Ayush Mittal
2021201030

iNLP Assignment 3

1. Theory Questions

1.1 Negative Sampling

Question - **Explain negative sampling. How do we approximate the word2vec training computation using this technique?**

Answer - Negative sampling is a technique to improve the accuracy of model predictions and make the model's training efficient. In CBOW the aim is to predict the target word for the given context.

In the training of CBOW, we train the model with true as well as with false target words, to make the model able to distinguish between the words, which word is more likely to appear in the given context and which is not. The objective of the training is to maximise the probability of true target words and minimise the probability of negative words. The number of negative words for each true target word is hyper-parameters, and choosing 5 to 20 negative samples gives better results(given in the articles referred).

In CBOW without negative sampling, we give the target word and context words to the model, and the model returns us an embedding for the target word and the mean of embeddings of words in context. Here we treat the current task as a multiclass classification problem, need to use computationally heavy softmax function, softmax returns probability for each class to be the target.

When using negative sampling, we reduce the multiclass classification problem to predicting the probability of word present in given contest, as described above for positive sample the probability should be high, close to 1 and close to 0 for negative samples. So in place of computationally heavy softmax function - classifying among |V| classes, we have turned this probelm into |V| binary classification problems. Not only this we only train model for one positive and some negative samples, this make the training further simpler and computationally efficient and fast. Loss will we calculated for 1 + n samples and therefore weights corresponding to them will be trained(not all).

$$J_t(\theta) = \log \sigma \left(u_o^T v_c \right) + \sum_{j \sim P(w)} \left[\log \sigma \left(-u_j^T v_c \right) \right]$$

The above loss function is used, where the first term is used to maximise the probability of a true sample and second term is used to minimise the probability of negative samples.

$$\left(u_o^T v_c \right)$$

This term represents the similarity between the output vector for target word and context. For negative words, we have taken the negative of this term.

$$\log \sigma \left(u_o^T v_c \right)$$

This shows we took sigmoid to get the probabilities (for similarity) and then took the log of it.

While implementing the model I also did using binary cross entropy loss with sigmoid, but it was not giving good results.

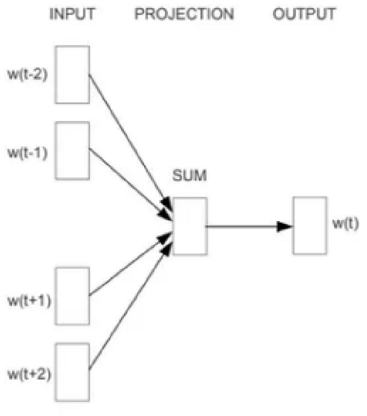
1.2 Semantic Similarity

Question - **Explain the concept of semantic similarity and how it is measured using word embeddings. Describe at least two techniques for measuring semantic similarity using word embeddings.**

Answer - Semantic similarity means the similarity between two texts based on their meaning. This is important in various fields like text classification, question-answering, information retrieval etc. Semantic similarity helps the model to take a better decision and make a better prediction.

Word embeddings are one of the best methods for measuring semantic similarity between text. Word embeddings are dense vectors(one corresponding to each word in vocablury) to capture the semantic meaning of words, for finding the similarity between two words we see how close these vector representations are. For example we did word2vec embedding for capturing the contextual meaning of the words, it is of two types - continuous bag of word - in which we predict the word for given context and other is skip gram - in this we predict the context given a word. In both of them we aim to train the word embeddings, such that those words which are more contextually close have similar representation(embedding).

CBOW

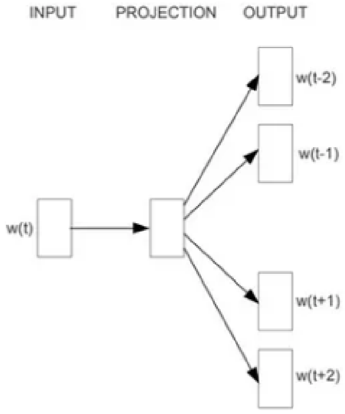


CBOW

This images show how we train model in CBOW, we pass the contest and model predicts the word which can be present in this context. For the correct word the probability should be 1 and 0 for other words, After training the models we get the embeddings which contains the information about the context in which the word would be present.

So words which have similar embeddings, mean they have more probability to be present in same type of context and hence more contextually similar.

Skip-Gram



Skip-gram

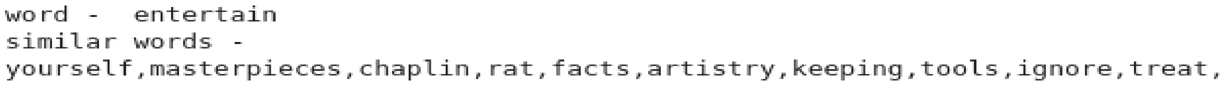
This image shows the idea behind skip gram word2vec technique. Here model is trained to predict the context for the given word. For the correct word the provavlity should be 1 and 0 for others. Since embedding are trainded such that embeddings contains the information about which context it would be present. So words which are present as context of same word have same embedding. This mean the words which have similar words embeddings are more semantically close.

When we have the embeddings there are ways to find the similarity or distance between them. Some of them are cosine similarity and euclidean similarity. Cosine similarity is one minus cosine distance. Cosine similarity is the measure of similarity between two non-zero vectors specified in an inner product space is called cosine similarity. The cosine of the angle between the vectors, or the dot product of the vectors split by the product of their lengths, is what is referred to as cosine similarity.

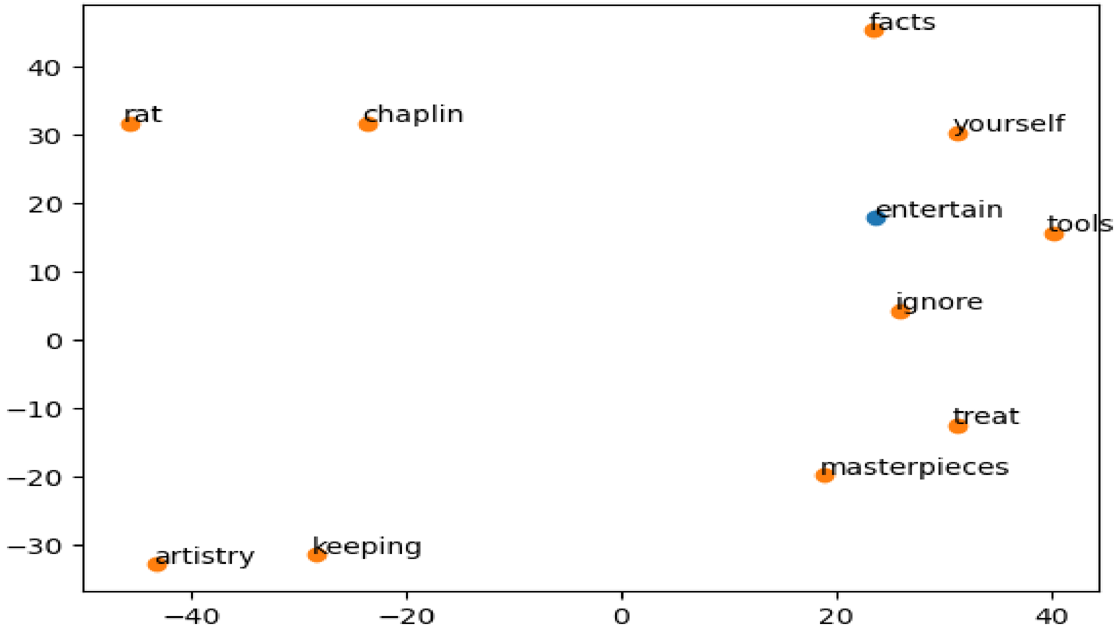
2. Analysis and Results

2.1 Top 10 for CBOW word2vec

Entertain-

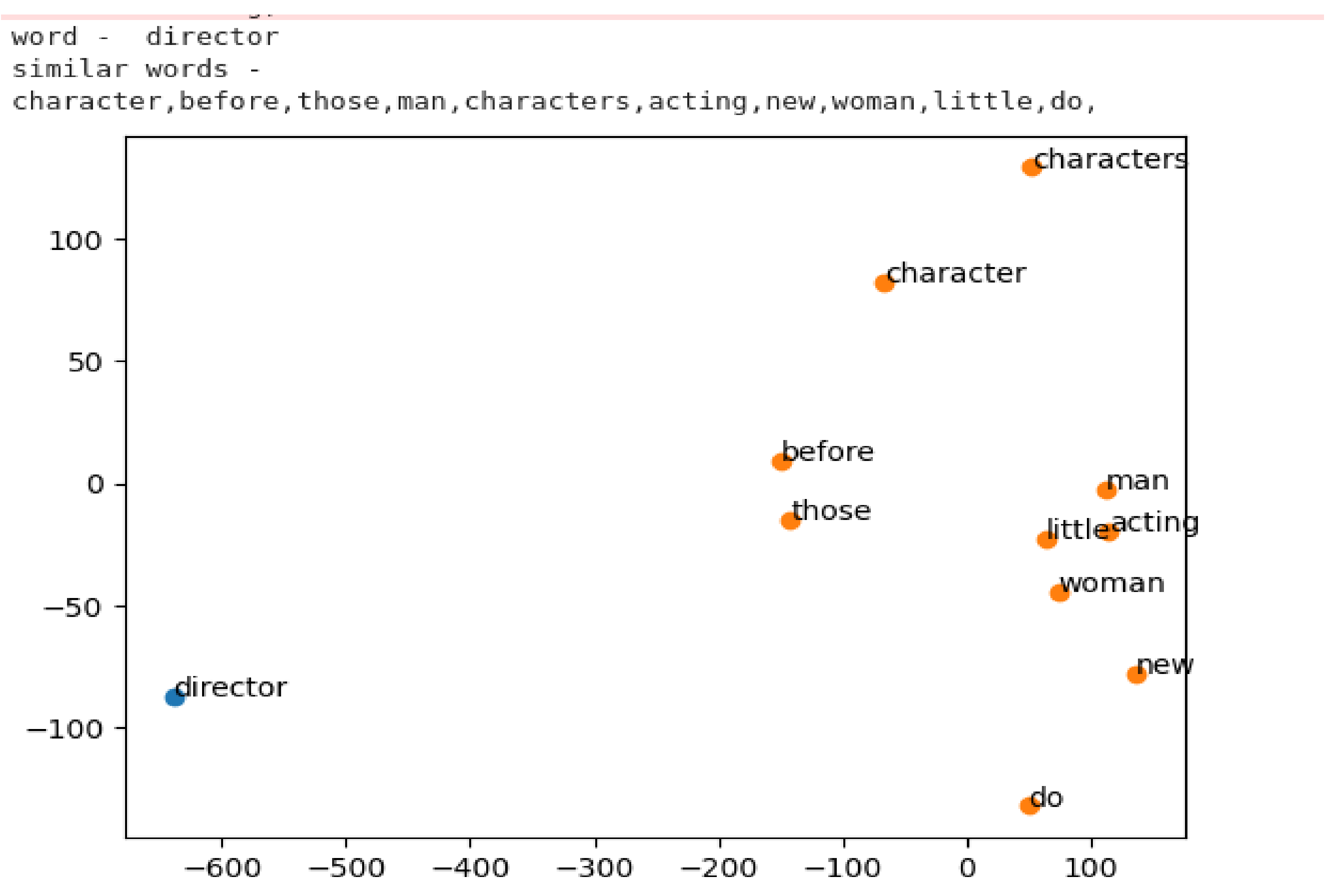


```
word -  entertain
similar words -
yourself,masterpieces,chaplin,rat,facts,artistry,keeping,tools,ignore,treat,
```



Correct closely related words **- Chaplin, treat, masterpiece, artistry**

Director -

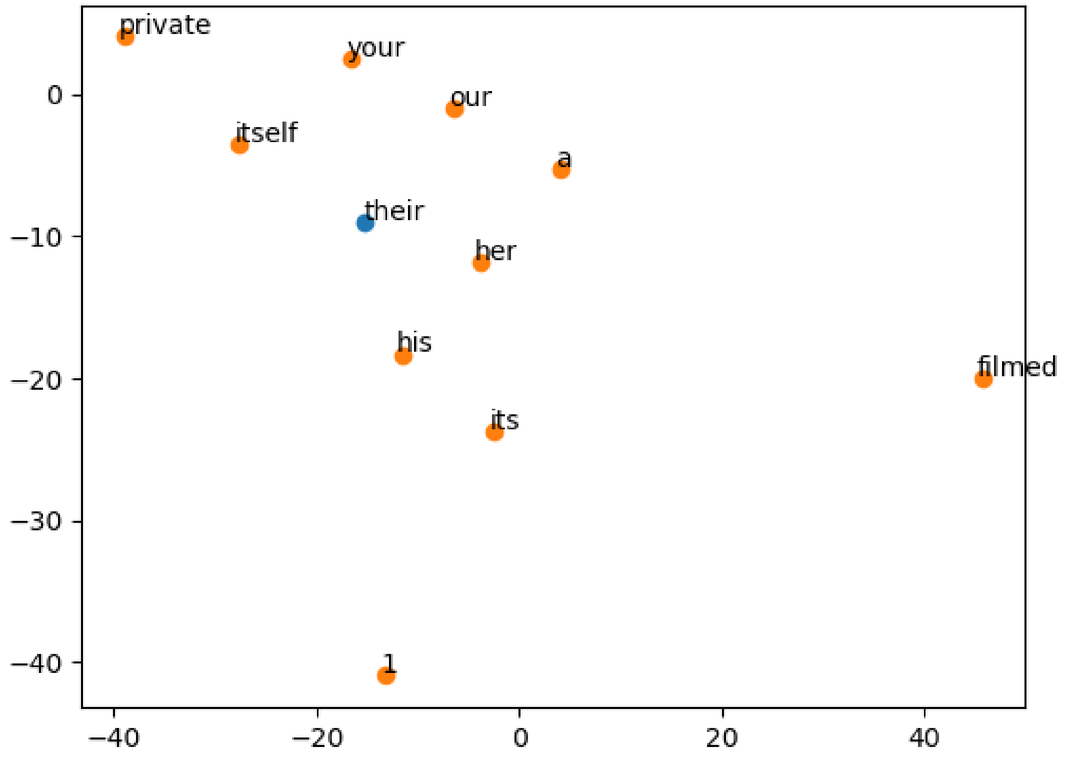


```
word -  director
similar words -
character,before,those,man,characters,acting,new,woman,little,do,
```

Correct closely related words **- man, woman, acting, character, characters**

Their -

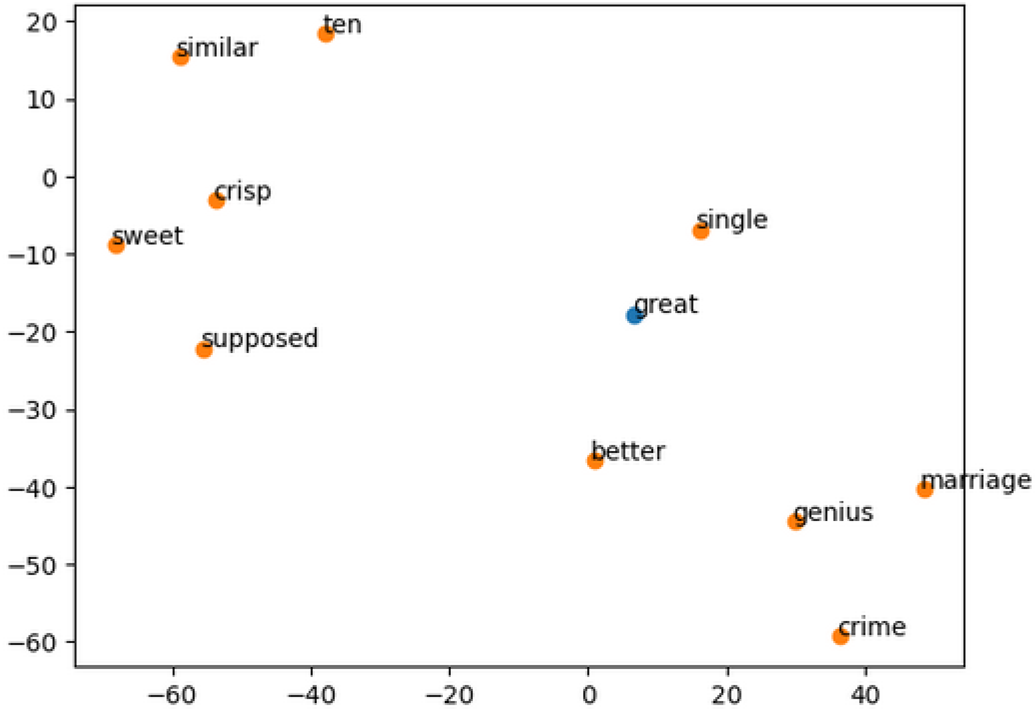
```
word -  their
similar words -
his,her,your,our,its,filmed,itself,a,private,1,
```



Correct closely related words **- your, our, itself, her, his, its**

Great -

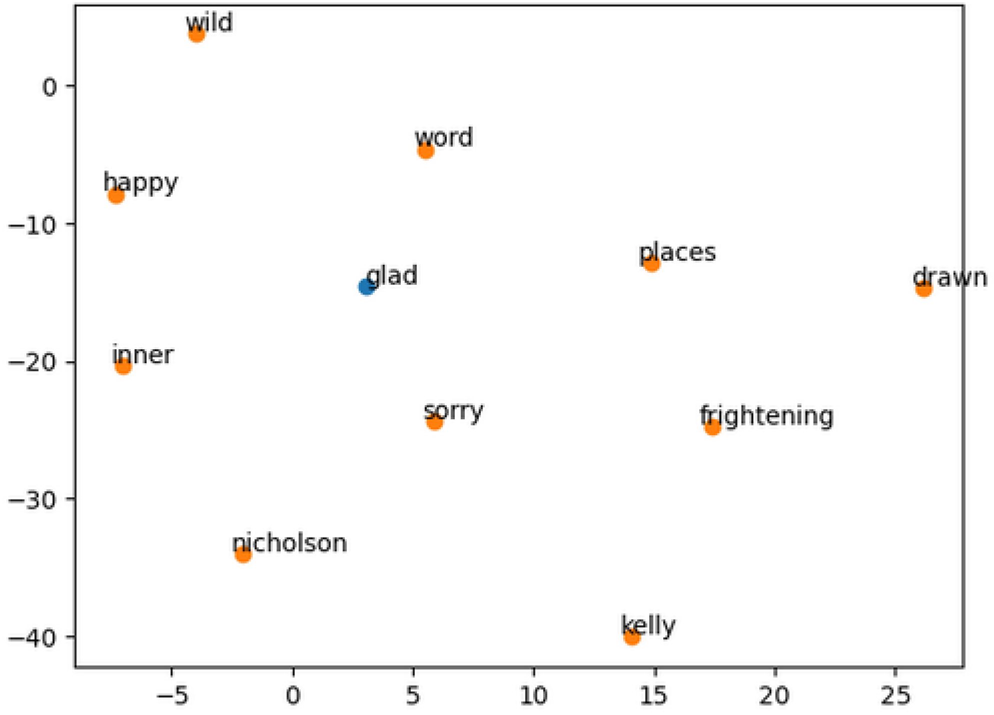
```
word -  great
similar words -
single,supposed,crime,sweet,ten,similar,crisp,genius,marriage,better,
```



Correct closely related words - **Genius, better, sweet, single**

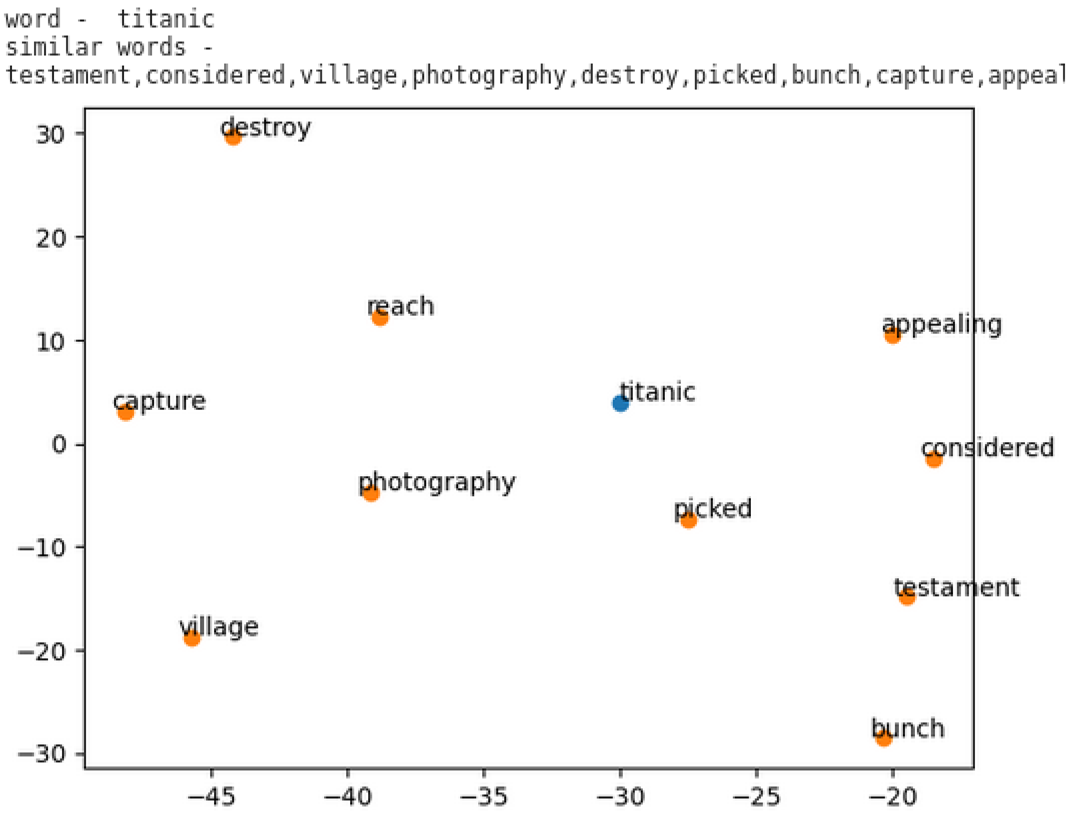
Glad -

```
word -  glad
similar words -
places,nicholson,inner,frightening,happy,drawn,sorry,kelly,word,wild,
```



Correct closely related words - **Happy, frightening, sorry**

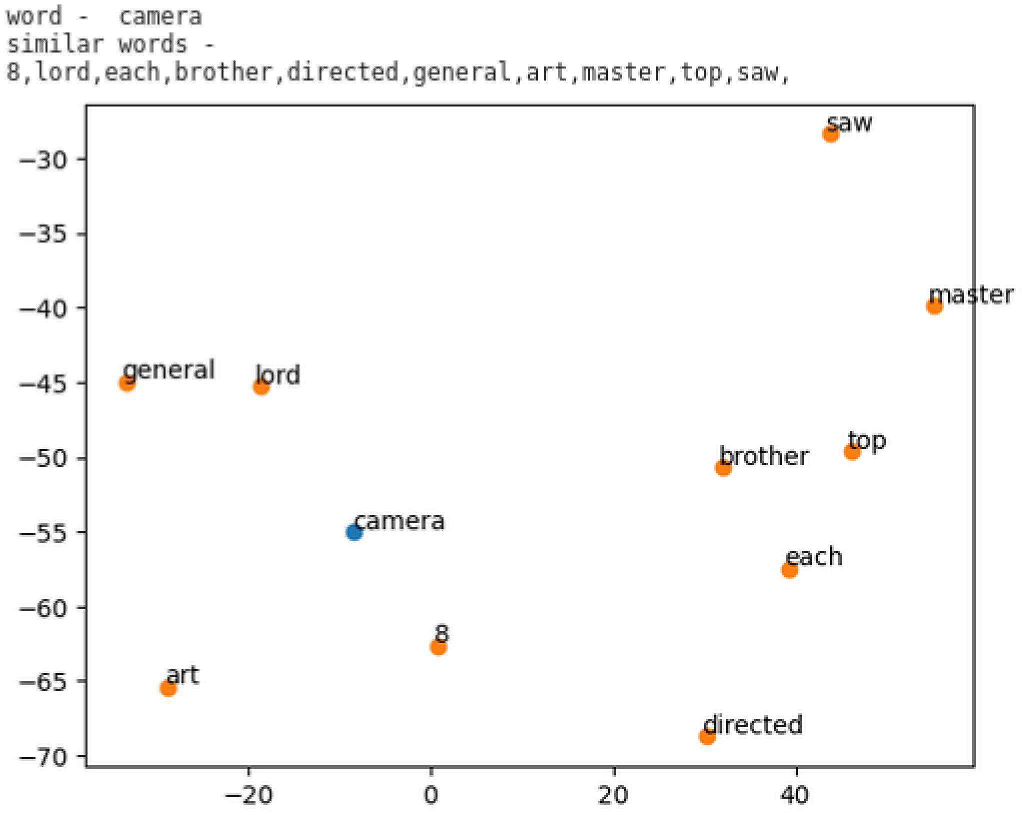
Titanic -



```
word -  titanic
similar words -
testament,considered,village,photography,destroy,picked,bunch,capture,appeal
```

Correct closely related words - **destroy**

Camera -



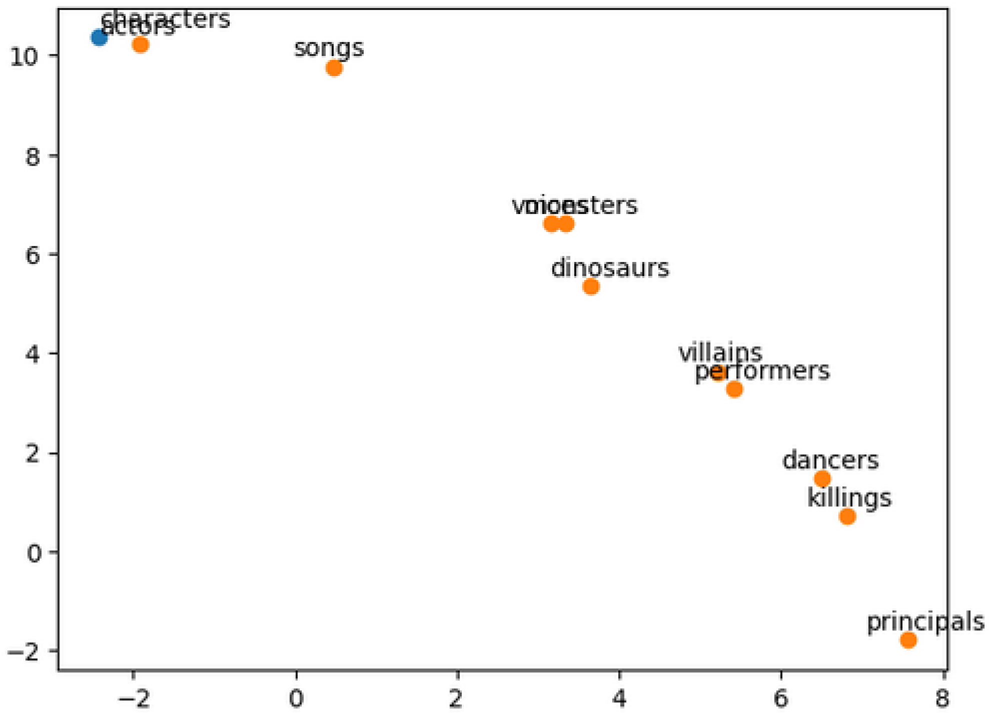
```
word -  camera
similar words -
8,lord,each,brother,directed,general,art,master,top,saw,
```

Correct closely related words - **Directed, art, saw**

2.2 Top 10 - Co-occurrence + SVD

Characters -

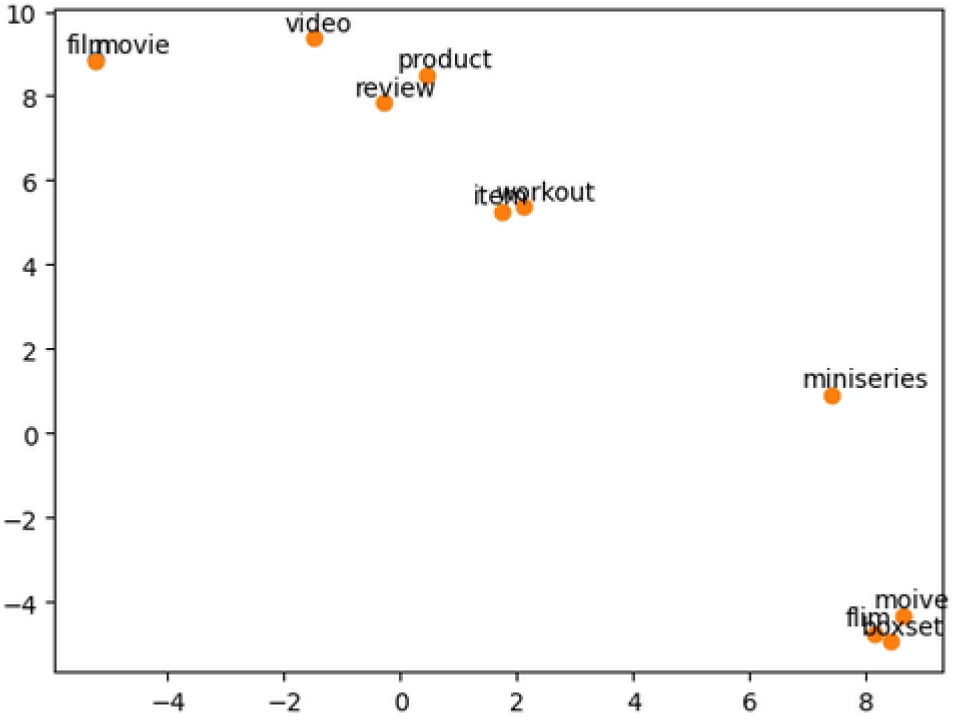
```
word -  characters
similar words -
songs,dancers,actors,voices,performers,monsters,killings,principals,dinosaur
```



Correct closely related words - **Dancers, actors, performers, monsters, principal**

Movie -

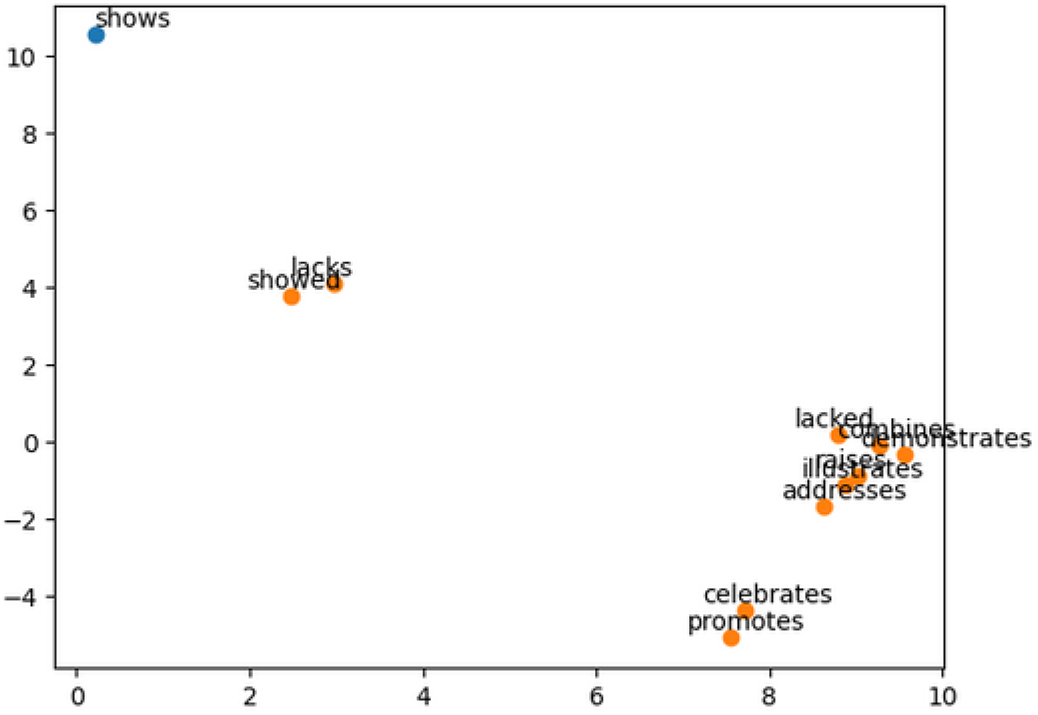
```
word -  movie
similar words -
moive,flim,film,item,boxset,miniseries,video,product,review,workout,
```



Correct closely related words - **Film, flim, miniseries, video, product, review, movie**

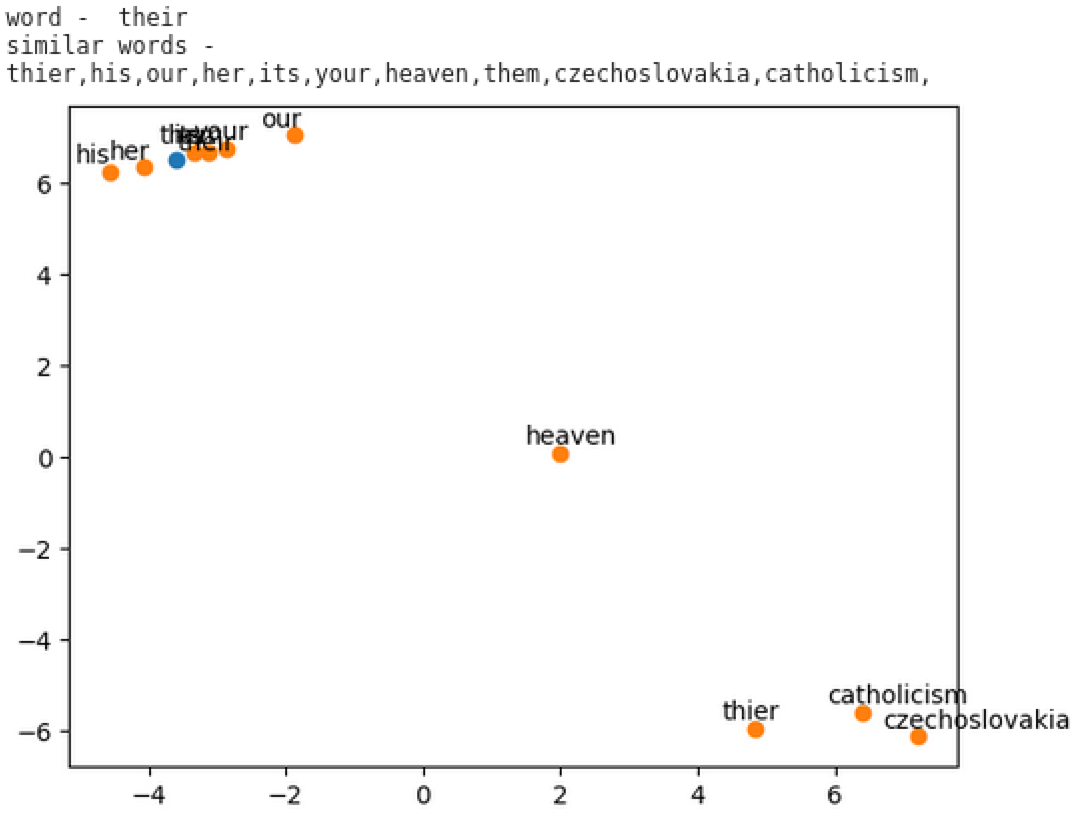
Shows -

```
word -  shows
similar words -
showed,demonstrates,raises,illustrates,promotes,lacked,addresses,celebrates,
```



Correct closely related words - **Showed, demonstrates, illustrates, addresses**

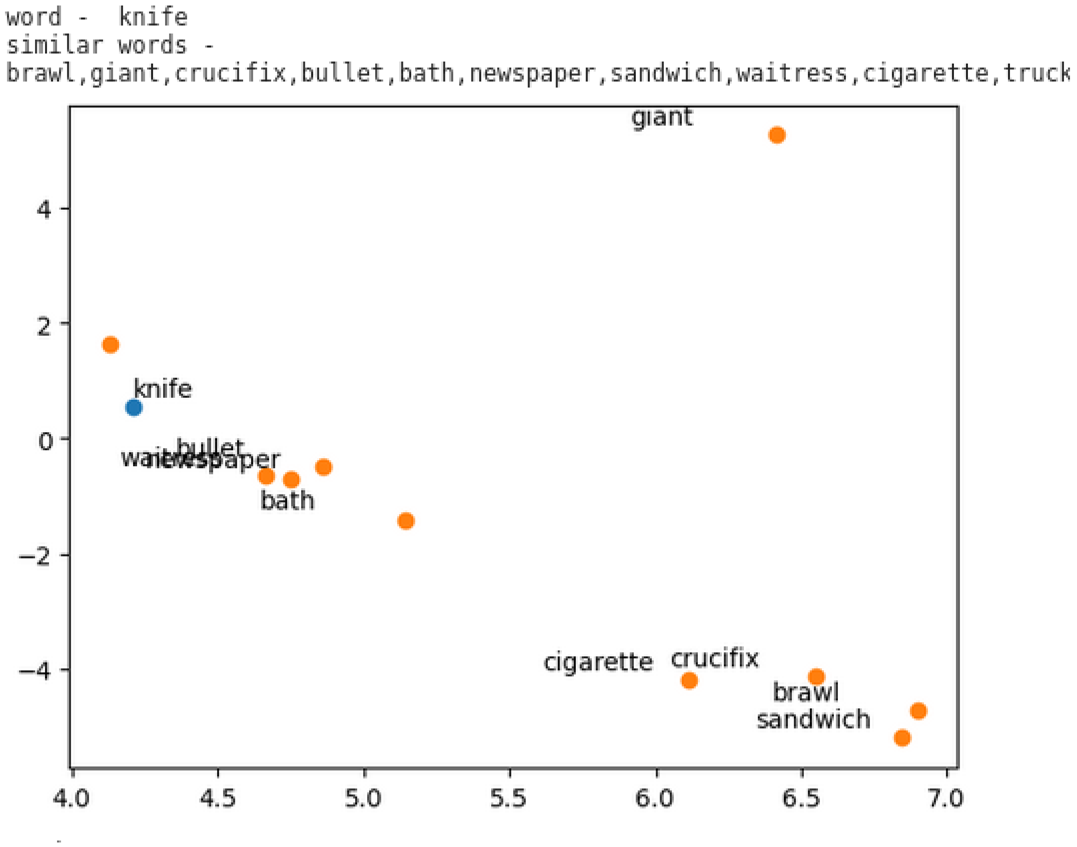
Their -



```
word -  their
similar words -
thier,his,our,her,its,your,heaven,them,czechoslovakia,catholicism,
```

Correct closely related words - **his, our, her, its, your, them**

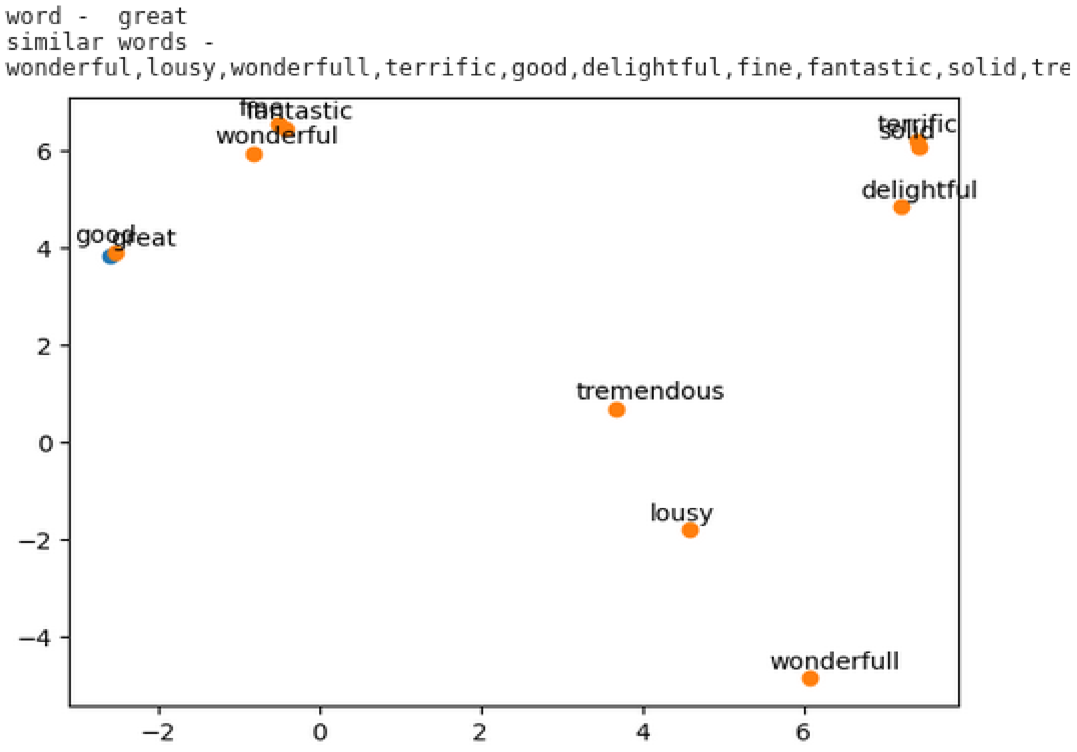
Knife -



```
word -  knife
similar words -
brawl,giant,crucifix,bullet,bath,newspaper,sandwich,waitress,cigarette,truck
```

Correct closely related words - **Brawl, giant, bullet, sandwich, waitress, cigarette**

Great -



```
word -  great
similar words -
wonderful,lousy,wonderfull,terrific,good,delightful,fine,fantastic,solid,tre
```

Correct closely related words - **wonderful, lousy, wonderfull, terrific, good, delightful, fine, fantastic, solid, tremendous**

2.3 Comparison with downloaded word2vec -

Close word of Titanic form downloaded embeddings -

1. unsinkable
2. athenia
3. lusitania
4. h.m.s.
5. lammiman
6. sinking
7. queenstown
8. samtampa
9. scharnhorst
10. Shipwrecked

Close word of Titanic form our CBOW word2vec-

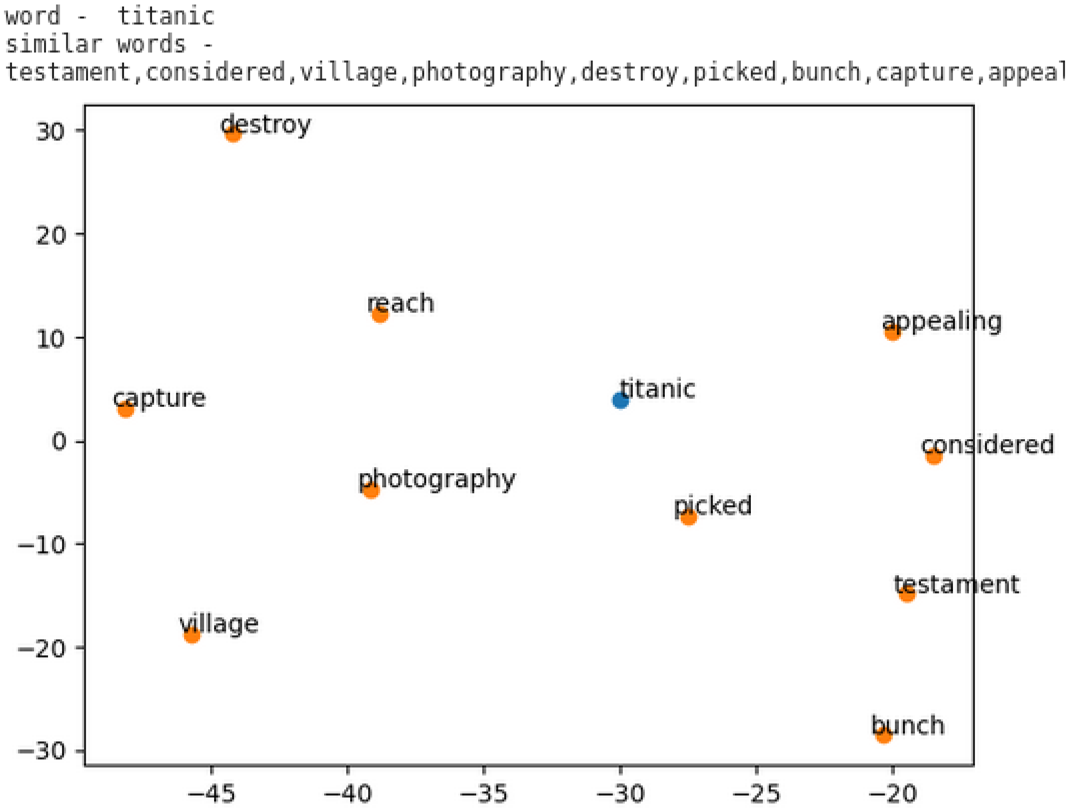
1. Testament
2. Considered
3. Village
4. Photography
5. Destroy
6. Picked
7. Bunch
8. Capture
9. Appealing
10. Reach

Close word of Titanic form our co-occurrence svd model -

1. Apollo
2. F
3. Conan
4. Seventh
5. Magic
6. Pink
7. Queen
8. Soup
9. Trilogy
10. Kindred

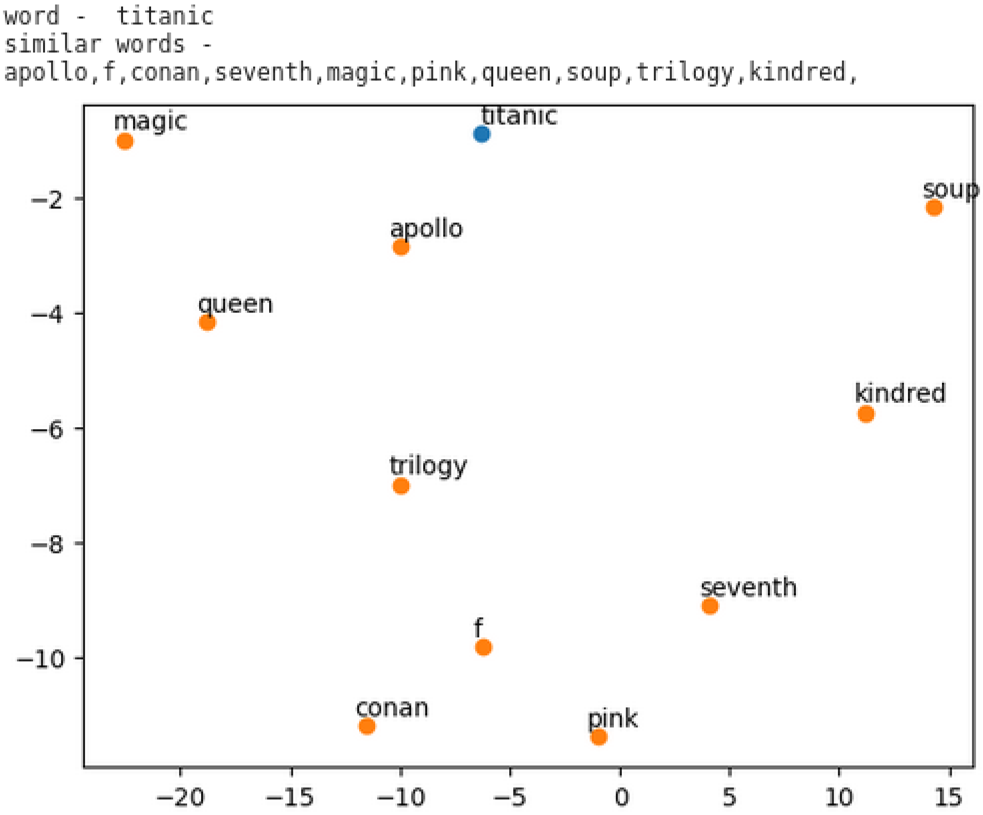
Since we have used starting 100K sentences and only 5 epochs, the accuracy of our model is not very good for all the words, although some words like Destroy, capture, Reach, testament are some what close to ship and movie.

CBOW predictions -



```
word -  titanic
similar words -
testament,considered,village,photography,destroy,picked,bunch,capture,appeal
```

SVD Co-occurrence predictions -



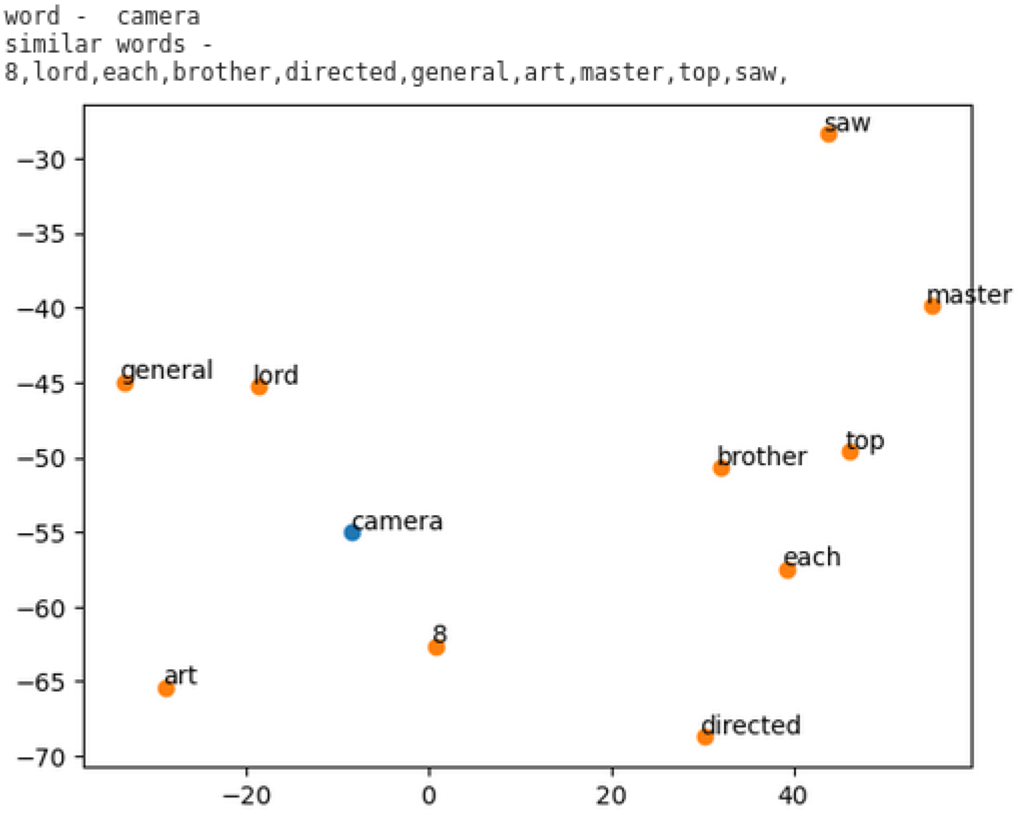
```
word -  titanic
similar words -
apollo,f,conan,seventh,magic,pink,queen,soup,trilogy,kindred,
```

Similar are the results for “camera”

From downloaded embeddings -

1. viewfinder
2. camcorder
3. zoom::lens
4. wide-angle::lens
5. telecine
6. darkroom
7. projector
8. slr
9. 35mm
10. closeup

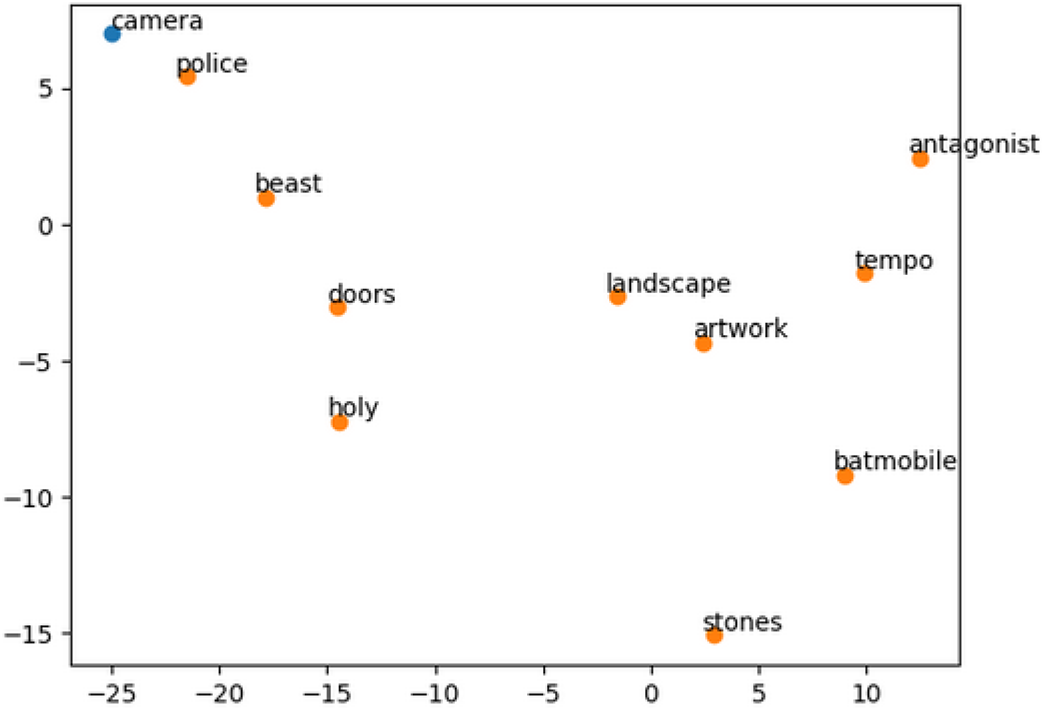
From CBOW -



```
word -  camera
similar words -
8,lord,each,brother,directed,general,art,master,top,saw,
```

From SVD co-occurrence -

```
word -  camera
similar words -
holy,batmobile,beast,doors,tempo,antagonist,police,artwork,stones,landscape,
```



Some of the words like landscape, batmobile are similar to camera.

Conclusion -

It can be clearly seen that out models are giving good results, CBOW on further training will outperform SVD model. On training our model on bigger corpus and for more epochs the result will be for sure better, but due to limitations of resources, need to limit the training with corpus of 100K reviews and 5 epochs.