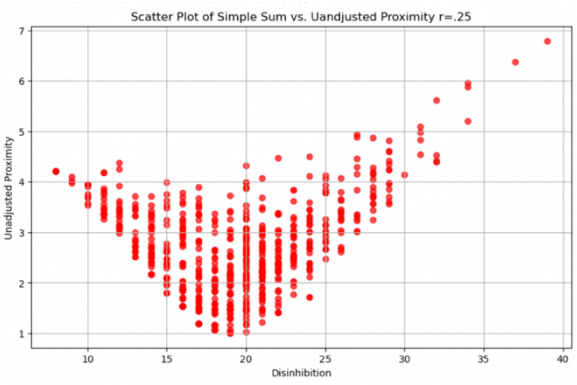
Convex Hull Applications to Likert Scale Psychometrics

Nigel Guenole

Convex hulls were proposed for natural language psychometrics by Guenole et al. (2024). They suggested viewing embeddings as coordinates, forming convex hulls, and interpreting item distance from the centroid as discrimination and free response distance as a person score. This by itself requires that the centroid represents an extremity. If it does not, how do we recover item discrimination and trait scores?

The challenge to resolving this in an artificial intelligence (AI) context is that it is not straight-forward to interpret 'high' on a transformer embedding, aside from reading the text. Scoring might not differentiate text equidistant in different directions from the convex hull centroid (*n.b.* this only matters in the unsupervised mode). Here we switch to a real data Likert scale context where it is clear what low and high mean (i.e., all 1s or 5s).

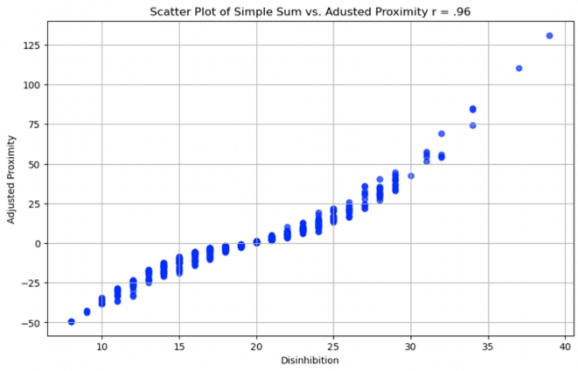
Figure 1. Unadjusted proximities (r=.25)



Proposal. Define a referent vector for low on the trait (i.e., all 1s). Calculate the distance between the real response vector and the referent vector. Calculate the dot product between the extreme vector and the difference vector. Multiply the proximity by the dot product. Check the correlation with familiar CTT sums.

The proximity to the construct centroid now incorporates a directional component indicating whether the response is moving toward or away from the extreme vector and a weight reflecting the degree to which the difference vector direction aligns with the extreme vector.

Figure 2. Adjusted proximities (r=.96)



Method & Results. Following these procedures with 671 human responses to the eight item G50 disinhibition scale shows the centroid is not extreme [3, 2, 2, 3, 2, 2, 2, 3]. Without the adjustment, the correlation between the proximities and the simple sum of all items is ***.25***, with the adjusted proximities, it is ***.96***. We recover the real sum scores.

Let's return attention to the natural language and A.I embedding context where low and high on an embedding means is less clear. We propose estimating centroid proximities and reading the extreme text to identify low and high poles and setting the referent accordingly.

Conclusion. We proposed a solution to differentiate vectors equidistant in different directions from a convex hull centroid in a context where 'extreme' has a clear interpretation. We will now return to apply this to AI embeddings. We also showed convex hulls have applications to rating scale psychometrics.

References

Guenole, N., Samo, A., Sun, T., & D'Urso, E. D. Convex Hull Applications to Natural Language Psychometrics. OSF. https://osf.io/t97hx