Decision- Revise and Resubmit (28-Sep-2018)

From: rgoldsto@indiana.edu
To: matthew.andreotta@research.uwa.edu.au
CC: jonesmn@indiana.edu
Subject: Behavior Research Methods - Decision on Manuscript ID BR-BD-18-018
Body: September 28, 2018

Dear Dr. Andreotta,

I am writing in regard to the manuscript "Performing qualitative analyses on social media data sets: an application to climate change commentary on Twitter," that you co-authored with Drs. Hurlstone, Nugroho, Boschetti, Farrell, Walker, and Paris, and submitted to Behavioral Research Methods (BRM) as a part of the special issue "Beyond the Lab: Using Big Data to Discover Principles of Cognition." I had continued difficulty finding three reviewers willing to review your manuscript, and so eventually I decided to make do with only one review. So, I have supplemented this review with my own reactions that are somewhat more detailed than usual. The reviewer was impressed with the manuscript and recommends that BRM invite a revision of the manuscript. I agree with their assessment. Accordingly, I am classifying this manuscript as "revise and resubmit." In what follows, I describe the reviewers' and my major reservations, and what are the likely prospects for this work at BRM.

Given BRM's emphasis on providing tools to behavioral science researchers, any efforts to include such tools would greatly increase the positive impact of your work. Can you provide technologies in this manuscript that can be used by others? In particular, can you provide scripts/code for your harvesting and analysis of the Twitter data? Even though the NMijF algorithm has been presented elsewhere, seeing example scripts for how it is incorporated into an analysis pipeline will be instructive for BRM's readers. Code related to your topic alignment algorithm and data cleaning would also help promote future research using Twitter for behavioral sciences.

I very much appreciated your description of an effective workflow involving Twitter specifically and qualitative-quantitative analysis in general. Your work could stand as an exemplary model for how to conduct this kind of complete analysis. But in this context, I was surprised by your admission, 'Although we provide anecdotal evidence that researcher observation of corpus-level representations (topics) facilitated our thematic analysis, we did not empirically evaluate this effect.' I appreciate your honesty, but this begs the question of how automatic machine learning methods should be combined with qualitative analyses. If a social scientist were not armed with the topic analysis, would they have ended up with a different set of 5 themes? How are the 205 (or 410) topics related to the 5 identified themes? Did you end up manually assigning themes to a subset of the tweets? If so, then it would seem relatively straightforward to determine whether tweets that belong to the same topic also belong to the same theme (the opposite won't usually be true because numberTopics >> numberThemes). If far fewer topics had been used, would they have lined up better with the 5 qualitatively extracted themes? Given that one of the most striking innovations of your work is that it incorporates machine learning into a qualitative analysis pipeline, it seems important to be more precise in how the machine learning analysis actually did fit in the process of extracting themes.

At the end of the paper, I admit that I'm still not sure what was gained by the topic extraction exercise. Does the machine learning approach help us understand the structure of the discourse on climate change, or more generally, how topics change over time, or are (in)consistent over time? Ideally, you would be able to describe ways in which the revealed topic structure informed your, or an expert's, understanding, or correlated with some ground truth. If that is not forthcoming, at least some suggestion should be given for how a robust, automatically extracted topic structure could potentially be used.

The division of the corpus into smaller batches introduces many complexities that you should discuss. This binning introduces possible confounds – you don't really know whether a mismatch between topics over time is caused by actual changes to topics over time (e.g. the kinds of things on people's minds), or just noisy and irreproducible topics.

If you broke the tweets into the same number of smaller batches that were nonetheless distributed over the entire examined period of time, would the alignability of topics be similar? If the alignability were higher in this case, this would be consistent with the actual topics changing over time. Are your batches overlapping or disjoint? I'm assuming that they are disjoint, although that would probably be expected to decrease the alignability of topics. Did you consider using a topic model algorithm that allows for dynamic topics that change over time, such as the model by Blei and Lafferty (2006) and the large number of models that it inspired?

I understand that NMijF incorporates not only the lexical similarities of different tweets but also the similarities of their tweeters, but exactly how this was done was fairly opaque. Do you use the network relations between tweeters, their demographic similarities, their direct response relations, something else, or some combination of these?

Is the alignment algorithm only applied to temporally CONSECUTIVE batches? If not then it seems like there will be problems with transitivity. That is, Topic A might align with B (at a different time batch), and B with C, but A doesn't align with C. The alignment algorithm also seems rather "greedy." Is that a problem in terms of coming up with globally harmonious alignments?

You took the sound step of trying to empirically choose between the 205 and 410 topic solutions by asking subjects to rate them. This is, however, obviously labor intensive. What are the prospects for an automatic determination of the ideal number of topics? Can cross-validation be used to determine how consistent topics are?

What was the relation between the themes (or topics) you identified and social network? Like Adamic's classic work, do you find that social network proximity is a strong determinant of discussed theme similarity?
Some minor sentence-level issues to search the document for:
"existing inquires of Twitter activity are" -> "inquiries"
"This corroborates with the ephemeral and dynamic attributes of Twitter activity" – "accords" instead of "corroborates"?

The designation of "Revise and Resubmit" is often times ambiguous so let me try to be clear about what I feel are the likely prospects for this work. To me, the most important changes to make are the ones described in the second, third, fourth, and fifth paragraphs above. I ask that you also address the bulk of the additional comments made by the reviewer (particularly their suggestions for: abridging some of the repetitious discussion, adding additional information about your methods, adding results from the human rating study, adding information about the corpus, and pointing out limitations both to the external validity of Twitter and the machine learning algorithms), or describe why you have chosen not to when that is the case. Overall, the amount of change being requested is moderate but the pathway to providing these changes seems relatively straightforward. Given the strength of this research and the strong possibility that this paper could become a noteworthy example of how to wed machine learning and qualitative analyses, I hope that you do take up the challenge of preparing a revision of this work. To keep with the timeline for the special issue, I request your revision by December 1, 2018. In any case, thank you for considering BRM for this exciting line of research.

Sincerely,
Robert Goldstone, Co-editor of the "Beyond the Lab" Special Issue of Behavioral Research Methods

Reviewer(s)' Comments to Author:
Reviewer: 1

Comments to the Author
Andreotta and colleagues propose a framework for using machine learning methods to choose samples for more detailed qualitative analysis. They illustrate its use through topic modeling and thematic analysis of the discourse surrounding climate change in Australia during 2016. The chosen discourse topic is important, the methods are generally rigorous and the results interesting. The presentation of the material is clear, if somewhat verbose.
The combination of quantitative and qualitative methods would interest a broad audience. The process for determining the best number of topics is more rigorous than many studies. The application of non-negative inter-joint factorization (NMijF) is a departure from common methods in the field and more suitable for sparse data. It would be interesting to extend the time-frame of the analysis to make greater use of the model's sensitivity to temporal and user data, an attribute missing from many topic modeling studies. The use of topical alignment algorithms is also of methodological interest to behavioral scientists.

General points:
* Throughout the paper, discussion can be shortened and the authors can refrain from repeating points (e.g. the opportunities that social media provide for research). Informal investigation of the corpus is common practice and its usefulness is not an important point to belabor.
* Some claims in the text need qualifications:
1. The external validity of corpus studies on Twitter is discussed at some length. But it is important to note that the results are limited to only people who are active on social media and specifically Twitter. There is evidence that Twitter demographics are heavily biased (Mislove et al., 2011). Similar biases may exist regarding diversity of opinions, partly induced by Twitter's policies regarding acceptable material.
2. The machine learning techniques have their own shortcomings too, which aren't mentioned. For instance, NMijF ignores word order information.
3. Claims about the overall popularity of climate change discourse (page 32) can only be made in the presence of data showing what percentage of tweets are relevant to climate change in general, as well as the time-course of this popularity. Even information about the size of the corpus is missing, let alone information relevant to the mentioned claim.
* Information about the distribution of topic and word coefficients is missing too. How confident was the model on average in assigning tweets to a topic ? How different were topics in general in terms of assigned weights to high-probability words? These qualities can be important for interpreting the results and providing them as supplemental pieces of information can buttress the conclusions. Relatedly, supplemental material mentions gathering confidence ratings from human raters during phase 2. However, I couldn't find any report of average statistics for these ratings.

Minor points:
* The equation for similarity in referenced people (page 3 of the supplemental material) includes typographical mistakes.
* The specific machine learning methods used can be mentioned in the abstract and keywords to ensure researchers focused in those aspects of the methodology can find the paper
* I found the analysis generally convincing. However, providing open access to data and code can be helpful for reproduction of the results. This can be especially important given that a general methodological framework is being proposed.
* Was the side on which the two topic presentations were displayed for human rating randomized?

Reference:
Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 11(5th), 554-557.

Date Sent: 28-Sep-2018

Confirmation of resubmission (11-Dec-2018)

Preview (BR-BD-18-018)

From: journals@psychonomic.org

To: matthew.andreotta@research.uwa.edu.au

CC:

Subject: Behavior Research Methods - Manuscript ID BR-BD-18-018.R1

Body: 11-Dec-2018

Dear Mr. Andreotta:

Your manuscript entitled "Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis" has been successfully submitted online and is presently being given full consideration for publication in the Behavior Research Methods.

Your manuscript ID is BR-BD-18-018.R1.

Please mention the above manuscript ID in all future correspondence or when calling the office for questions. If there are any changes in your street address or e-mail address, please log in to Manuscript Central at https://mc.manuscriptcentral.com/brmic and edit your user information as appropriate.

You can also view the status of your manuscript at any time by checking your Author Center after logging in to https://mc.manuscriptcentral.com/brmic.

Thank you for submitting your manuscript to the Behavior Research Methods.

Sincerely,
Behavior Research Methods Editorial Office

Date Sent: 11-Dec-2018

Decision – Accepted (11-Jan-2019)

Preview (BR-BD-18-018)

From: rgoldsto@indiana.edu

To: matthew.andreotta@research.uwa.edu.au

CC: jonesmn@indiana.edu

Subject: Behavior Research Methods - Decision on Manuscript ID BR-BD-18-018.R1

Body: January 11, 2019

Dear Dr. Andreotta,

I am writing in regard to the manuscript "Analyzing social media data: A mixed-methods framework combining computational and qualitative text analysis," that you co-authored with R. Nugroho, M. Jurlstone, F. Boschetti, S. Farrell, I. Walker, and C. Paris, and submitted to Behavioral Research Methods (BRM) as a part of the special issue "Beyond the Lab: Using Big Data to Discover Principles of Cognition." Looking over your manuscript and description of your revisions, I decided not to send your manuscript out for review again. Instead, I read the revision over and am happy with the changes you made and your responses to the previous set of reviews. I am particularly grateful that you are providing the community with scripts for the topic alignment algorithm and extracting data samples for thematic analysis. I also appreciated your additional justification for selecting NMijF over other machine learning text methods, your revisions to the algorithm which produced more coherent topics, and your thoughts on how the outputs of machine learning algorithms can best be incorporated into a pipeline involving qualitative analysis. Accordingly, I am accepting this manuscript for publication in our special issue of BRM. Thank you for your efforts in revising this compelling example of combining machine learning into a qualitative analysis of rich text. I believe that it will serve as an excellent model for others pursuing similar investigations, and the specific tools that you have made available will greatly facilitate other researchers' efforts. Congratulations on this fine work.

If you think your article may be of interest to a broader scientific audience, please email Dr. Stephan Lewandowsky, stephan.lewandowsky@bristol.ac.uk , the Digital Content Editor of the Psychonomic Society. Dr. Lewandowsky may decide to do a featured blog post on your article for the society website. Furthermore, if your article might appeal to a general lay-person audience, Dr. Lewandowsky will forward it to the press team at Springer for a possible press release.

Behavior Research Methods is subscribed to by most major universities and research institutions. In addition, the journal has a variety of options to share your work to readers without subscriptions, including:

1. Open Choice: Open Choice allows you to publish open access in the majority of Springer's subscription-based journals.
2. SharedIt: Once published, you can share a link to your article that will make it available to non-subscribers for online viewing.
3. Pre-publication manuscripts may be self-archived using institutional archives or your own web site; the final, published version of your article may be self-archived one year following publication.

Behavior Research Methods is a journal of the Psychonomic Society. If you are not already a member of the society, we encourage you to join. For more details about the benefits (and low cost) of membership, please visit http://www.psychonomic.org/

Many thanks for submitting your fine work to the journal, and for all your effort throughout the revision process to produce an excellent final paper.

I look forward to seeing this paper in print.

Sincerely,
Dr. Robert Goldstone
Action Editor, Behavior Research Methods

Editor:
Michael N. Jones - Indiana University

Reviewer comments to Author (if any):

Date Sent: 11-Jan-2019