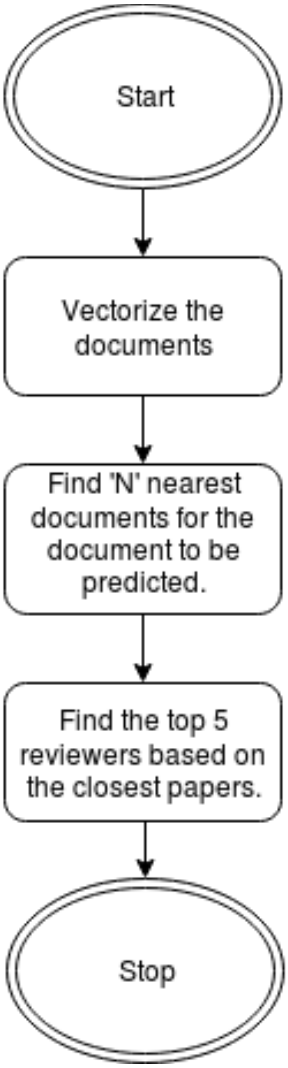
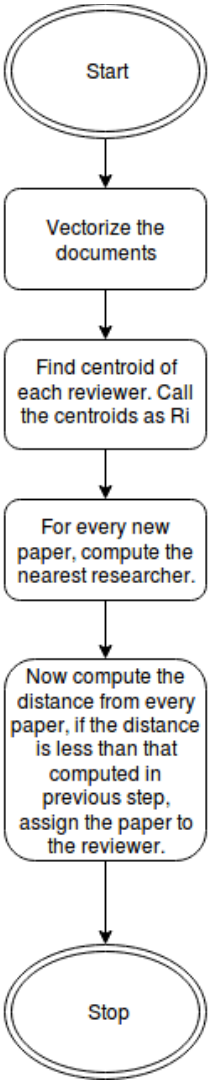
Expert Systems

Comparing various algorithms

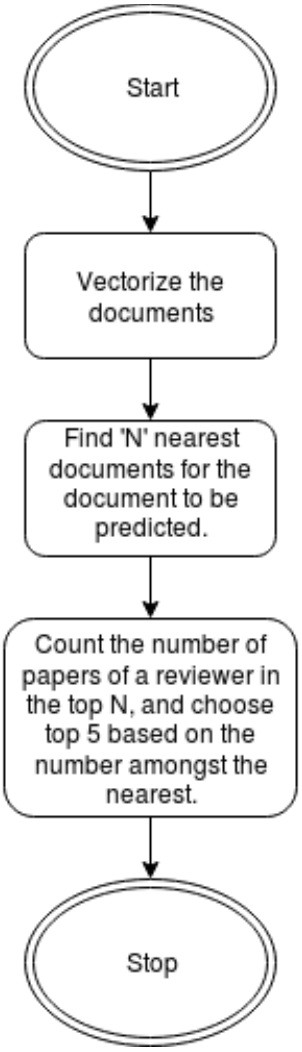
Following flow diagrams depict various algorithms which we have tried and you would find the quantitative and qualitative analysis of the performance of various algorithms in following sections.



Algo 1



Algo 2



Algo 3

Algorithm 1:

This algorithm takes into account the papers that are the nearest to the given paper. It selects the reviewers who have written and published similar papers in past. A drawback of this algorithm is that it doesn't take into considerations the number of papers a reviewer has written which are similar to the paper to be reviewed. Algorithm 3 cleverly deals with this issue.

Algorithm 2:

This algorithm is computationally very intensive. Despite being computationally intensive, the results, are expected to be highly inaccurate. Firstly the algorithm needs to find centroids of

reviewers. Logically, the term centroid of a reviewer has no physical significance. This term eliminates the possibility that a reviewer might have multiple fields of interest. The algorithm, as expected, takes a lot of time to evaluate the output, because of distance calculations from every paper in the training set.

Algorithm 2 utilises a property which makes no sense: centroids of reviewers. The problem here is that if we try and find the centroids of reviewers, we are straight away discarding the possibility of a reviewer to be interested in multiple domains. In fact, any algorithm which requires centroids of reviewers is, in fact, baseless and more prone to errors.

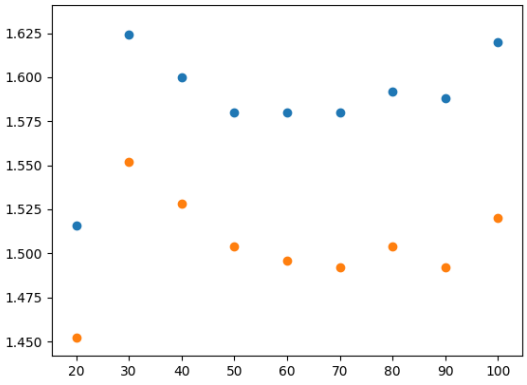
Algorithm 3:

This algorithm not only acknowledges the papers which are close to a given paper but also consider the number of papers a particular reviewer has in the vicinity of the paper. This, I think could give us the best possible reviewers for a particular paper.

Quantitative analysis

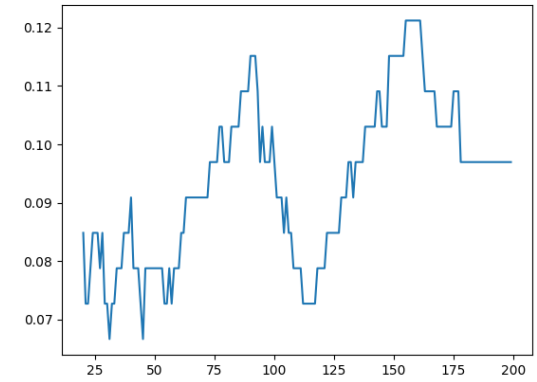
We have implemented Algorithm 1, 2 and 3. Our implementation of algorithm 2 was without library functions from sk-learn, so the computations were quite inefficient. A comparison of quantitative parameters for the same with other 2 would be inappropriate. The other 2, namely 1 and 3 are quantified as below.

Time



N ---->

Accuracy



N ---->

The blue dots correspond to algorithm 3, the orange dots correspond to algorithm 1. As is evident from the steps involved, 3 is a mere extension of 1 and it takes an offset amount of time more than algorithm 1.

As could be seen from the graph, the time required is the minimum in the range of N in 50 – 70. Increasing the value of N only adds up to the time required to compute the number of instances of a particular author.

It has been found empirically that the accuracy is the maximum for N in range 157 to 162 for this dataset. The accuracy is found to show a sinosoidal nature. The two maximas for accuracy occur at 70 and 160, the latter being the global maxima.

Contributions

Everyone has contributed equally.