paradigmatic task with or without the *w+c* option (compare the solid lines). In fact, the performance in the paradigmatic task was slightly enhanced too. Putting this together with what we saw above regarding SGNS performance in the syntagmatic task brings us to an interesting conclusion about the “optimal parameter setting” for this model: using the *w+c* option is a good choice adding to the robustness of SGNS, particularly when unsure of which type of similarity inference we would like the model to perform at the end. The SVD model, on the other hand, does not show the capability to learn both tasks at the same time; it gets better in one at the expense of the other. In the next section we try to explain this difference by looking into the way the two models distribute words within the high dimensional vector space.

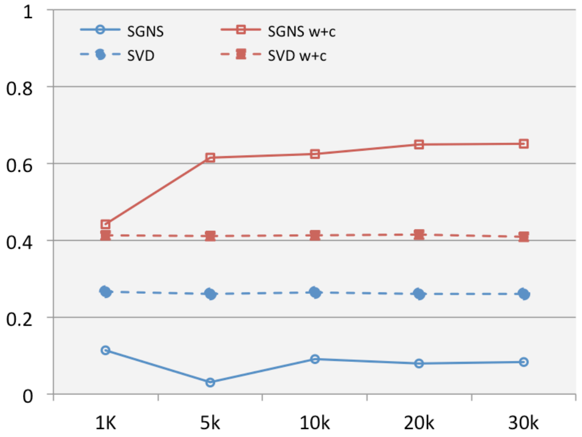

Figure 1. Accuracy of SGNS and SVD with word only vs. word+context vectors trained on corpuses of different sizes (1K to 30K sentences) in the syntagmatic task.

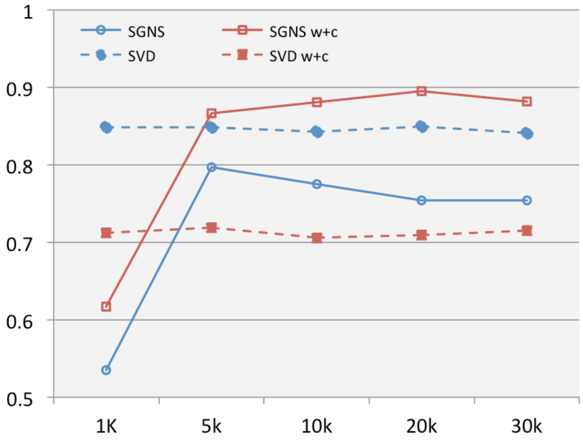

Figure 2. Accuracy of SGNS and SVD with word only vs. word+context vectors trained on corpuses of different sizes (1K to 30K sentences) in the paradigmatic task.

### 3.4 Metric Space Expansion/Compression

The above experiments showed a lower ceiling for SVD performance compared to SGNS in both tasks when sufficient data was available to the models and the parameter space was thoroughly explored. In order to explain this observation, we took a closer look at the vectors generated by each model and specifically examined the range of the similarity scores of all word pairs in the vocabulary. We found that SVD generated numerically closer vectors compared to SGNS. This results in a smaller range of similarity scores: totally interchangeable words, such as *man* and *woman* get a cosine similarity score close to 1.0; completely different words (that neither appear in a sentence together, nor share similar contexts) such as *glass* and *chase* get a negative similarity score typically close to 0.0, or around -0.5 in a best case scenario.

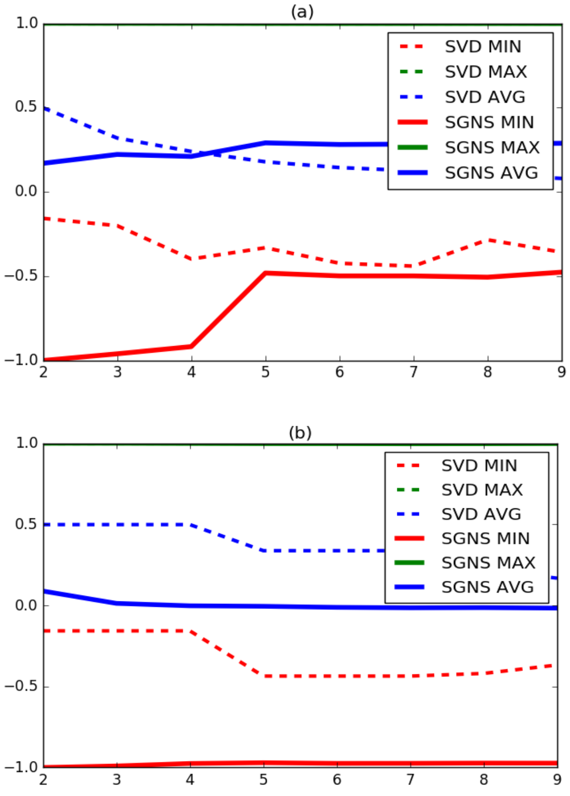

Figure 3. Spectrum of similarity scores between words in SVD and SGNS (10K corpus, neg = 1, eig = 0, dim = 2 to 9 on the x-axis): (a) with *w* and, (b) with *w+c* post-processing.

Figure 3 depicts the minimum, maximum and average similarity scores obtained for all word pairs from the vocabulary through repeated experiments on a 10K corpus by manipulating the dimensionality (x-axis). It is almost the same for SGNS and SVD when the word-only post-