Perspectives to Computational Analysis
Assignment 9

The paper 'The Geometry of Culture: Analyzing Meaning through Word Embeddings' (Kozlowski et al, 2018) is very clear in defining its purpose and research question. It serves primarily to introduce a particular technique of analysis- Neural-Network Word Embedding models- into the scholarship of sociology. The authors also seek to evaluate its suitability to understand nuances of macro-level semantic meaning that in the authors' own terms are "consistent with contemporary theories of identity and culture."(Kozlowski et al, 2018, p. 1). They illustrate this idea with applications to the specific cultural concepts of race, class and gender across time (the 20th Century) and geographies (Britain and the United States).

In this vein, the research question of this paper would be: "Does the method of Neural Network Word Embedding Models provide significant improvements over existing methods of text analysis in the 'sociological analysis of culture' (Kozlowski et al, 2018, p.1) across time and geographies in a manner that is both internally sound and externally applicable to real-world analyses of key concepts such as class, gender and race?"

The authors then provide diverse, meticulously organized and rigorous methods to compellingly address all the components of the above question. Since the authors are arguing for the use of a new research method, they have before them the task of establishing its benefits over existing methods, followed by its validity across diverse settings. They excel on both fronts.

For the first section, the authors put forth seven arguments: revealing both hidden biases and cultural associations (Kozlowski et al, 2018, p.16) the intersectionality (and not individual existence) of multiple cultural categories (ibid, p. 20) the structuralist insight that culture may be constructed into "dimensions based on binary opposition" (ibid, p.19), representing heterogeneous voices (ibid, p. 21), passive surveying and coverage of previous historical periods (ibid, p. 22), and high dimensionality (ibid, p. 23). Each is backed by citations and a clear elaboration of pros and cons. Collectively, they provide compelling reasons to add this tool to the arsenal of research techniques in sociology.

In the second section, the results are again very convincing. For the year of publication, the analysis is conducted on not one, but four models- Google News, Common Crawl, Wikipedia, and Google n-grams. The data for their historical analysis- Google n-grams- is publicly available and hence allows for reproducibility.

The correlation between ratings on the concept of gender across both a survey and the four proposed word embeddings model reaches impressive correlations (between 0.78-0.97). The authors also achieve high performance in other intuitively appealing empirical tests- such as demonstrating the accuracy of the gender dimension for correctly classifying the gender of popular first names. The fact that they succeed in doing so across all the decades of the 20th century adds more credence to their claims (ibid, p. 28).

Fortunately, the methods they use to arrive at these results are both appropriate and sufficient. They begin by addressing how qualitative and quantitative textual analysis -and more specifically,

word embedding models- have hitherto been harnessed in social science research, and elaborate on the inherent challenges of this approach. An entire section- 'Formal Text Analysis in the Sociology of Culture'- has been devoted to this background research (ibid, p.5). They maintain a neutral and never dismissive tone throughout.

When they move to explaining the mechanics of their model, they provide sufficient mathematical details to convey the core concepts, and direct the reader to relevant readings for further details. For example, they outline their innovation by merging terms from both linear algebra and sociology: "To identify the connotation of a word on a cultural dimension, we calculate the orthogonal projection of the normalized word vector onto the cultural dimension of interest. Because vectors are normalized, the projection of a word vector onto a "cultural dimension" is equivalent to their cosine similarity. (Kozlowski et al, 2018, p.14).

Having described their models' working, they test them on 4 models trained on diverse corpora. The uniformly high performance suggests that their success has not been an anomaly and is based on a sound theoretical footing. They also test their correlations for statistical significance. Here as well, the consistent performance continues. The authors sufficiently introduce the specific techniques used- such as bootstrapping and subsampling (ibid, p. 30). As shown in Table 3, the percentage of correctly classified statistically significant differences reveal high figures except in the case of 'Occupations'(ibid, p. 60). In spite of their successes, the researchers also demonstrate intellectual humility by highlighting (through their Principal Components Analysis) that no one dimension of the vector space represents more than a fraction (3.65%) of its total variance (ibid, p. 31).

Simply presenting their results in isolation would not make a strong case for their methods. The authors turn to a secondary source- survey data- to cross-verify their findings. Though they rely on Amazon Mechanical Turk- whose participants need not represent the general demographics of the US- they make the necessary post-stratification updates for this purpose. They also detail their survey design and remuneration schedule which allows checking for any potential biasing factors (ibid, p. 70).

In an ideal situation, the researchers would have had access to similar survey data for respondents in past decades. Moreover, these techniques do not provide any causal explanations for the associations between culturally relevant terms. However, this was never claimed, and hence, cannot be held against the authors.

The inclusion of a section on the differences in the same language across different countries (the US and Britain) helps identify subtle cultural nuances. Not stopping at their quantitative results, they also delve into historical theories on the gendered perceptions of colonization (ibid, p. 47). Though this theory was not initially enunciated in the research question, its inclusion towards the end of the paper lends the authors' method an added degree of credibility and sets the stage for future work.

They also situate their method very clearly in the midst of existing literature in terms of both the subject areas that they are choosing to explore (linguistics, anthropology, sociology, political science, etc) as well as methods (the technical dimensions in terms of computer science and machine learning). The authors also use a number of supplementary sources that are to a large extent clearly cited. These includes lists of names most likely to be evocative of race (ibid, 25).

With enough citations to fill their 7 page bibliography, the authors for the most part have been very comprehensive in their approach.

On the whole, the approach to citations appears to be evenly split between two objectives. The first is to build a broader narrative by enumerating relevant authors known in their respective domains. The second- and less frequent one- is to cite specific findings of those authors that would require a detailed reading of the cited research. While this does not necessarily imply a criticism, the arguments in the paper would prove more compelling with the latter approach.

Furthermore, the authors do, on very rare occasions, commit small errors in the inclusion and exclusion of certain sources- both on the subject area and technical fronts.

In certain cases, they have included citations whose authors may be known for their contributions in a specific field. However, the specific work does not contribute to the argument being made in this paper. For example, when speaking of the context in which words are found, they authors speak of "practice-oriented theories of language in which word meanings are always understood through usage (Austin 1962; Searle 1969)" (Kozlowski et al, 2019, p. 18). The cited papers for these writers focus on area of linguistic anthropology called Speech Act Theory. To that extent, their inclusion seems germane. However, the cited papers speak of 'performative' aspects of certain verbs and constructions, which not only 'tell', but perform a social function by their mere use (Austin, 1962). Searle's (1969) work further analyzes some of the types of speech acts outlined by Austin, such as illocutionary acts. None of this is relevant to the stated idea that the meaning of words is determined by context. In the authors' defence, they do make adequate reference to the linguist Saussure (1916), who was the authority on this aspect of word meaning.

Similarly, in the technical domain, the authors (Kozlowski et al, 2018, p.49) make another unnecessary citation. They quote Lev, Klein and Wolf (2015) when they suggest extensions of their methods to other languages or genres. However, a closer inspection reveals that this second highly technical research paper does not even touch on these issues. In fact, the word 'genre' is entirely absent from its text.

Nonetheless, some citations that should have been included are not found here. The paper does look into changes of meanings of words over time- like 'gay'(Kozlowski et al, 2018, p. 10)- but could have delved more into linguistic anthropology literature within this area. Given the centrality of 'meaning' to the author's proposed understanding of cultural context, philosophers such as Putnam (1975) could have cast light on the fundamental difficulties of ascribing a definite semantic meaning to even material objects, which becomes only more challenging in the realm of social abstraction.

There are no strictly grammatical or spelling-related errors. There are however, some minor issues with the formatting and delineation of the paper's layout. For example, in the last paragraph of the introduction, the authors mention: "The following analyses proceed in four steps. " (Koszlowski et al, 2018, p. 4). However, they proceed to enumerate only three.

More critically, in this particular version of the paper, almost all the figures are missing from the main text (with some space provided to include them). They are included separately in the index.

However, the resulting back-and-forth for the reader makes any understanding of the already complex concepts even more challenging.

All the above issues notwithstanding, the paper thus establishes the utility of their method to their stated scope of research. Having thus planted the seeds of possibility, the authors demonstrate cognizance of its potential future extensions in two areas.

Firstly, they themselves suggest that "any number of meaningful cultural categories can be interrogated using the method we outline." (Kozlowski et al, 2018, p. 37). They then provide an example through the exploration of political ideology in the American context. Similar projects could be undertaken on a number of stereotypes and extant belief systems. Not far from the scope of this paper, associations between professions and class could be tracked over the decades (especially since this correlation was found to be among the weakest). For example, terms related to technology- such as engineer (already checked for gender associations in their work) could be analyzed to help understand the trajectory of the importance of both hardware and software-related roles before arriving at the present age of data. Other sensitive connotations such as the linking of 'terrorism' or 'extremism' with 'Islam' could furnish quantitative evidence of the common perceptions writ large in English-speaking society.

Secondly, the authors also add that "analysts could use word embedding models to compare the cultural systems represented between literary genres, texts produced by different authors, or texts written in different languages" (Kozlowski et al, 2018, p. 49). Similar work could be taken up relying on the same Google N-grams source as the authors- albeit this time in other available languages such as Spanish, Russian, French and simplified Chinese, etc. Like in this paper's comparison of the UK and USA in terms of colonial perception, Spanish usage could be contrasted between Spain and its former colonies in Latin America. Furthermore, the vector scores would help quantify cultural perception *across* languages to evaluate the contention in Whorf (1956) on linguistic relativity and the representation of similar terms across cultures- a theme also touched on in this paper (Kozlowski et al, 2018 p. 5).'

Interestingly, this method may also find applications in the world of behaviour change communication in public policy and international development. For example, Africa's Voices (2017) manually coded a large volume of SMS data received as feedback from listeners to radio broadcast on health practices in Somalia. They uncovered a strong association between the local language terms for 'polio' and 'faith' (since respondents claimed that devout Muslims would not be affected by the disease). These helped provide advice to the project partner- UNICEF – on how to reposition its vaccination interventions.

As the authors acknowledge in their overview of qualitative text analysis, such coding methods-even when supported by inter-coder reliability analysis- could suffer from drawbacks when dealing with complicated or nuanced concepts (Kozlowski et al, 2018, p. 5). Much of them could be remedied through the suggested Word Embeddings approach. The only major obstacle to such an application would be the availability of a similar quantum of data as the authors had access to in English. Nonetheless, the creation of such corpora may soon be facilitated even for low-resource languages such as Somali through the use of Amazon Mechanical Turk (Callison-Burch and Dredze, 2010).

In summary, this research paper provides a cogent and compelling answer to its proposed research question, correctly harnessing a wide range of methods and data sources. Apart from only minor drawbacks in its choice of citations, it sets the stage for promising lines of inquiry using large-scale text data for the analysis of culture in sociology and beyond.

REFERENCES

Africa's Voices, 2017. 'Engaging Somali Voices- Strengthening Health Programming by Gathering and Analyzing Beliefs'. Africa's Voices, retrieved on Dec 14, 2018 from:
https://www.africasvoices.org/engaging-somali-voices/

Austin, John L. 1962. *How to Do Things with Words* . Clarendon Press.

Callison-Burch, Chris and Dredze, Mark. 2010. 'Creating speech and language data with Amazon's Mechanical Turk', CSLDAMT '10 Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, p. 1-12.
https://aclanthology.info/pdf/W/W10/W10-0701.pdf

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.

Lev, Guy, Benjamin Klein, and Lior Wolf. 2015. "In Defense of Word Embedding for Generic Text Representation." In *Natural Language Processing and Information Systems* , 35–50. Springer International Publishing

Putnam, Hilary. 1975. The Meaning of Meaning. In *Mind, Language and Reality: Philosophical Papers*, *vol. 2*. Cambridge University Press.

Saussure, Ferdinand de. 1916. *Course in General Linguistics* . Columbia University Press**.**

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language* . Cambridge University Press.

.