Topic Recommender

Parameter

Topic cutoff = 10

Num of neighbours = 5

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k10_n5/")
sr_r1 <-  read.csv("SR_Round1", sep="\t", header = F)
sr_r2 <-  read.csv("SR_Round2", sep="\t", header = F)
sr_r3 <-  read.csv("SR_Round3", sep="\t", header = F)
sr_r4 <-  read.csv("SR_Round4", sep="\t", header = F)
sr_r5 <-  read.csv("SR_Round5", sep="\t", header = F)
sr_r6 <-  read.csv("SR_Round6", sep="\t", header = F)
sr_r7 <-  read.csv("SR_Round7", sep="\t", header = F)
sr_r8 <-  read.csv("SR_Round8", sep="\t", header = F)
sr_r9 <-  read.csv("SR_Round9", sep="\t", header = F)
sr_r10 <-  read.csv("SR_Round10", sep="\t", header = F)
sr_tot <- rbind(sr_r1, sr_r2, sr_r3, sr_r4, sr_r5, sr_r6, sr_r7, sr_r8, sr_r9, sr_r10)
sr_tot$V3 <- NULL
sr_tot <- sr_tot %>% group_by(sr_tot$V1) %>% summarise(mean = mean(V2))

sr_tot <- sr_tot %>% rename(k = `sr_tot$V1`)
sr_tot
```

```
## # A tibble: 20 x 2
##        k  mean
##    <int> <dbl>
##  1     1 0.357
##  2     2 0.479
##  3     3 0.546
##  4     4 0.592
##  5     5 0.622
##  6     6 0.647
##  7     7 0.664
##  8     8 0.683
##  9     9 0.698
## 10    10 0.707
## 11    11 0.720
## 12    12 0.722
## 13    13 0.721
## 14    14 0.709
## 15    15 0.688
```

```
## 16    16 0.669
## 17    17 0.630
## 18    18 0.583
## 19    19 0.527
## 20    20 0.471
```

```
sr_tot$mean %>% summary()
```

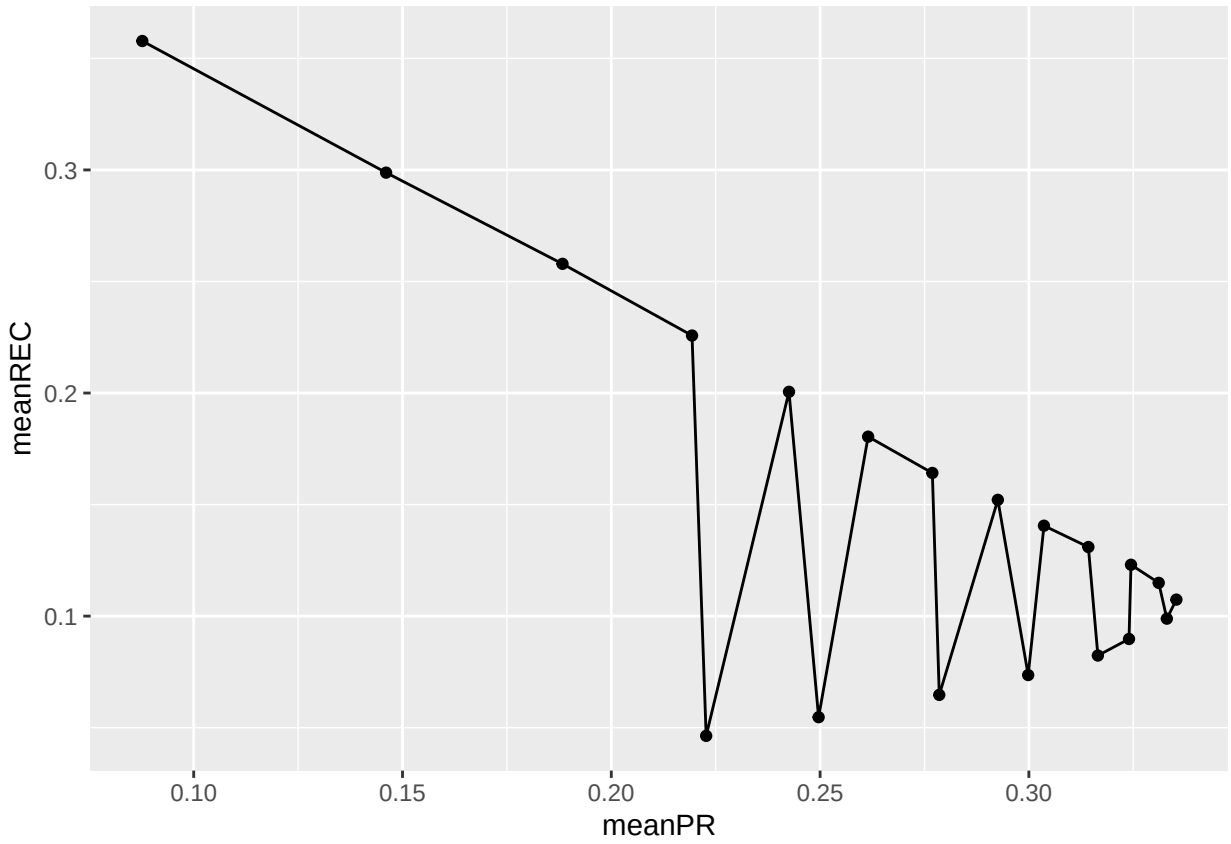
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3567  0.5740  0.6557  0.6218  0.7000  0.7221
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k10_n5/")
pr_r1 <-  read.csv("PRC_Round1", sep="\t", header = F)
pr_r2 <-  read.csv("PRC_Round2", sep="\t", header = F)
pr_r3 <-  read.csv("PRC_Round3", sep="\t", header = F)
pr_r4 <-  read.csv("PRC_Round4", sep="\t", header = F)
pr_r5 <-  read.csv("PRC_Round5", sep="\t", header = F)
pr_r6 <-  read.csv("PRC_Round6", sep="\t", header = F)
pr_r7 <-  read.csv("PRC_Round7", sep="\t", header = F)
pr_r8 <-  read.csv("PRC_Round8", sep="\t", header = F)
pr_r9 <-  read.csv("PRC_Round9", sep="\t", header = F)
pr_r10 <-  read.csv("PRC_Round10", sep="\t", header = F)
pr_tot <- rbind(pr_r1, pr_r2, pr_r3, pr_r4, pr_r5, pr_r6, pr_r7, pr_r8, pr_r9, pr_r10)
pr_tot <- pr_tot %>% group_by(pr_tot$V1) %>% summarise(meanPR = mean(V2), meanREC = mean(V3)) %>% renam
pr_tot
```

```
## # A tibble: 20 x 3
##        k meanPR meanREC
##    <int>  <dbl>   <dbl>
##  1     1 0.0877  0.358
##  2     2 0.146   0.299
##  3     3 0.188   0.258
##  4     4 0.219   0.226
##  5     5 0.243   0.201
##  6     6 0.262   0.180
##  7     7 0.277   0.164
##  8     8 0.293   0.152
##  9     9 0.304   0.141
## 10    10 0.314   0.131
## 11    11 0.324   0.123
## 12    12 0.331   0.115
## 13    13 0.335   0.107
## 14    14 0.333   0.0989
## 15    15 0.324   0.0898
## 16    16 0.317   0.0823
## 17    17 0.300   0.0736
## 18    18 0.279   0.0647
## 19    19 0.250   0.0547
## 20    20 0.223   0.0463
```

```
pr_tot %>% ggplot(aes(x=meanPR, y=meanREC)) + geom_line() + geom_point()
```



Precision summary

```
pr_tot$meanPR %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08769 0.23759 0.28555 0.26740 0.31839 0.33532
```

Recall summary

```
pr_tot$meanREC %>% summary()
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04633 0.08790 0.12700 0.14823 0.18548 0.35779
```