Mukul Kumar | 2017350

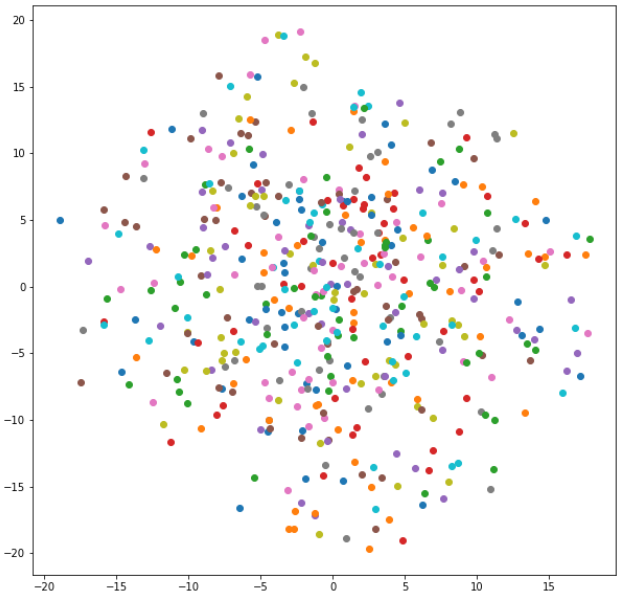
Problem 1:

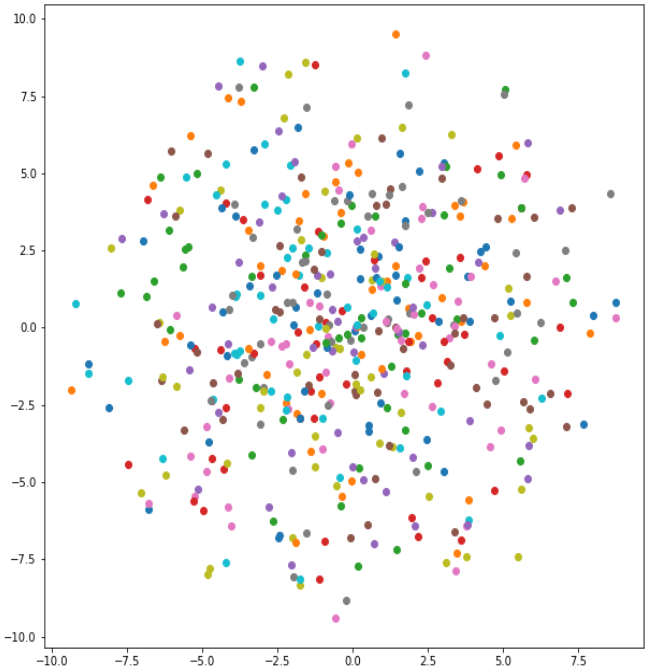
Word2Vec is basically used to get vector representation of a word, this representation also depends on the context that is, the words that occur frequently with the given word.

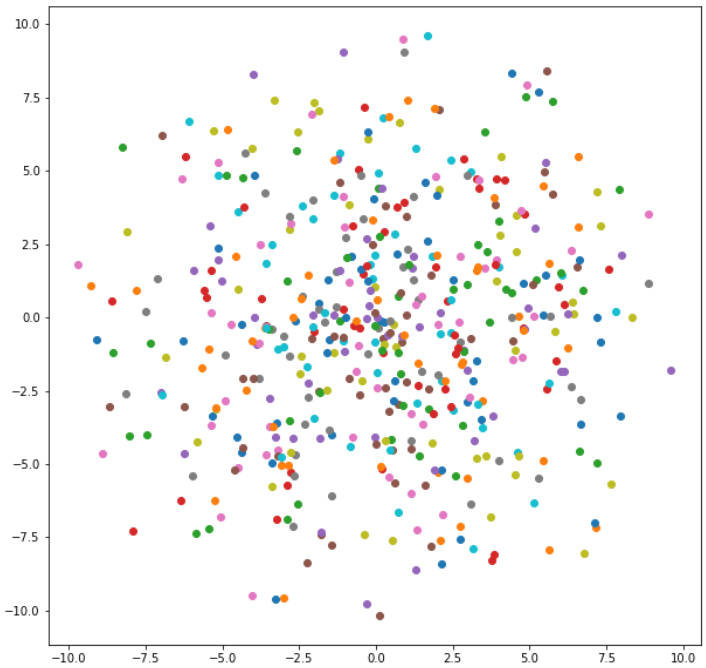
Word2Vec can be trained using two techniques skip-gram and Cbow. We have used skip-gram here, what it does it, it tries to predict the context words given a word.

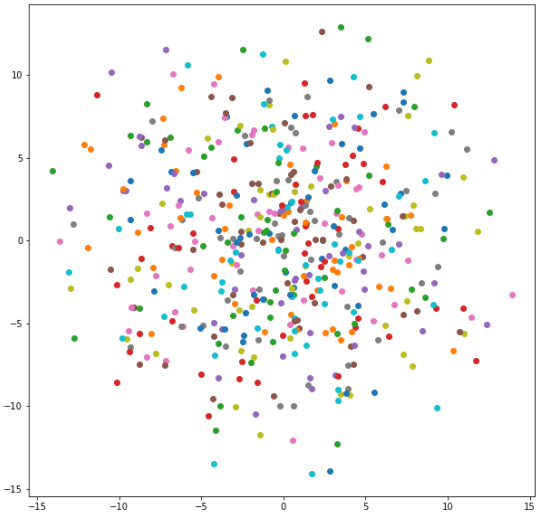
For faster and better results, there is a technique of using subsampling and negative sampling in the training of word2vec.

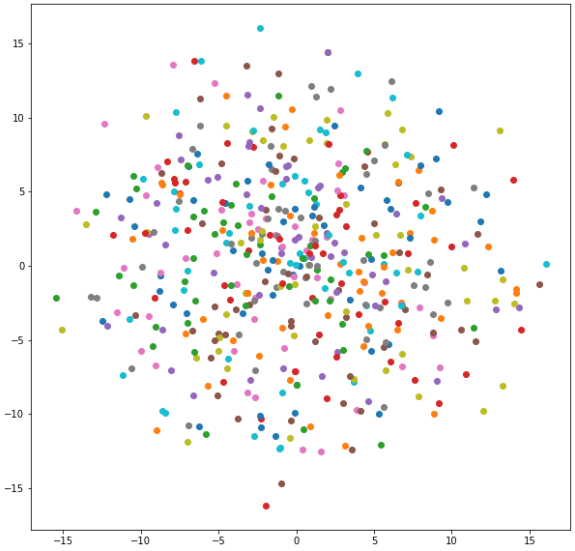
Subsampling removes very frequently occurring words because they rarely add meaning to the process and the context of a given word. Whereas negative sampling uses the negative examples, which are the example not associated with the context of a given word, to ensure that only the related words should get the update, making the training process faster.




10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5


10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0





Problem 2:
Baseline MAP: 0.49176786896815833
With Relevance feedback: 0.5859090029055217

Relevance feedback with query expansion: 0.584671433541388

The alpha-beta values used were 0.8 and 0.2 respectively.
The values were found to be almost similar.