Assignment 3: Word Vectorization Report

Aadya Ranjan

Results Overview:

Skip-gram does a better job than SVD at distinguishing between different words, leading to more accurate classifications. In terms of accuracy and F1 score, Skip-gram consistently beats SVD across various window sizes. When looking at confusion matrices, Skip-gram consistently performs better at correctly classifying different instances.

Hyperparameter Testing:

In the experiment, three window sizes was tested (2, 4, and 6) to analyse how they influenced the performance of both SVD and Skip-gram models. The window size refers to the number of surrounding words considered during training. By varying this parameter, it was aimed to understand how different levels of contextual information affected the models' ability to generate accurate word embeddings. This investigation helps determine the optimal window size for each model, improving their effectiveness in tasks such as classification or language understanding.

Window Size Impact:

The window size has a significant impact on how these models understand the meaning of words. A smaller window size means that the model only considers nearby words, which might limit its ability to capture broader context and semantic relationships. Conversely, a larger window size allows the model to take into account more words around the target word, providing richer context and potentially capturing more nuanced semantic connections.

So, when we conduct experiments to test different window sizes, we're essentially trying to understand how changing this parameter affects the performance of SVD and Skip-gram.

For SVD, increasing the window size from 2 to 6 improves its performance in various measures like accuracy and recall. This means SVD benefits from looking at more words around the target word.

Similarly, Skip-gram does better with larger window sizes. With a window size of 6, it performs the best, showing that it benefits from considering more words for understanding context.

Best Configuration:

Skip-gram tends to capture more semantic relationships between words. It achieves this by learning word embeddings that represent semantic similarities between words based on their contexts in a corpus. In contrast, SVD relies primarily on co-occurrence statistics, which might not capture as much semantic richness.

Among the setups tested, Skip-gram with a window size of 6 gives the best results overall. It has the highest accuracy, F1 score, precision, and recall on both training and testing data. This means using a wider context helps Skip-gram understand words better for classification tasks.

Possible Shortcomings :

SVD Limitations:

SVD heavily depends on co-occurrence statistics, which may not fully capture semantic relationships between words. SVD's matrix factorization approach

may lead to the loss of contextual information, especially in complex linguistic contexts and Scalability Issues.

Skip-gram Limitations:

Training Skip-gram models can be computationally intensive, particularly with large datasets and complex neural network architectures. Skip-gram may struggle with rare words or phrases lacking sufficient context in the training data, potentially leading to less accurate word embeddings for such terms.

Conclusion:

In conclusion, Skip-gram is more effective than SVD for classifying words in tasks like these. It performs consistently better across different measures, especially when given more context to work with. However, it's important to consider factors like how much data you have and how much computing power you can use when choosing between these methods.

1 .SVD

1st window size = 1

Train Metrics:
Train Accuracy: 78.18 %
Train Precision: 0.7856
Train Recall: 0.7837
Train F1 Score: 0.7818

Confusion Matrix (Train):
[[4323 317 263 309]

[51 4866 10 55]
[219 47 3938 728]
[149 41 241 4577]]

Test Metrics:
Test Accuracy: 74.54 %
Test Precision: 0.7524
Test Recall: 0.7445
Test F1 Score: 0.7451

Confusion Matrix (Test):
[[1389 90 137 208]
[103 1627 30 66]
[102 46 1274 395]
[79 83 208 1515]]

2nd Window size = 3

Train Metrics:
Train Accuracy: 80.39%
Train Precision: 0.8048
Train Recall: 0.7939
Train F1 Score: 0.7955

Confusion Matrix (Train):
[[4556 241 282 127]
[347 4514 11 60]
[196 36 3883 608]
[231 52 271 4490]]

Test Metrics:
Test Accuracy: 76.36 %
Test Precision: 0.7648
Test Recall: 0.7706
Test F1 Score: 0.7718

Confusion Matrix (Test):
[[1491 62 139 84]
 [193 1487 39 74]
 [142 22 1373 402]
 [138 35 149 1528]]

3rd Window size = 5

Train Metrics:
Train Accuracy: 85.80 %
Train Precision: 0.8519
Train Recall: 0.8580
Train F1 Score: 0.8587

Confusion Matrix (Train):
[[4533 240 255 242]
 [68 5829 21 32]
 [137 26 3478 449]
 [145 18 136 3896]]

Test Metrics:
Test Accuracy: 81.12 %
Test Precision: 0.8147
Test Recall: 0.8112
Test F1 Score: 0.8117

Confusion Matrix (Test):

[[1390 66 123 201]
 [98 1630 60 112]
 [110 45 1430 315]
 [84 76 253 1487]]

2. Skip-gram

1st Window size = 1

Train Metrics:
Train Accuracy: 96.29 %
Train Precision: 0.9631
Train Recall: 0.9630
Train F1 Score: 0.9627

Confusion Matrix (Train):
[[29432 256 207 105]
[35 29888 39 38]
[145 17 29430 408]
[169 29 481 29321]]

Test Metrics:
Test Accuracy: 76.39 %
Test Precision: 0.7641
Test Recall: 0.7639
Test F1 Score: 0.7639

Confusion Matrix (Test):
[[1642 77 103 78]
[50 1779 43 28]
[94 26 1589 191]
[86 37 221 1556]]

2nd Window size = 3

Train Metrics:
Train Accuracy: 97.26 %
Train Precision: 0.9745
Train Recall: 0.9736
Train F1 Score: 0.9706

Confusion Matrix (Train):
[[29477 250 178 95]
[34 29938 8 20]
[167 44 29469 320]
[193 91 688 29028]]

Test Metrics:
Test Accuracy: 81.89 %
Test Precision: 0.8184
Test Recall: 0.8189
Test F1 Score: 0.8184

Confusion Matrix (Test):
[[1655 77 90 78]
[43 1793 31 33]
[99 30 1586 185]
[122 52 232 1494]]

3rd Window size = 5

Train Metrics:
Train Accuracy: 97.38 %
Train Precision: 0.9739
Train Recall: 0.9737
Train F1 Score: 0.9737

Confusion Matrix (Train):

```
[[29516   160   171   153]
 [   94 29845    18    43]
 [  139    32 28989   840]
 [  100    29   171 29700]]
```

Test Metrics:

Test Accuracy: 0.8237
Test Precision: 0.8243
Test Recall: 0.8237
Test F1 Score: 0.8237

Confusion Matrix (Test):

```
[[1635   78   94   93]
 [  78 1757   25   40]
 [ 109   23 1536  232]
 [  89   35  140 1636]]
```