Towards automatic data extraction from clinical research reports: a case study of a systematic review of oral pain relief

Linh Hoang¹, Tanja Bekhuis², Jodi Schneider¹

¹ School of Information Sciences, University of Illinois at Urbana-Champaign ² Department of Biomedical Informatics, University of Pittsburgh

1. Introduction

Systematic Review:
- Systematic review is a type of literature review. In healthcare, to answer a particular clinical question, all available evidence is synthesized into a single systematic review.
- The systematic review process includes a series of steps. Data extraction from clinical research reports is one of the most time-consuming steps.

2. Related Work

RobotReviewer [Wallace et al., 2016]
- RobotReviewer is an automatic data extraction system. It uses machine learning and natural language processing to extract data from clinical research reports.
- RobotReviewer extracts 3 key elements (Participant, Intervention & Outcome) from the full-texts of the clinical research reports.

3. Research Questions and Methods

Questions:
- How does RobotReviewer’s data extraction compare to systematic reviewers’ data extraction?
- How does RobotReviewer’s data extraction compare to a single novice reviewer’s data extraction?

Methods:
An in-depth case study of a single systematic review, a Cochrane Review about oral pain relief [Bailey et al., 2013] which synthesizes 6 clinical research reports.
- Manually extract data elements from the 6 included reports.
- Run RobotReviewer on the 6 included reports.
- Compare the novice’s manual extraction and the RobotReviewer’s extraction with the published review as a gold standard.

4. Data Extraction

Data Extraction in Systematic Review:
- Extract Data: Reviewer 1
- Extract Data: Reviewer 2
- Published Review

Long-term Goal:
- Extract Data: RobotReviewer 1
- Extract Data: RobotReviewer 2
- Published Review

PROPOSED SEMI-AUTOMATION: A semi-automated system could support a single reviewer during data extraction. Differences in information extracted by a human reviewer and a computerized system could be displayed. The reviewer decides on the consensus version.

5. Results of Case Study

Compare the extraction results.

Calculate information retrieval metrics.

Precision: The percentage of data elements that are correctly identified in RobotReviewer’s extraction or the novice’s manual extraction.

Recall: The percentage of data elements that are correctly identified, comparing with the published review.

F-Measure: A weighted average of Precision and Recall.

6. Discussion

Data Extraction Results:
- Identify potential features for automating data extraction.
- Develop hypotheses about which features could be used to automate data extraction.

Case Study Results:
- RobotReviewer’s extraction results are inconsistent.
- Performance measures for the novice’s manual extraction are not as high as expected. However, they are consistent for the 6 studies included in the systematic review (P, R, F are all ranged from 0.55 – 0.8).

7. Limitations
- Small sample size.
- 3/6 articles from the same author, 4/6 articles were published in the same journal.
- The novice’s manual extraction was not completely independent: data extraction from the first paper was done after looking at the published review.
- The evaluation was based on a list of data elements which were created by the 1st author.

8. Next Steps
- Increase the sample size by examining more systematic reviews and the clinical research reports they include.
- Consider other metrics for assessing data extraction quality.
- Draw on the PICO ontology [The Cochrane Collaboration, 2014-] to update the list of data elements to be extracted.
- Conduct an error analysis and study RobotReviewer’s code to understand what works and where it goes wrong.
- Interview systematic reviewers to understand their expectations for automation.

References