

On the unreasonable shallowness of natural language

AKA Adding up word embeddings works far too well

Lyndon White

School of Electrical, Electronic and Computer Engineering
The University of Western Australia

We like to think language is very complicated

This is what we do, and complicated models
make us feel good and publish well.

and sometimes it is

And so we need the complicated models.

but sometimes it isn't

And so simple methods obviously work

and often it looks like it is complicated but
isn't

and so using complicated methods is a
mistake leading to worse performance.

SkipGram etc. are basically iterative algorithms for weighted collocation matrix factorization.

When trying to factorize very large matrices numerical linear algebraists often use iterative methods.