Referee report with extension

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.

Kozlowski et al. (2018)'s paper seeks to introduce a novel massive text analysis method—neural-network word embedding models— to cultural sociologists. Since the focus of this paper is methodology, the authors do not define a central substantive sociological puzzle. Instead, the paper mainly addresses two interrelated methodological questions: (1) can the neural-network word embedding models contribute to decode and analyze cultural associations, categories and processes? (2) how can sociologists make use of the neural-network word embedding models to answer macro-cultural inquiries?

In terms of the first question, Kozlowski et al. (2018) argue that the neural-network word embedding models break through the limitations of previous qualitative and quantitative text analysis methods. They serve as very promising tools to be operated on large bodies of texts to analyze cultural associations and categories in a more comprehensive and nuanced manner. To back up this argument, the authors first theoretically explain how the word embedding models prevail over pervious methods and how the models in nature well suit the needs to analyze cultural associations and categories. They then empirically test the validity of the word embedding models by comparing survey results and results produced by the word embedding models. The whole procedure looks appropriate under the standard of social science research. In general, the authors quite compellingly prove their points and answer the first question, though there are a few caveats that will be discussed in the following paragraphs.

In the theoretical explanation part, the authors are doing well on putting their paper in the context of the broader literature. They comprehensively engage with literatures on both methodological and theoretical sides. Kozlowski et al. (2018) first discuss the limitations of the previous formal

text analysis methods in cultural studies. Based on this review, they then describe how the word embedding models work differently through representing “the semantic relations between words as geometric relationships between vectors in a high dimensional space.” (ibid. p.1) According to Kozlowski et al. (2018), compared with previous methods, the major advantage of the word embedding models is their ability to “distill vast collections of text into a singular representation while still preserving much of the richness and complexity of the semantic relations for systematic interpretation.” (p.48) That is to say, the word embedding models can deal with large bodies of texts that could not be handled by previous methods. Moreover, the word embedding models can more comprehensively capture the complexity, subtlety and multifarious associations demonstrated in a given corpus and the underlying linguistic and cultural system.

After engaging with the methodology literatures, Kozlowski et al. (2018) proceed to explain the potentials of the word embedding models in sociocultural studies. They first explain how dimensions derived from the word embedding models correspond to established cultural meaningful dimensions in our daily lives. Given this premise, they propose to study the patterns of cultural associations and interrelations between cultural categories through “examining the position of words arrayed upon the salient cultural dimensions of a word embedding” and “calculating the angles between the cultural dimensions.”(ibid. p.48-9) In the article, they demonstrate well how their proposed ways of using the word embedding models build upon and beyond previous empirical studies applying the models in other disciplines or for other purposes. Kozlowski et al. (2018) then explicate the structural connection between the underlying logics of word embedding models to the assumptions of a range of cultural theories. They present a quite exhaustive literature review in this section and each part is convincing, though a bit disorganized and messy. It would be better if the authors review these different cultural theories in a systematic order that implies their inner relations rather than just list all of them one by one.

In the empirical test part, it is definitely appropriate to examine the validity of the word embedding models by comparing the results produced by these models with that by other methods. But the comparison made by the authors in this paper is problematic in some aspects. To start with, it is not appropriate to compare cross sectional survey data that only reflect the

situation at a recent time point with cultural patterns aggregated from a much longer period of time through each of the 4 corpus. Besides, it is questionable to focus on binary cultural classifications, especially for the race category. In gender theories, it is acceptable to some extent to depict gender with a one dimensional spectrum ranging from femininity to masculinity. Yet the cultural categories based on race clearly have more than one dimension. Just considering the questions of where to place the Hispanics and Asians, etc. on the spectrum used by the authors, we can easily find this one dimensional race spectrum ranging from black to white unreasonable.

As for the second question, Kozlowski et al. (2018) simply answer it through demonstrating two cases, of which one traces the changing relations between words in a given culture, and the other compares the meanings of the same cultural markers in different cultural systems. The authors also talk a bit about other possible applications in the conclusion and discussion section. To me, the best answer to this type of question on how to apply certain method in certain field should be an exhaustive and critical discussion of related empirical studies and a proposal of potential development/extension based on the discussion. Therefore, personally I can't say with full certainty that their methods of answering the second question are sufficient, and that they compellingly answer the question. Yet considering that the authors are actually pioneering a new field, the way they answer the second question is perfectly fair and reasonable. For the same reason, the two cases they present serve as illustrations to methodologically inspire and guide future studies, though the cases per se also point to interesting substantive research questions. As a result, the two cases are not well elaborated as independent empirical studies in response to substantive research puzzles and hence are not well placed in the context of broader literature on the research puzzles. Of course, it seems unreasonable to add individual literature reviews for the two cases in this already 73 pages long article.

From my point view, the authors can consider dividing this paper into two articles that respectively address the two research questions identified at the beginning of my essay. The division can make their papers better qualify as journal articles with regard to length. More importantly, in the second article they will be able to add the necessary citations regarding the substantive puzzles embodied in the two cases. In this way, they can better engage with existing

social and cultural theories and studies when explaining their novel applications. As a result, they can better address their targeted audiences (the cultural sociologists) and better illustrate the contribution the word embedding models can make in terms of answering sociological questions.

The article has a few style and grammatical problems. The Tables and figures are extremely important elements of this article and are great supportive visualizations to help get the message across. Yet the fact that all of them are appended at the end of the article largely undermines the readability. I would suggest to present the figures and tables in the main body of the article as the authors discuss them. Besides, there are several grammatical errors in the article:

(1) The following sentence misses an "of" between "end" and "the": "…we find that traditionally feminine occupations such as "nurse" and "nanny" are positioned at one end the dimension.." (p. 4)

(2) The following sentence misses a "to" between "able" and "capture": "word embedding models depict relations as presented in texts produced by a given social group, and hence word embedding models are able capture precisely these visions of social space from a given perspective." (p.21)

(3) The following sentence misses a "is" between "steak" and "more": "if "steak" is significantly more masculine than "salad" in the survey, we test if steak more masculine than salad in the embedding, and then calculate the percentage of all such pairs of words that are correctly matched." (p.35)

Kozlowski et al. (2018)'s paper only present the application of the word embedding models to texts in English. In the paper, they encourage future researchers to apply the method to texts in other languages to analyze other cultural systems. Therefore, I would like to extend Kozlowski et al. (2018)'s research by performing their method on Chinese texts to examine a question inspired by one of their empirical cases. Specifically, I want to examine how Chinese people's perceptions regarding gender roles and relations shift since the establishment of the People's Republic of China in 1949. I will first train separate word embedding models on corpus

composed of news, books and other publications in China from each decade since the 1950s. Then I will delineate the broad historical change regarding gender associations in the past seventy years in China through projecting the words along the gender dimensions in each embedding.