

## ASSIGNMENT-03

## ANALYSIS

- **Size of vocab after preprocessing:** 58126
  - **Total number of documents :** 5000
  - Data is extracted from 5 folders namely - [ 'comp.graphics', 'sci.med', 'talk.politics.misc', 'rec.sport.hockey', 'sci.space' ] containing 1000 files each.
  - Query is modified by relevant and non-relevant documents marked by a user in current iteration.
1. Index containing tf\_idf value for each term i.e document vector having tf\_idf values if term present otherwise 0.

[illegible]

**Query:** *Pretty good opinions on biochemistry machines*  
**top k :** 100

**Ground truth** : Documents inside folder sci.med

- **Normal query results:**

Enter number of relevant docs:

0,4,5,6,7,9,10,11,16,20,21,22,23,24,25,30,60,63,64,73,80,87,92,93,99

MAP: 0.3944488704503613

```

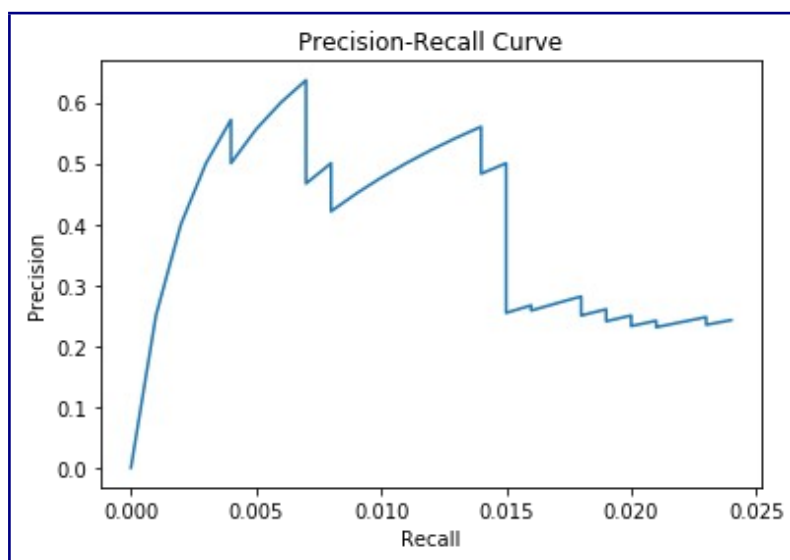
84 3112 rec.sport.hockey 54726
85 3386 rec.sport.hockey 53779
86 2756 talk.politics.misc 178386
87 1582 sci.med 59184
88 618 comp.graphics 38869
89 4415 sci.space 61068
90 4619 sci.space 60897
91 103 comp.graphics 38908
92 1081 sci.med 59564
93 1011 sci.med 59115
94 2559 talk.politics.misc 178554
95 2995 talk.politics.misc 176902
96 2785 talk.politics.misc 178905
97 3873 rec.sport.hockey 53827
98 212 comp.graphics 38878
99 1819 sci.med 59017
Enter number of relevant docs: 0,4,5,6,7,9,10,11,16,20,21,22,23,24,25,30,60,63,64,73,80,87,92,93,99
MAP: 0.3944488704503613

```

Relevant documents: 25

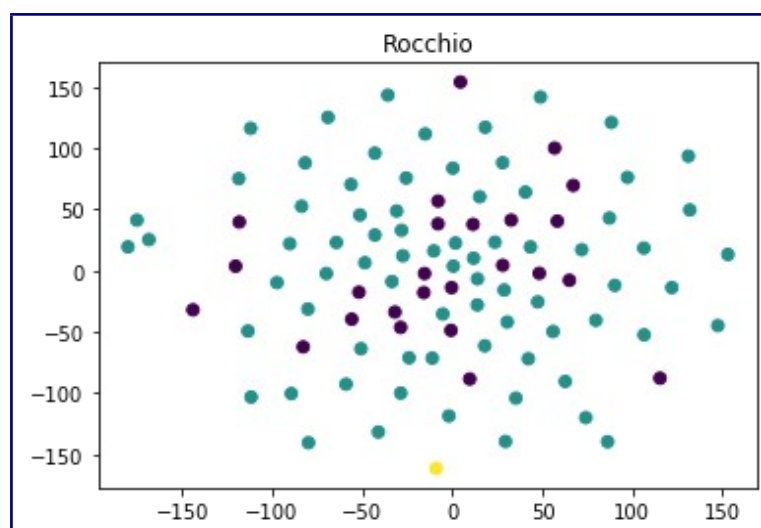
Non-Relevant documents: 75

### PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

### TSNE



- Non relevant documents vector
- Relevant documents vector
- Query Vector

- **Modified Query Results:**

Marked Relevant documents by user as '\*':

```
iteration 1
Top docs:
0 1856 sci.med 59429 *
1 774 comp.graphics 38965
2 1826 sci.med 59116 *
3 1949 sci.med 59117 *
4 1429 sci.med 59393 *
5 1081 sci.med 59564 *
6 1864 sci.med 59060 *
7 1594 sci.med 59301 *
8 1481 sci.med 58082 *
9 1765 sci.med 59233 *
10 1139 sci.med 58987 *
11 1973 sci.med 59083 *
12 1653 sci.med 59400 *
13 1011 sci.med 59115 *
14 1409 sci.med 59504 *
15 1640 sci.med 59488 *
16 1853 sci.med 58801 *
```

### Iteration-01

Enter number of relevant docs:

0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21,22, 23, 24, 25, 26, 28, 29, 31,32, 33, 34, 35, 36, 37, 38, 39, 43, 45, 47, 48,49, 51, 52, 54, 58, 59, 60, 62, 63, 66, 70, 71, 72, 76, 77, 78, 83, 84, 85, 88, 89, 90, 92, 94, 95, 96, 97, 98, 99

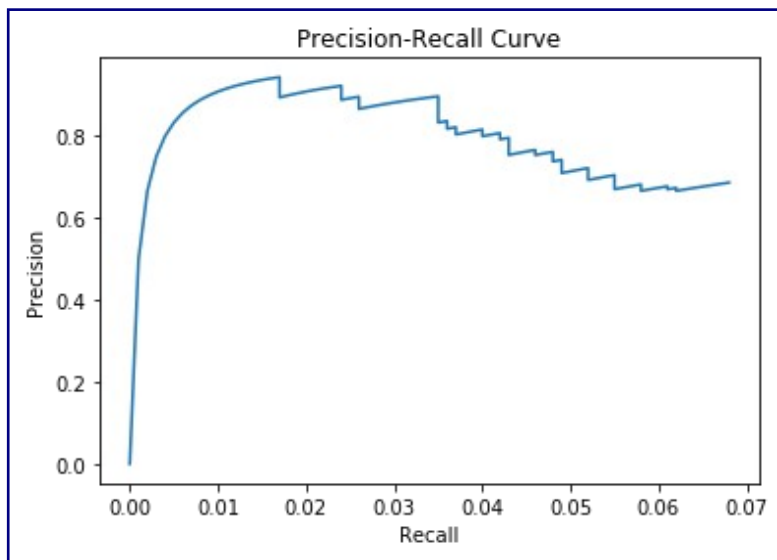
MAP: 0.7936849629161259

```
91 2553 talk.politics.misc 179100
92 1310 sci.med 59286
93 2049 talk.politics.misc 178293
94 1546 sci.med 58820
95 1811 sci.med 59303
96 1039 sci.med 58910
97 1201 sci.med 59063
98 1539 sci.med 59160
99 1850 sci.med 58984
Enter number of relevant docs: 0,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,20,21,22,23,24,25,26,28,29,31,32,33,
34,35,36,37,38,39,43,45,47,48,49,51,52,54,58,59,60,62,63,66,70,71,72,76,77,78,83,84,85,88,89,90,92,94,95,96,97,9
8,99
[1856, 1826, 1949, 1429, 1081, 1864, 1594, 1481, 1765, 1139, 1973, 1653, 1011, 1409, 1640, 1853, 1970, 1220, 148
6, 1690, 1709, 1602, 1866, 1582, 1215, 1888, 1911, 1819, 1288, 1540, 1965, 1509, 1893, 1071, 1871, 1725, 1189, 1
632, 1790, 1659, 1084, 1326, 1875, 1684, 1847, 1829, 1250, 1928, 1338, 1406, 1041, 1190, 1053, 1365, 1431, 1859,
1115, 1485, 1174, 1108, 1506, 1788, 1310, 1546, 1811, 1039, 1201, 1539, 1850] [2049, 774, 651, 13, 538, 3357, 41
4, 2336, 2986, 819, 3508, 829, 718, 721, 3794, 4824, 2136, 861, 354, 996, 3556, 2918, 2281, 2539, 2157, 4846, 40
78, 2547, 244, 2553, 382]
MAP: 0.7936849629161259
```

Relevant documents: 69

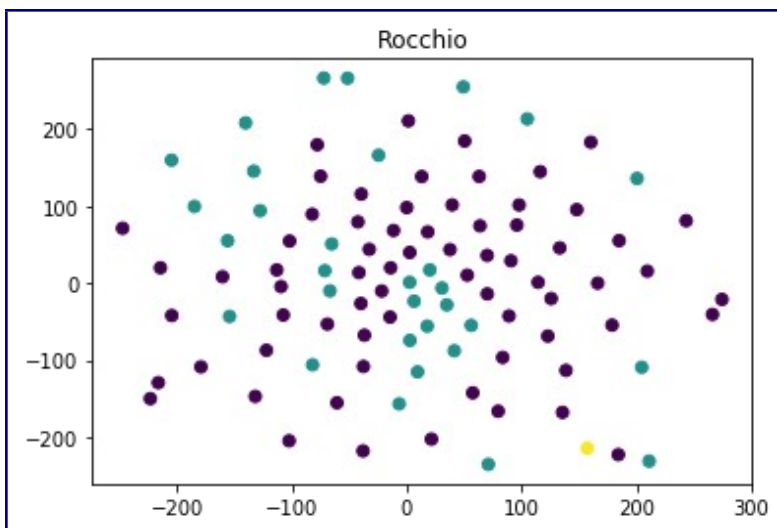
Non-Relevant documents: 31

## PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non relevant documents vector
- Relevant documents vector
- Query Vector

## Iteration-02

Enter number of relevant docs:

0, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 29, 30, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 59, 60, 62, 63, 64, 66, 67, 68, 70, 72, 73, 75, 77, 78, 79, 86, 88, 89, 90, 91, 93, 94, 95, 97, 98

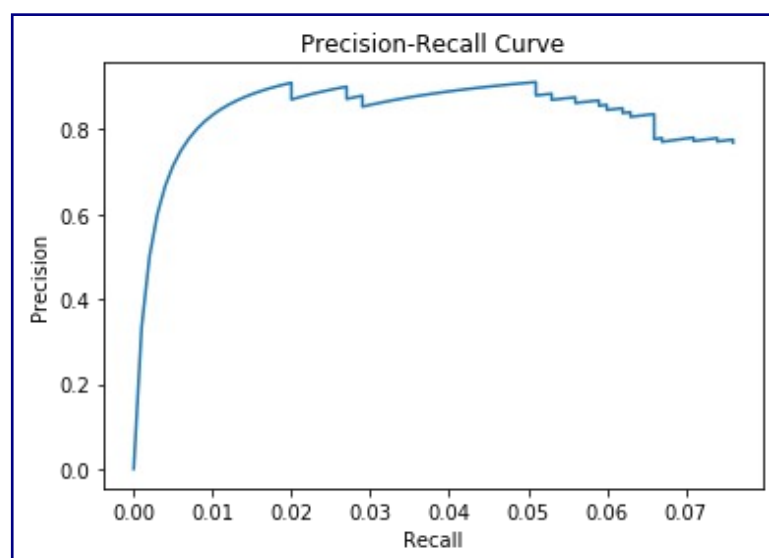
MAP: 0.8280720679368853

```
95 1201 sci.med 59063 *
96 3508 rec.sport.hockey 53677
97 1238 sci.med 59470
98 1231 sci.med 59276
99 2157 talk.politics.misc 178311
Enter number of relevant docs: 0,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,24,25,26,27,28,29,30,32,3
3,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,59,60,62,63,64,66,67,68,70,72,73,75,77,78,7
9,86,88,89,90,91,93,94,95,97,98
[1856, 1826, 1081, 1949, 1429, 1864, 1594, 1765, 1481, 1139, 1653, 1973, 1011, 1409, 1640, 1970, 1486, 1853, 122
0, 1690, 1602, 1709, 1215, 1582, 1866, 1888, 1911, 1288, 1819, 1540, 1965, 1071, 1871, 1189, 1893, 1509, 1632, 1
790, 1659, 1326, 1829, 1847, 1725, 1250, 1928, 1875, 1084, 1190, 1406, 1338, 1684, 1431, 1174, 1108, 1115, 1859,
1053, 1310, 1485, 1041, 1365, 1539, 1788, 1546, 1850, 1506, 1051, 1039, 1359, 1713, 1389, 1375, 1619, 1883, 120
1, 1238, 1231] [2049, 774, 651, 13, 414, 2336, 2986, 819, 3508, 829, 718, 4824, 861, 354, 996, 2918, 2539, 2157,
4846, 4078, 2547, 244, 382]
MAP: 0.8280720679368853
```

Relevant documents: 77

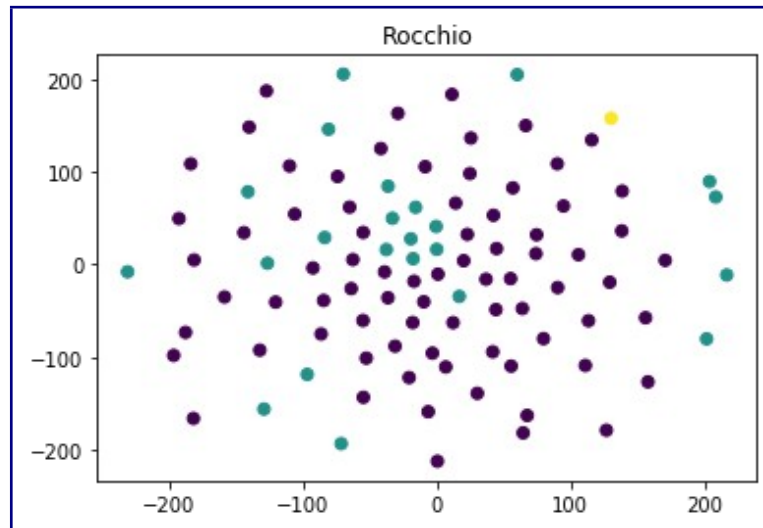
Non-Relevant documents: 23

PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non relevant documents vector
- Relevant documents vector
- Query Vector

### Iteration-03

Enter number of relevant docs:

0, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 72, 73, 74, 75, 76, 78, 79, 80, 81, 83, 84, 85, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 98, 99

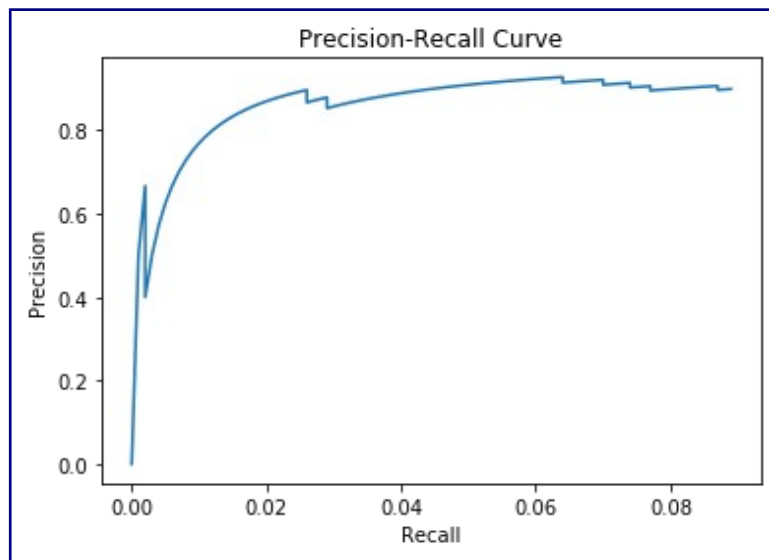
MAP: 0.8548783820401292

```
91 1042 sci.med 58760
92 1665 sci.med 59277
93 1201 sci.med 59063 *
94 1182 sci.med 59131
95 1034 sci.med 59261
96 1568 sci.med 59246
97 2986 talk.politics.misc 178382
98 1733 sci.med 59347
99 1808 sci.med 59168
Enter number of relevant docs: 0,2,3,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,31,32,33,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,71,72,73,74,75,76,78,79,80,81,83,84,85,87,88,89,90,91,92,93,94,95,96,98,99
[1856, 1081, 1949, 1765, 1429, 1139, 1481, 1653, 1864, 1594, 1011, 1409, 1486, 1970, 1973, 1640, 1220, 1853, 1215, 1602, 1582, 1888, 1709, 1911, 1690, 1288, 1866, 1540, 1965, 1071, 1819, 1189, 1871, 1829, 1893, 1326, 1250, 1847, 1659, 1790, 1190, 1928, 1632, 1108, 1431, 1115, 1406, 1509, 1310, 1875, 1084, 1338, 1174, 1539, 1546, 1859, 1725, 1485, 1051, 1850, 1713, 1389, 1053, 1788, 1684, 1365, 1883, 1041, 1359, 1619, 1375, 1238, 1574, 1506, 1231, 1039, 1812, 1798, 1881, 1332, 1311, 1179, 1042, 1665, 1201, 1182, 1034, 1568, 1733, 1808] [2336, 2049, 1826, 354, 774, 2986, 651, 2735, 2547, 382]
MAP: 0.8548783820401292
```

Relevant documents: 90

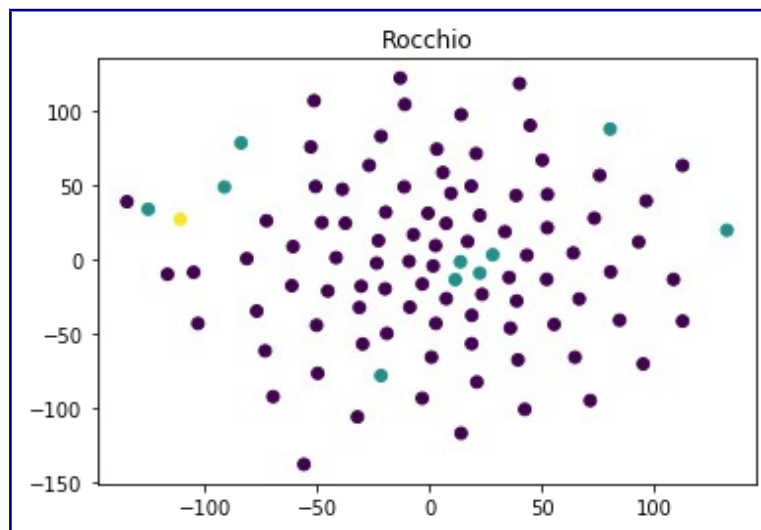
Non-Relevant documents: 10

## PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non relevant documents vector
- Relevant documents vector
- Query Vector



## Iteration-04

Enter number of relevant docs:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 76, 77, 78, 79, 80, 81, 82, 83, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99

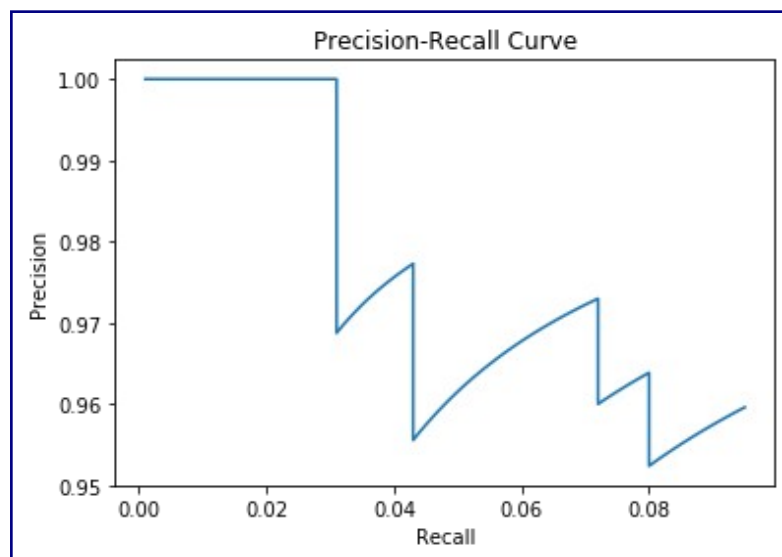
MAP: 0.9660819279184428

```
91 1034 sci.med 59201 *
92 1808 sci.med 59168 *
93 1129 sci.med 59181
94 1117 sci.med 59199
95 1736 sci.med 59234
96 1948 sci.med 59133
97 1684 sci.med 59529 *
98 1901 sci.med 59189
99 1664 sci.med 58883
Enter number of relevant docs: 0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,33,34,35,36,37,38,39,40,41,42,43,44,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,76,77,78,79,80,81,82,83,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99
[1856, 1826, 1081, 1765, 1139, 1653, 1481, 1949, 1970, 1486, 1220, 1215, 1409, 1011, 1640, 1429, 1973, 1888, 1594, 1288, 1582, 1709, 1853, 1602, 1864, 1911, 1965, 1540, 1071, 1829, 1871, 1189, 1190, 1847, 1928, 1690, 1250, 1108, 1866, 1819, 1326, 1790, 1115, 1659, 1893, 1310, 1632, 1546, 1539, 1406, 1431, 1389, 1850, 1859, 1338, 1084, 1051, 1713, 1174, 1883, 1574, 1238, 1485, 1875, 1798, 1619, 1375, 1788, 1359, 1332, 1053, 1509, 1812, 1365, 1733, 1665, 1179, 1182, 1696, 1311, 1601, 1725, 1476, 1039, 1231, 1881, 1563, 1034, 1808, 1129, 1117, 1736, 1948, 1684, 1901, 1664] [354, 2547, 382, 2735]
MAP: 0.9660819279184428
```

Relevant documents: 96

Non-Relevant documents: 04

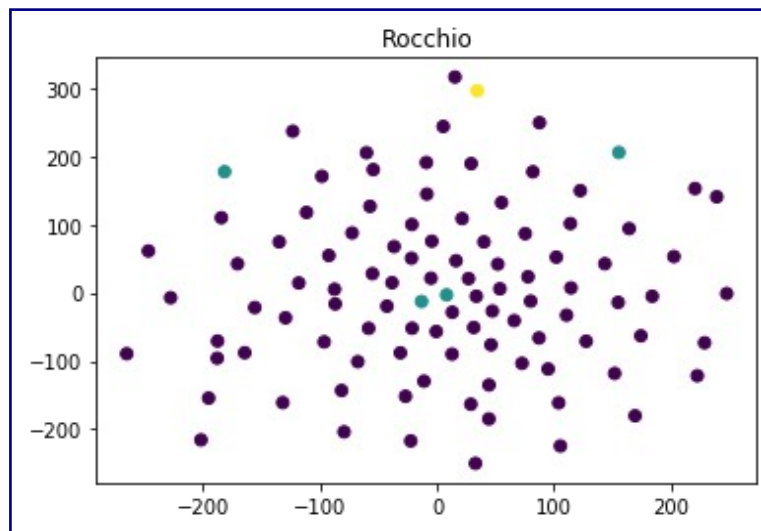
### PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

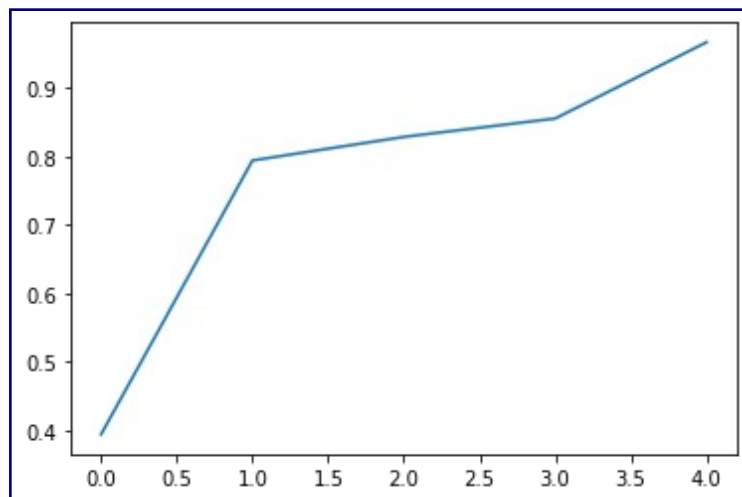


## TSNE



- Non relevant documents vector
- Relevant documents vector
- Query Vector

## Map



**Query :** *Scientific tools for preserving rights and body.*

**top k :** 100

**Ground Truth :** Documents inside folder talk.politics.misc

- **Normal query results:**

*Enter number of relevant docs:*

13,18,26,30,31,50,56,57,59,64,72,77,78,79,84,90,91,97,98

MAP: 0.15091033968743253

```

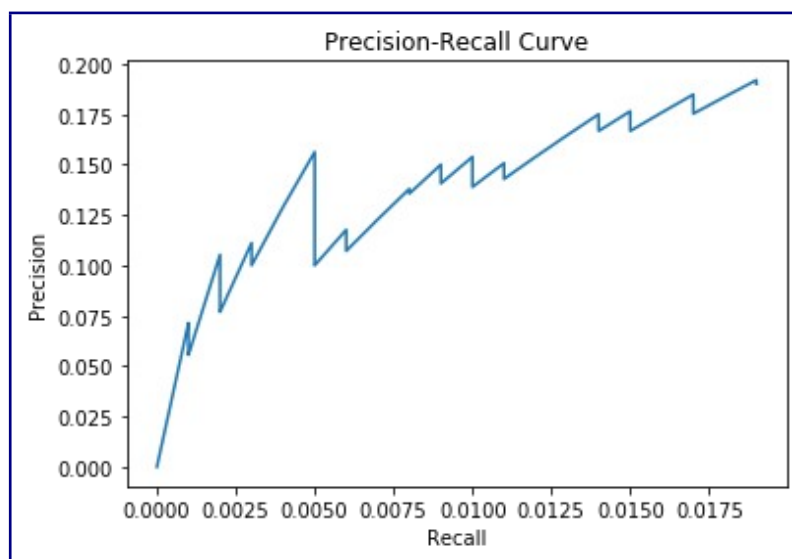
83 1299 sci.med 58941
84 2191 talk.politics.misc 178713
85 1739 sci.med 59587
86 1716 sci.med 58867
87 777 comp.graphics 39736
88 1897 sci.med 59023
89 4283 sci.space 61425
90 2161 talk.politics.misc 178877
91 2195 talk.politics.misc 178776
92 1956 sci.med 58938
93 4210 sci.space 61273
94 949 comp.graphics 38912
95 4433 sci.space 61278
96 4075 sci.space 61413
97 2669 talk.politics.misc 178879
98 2133 talk.politics.misc 176858
99 371 comp.graphics 38845
Enter number of relevant docs: 13,18,26,30,31,50,56,57,59,64,72,77,78,79,84,90,91,97,98
MAP: 0.15091033968743253

```

Relevant documents: 19

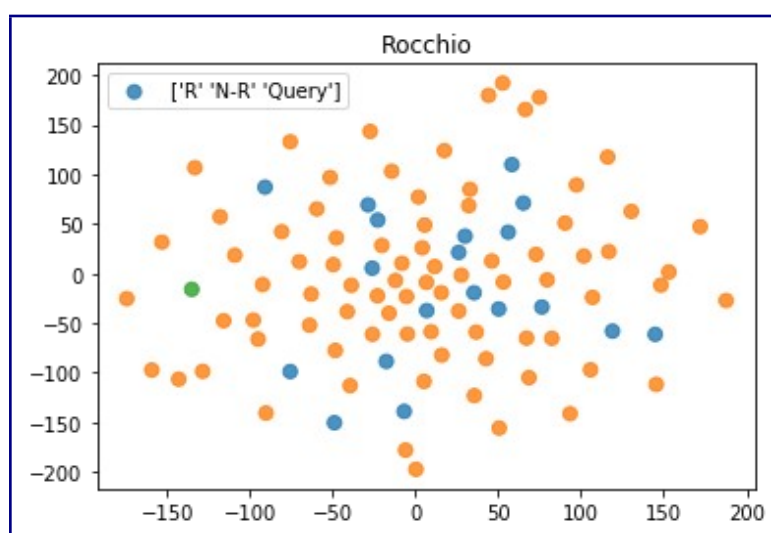
Non-Relevant documents: 8

### PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

### TSNE



- **Non-Relevant documents vector**
- **Relevant documents vector**
- **Query Vector**
- **Modified Query Results:**

*Marked Relevant documents by user as “\*”:*

```
iteration 1
Top docs:
0 1024 sci.med 59589
1 2526 talk.politics.misc 178540 *
2 1855 sci.med 59434
3 4079 sci.space 61335
4 1757 sci.med 59137
5 1802 sci.med 58569
6 1975 sci.med 58900
7 4129 sci.space 61307
8 275 comp.graphics 38375
9 243 comp.graphics 38852
10 4239 sci.space 60820
11 4914 sci.space 61385
12 4493 sci.space 60809
13 4411 sci.space 61079
14 4428 sci.space 60797
15 2138 talk.politics.misc 178738 *
16 4638 sci.space 61357
```

## Iteration-01

Enter number of relevant docs:

1, 15, 20, 21, 23, 28, 32, 37, 41, 42, 44, 45, 47, 52, 58, 59, 60, 61, 62, 74, 83, 84, 90

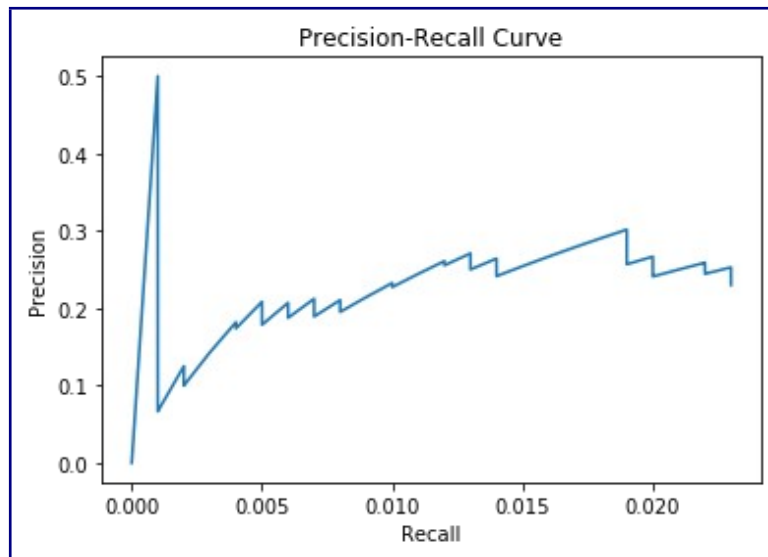
MAP: 0.27460739357221425

```
89 4973 sci.space 61145
90 2372 talk.politics.misc 178724
91 1259 sci.med 58941
92 4951 sci.space 60872
93 4025 sci.space 60229
94 4433 sci.space 61278
95 4754 sci.space 61487
96 4210 sci.space 61273
97 4075 sci.space 61413
98 1739 sci.med 59587
99 777 comp.graphics 39736
Enter number of relevant docs: 1,15,20,21,23,28,32,37,41,42,44,45,47,52,58,59,60,61,62,74,83,84,90
[2526, 2138, 2824, 2083, 2274, 2316, 2749, 2956, 2705, 2798, 2614, 2763, 2479, 2553, 2161, 2669, 2195, 2191, 213
3, 2913, 2884, 2985, 2372] [1024, 1026, 1795, 775, 777, 1802, 269, 1039, 1554, 275, 4887, 4638, 31, 4129, 1839,
4914, 1332, 4404, 4411, 1084, 1342, 575, 1855, 1603, 1094, 4428, 4433, 4434, 4183, 1367, 4951, 4699, 1371, 1376,
1383, 4973, 4210, 898, 1925, 138, 4493, 1934, 4239, 4754, 1691, 1435, 1190, 1711, 433, 1969, 4019, 1975, 4025, 4
283, 4284, 4540, 961, 1731, 713, 1739, 1235, 1751, 1757, 1507, 1765, 4838, 232, 1770, 1259, 4075, 4079, 4338, 24
3, 756, 1526, 1274, 1787]
MAP: 0.27460739357221425
```

Relevant documents: 23

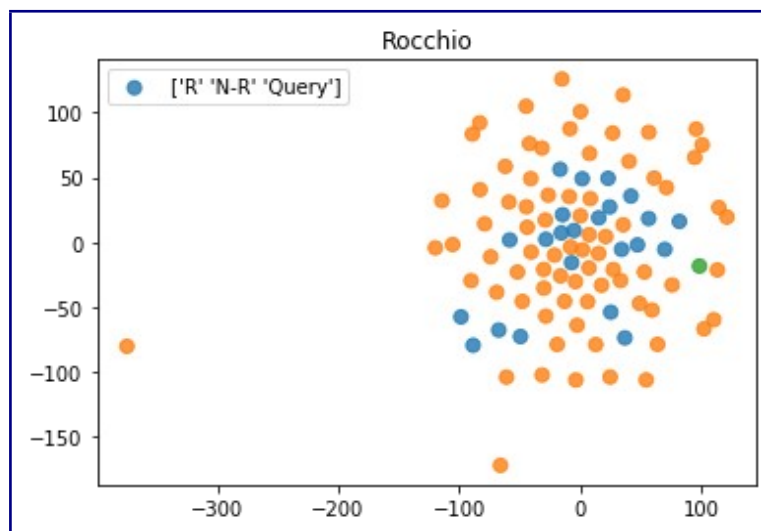
Non-Relevant documents: 77

## *PR-Curve*



- PR curve is a plot taking relevant and retrieved as ground truth.

## *TSNE*



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector

## Iteration-02

Enter number of relevant docs:

1, 8, 12, 13, 14, 19, 21, 23, 24, 26, 28, 31, 32, 33, 34, 35, 37, 39, 41, 42, 44, 50, 56, 71, 73, 82, 83, 87, 90, 92, 93, 95, 97

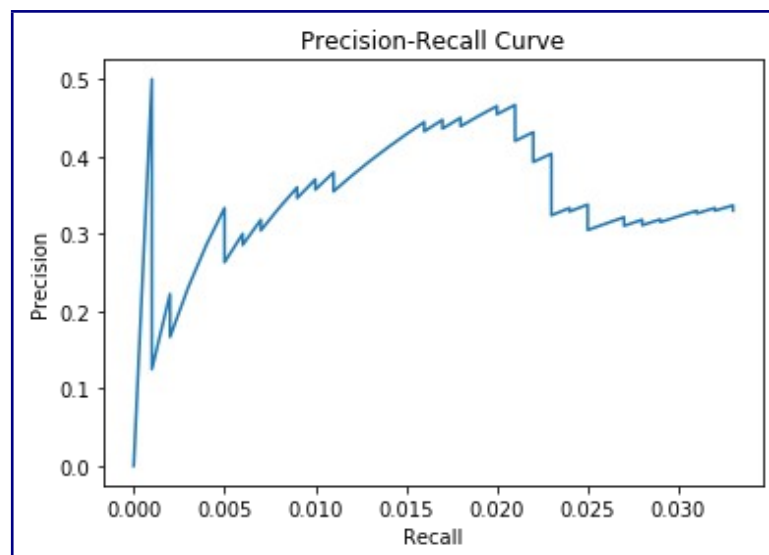
MAP: 0.39158040418568274

```
09 4433 sci.space 61278
90 2568 talk.politics.misc 177014
91 1367 sci.med 58880
92 2993 talk.politics.misc 176859
93 2650 talk.politics.misc 178329
94 4404 sci.space 60938
95 2329 talk.politics.misc 178519
96 4640 sci.space 61094
97 2818 talk.politics.misc 178698
98 1839 sci.med 58958
99 232 comp.graphics 38272
Enter number of relevant docs: 1,8,12,13,14,19,21,23,24,26,28,31,32,33,34,35,37,39,41,42,44,50,56,71,73,82,83,87,90,92,93,95,97
[2526, 2138, 2824, 2083, 2274, 2316, 2749, 2956, 2798, 2705, 2614, 2479, 2669, 2195, 2161, 2133, 2553, 2763, 2191, 2913, 2372, 2884, 2985, 2472, 2762, 2786, 2743, 2600, 2568, 2993, 2650, 2329, 2818] [1024, 1026, 1795, 898, 1925, 775, 1802, 4493, 1934, 1039, 4239, 1554, 4754, 275, 4887, 1691, 1435, 4638, 31, 4640, 4129, 1190, 4395, 1711, 1839, 433, 4914, 1969, 1332, 4404, 4019, 1975, 4283, 1084, 4284, 4411, 575, 1342, 1855, 4540, 961, 1094, 713, 4428, 4433, 4434, 1235, 4183, 1751, 1367, 4699, 1371, 1757, 1376, 1507, 1765, 4838, 1383, 232, 1770, 4079, 4338, 243, 756, 1526, 1274, 1787]
MAP: 0.39158040418568274
```

Relevant documents: 33

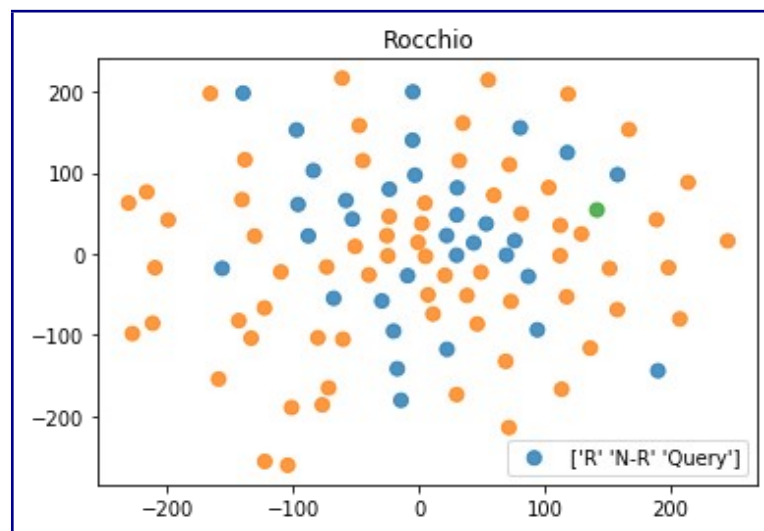
Non-Relevant documents: 67

### PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector

### Iteration-03

Enter number of relevant docs:

0, 5, 7, 8, 10, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 33, 34, 35, 38, 40, 45, 47, 48, 50, 52, 53, 54, 60, 63, 66, 67, 70, 73, 75, 76, 78, 79, 80, 81, 84, 89, 92

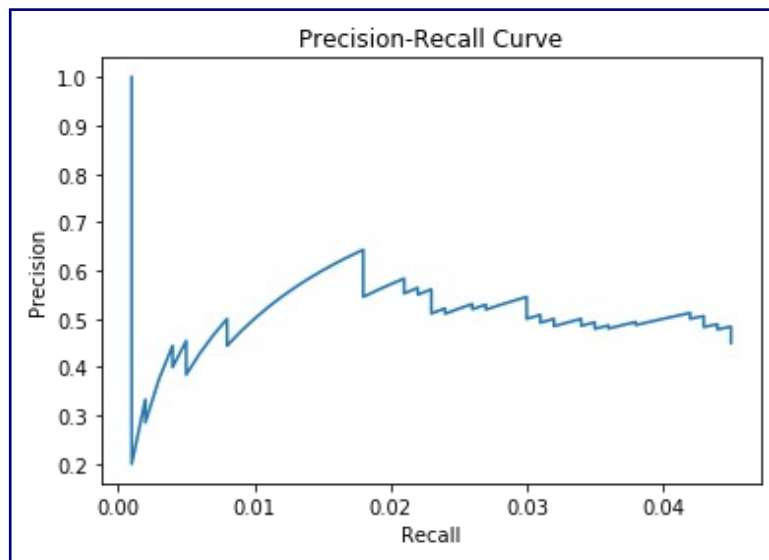
MAP: 0.48437564457629223

```
89 2043 talk.politics.misc 177021
90 1371 sci.med 58840
91 1342 sci.med 59596
92 2694 talk.politics.misc 179114
93 433 comp.graphics 38484
94 4433 sci.space 61278
95 4283 sci.space 61425
96 756 comp.graphics 38409
97 1332 sci.med 58108
98 1094 sci.med 59372
99 4139 sci.space 61110
Enter number of relevant docs: 0,5,7,8,10,13,14,15,18,19,20,21,22,23,24,25,26,27,33,34,35,38,40,45,47,48,50,52,5
3,54,60,63,66,67,70,73,75,76,78,79,80,81,84,89,92
[2526, 2138, 2824, 2083, 2274, 2316, 2798, 2749, 2956, 2705, 2614, 2195, 2669, 2913, 2133, 2161, 2479, 2372, 255
3, 2191, 2763, 2884, 2985, 2472, 2762, 2743, 2600, 2786, 2650, 2993, 2818, 2949, 2075, 2568, 2072, 2321, 2329, 2
085, 2986, 2336, 2975, 2213, 2878, 2043, 2694] [1024, 1026, 1795, 775, 1802, 4493, 1934, 1039, 4239, 4754, 275,
1691, 1435, 4638, 4129, 1190, 4139, 4395, 1711, 433, 4914, 4019, 1332, 1975, 4283, 4284, 1084, 4411, 1855, 1342,
4540, 961, 1094, 713, 4428, 4433, 4434, 1235, 4183, 1751, 4699, 1371, 1757, 1376, 1507, 1765, 4838, 1383, 1770,
4079, 243, 756, 1526, 1274, 1787]
MAP: 0.48437564457629223
```

Relevant documents: 45

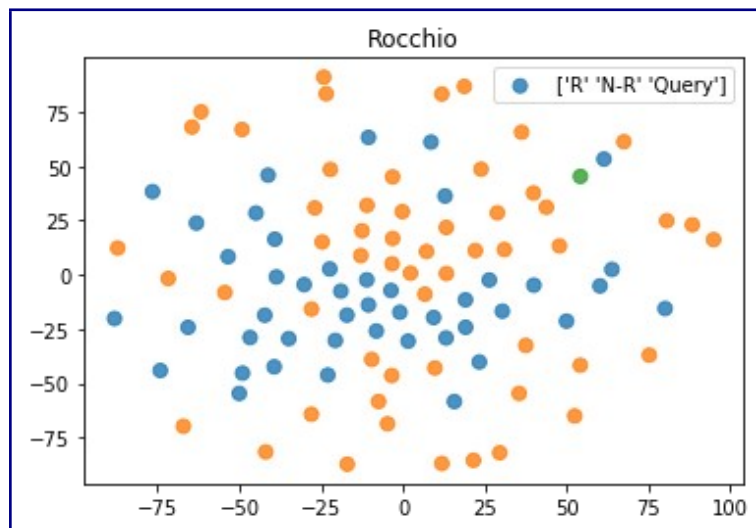
Non-Relevant documents: 55

## PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector



## Iteration-04

Enter number of relevant docs:

0, 4, 6, 7, 8, 9, 10, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 27, 28, 29, 30, 33, 34, 35, 36, 37, 39, 42, 44, 45, 46, 48, 51, 52, 53, 57, 58, 60, 61, 64, 65, 66, 68, 70, 73, 76, 79, 85, 86, 87, 88, 89, 90, 92, 94, 95, 96, 97, 98, 99

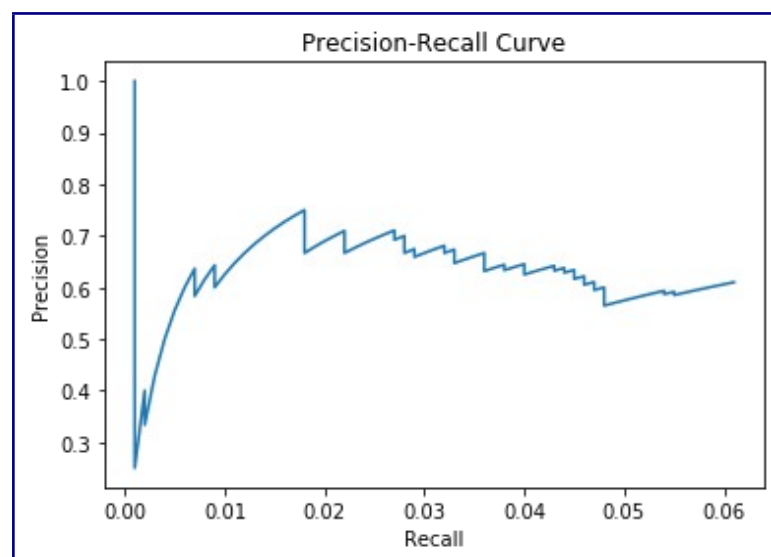
MAP: 0.6106948895469243

```
90 2918 talk.politics.misc 177004
91 4139 sci.space 61110
92 2628 talk.politics.misc 178341
93 1795 sci.med 59371
94 2464 talk.politics.misc 178298
95 2177 talk.politics.misc 178584
96 2168 talk.politics.misc 178406
97 2111 talk.politics.misc 178903
98 4433 sci.space 61278
99 2369 talk.politics.misc 176968
Enter number of relevant docs: 0,4,6,7,8,9,10,12,13,15,16,17,18,19,20,21,22,23,27,28,29,30,33,34,35,36,37,39,42,
44,45,46,48,51,52,53,57,58,60,61,64,65,66,68,70,73,76,79,85,86,87,88,89,90,92,94,95,96,97,98,99
[2526, 2824, 2138, 2083, 2274, 2798, 2316, 2749, 2705, 2372, 2913, 2195, 2956, 2614, 2133, 2161, 2669, 2479, 255
3, 2191, 2884, 2763, 2472, 2762, 2743, 2985, 2600, 2650, 2993, 2949, 2786, 2075, 2818, 2321, 2072, 2986, 2568, 2
878, 2336, 2975, 2329, 2085, 2043, 2213, 2694, 2261, 2102, 2810, 2346, 2286, 2547, 2351, 2173, 2918, 2628, 2464,
2177, 2168, 2111, 4433, 2369] [1024, 1026, 1795, 1802, 4493, 1934, 1039, 4239, 4754, 275, 4638, 4129, 4139, 439
5, 4914, 4019, 1975, 4411, 4284, 1084, 4540, 1855, 961, 4428, 4434, 1235, 1751, 4183, 4699, 1757, 1376, 1765, 48
38, 1383, 4079, 243, 1526, 1274, 1787]
MAP: 0.6106948895469243
```

Relevant documents: 61

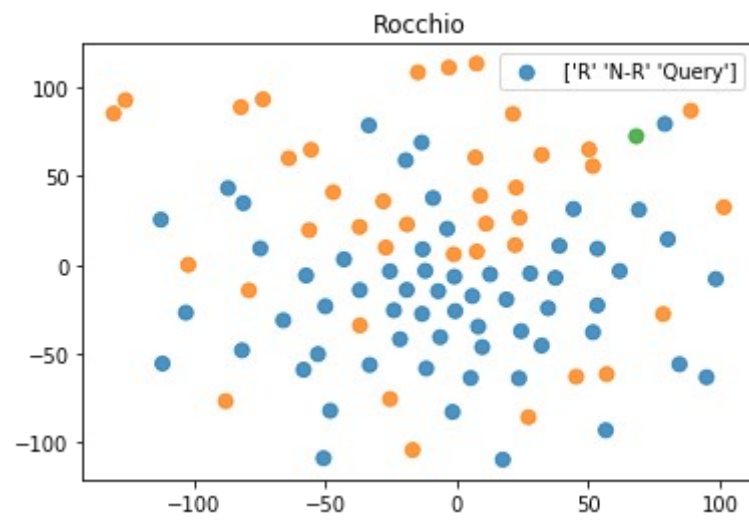
Non-Relevant documents: 39

### PR-Curve



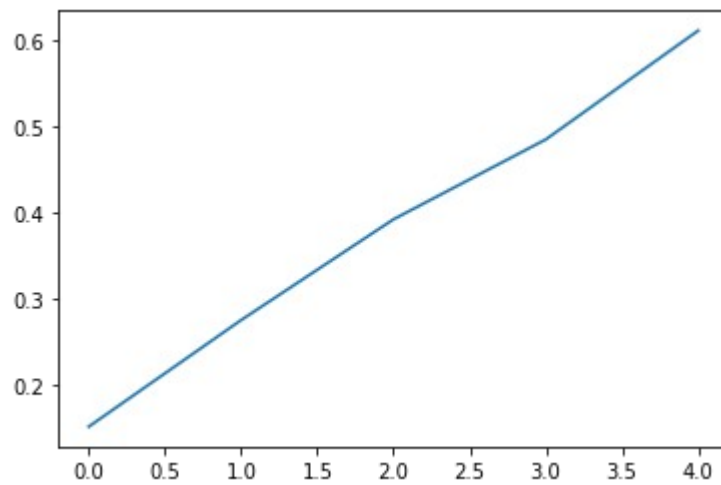
- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector

## MAP



**Query:** *Frequently asked questions on State-of-the-art visualisation tools*  
**top k :** 100

**Ground Truth:** *Documents inside folder sci.med*

- **Normal query results:**

Enter number of relevant docs:

0,3,4,6,9,12,23,38,39,40,42,45,46,47,54,67,71,72,73,74,75,77,78,80,81,  
84,85,91,99

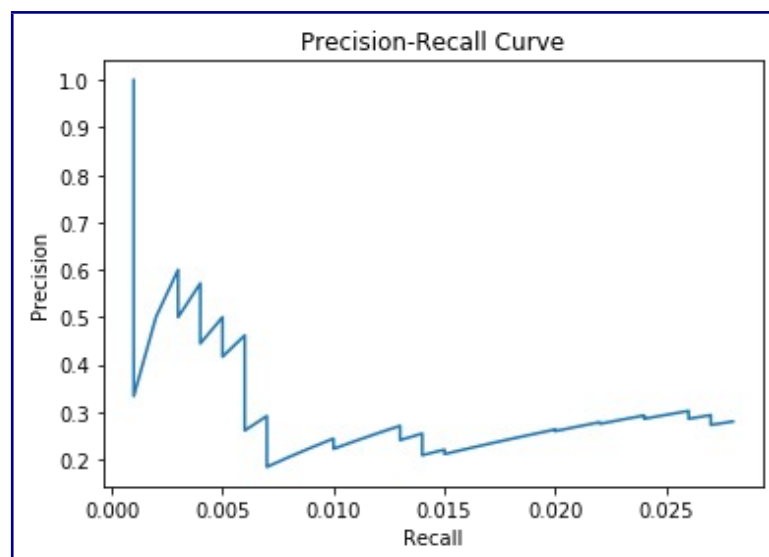
MAP: 0.280203245055495

```
83 3073 rec.sport.hockey 54130
84 1367 sci.med 58880
85 1584 sci.med 59185
86 2425 talk.politics.misc 178550
87 758 comp.graphics 38824
88 2600 talk.politics.misc 178695
89 294 comp.graphics 38434
90 4973 sci.space 61145
91 1257 sci.med 58844
92 3528 rec.sport.hockey 53908
93 508 comp.graphics 38579
94 4292 sci.space 60846
95 815 comp.graphics 39065
96 2336 talk.politics.misc 179058
97 4328 sci.space 61074
98 635 comp.graphics 38733
99 1598 sci.med 59079
Enter number of relevant docs: 0,3,4,6,9,12,23,38,39,40,45,46,47,54,67,71,72,73,74,75,77,78,80,81,84,85,91,99
MAP: 0.2723270987821137
```

Relevant documents: 29

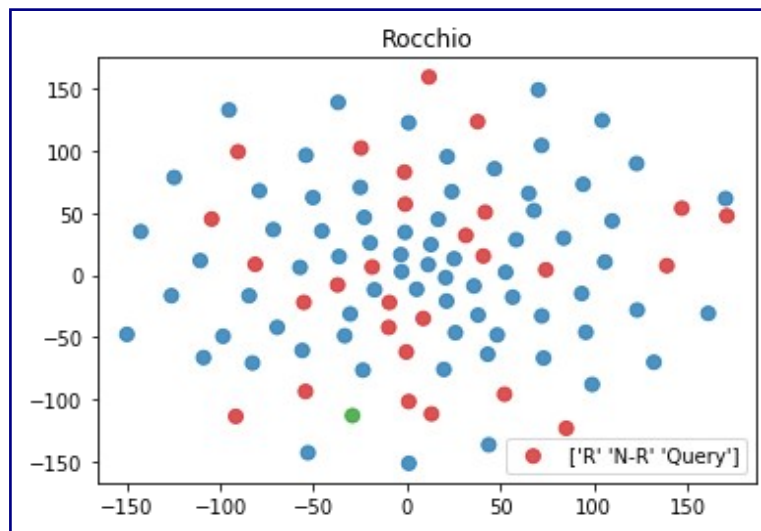
Non-Relevant documents: 71

### *PR-Curve*



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector

### Modified Query Results:

Marked Relevant documents by user as '\*':

```
iteration 1
Top docs:
0 1855 sci.med 59434 *
1 3073 rec.sport.hockey 54130
2 4079 sci.space 61335
3 1724 sci.med 59324 *
4 775 comp.graphics 37919
5 1435 sci.med 58052 *
6 961 comp.graphics 37920
7 1934 sci.med 58053 *
8 1802 sci.med 58569 *
9 490 comp.graphics 38962
10 2526 talk.politics.misc 178540
11 1660 sci.med 59370 *
12 4239 sci.space 60820
13 4493 sci.space 60809
14 4428 sci.space 60797
15 275 comp.graphics 38375
16 243 comp.graphics 38852
17 4411 sci.space 61070
```

### Iteration-01

Enter number of relevant docs:

0, 3, 5, 7, 8, 11, 18, 26, 29, 30, 38, 39, 40, 44, 45, 47, 48, 49, 50, 51, 52, 53, 54, 58, 59, 65, 66, 68, 70, 77

MAP: 0.34741462856449307

```

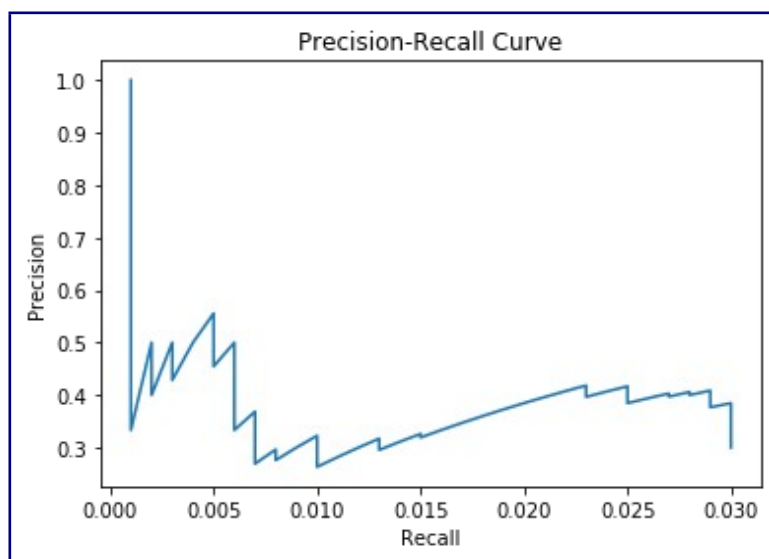
90 817 comp.graphics 38311
91 2600 talk.politics.misc 178695
92 932 comp.graphics 38853
93 203 comp.graphics 38376
94 2585 talk.politics.misc 178761
95 294 comp.graphics 38434
96 906 comp.graphics 38778
97 4292 sci.space 60846
98 758 comp.graphics 38824
99 3528 rec.sport.hockey 53908
Enter number of relevant docs: 0,3,5,7,8,11,18,26,29,30,38,39,40,44,45,47,48,49,50,51,52,53,54,58,59,65,66,68,7
0,77
[1855, 1724, 1435, 1934, 1802, 1660, 1466, 1650, 1028, 1088, 1584, 1958, 1908, 1839, 1414, 1527, 1080, 1170, 159
8, 1198, 1338, 1133, 1286, 1678, 1201, 1770, 1923, 1635, 1367, 1257] [3073, 2054, 2695, 775, 137, 522, 138, 423
3, 777, 2316, 4239, 2191, 529, 906, 275, 4493, 4373, 2961, 920, 2585, 2205, 544, 2336, 548, 164, 294, 2424, 260
0, 932, 2988, 2479, 4401, 4914, 817, 433, 4019, 4411, 4284, 961, 4292, 4677, 836, 711, 199, 713, 4936, 203, 442
8, 4941, 3528, 82, 4434, 597, 350, 2526, 352, 4705, 2913, 232, 106, 490, 508, 4079, 243, 756, 758, 760, 2425, 28
10, 892]
MAP: 0.34741462856449307

```

Relevant documents: 30

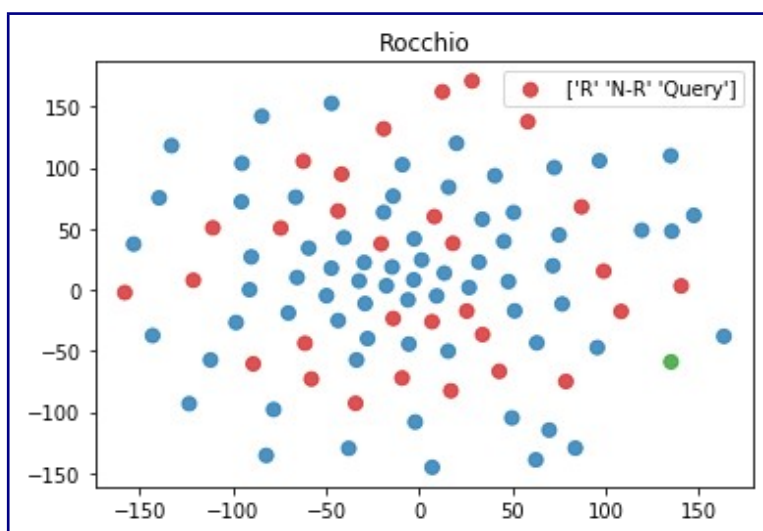
Non-Relevant documents: 70

### PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

### TSNE



- **Non-Relevant documents vector**
- **Relevant documents vector**
- **Query Vector**

## Iteration-02

Enter number of relevant docs:

0, 2, 4, 7, 8, 9, 12, 13, 14, 17, 18, 22, 24, 27, 28, 30, 31, 32, 34, 38, 39, 41, 42, 44, 47, 46, 53, 54, 56, 61, 79, 84, 89, 93

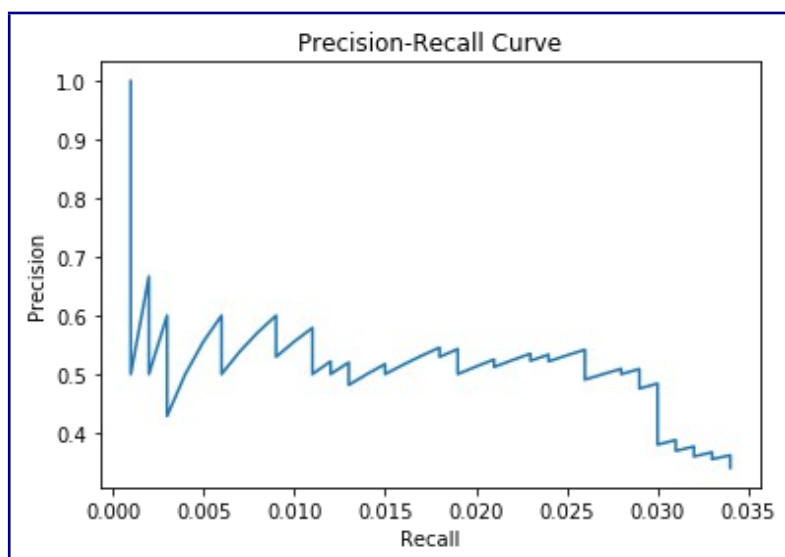
MAP: 0.4850877659457784

```
90 100 comp.graphics 38692
91 760 comp.graphics 38627
92 548 comp.graphics 38820
93 1588 sci.med 59481
94 777 comp.graphics 39736
95 2479 talk.politics.misc 179047
96 653 comp.graphics 38848
97 920 comp.graphics 38497
98 529 comp.graphics 38822
99 294 comp.graphics 38434
Enter number of relevant docs: 0,2,4,7,8,9,12,13,14,17,18,22,24,27,28,30,31,32,34,38,39,41,42,44,47,46,53,54,56,
61,79,84,89,93
[1855, 1724, 1435, 1934, 1660, 1802, 1466, 1650, 1584, 1958, 1028, 1598, 1088, 1080, 1923, 1908, 1414, 1839, 119
8, 1286, 1527, 1133, 1338, 1678, 1201, 1170, 1367, 1770, 1635, 1257, 1161, 1460, 1788, 1588] [3073, 2054, 2695,
775, 137, 522, 4233, 138, 653, 777, 4239, 2191, 529, 2316, 275, 4493, 4373, 2961, 920, 2205, 544, 2336, 548, 16
4, 294, 2424, 2600, 932, 2988, 2479, 4401, 4914, 433, 4019, 4411, 4284, 961, 836, 4677, 711, 199, 713, 4936, 20
3, 4428, 4941, 906, 82, 4434, 597, 350, 2526, 352, 4705, 2913, 232, 106, 490, 508, 4079, 243, 756, 760, 2425, 28
10, 892]
MAP: 0.4791653749368101
```

Relevant documents: 34

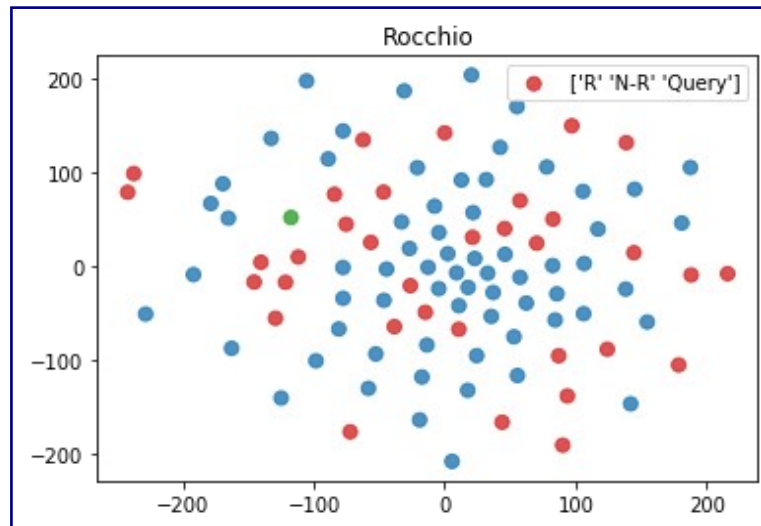
Non-Relevant documents: 66

## *PR-Curve*



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- **Non-Relevant documents vector**
- **Relevant documents vector**
- **Query Vector**

### Iteration-03

Enter number of relevant docs:

0, 2, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, 19, 20, 21, 22, 23, 24, 25, 30, 31, 32, 33, 34, 36, 42, 44, 45, 50, 51, 53, 67, 68, 72, 79, 82, 83, 89, 90, 92, 93, 96, 97, 99

MAP: 0.5857219686360018

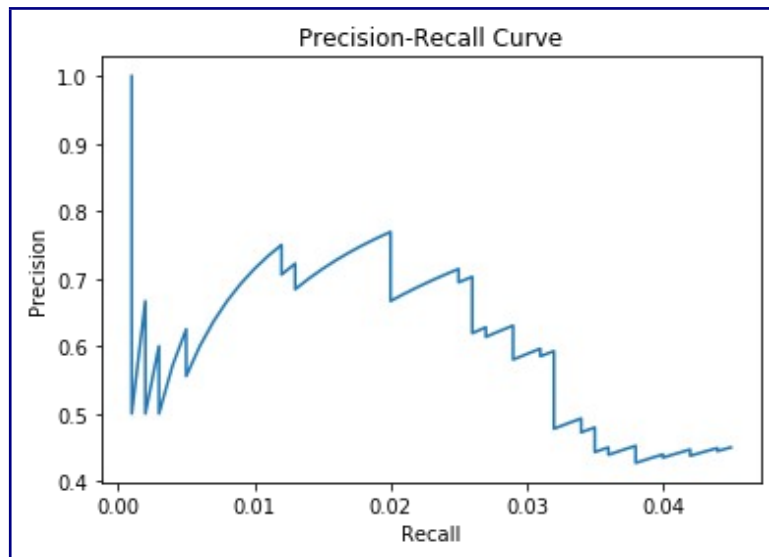
```
90 1506 sci.med 58947
91 433 comp.graphics 38484
92 1888 sci.med 59183
93 1095 sci.med 58970
94 2054 talk.politics.misc 178343
95 2600 talk.politics.misc 178695
96 1619 sci.med 59432
97 1883 sci.med 58907
98 352 comp.graphics 38648
99 1034 sci.med 59261
Enter number of relevant docs: 0,2,4,6,7,9,10,11,12,13,14,15,17,19,20,21,22,23,24,25,30,31,32,33,34,36,42,44,45,
50,51,53,67,68,72,79,82,83,89,90,92,93,96,97,99
[1855, 1724, 1435, 1934, 1660, 1584, 1650, 1958, 1802, 1466, 1598, 1923, 1028, 1080, 1088, 1198, 1414, 1908, 128
6, 1839, 1133, 1201, 1338, 1678, 1170, 1527, 1367, 1770, 1635, 1257, 1460, 1161, 1639, 1788, 1588, 1850, 1394, 1
321, 1020, 1506, 1888, 1095, 1619, 1883, 1034] [3073, 2054, 2695, 775, 137, 522, 4233, 2316, 906, 4493, 4239, 21
91, 2961, 275, 4373, 2205, 544, 2336, 164, 932, 2600, 2988, 4401, 4914, 433, 4019, 4411, 4284, 961, 836, 4677, 7
11, 199, 713, 4936, 203, 4428, 4941, 4434, 597, 350, 2526, 352, 4705, 2913, 232, 490, 508, 4079, 243, 756, 2424,
2425, 2810, 892]
MAP: 0.5780402720257919
```

Relevant documents: 45

Non-Relevant documents: 55

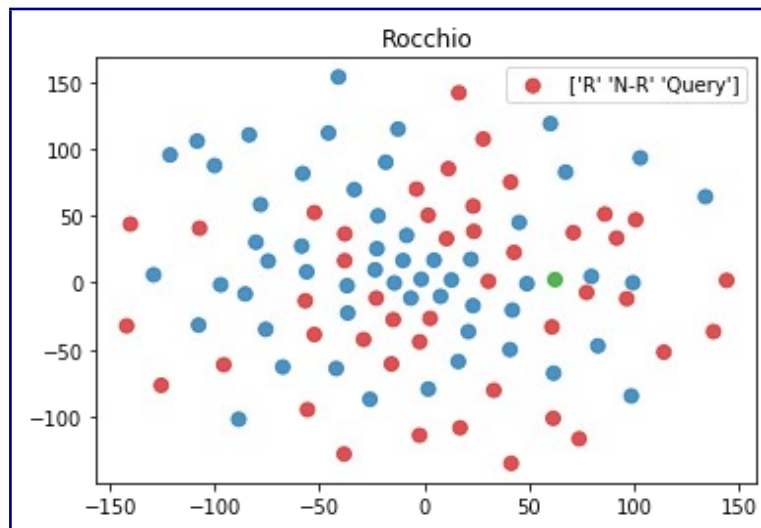


## PR-Curve



- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- **Non-Relevant documents vector**
- **Relevant documents vector**
- **Query Vector**

## Iteration-04

Enter number of relevant docs:

0, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 33, 34, 35, 38, 43, 45, 47, 52, 53, 54, 57, 59, 63, 65, 66, 67, 72, 74, 76, 80, 85, 90, 92, 97, 99

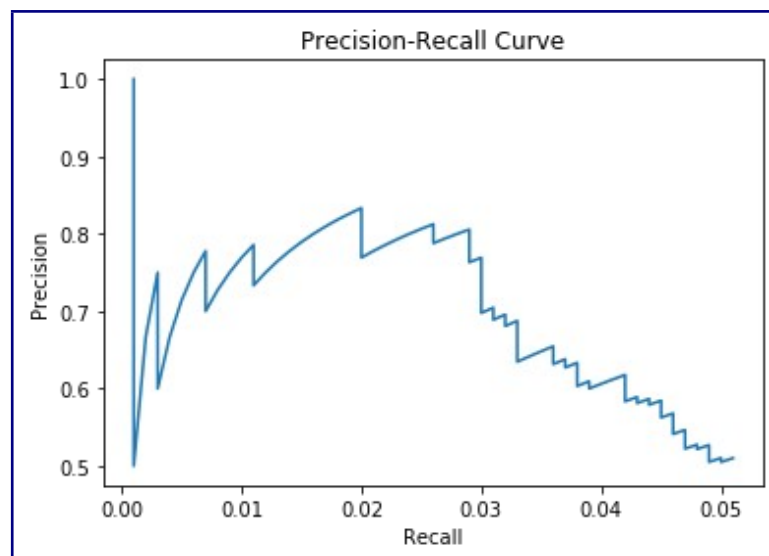
MAP: 0.683057937530745

```
90 1534 sci.med 59087
91 836 comp.graphics 37926
92 1476 sci.med 59430
93 892 comp.graphics 38906
94 4233 sci.space 61241
95 597 comp.graphics 38496
96 2547 talk.politics.misc 178786
97 1365 sci.med 58578
98 2605 talk.politics.misc 178653
99 1719 sci.med 58897
Enter number of relevant docs: 0,2,3,5,6,7,8,10,11,12,13,15,16,17,18,19,20,21,22,23,26,27,28,29,30,31,33,34,35,3
8,43,45,47,52,53,54,57,59,63,65,66,67,72,74,76,80,85,90,92,97,99
[1855, 1724, 1584, 1435, 1650, 1958, 1660, 1934, 1598, 1466, 1923, 1028, 1802, 1080, 1198, 1088, 1414, 1201, 128
6, 1908, 1839, 1133, 1338, 1170, 1678, 1527, 1460, 1367, 1635, 1770, 1161, 1639, 1257, 1394, 1020, 1788, 1588, 1
850, 1095, 1888, 1672, 1321, 1883, 1619, 1034, 1506, 1946, 1534, 1476, 1365, 1719] [3073, 2695, 775, 137, 522, 4
233, 2316, 906, 4493, 4239, 2191, 2961, 275, 4373, 2205, 544, 2336, 164, 932, 2988, 2605, 4401, 4914, 4019, 441
1, 4284, 961, 836, 4677, 711, 4936, 203, 4428, 4941, 597, 350, 2526, 4705, 2913, 232, 490, 508, 4079, 243, 2547,
2424, 2425, 2810, 892]
MAP: 0.6737922199438304
```

Relevant documents: 51

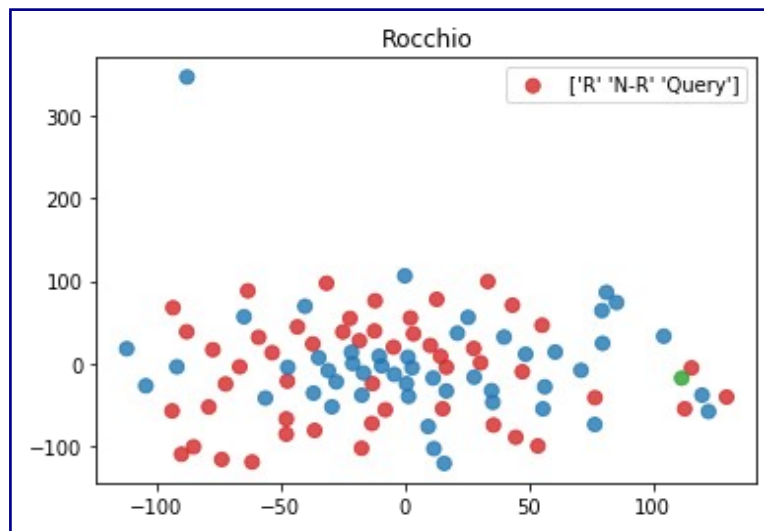
Non-Relevant documents: 49

### PR-Curve



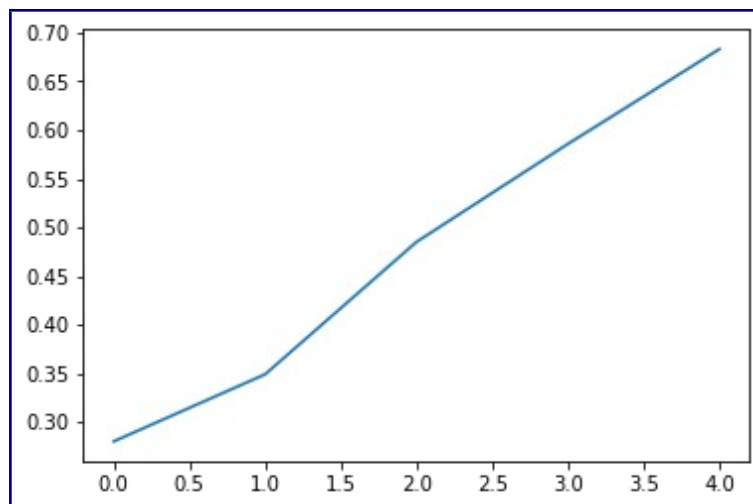
- PR curve is a plot taking relevant and retrieved as ground truth.

## TSNE



- Non-Relevant documents vector
- Relevant documents vector
- Query Vector

## Map



### **MAP For all iterations of 3 Queries:**

Query	Initial	Iteration 1	Iteration2	Iteration 3	Iteration 4	MAP of single query
Query 1	0.39444887 04503613	0.793684962 9161259	0.828072067 9368853	0.854878382 0401292	0.9660819279 184428	0.7674332422 52389
Query 2	0.150910339 68743253	0.27460739 357221425	0.39158040 418568274	0.48437564 457629223	0.610694889 5469243	0.3824337343 137092
Query 3	0.28020232 45055495	0.34909650 7523992	0.48508776 59457784	0.58572196 86360018	0.683057937 530745	0.4766333008 284134
MAP among different queries	0.27518717 821444777	0.47246295 46707774	0.56824674 60227821	0.64165866 5084141	0.753278251 6653707	

### **Observations:**

- After each iteration query vector tends to moves towards relevant documents.
- For all 3 queries, Average Precision tends to increases in each iteration after the user feedback.
- For all the modified queries also, Average Precision increases.
- MAP for each iteration also increases with number of iterations in this case. Too many iterations may result in decrease of MAP value but that is not the case here.
- Best results are shown for first query.
- With each iteration precision also increases after user feedback. Results are shown using PR-Curve.