A Natural Language Processing Approach to Content Analysis of CSR Survey Responses

Center for Scientific Review

Charles Dumais, M.S., CSR; Mary Ann Guadagno, Ph.D., CSR; Calvin A. Johnson, Ph.D., CIT; Daniel Russ, Ph.D., CIT

Peer Review is Fundamental to the NIH Mission

CSR Mission

To see that NIH grant applications receive fair, independent, expert, and timely reviews – free from inappropriate influences – so NIH can fund the most promising research.

The NIH Center for Scientific Review Evaluates the Majority of NIH Grant Applications

NIH will review over 81,000 grant applications in fiscal year 2015.

The Center for Scientific Review:

- Receives all NIH grant applications
- Reviews 75% of them while the other NIH Institutes and Centers review the rest
- Recruits 17,000 reviewers a year
- Holds 1,500 review meetings a year
- Manages the process with 247 Scientific Review Officers

Background on Quick Feedback for CSR - Pilot Surveys

CSR has recently conducted a series of pilot studies using short surveys with stakeholders to assess the utility of asking about their CSR study section meeting experience. Stakeholders include: NIH Scientific Review Officers (SROs), NIH Program Officers (POs), NIH Institute and Center Directors, Applicants, and Reviewers.

This poster focuses on a pilot survey conducted with one key stakeholder group - Reviewers. For 70 years, NIH has recruited groups of these external scientific experts to help evaluate grant applications from scientists across the country and abroad. CSR wants to obtain quick feedback from Reviewers about their study section meeting experience with minimum burden and maximum input for CSR senior managers and Scientific Review Officers (SROs).

Evaluating the CSR Peer Review Process

Evaluation Objectives

- To determine the extent to which current CSR best practices for peer review are optimal for achieving its' mission.
- To identify areas of success and improvement in the quality of peer review.

Primary Intended Users – senior managers and SROs at CSR and NIH who make critical decisions about the utility and efficacy of current best practices and quality of peer review.

Timeliness of the Evaluation

Why now?

- 5-years since the 2010/2012 NIH survey examining changes to the NIH peer review process –
  - utility and fairness of review using shortened applications
  - bulleted format
  - narrative expression of the overall impact statement
  - 9-point rating scale
  - use of criterion scores
- Need to gain a comprehensive and systematic view of stakeholder perceptions of the quality of peer review during a time of social and economic change as evidenced by
  - Shrinking pay lines.
  - Increased competition for research dollars.
  - Economic and personal burden of peer review to all stakeholders.

Evaluating the CSR Peer Review Process

Quick Feedback for CSR from Reviewers	**Pilot 1** - February/March 2014 **Pilot 2** - May/June 2014 **Pilot 3** - September/October 2014 **Pilot 4** - March/April 2015
Quick Feedback for CSR from Program Officers	**Pilot 1** - January 2015
Evaluation of the Peer Review Process from Stakeholders	**Focus Groups** – SROs, POs, Reviewers, Applicants **Personal Interviews** – NIH Institute and Center Directors
Other Studies	CSR Leadership Program Surveys CSR New Chair Training Surveys

Focus on Pilot 3: Quick Feedback for CSR from Reviewers

Approach

Likert Type Scales – Four agreement statements, measured from strongly agree to strongly disagree, using a 7-point Likert-type scale, and providing an open-ended text box for reviewers to share additional comments. Estimated time to completion was 5 minutes.

Quality of Prioritization The panel was able to prioritize applications according to their impact/scientific merit.						
○ 1 Strongly agree	○ 2	○ 3	○ 4 Neutral	○ 5	○ 6	○ 7 Strongly disagree
Collective Expertise The roster of reviewers was an appropriate assembly of scientific expertise for the set of applications in the meeting.						
○ 1 Strongly agree	○ 2	○ 3	○ 4 Neutral	○ 5	○ 6	○ 7 Strongly disagree
Assignment of Applications to Reviewers Assignment of applications to reviewers made appropriate use of their broad expertise.						
○ 1 Strongly agree	○ 2	○ 3	○ 4 Neutral	○ 5	○ 6	○ 7 Strongly disagree
Quality of Discussion The nature of the scientific discussions supported the ability of the panel to evaluate the applications being reviewed.						
○ 1 Strongly agree	○ 2	○ 3	○ 4 Neutral	○ 5	○ 6	○ 7 Strongly disagree

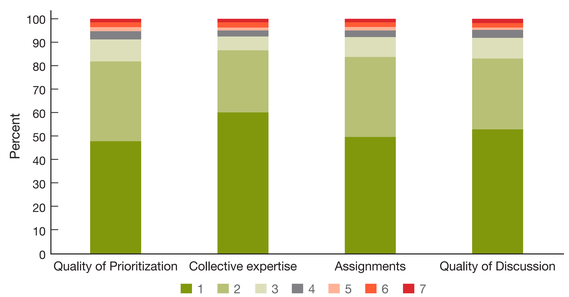
Results

2,990 Reviewers in 167 Chartered CSR Study Sections responded with an overall response rate of 64%. Response rates varied by study section from 20% to 100%.

From Likert-type scales

The overall reviewer feedback was quite favorable.

CSR Reviewer Quick Feedback Survey (January 2015)



Likert type scale where 1= very strongly agree to 7= very strongly disagree.
Overall NIH response rate = 64.2% (n=2990)

Content Analysis - From Manual Annotation to a Natural Language Processing Approach

Content analysis is a widely used qualitative research technique to interpret meaning from the content of text data. In the early stages of this effort, the number of respondents providing general comments was small enough to manually establish a taxonomy of categories and evaluate the specific polarity of each comments.

A manual Content Analysis was conducted on the first pilot study by two senior CSR staff members, a Scientific Review Officer and Program Analyst (105 comments were manually annotated representing about 40% of all reviewers who answered the survey).

10 categories of key features of the peer review process were identified as consistently mentioned by reviewers:

SRO Quality	Study Section Chair Quality	Roster Quality	Scoring	Assignments
Quality of Discussion	Meeting Management	Meeting Format	Meeting Logistics	Overall quality of Peer Review

Comments with positive (i.e., opportunities) and negative (i.e., challenges) sentiment were identified and counted within the 10 categories. A category was added to capture needs and/or suggestions for improving the peer review process. Emerging themes were interpreted based upon a summative analysis of sentiment counts and comparisons.

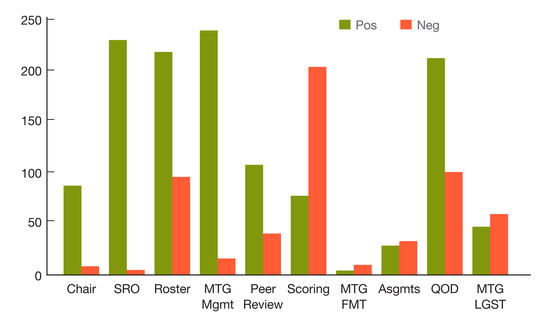
Automated Approach to Content Analysis using Natural Language Processing

In subsequent studies, as the respondent population has increased, manual content analysis has become infeasible. In order to analyze thousands of textual comments, CSR is collaborating with CIT to use innovative, computational linguistics to automate the process of capturing and categorizing stakeholder responses as well as assessing the sentiment expressed in these responses.

Initially, we trained a set of linguistic models using a set of manually annotated text fragments from the first pilot study. Pilot 1 results were used to train a set of linguistic models that were applied to Pilot 2 (211 comments representing about 82% of all reviewers who answered the survey). The results of Pilot 2 were annotated to refine the linguistic models, thereby improving their predictive accuracy and reliability in subsequent studies. When the comment text presents as a need or suggestion we have trained the software to capture these important issues as well. We are currently performing cluster-based analyses to understand the dominant needs and suggestions identified by the respondents.

A summary of Pilot 3 text analysis results (2252 comments representing 75% of all reviewers who answered the survey) suggests that reviewers are generally happy with the process, and are particularly likely to provide positive comments about the SRO, Chair, and meeting management. Reviewers were most likely to provide negative comments about scoring, the quality of discussion, and the roster, although the negative comments are still in the minority in quality of discussion and roster. The categories meeting logistics, meeting format and assignments each had few responses yet those responses were largely negative.

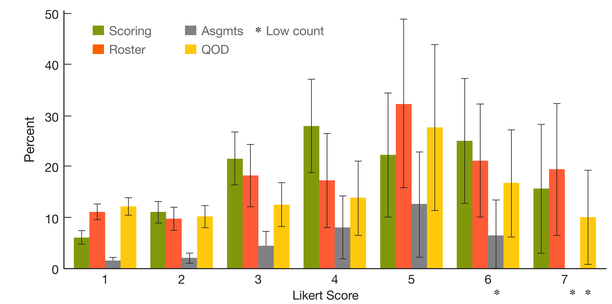
Pilot 3 Summary



Correlation of Likert Scores with Comments

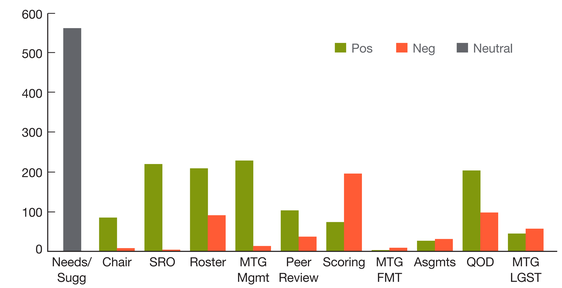
Our analysis of the Likert scores has shown that reviewers are more likely to provide text comments when they give a Likert score of four or greater (i.e., a poor Likert score). We suspect that the low-response categories are particularly affected by this behavior, namely that reviewers who are satisfied are not likely to provide a comment. Additionally, the outcome of this analysis provides guidance on which questions elicit less commentary and therefore could potentially be replaced by another more informative question.

Percent of Likert Scores with Comments



When the Needs/Suggestions category is included in the chart, the Needs/Suggestions category captures far more of the responses than the 10 key peer review process categories. There is some overlap between these categories, particularly between needs/suggestions and the 10 negative categories. Comments regarding needs and suggestions for improving peer review are not necessarily correlated with a poor Likert score, suggesting that reviewers are generally happy with the process but are interested in improving it.

Pilot 3 Summary with Needs and Suggestions



Impact for CSR

This automated approach to content analysis will allow CSR to evaluate thousands of records of stakeholder sentiment with accuracy and speed on a broad range of issues.

The resulting natural language processing model will also capture immediate stakeholder needs and suggestions, demonstrating the potential use of computational linguistics in informing NIH leadership about critical stakeholder needs in their own language, from their own perspective.

Learn More

See How Real Reviewers Work!

www.csr.nih.gov/video/video.asp

www.csr.nih.gov

References

Xiong, W., and Litman, D. S., C., 2012, "Natural Language Processing techniques for researching and improving peer feedback," Journal of Writing Research, pp. 155-176.

Friedman, C., Rindflesch, T., and Corn, M., 2013, "Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine," Journal of Biomedical Informatics, 46(5), pp. 765-773.

Nandy, B., and Sarvela, P., 1997, "Content analysis reexamined: A relevant research method for health education," American Journal of Health Behavior, 21(3), pp. 222-234.

Verspoor, K., Sanfilippo, A., Elmore, M., and MacKerrow, E., "Deploying Natural Language Processing for Social Science Analysis."

Xiong, W., Litman, D., V, A., Kay, J., and Mostow, J., 2010, "Identifying Problem Localization in Peer-Review Feedback," Intelligent Tutoring Systems, Part Ii, 6095, pp. 429-431.

We thank Huong Tran for assistance assembling materials and references for this poster.