libraries*, *41*(3),https://doi.org/10.6017/ital.v41i3.14967