Assignment #3

CSE 447: Natural Language Processing
Eric Boris: 1976637
Collaborators: Allyson Ely, Ardi Madadi, Rahul Ram

Word Vector Embeddings

3.1 Results of using cosine similarity to find the most similar word to the given word and the cosine similarity between the two words.

Given Word	Most Similar Word	Cosine Similarity
dog	cat	0.9218
whale	whales	0.8987
before	after	0.9512
however	although	0.9801
fabricate	fabricating	7595

3.2 Results of using vector addition to compute the top most likely candidates (c) to complete the given analogies and the cosine similarities (cos sim) for each.

Given Analogy	(c1, cos sim)	(c2, cos sim)	(c3, cos sim)
dog : puppy :: cat :	puppies, 0.7629	scaredy, 0.7438	kitten, 0.7406
speak : speaker :: sing :	sang, 0.6226	nateq, 0.6217	lyricist, 0.6019
France : French :: England :	scottish, 0.8679	english, 0.8374	welsh, 0.8057
France : wine: :: England :	orchard, 0.6624	tasting, 0.6325	tea, 0.6155

3.3-3.5 Results: Training a text sentiment classifier using a GRU-based RNN model. The model uses an UNK threshold of 5, batch size of 128, 8 epochs, 256 hidden dimensions, and 2 hidden layers. Results from the model without fine tuning:

Dataset	Accuracy (%)	F1 Score	Runtime (secs)
Training	0.829	0.8288	298
Test	0.8186	0.8174	11

Results from the model with fine tuning:

Dataset	Accuracy (%)	F1 Score	Runtime (secs)
Training	0.823	0.8929	242
Test	0.8935	0.8935	11