

The data set is constructed by the following contents:

*“orange juice
apple juice
eat apple
eat orange
drink juice
drink water” * 100*

*“save water” * 1*

	orange	apple	juice	eat	drink	water	save
orange	0	0	1	1	0	0	0
apple	0	0	1	1	0	0	0
juice	1	1	0	0	1	0	0
eat	1	1	0	0	0	0	0
drink	0	0	1	0	0	1	0
water	0	0	0	0	1	0	1
save	0	0	0	0	0	1	0

Thus, we make the word “save” as an rare word, while all others represent frequent words.

I hope to find the reason why rare words get poor embeddings. I conduct several experiments using the same initialisation while only the different iterations.

As you can see, there are only 7 words, and I use 2-dimension embeddings for the sake of convenience sparing the effort to use PCA or t-SNE in order to visualisation.

We can expect that rare word “save” embedding is closest to word “drink”, for they share the common context “water”, but they may not completely overlap to the extent of “apple” and “orange” who share the exactly same contexts (“drink” is constrained by the context word “juice”).

After 100 iterations, the whole picture runs well and is ready to be used except the word “save” which rarely moves, but after enough iterations, like 1000 or more, the “save” embedding is getting well. Thus, the main problem of rare word embedding is lack of moving.















