<Assignment 9 - Referee Report>

Donghyun Kang

Article: Kozlowski, Austin C., Matt Taddy, and James A. Evans, "The Geometry of Culture: Analyzing Meaning through Word Embeddings," under review, Knowledge Lab, University of Chicago, https://arxiv.org/pdf/1803.09288.pdf 2018.

In this paper, the authors attempt to introduce word embedding models - contemporary computational techniques for a large textual data- - to the community of social scientists and connect the methodological breakthrough to a sociological tradition in which the cultural meanings are explored. Their research question is mainly regarding how the word-embedding models can reveal the underlying structure of cultural meanings for the concepts such as "gender", "race", and "class." This means that they do not particularly raise "why" type of questions but it does not mean that the authors should have proposed questions with respect to causality, considering the purpose of this paper. Rather, by pointing out the limitations of other computational methods such as "semantic network analysis" and "topic modeling" and emphasizing the connection between word embedding models and sociological tradition of text analysis, I think in general the authors successfully demonstrate the usefulness of the word embedding models, which were developed by computer scientists and engineers, for the sociological/social scientific research.

The paper can be broken down into more or less three parts: 1) a review of relevant literature of studying culture, 2) a brief introduction of word embedding models, and 3) the showcases of applying word-embedding models to the study of culture. The first part includes the formal approaches of cultural meanings such as "semantic network analysis" and "topic modeling" and the theoretical connections between the word embedding models and traditional cultural theories such as structuralism, relationalism, and the intersectionality of cultural categories. I would first like to suggest that the authors need to cite more recent work from Nelson (2017) and Nelson et al. (2017) in which the methodological guidelines for applying topic models for sociological research are discussed. This is because the authors attempt to display the capacity of word embedding models by pointing out the methodological limitation of topic models as following: "both networks and topic models are ill-suited for representing the

multifarious associations and cultural valances that characterize all words in corpus…investigation into the relations between cultural categories…is also beyond the scope of prior methods." (p. 7) The authors' intention to showcase the novel computational methods is sufficiently understandable but I think it would be unfair if the authors do not mention progress made by other researchers who try to refine the limitation of any prior methods. Reversely, I think the limitations of other tools for text analysis could be more elaborated, for example in the Appendix. This does not mean that the description of other tools is seriously insufficient. However, especially given that the topic modeling has been popular for the past years, I believe presenting a few concrete examples that demonstrate what topic models are not able to do would be helpful for readers to clearly understand the limitation of prior methods, which will support the authors' claim regarding the word embedding models.

As for the theoretical connections between the word embedding models and cultural theories, I think the authors widely cover the relevant literature. As mentioned above, the theoretical connection between intersectionality and word embedding models is well discussed. And I also believe that the authors' consideration of Evaluation, Potency, Activity (EPA) model is indeed relevant in terms of situating this paper in the broader academic discussion. However, I think the authors can strength their points by presenting a variety of types of relevant theories with an organized way, rather than enumerating them. I respect their effort to advocate the word embedding models in the realm of the cultural studies but I guess that by providing a more systematic summary of the related theories, the authors can better engage in and reach the audiences whom they target. Additionally, despite the fact that the strength of high-dimensionality of the word embedding models is emphasized to argue the superiority of the word embedding model for studying the "geometry of culture", I think the fact that the authors mainly focus on the antonyms should be more clearly discussed. I expect that those who study gender relationship would argue that the gender spectrum cannot be reduced to feminine-masculine dimension and the race scholars would reject the notion of the "white-black" spectrum. Even though the authors briefly mention the possible extension as "other word pairs such as *hispanic - white* or *hispanic - black* similarly capture meaningful semantic relations." (p. 14), I think this issue should be more directly discussed in the paper, at least in the discussion section.

Overall, the tables and figures in the paper are well designed and effectively deliver information. Especially, Figure 2 does a good job of presenting the concept of projection using

word embedding models and Figure 3, 4, 5, 6, 7 also well summarize the analyses conducted by the authors. But I think the upper and the bottom section of Figure 8 should be separated if they want to communicate the result more clearly and I find it hard to read the names on Figure B2 so I would like to recommend to authors to revise the two figures. Moreover, I found citation errors in the body on page 25. There are two papers from Mikolov et al. (2013) in the reference section but the authors do not differentiate the two papers in the text. Thus, this has to be corrected before publication.

As the authors put in the discussion section, the potential of further application of word embedding models in computational sociological research appears truly enormous and I think the authors convincingly show the strength of word embedding models, in spite of some issues that I layout above. Methodologically, extending the word embedding beyond the antonym seems not only required but also promising. This is first because the meaning structure of the words does not necessarily have to comprise of direct opposite words. I expect that the applicability of the embedding models can be widely expanded if an advanced technique also can capture the nature of multi-dimensional cultural objects such as literary genres or multiple academic disciplines.

Substantially, I think the "Sacred-Profane dichotomy" proposed earlier by sociologist Durkheim (1915) can be studied with the word embedding model as the authors show that current embedding models are good at capturing the relative positions of other concepts on the binary plane. This line extension will be interesting considering that the word embedding can reveal how the representation of groups/concepts/behaviors (nations, mental illness, homosexuality, etc.) shifted through the time along the sacred-profane spectrum. And I also think that the word embedding models have an affinity with what Jeffery Alexander and Philp Smith. For example, Alexander and Smith (1993) proposed a framework of "Democratic-Counter democratic" to understand the civic discourse of the United States, and I think it can be examined with the help of word embedding models as the authors demonstrate. This will be an interesting application combining theoretical insights with the contemporary methodology.

Moreover, I think studying the nature of interdisciplinary research fields using word embedding models can be another exciting extension. A recent study conducted by Leahey et al. (2017) revealed that interdisciplinary research is a high-risk and high-risk because it can receive more citations than field-specific research but is hard to produce. They measured the level of interdisciplinarity using the predetermined categories of research field in the Web of Science and

connected the interdisciplinarity with the normalized number of citations. I think the limitation of these types of analyses is that they do not take into consideration the contents of the research articles. Thus, I think computational text analysis techniques including word embedding models introduced by the current authors can help to better conceptualize the nature of interdisciplinarity instead of relying on the given classification of subjects. But this line of extension also requires an advanced methodology that can handle multi-dimensional objects beyond the antonyms, considering that the scope of interdisciplinary collaboration can be broader than combining two research disciplines.

In sum, despite several points that were discussed above, I believe this paper inspiringly addresses the usefulness of word embedding models in terms of answering the questions that have been explored by sociologists or broader scholarly community of studying culture.

References

Alexander, Jeffrey C., and Philip Smith. "The discourse of American civil society: A new proposal for cultural studies." *Theory and society* 22, no. 2 1993: 151-207.

Durkheim, Émile. *The Elementary Forms of the Religious Life, a Study in Religious Sociology*. London : New York :G. Allen & Unwin; Macmillan, 1915.

Leahey, Erin, Christine M. Beckman, and Taryn L. Stanko. "Prominent but less productive: The impact of interdisciplinarity on scientists' research." *Administrative Science Quarterly* 62, no. 1 2017: 105-139.

Nelson, Laura K. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research*, 2017 https://doi.org/10.1177/0049124117729703.

Nelson, Laura K., Derek Burk, Marcel Knudsen, and Leslie McCall. "The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods." *Sociological Methods & Research*, 2017. https://doi.org/10.1177/0049124118769114.