Analyzing Inclusion Criteria of 7000 Cochrane Systematic Reviews

Presenters: Jingyi Xie[1]& Xiaoru Dong[1]; Advisor: Prof. Jodi Schneider[2]; Mentor: Linh Hoang[2]

1 - Department of Statistics, College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign; 2 – School of Information Sciences, University of Illinois at Urbana-Champaign

ILLINOIS
School of Information Sciences

INTRODUCTION

- We are interested in building a machine learning classifier to distinguish sentences like these:

  "Only randomized trials were included…"
  ("Only RCTs")

  "Randomized and quasi-randomized controlled trials were included…"
  ("Others")

DEFINITIONS

- **Systematic review:** A particular kind of literature review.
  - **Inclusion criteria:** Characteristics that the prospective trials must have if they are to be included in the systematic review, such as types of a trial.
  - **Randomized controlled trial (RCT):** The people participating in the trial are randomly allocated to a group receiving treatment and a control group.
  - **Cochrane**: The group conducts systematic reviews of health care interventions publishes them in the Cochrane Library.
- **Weka:** A machine learning toolkit. Weka uses different algorithms to train a **classifier** based on our annotated data.

RESULTS

	Precision	Recall
"Only RCT"	81.8%	96.3%
Others	90.7%	62.4%
Weighted Avg.	85.0%	84.0%

Results from: Weka (RandomForest)

	Precision	Recall
"Only RCT"	80.9%	78.7%
Others	64.4%	67.3%
Weighted Avg.	74.9%	74.6%

Results from: Weka (Naive Bayes Classifier)

	Precision	Recall
"Only RCT"	87.4%	91.1%
Others	83.2%	77.1%
Weighted Avg.	**85.9%**	**86.0%**

Results from: Weka (J48)

FUTURE WORK

- Finishing annotations for the whole data set (7500 reviews).
- Right now each annotator works in different portions of the data set. We want to have two annotators annotate the whole data set independently and determine internal annotation agreement after that. This would help us to access the quality of the ground truth.
- Doing in-depth analysis to understand which algorithm works best for the data and why.
- Analyzing and doing features selection (e.g. which words are informative for the classification).

METHOD

Annotate data manually: ground truth
- 1756 lines till now have been analyzed (by two people)
- First stage: two people annotate same data to check the agreement

→ More than 90% of the annotated results are the same!

Propose features and implement codes to extract features
- Write Python code to count the words
- The frequency of some word can be a feature

→ "Only randomized trials were included…" **and controlled included only quasi-randomized randomized trials were**: 0 0 1 1 0 1 1 1

→ "Randomized and quasi-randomized controlled trials were included…" **and controlled included only quasi-randomized randomized trials were**: 1 1 1 0 1 1 1 1

Use Weka to run classification
- Cross validation: 10 times
- Algorithms "RandomForest" "Naïve Bayes Classifier" and "J48" are used

Evaluate the results by comparing to the ground truth
- Values of Precision and Recall from the three algorithms are compared

→ Seen from the "Result"

ACKNOWLEDGEMENTS

Linh Hoang is funded by National Library of Medicine: "Text Mining Pipeline to Accelerate Systematic Reviews in Evidence-based Medicine" (R01LM010817).

Thank you to Cochrane for providing Cochrane reviews as machine readable XML from which inclusion criteria were extracted.

REFERENCES

Weka
https://www.cs.waikato.ac.nz/ml/weka/

ILLINOIS