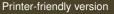SOIL Open Access Discussions EGU

Interactive comment on "Word embeddings for application in geosciences: development, evaluation and examples of soil-related concepts" *by* José Padarian and Ignacio Fuentes

Kristof Van Oost

kristof.vanoost@uclouvain.be

Received and published: 22 February 2019

This study discusses the potential of word embeddings to include descriptive information in geosciences. Although I found this an interesting approach, that could be particularly valuable for soil science, I struggled somewhat to see how this manuscript exactly demonstrates its potential. The conclusion that a domain-specific embeddings outperforms a general domain embeddings has been reported before. The development of the domain-specific GeoVec embedding is an interesting starting point/tool for new research and the linguistic relations captured by the model (as shown in figure 5-7) make sense. However, in my opinion, the study could have a much bigger im-

pact if it could describe and demonstrate how we can valorize the large amounts of descriptive information into quantitative soil science. The authors make some valuable suggestions in the 'Future work' section that could exactly do this.

Other comments: Section 3.2. Is the approach robust? It could be interesting to see how these specific pre-processing steps influence the model performance and relations. Figure 3: a relative scale is used. Can this be quantified? Section 5.2: why is there such a big difference in performance stabilization between geosciences and biomedical sciences? What does the color code represent in figure 7¿

———————————————————

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2018-44, 2019.