

CS6910 Fundamentals of Deep Learning

Assignment-3 Report

ME18B182 Vasudev Gupta

January 7, 2021

Contents

1	Training Word2Vec model	2
1.1	Cosine Similarity	2
2	Sentiment Analysis in movie reviews	2
2.1	Impact of variation in Embedding dimension	3
2.1.1	Accuracy graphs	3
2.1.2	Classification Accuracy	3
2.1.3	Confusion Matrix	4
2.2	Impact of variation in window sizes	5
2.2.1	Accuracy graphs	5
2.2.2	Classification Accuracy	5
2.2.3	Confusion Matrix	6
3	Conclusion	6
4	Appendix	6

1 Training Word2Vec model

Continuous bag of words (cbow), skip-gram, LSTM based models are trained to map words into certain dimension and generate their feature representations. Different embedding dimensions and window sizes are used for word2vec training. Please find link to weights at the end of the assignment.

1.1 Cosine Similarity

Cosine similarity is included in below table for some of the word-pairs.

Table				
Word-pairs	glove-300d	skip-gram-300d	cbow-300d	lstm-based-300d
king, emperor	0.483	0.542	0.396	0.0372
teacher, student	0.687	0.393	0.362	0.126
sleep, night	0.448	0.394	0.215	-0.0838
phone, mobile	0.648	0.646	0.342	0.046
computer, desktop	0.545	0.460	0.276	-0.016
husband, wife	0.864	0.624	0.506	0.113
lock, key	0.254	0.425	0.195	0.179
gift, water	0.0746	-0.0005	0.092	-0.0536
stair, chair	0.235	-0.1103	0.088	0.1557

Embedding obtained from skip-gram performed very similar to glove embedding for most of the pairs. Embedding from continuous bag of words performed slightly worse but still its considerable. But lstm-based embedding can't be considered good.

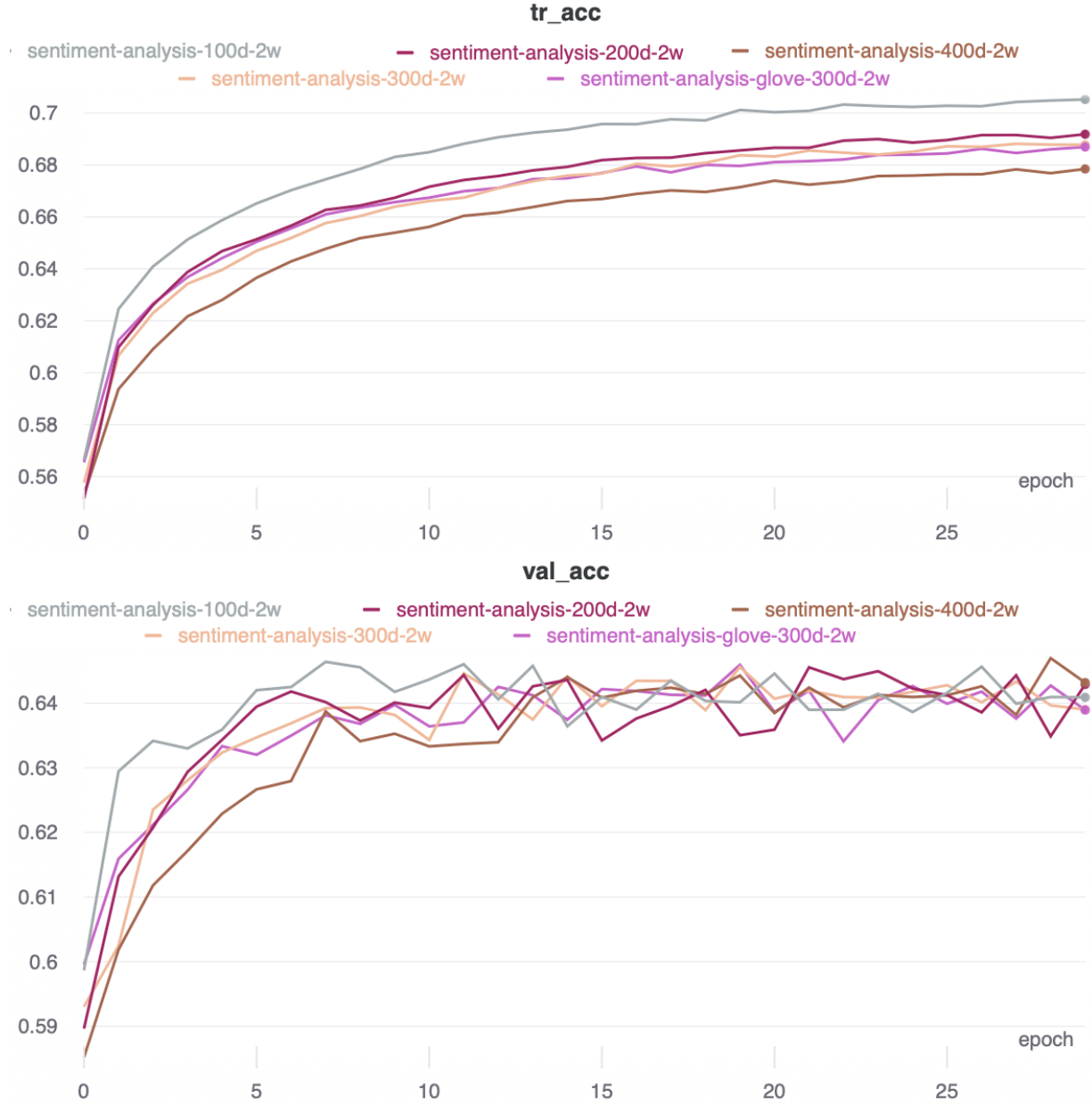
2 Sentiment Analysis in movie reviews

Embedding obtained in part-A are used to initialize classification model trained in this part. All models are trained to predict among 5 classes for around 30 epochs. Early stopping was implemented and only the best epoch is made available in 'weights' directory in part-B.

Results of different configurations are reported in following sections. **Note: Only best epoch for a particular model are reported.**

2.1 Impact of variation in Embedding dimension

2.1.1 Accuracy graphs



Above figure shows effect of embedding dimension in classification accuracy on training and validation dataset.

2.1.2 Classification Accuracy

Table-1	
Embedding dimension	Test Accuracy
100	64.6
200	64.55
300	64.56
400	64.7

Results are very similar in all the cases. The main reason can be embedding with bigger dimension is

capturing redundant information only and this classification task doesn't need that information and hence performance is not improving when embedding dimension is increased.

2.1.3 Confusion Matrix

	negative	somewhat negative	neutral	somewhat positive	positive
negative	314	575	158	14	0
somewhat negative	213	2029	1691	152	6
neutral	48	1076	9734	1049	30
somewhat positive	7	164	1953	2544	271
positive	0	13	149	773	446

Above shown is confusion matrix for the classification model trained with glove embedding of 300 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	292	552	203	12	2
somewhat negative	194	1774	1941	173	9
neutral	50	821	9863	1181	22
somewhat positive	7	99	1857	2720	256
positive	0	8	137	790	446

Above shown is confusion matrix for the classification model trained with skip-gram embedding of 100 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	241	643	156	19	2
somewhat negative	139	2008	1770	167	7
neutral	29	1013	9701	1165	29
somewhat positive	5	160	1837	2665	272
positive	2	13	129	805	432

Above shown is confusion matrix for the classification model trained with skip-gram embedding of 200 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	314	532	202	11	2
somewhat negative	232	1759	1937	154	9
neutral	45	825	10020	1011	36
somewhat positive	8	125	1965	2504	337
positive	0	10	155	723	493

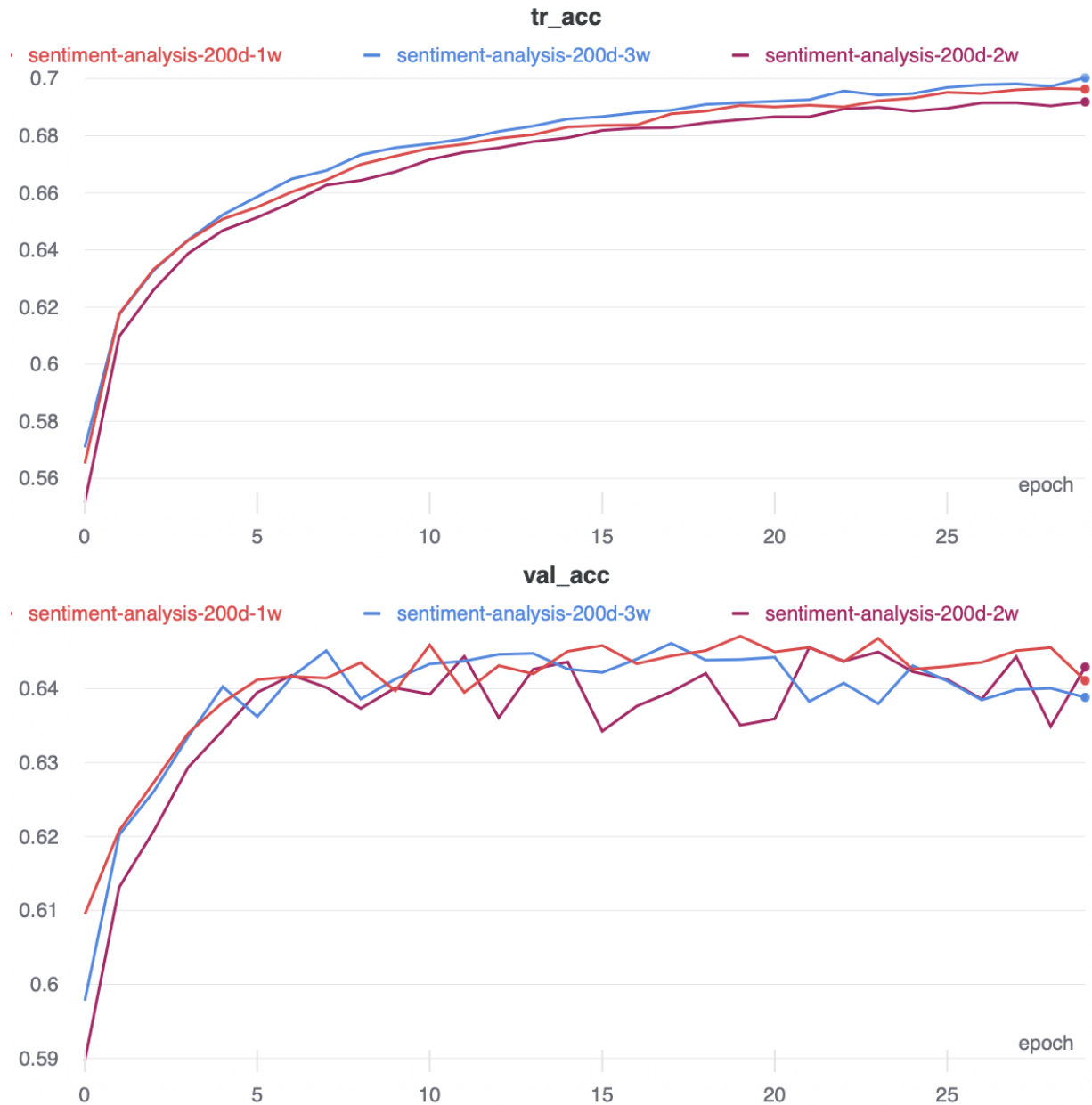
Above shown is confusion matrix for the classification model trained with skip-gram embedding of 300 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	360	498	192	10	1
somewhat negative	252	1942	1756	134	7
neutral	56	992	10073	782	34
somewhat positive	5	159	2186	2276	313
positive	0	15	176	708	482

Above shown is confusion matrix for the classification model trained with skip-gram embedding of 400 dimension

2.2 Impact of variation in window sizes

2.2.1 Accuracy graphs



Above figure shows effect of window size in classification accuracy on training and validation dataset.

2.2.2 Classification Accuracy

Table-2	
Window size	Test Accuracy
1	64.71
2	64.55
3	64.61

Results for window size more than 4 are not included because it was increasing training data for skip gram by significant amount. And it was impossible to train on such a big data on colab-gpu.

2.2.3 Confusion Matrix

	negative	somewhat negative	neutral	somewhat positive	positive
negative	326	545	165	24	1
somewhat negative	224	1937	1728	197	5
neutral	57	957	9734	1138	51
somewhat positive	9	152	1772	2634	372
positive	1	12	121	714	533

Above shown is confusion matrix for the classification model trained on dataset of window-size 1 with skip-gram embedding of 200 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	241	643	156	19	2
somewhat negative	139	2008	1770	167	7
neutral	29	1013	9701	1165	29
somewhat positive	5	160	1837	2665	272
positive	2	13	129	805	432

Above shown is confusion matrix for the classification model trained on dataset of window-size 2 with skip-gram embedding of 200 dimension

	negative	somewhat negative	neutral	somewhat positive	positive
negative	293	566	190	11	1
somewhat negative	197	1888	1884	119	3
neutral	39	898	10081	905	14
somewhat positive	6	153	2120	2480	180
positive	0	9	180	845	347

Above shown is confusion matrix for the classification model trained on dataset of window-size 3 with skip-gram embedding of 200 dimension

3 Conclusion

Token Embedding capture the meaning of words and it shows that model can understand the language by mapping it into list of numbers. This shows scope of deep learning based models into language understanding and many more language related tasks. Confusion matrices clearly shows that model is not much confused between completely opposite categories and hence its learning something. It's getting more confused among adjacent categories like positive somewhat positive or negative somewhat negative.

4 Appendix

Since weights size was big, I am adding them to my gdrive. Only best epochs are included in following link. Please get back to me incase you are unable to access the weights.

Link to weights: <https://drive.google.com/drive/folders/1MSBM0wMZDkHoma8nhPhK62scMems17q7?usp=sharing>