Analysis File

- ShaliniBhardwaj(MT19045)

Rochio algorithm is applied to see practical implementation of how feedback can improve the query and help in increasing **RECALL.**

Main motive of taking feedback is to take increase recall and it is observed through the implementation that precall also improves.

In order to have fair analysis the once the user marks the document as relevant are not considered for feedback next time and hence are not repetitively counted for rochio.As it is accommodated in query.

Usually Cosine does give relevant documents at the top 10 positions as seen in queries given below. Precision is high initially then it decreases with feedback it improves .This trend with certain variations is seen in the below queries.

Precision recall Curve

Precision and recall is calculated for all K documents retrieved at feedback iteration to see how query is improving .Here every timep(10%) of K are marked as relevant automatedly as per given ground truth folder by the user.

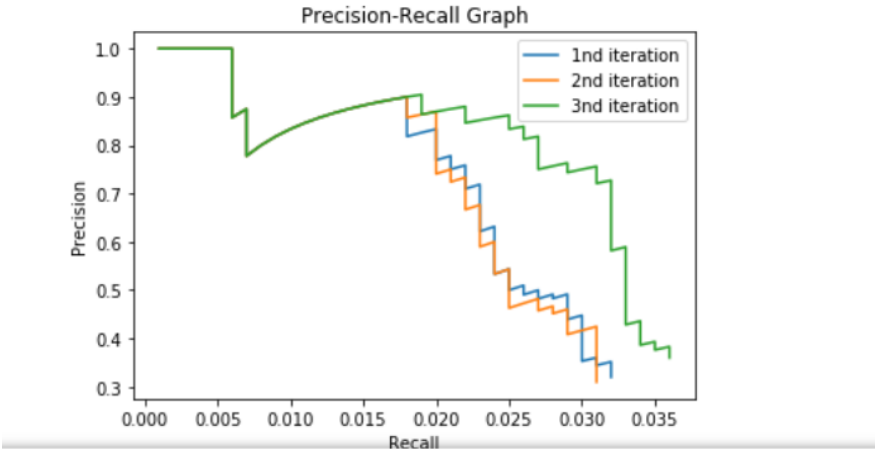
Query 1: Pretty good opinions on biochemistry machines

Relevant set 1: Documents inside folder sci.med

Run this query: using K=100

Hence p=10(10%ofK)

Here 3 iteration after feedback are performed and built.



Here we cane see variation in each iteration.

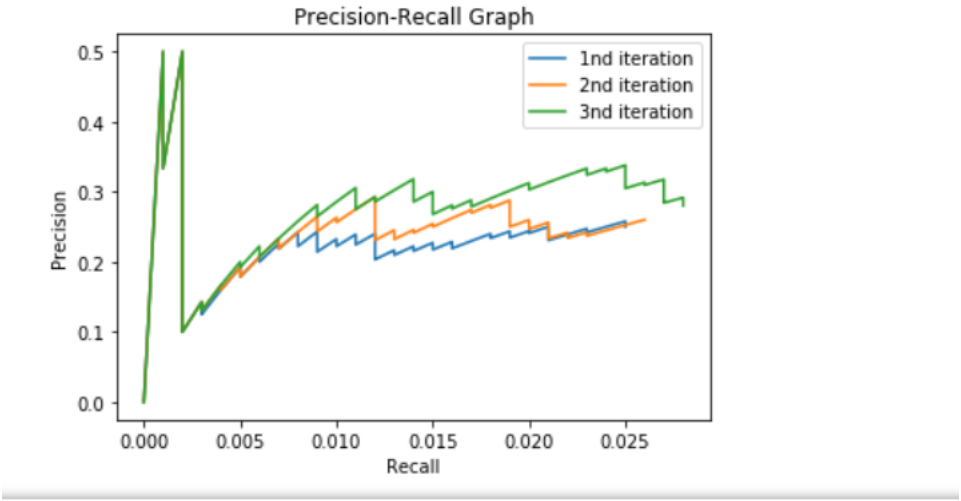
Initially for few top results cosine is giving all True positive result and precision value is seen as 1.As we can see the improvement taking place in next iterations that with each feedback iteration both precision is increasing. Hence we are able to give to the user more true positives.

In this query cosine is giving good results initially as terms present in query are more related to biology or medicine like biochemistry is most relevant to medicine folder than any political,sportsor space domain.Hence this query gives best result.

Query 2: Scientific tools for preserving rights and body

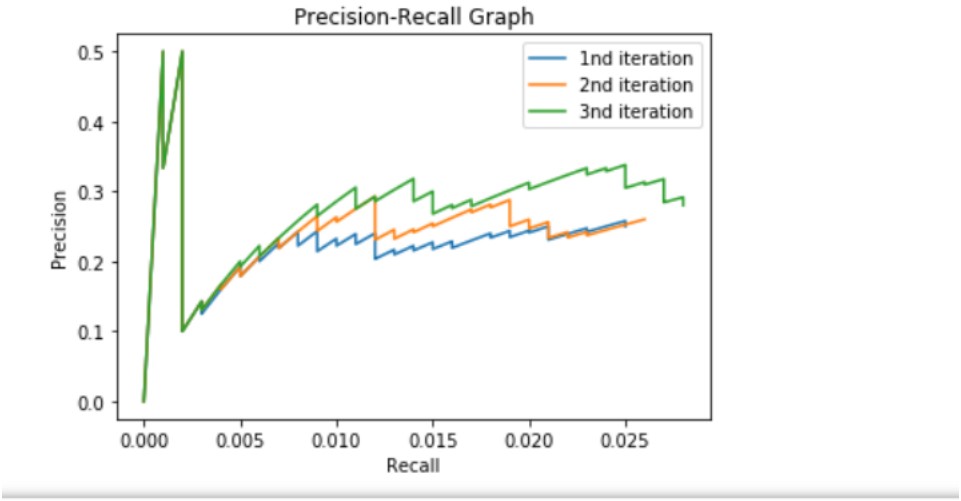
Relevant set 2: Documents inside folder talk.politics.misc

Similar to above patterns are seen just



Query 3: Frequently asked questions on State-of-the-art visualisation tools

Relevant set 3: Documents inside folder sci.med



Initially this query does not give good result for cosine values as it can be analysed that grounf truth for this folder is given as medical.But terms used in this query unlike the query1 of same folder has less medicinal term.Therefore it not easy to map query to its ground truth.But still with feedback of rochio algorithm it is improving.

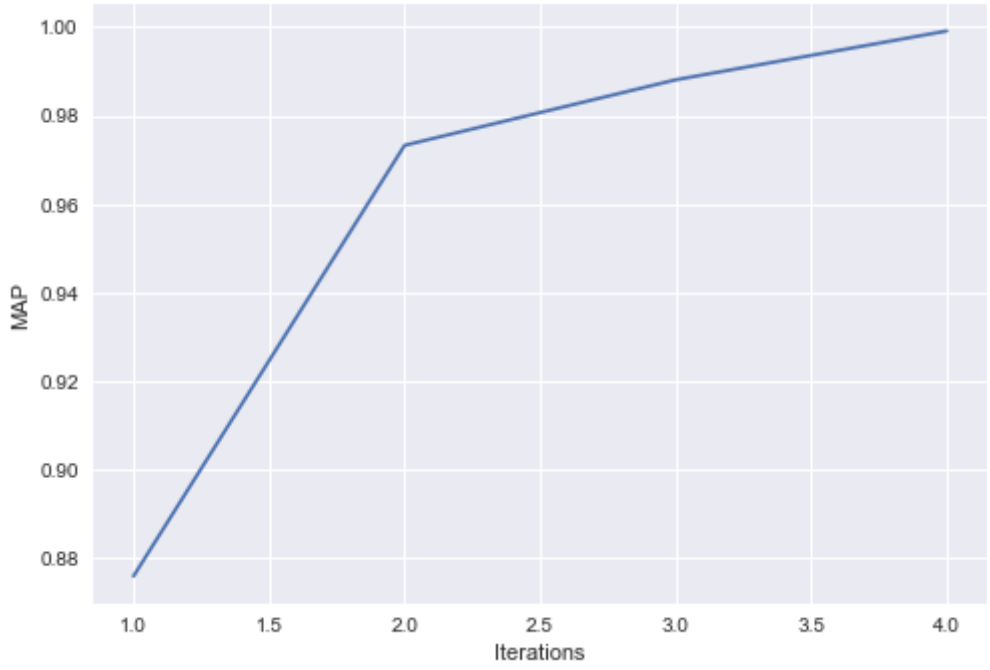
For different queries with different type of terms present same algo gave different graph .

Hence to analyse futher we will be evaluating on MAP.

MAP

MAP is mean average precision it gives equal weightage to all the queries. This gives us the better analysis over the queries.Because some queries may be easy others might be difficult.Hence we get average of all queries .Therefore this metric helps in analysing at iteration level comparison.

As after marking * after relevant feedback it can be seen the documents are occupying positions in top of retrieval.



Here we can see for initial iterations the average precision for all queries is less and as we move further

towards feedback iterations it reaches towards maximum possible.

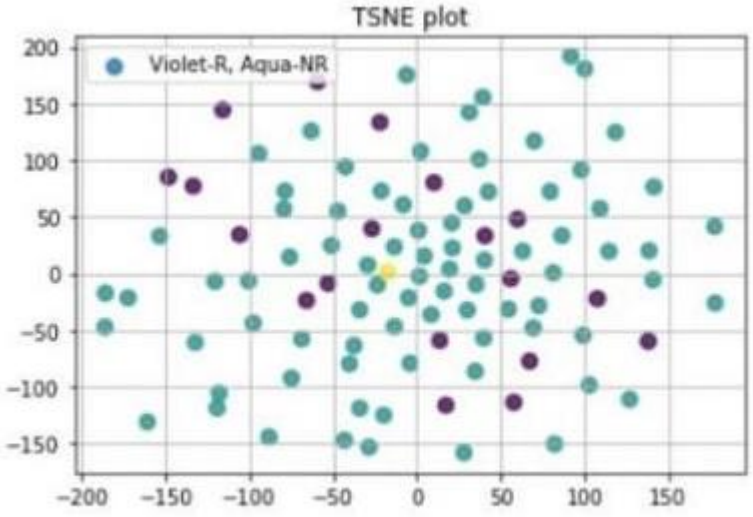
TSNE Plot

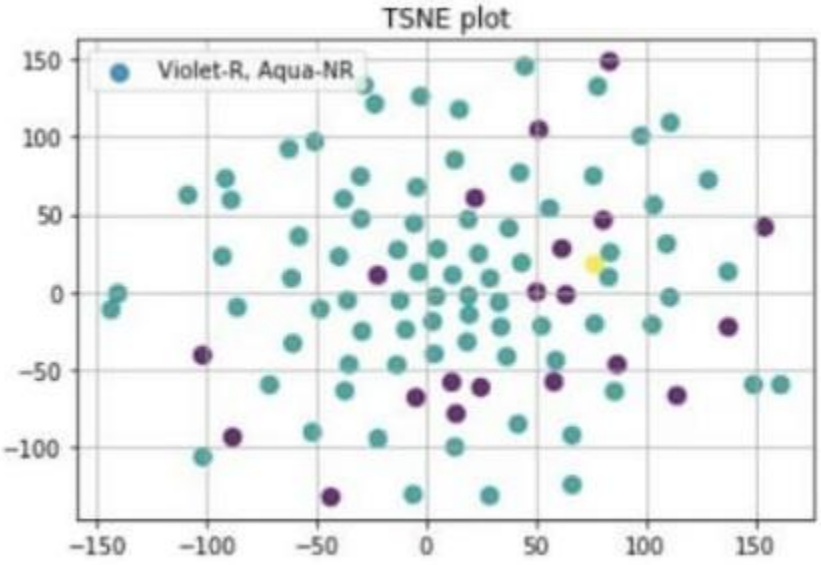
It is giving us better representation of query plot at every iteration on all the queries. We can see how the query is moving.It is always far from each other and improving by moving towards relevant documents as per the result seen before.

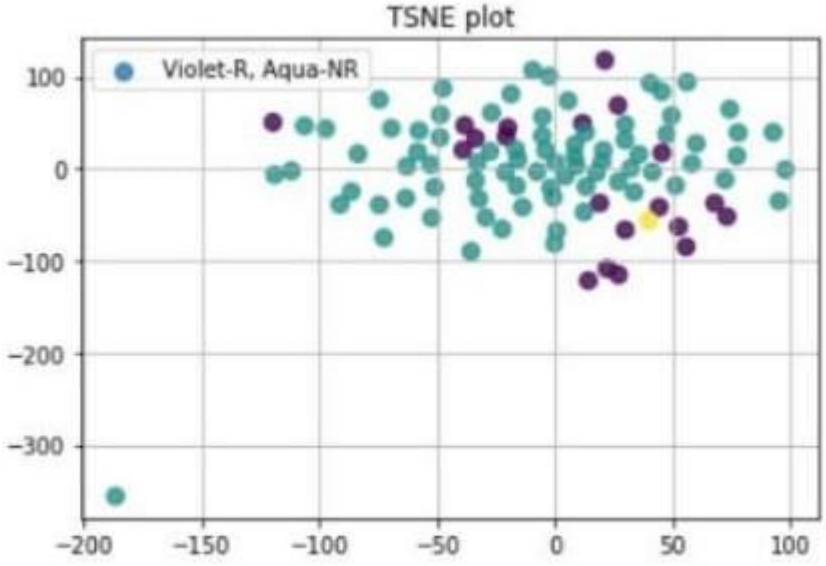
Different as query at different iteration are shown below.

Moreover how query is moving for relevant docs cluster is seen in the graph.

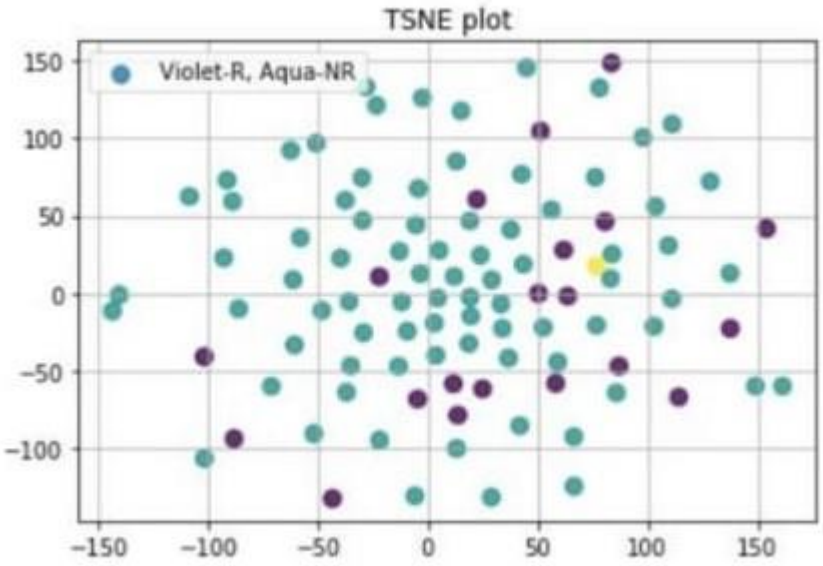
Query1



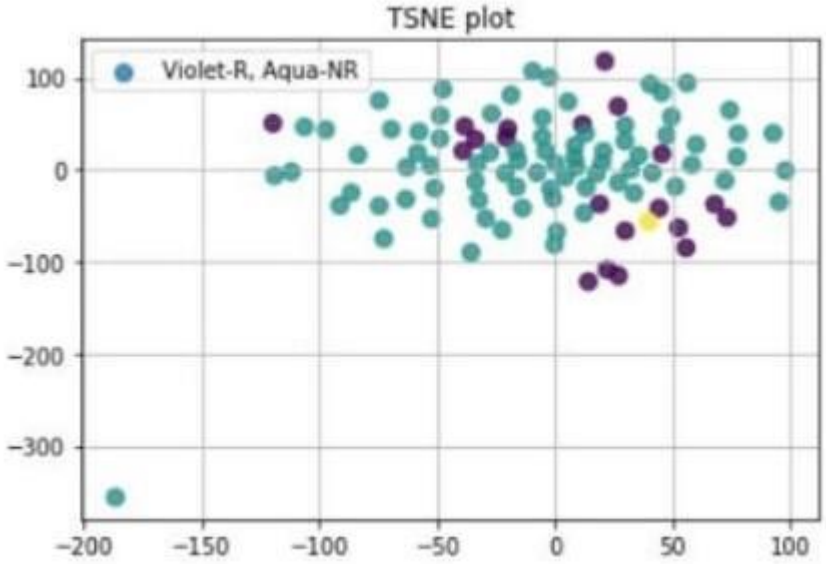

TSNE plot
Violet-R, Aqua-NR
150
100
50
0
−50
−100
−150
−100
−50
0
50
100
150


TSNE plot
Violet-R, Aqua-NR
100
0
−100
−200
−300
−200
−150
−100
−50
0
50
100

Query 2

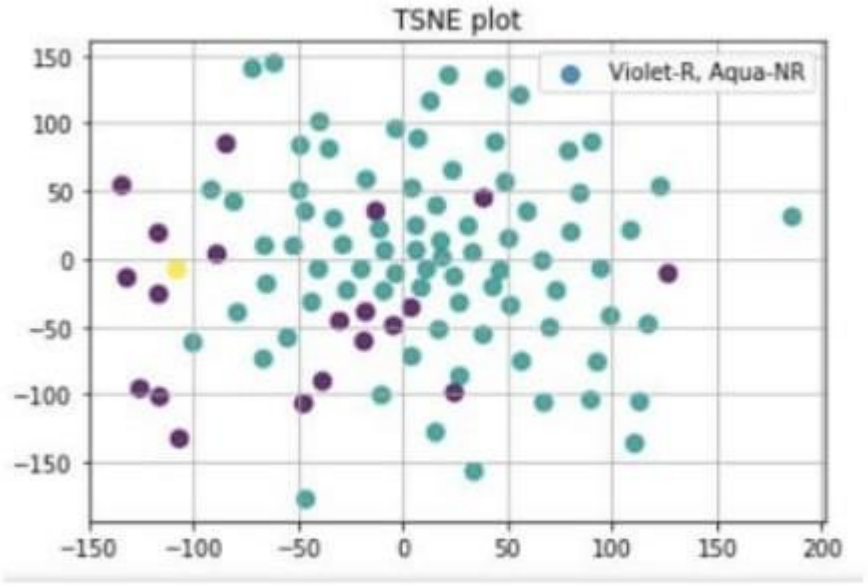


Iteration1

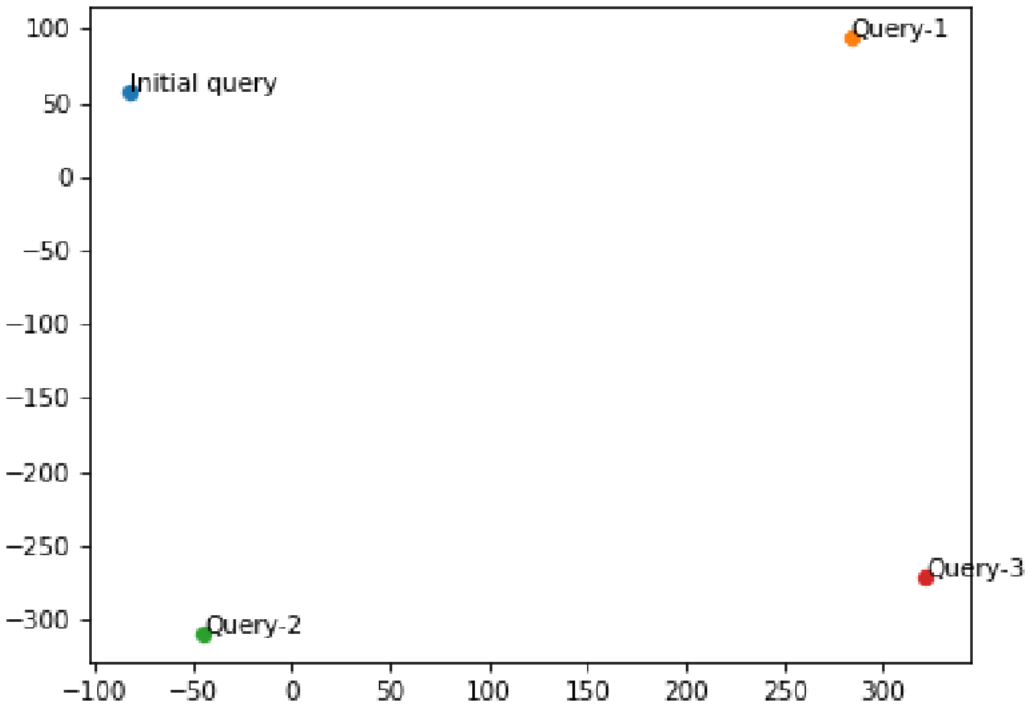


Iteration-2

Iteration 3

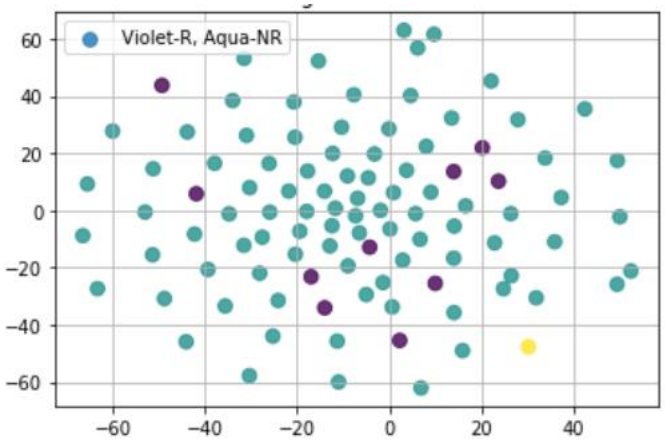


Here we can we initial queries after iterations of feedback are changing position and moving towards relevant doc cluster .Thus therefore more recall and precision is got and better results are provided to user.

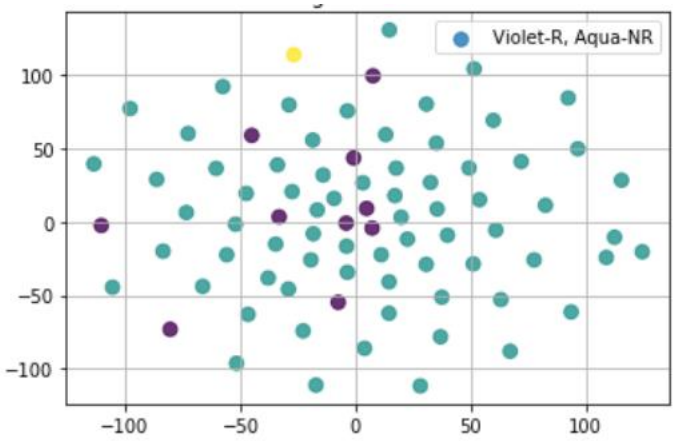


Query3

Iteration 1



Iteration 2



Iteration 3