Topic Recommender

Parameter

Topic cutoff = 20 Num of neighbours = 10 Sim function = topic-based ## Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n10/")
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 0.384
##  2     2 0.516
##  3     3 0.592
##  4     4 0.647
##  5     5 0.688
##  6     6 0.718
##  7     7 0.736
##  8     8 0.753
##  9     9 0.767
## 10    10 0.779
## 11    11 0.793
## 12    12 0.802
## 13    13 0.812
## 14    14 0.819
## 15    15 0.822
## 16    16 0.827
## 17    17 0.824
## 18    18 0.822
## 19    19 0.816
## 20    20 0.811
```

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

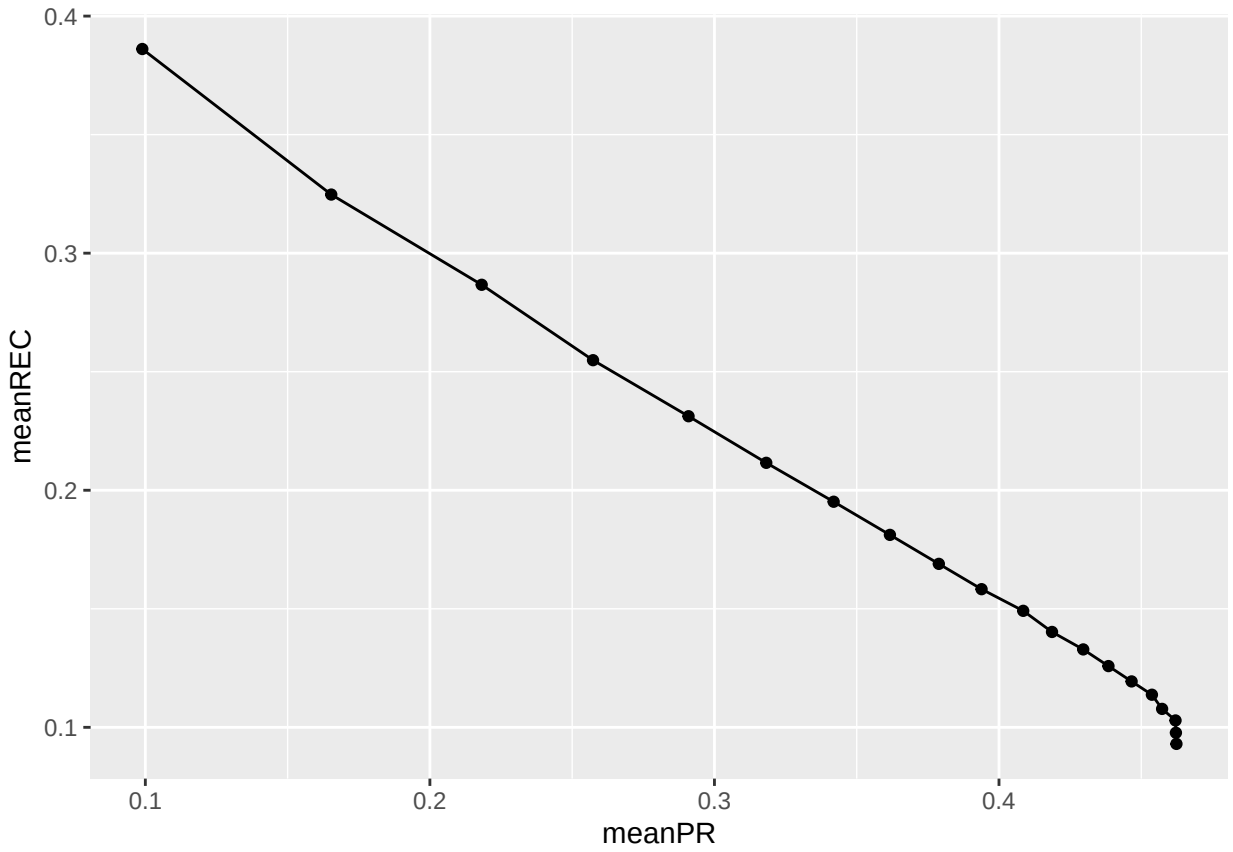
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3844  0.7104  0.7861  0.7364  0.8168  0.8272
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n10/")
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 0.0989  0.386
##  2     2 0.165   0.325
##  3     3 0.218   0.287
##  4     4 0.257   0.255
##  5     5 0.291   0.231
##  6     6 0.318   0.212
##  7     7 0.342   0.195
##  8     8 0.362   0.181
##  9     9 0.379   0.169
## 10    10 0.394   0.158
## 11    11 0.409   0.149
## 12    12 0.419   0.140
## 13    13 0.430   0.133
## 14    14 0.438   0.126
## 15    15 0.447   0.119
## 16    16 0.454   0.114
## 17    17 0.457   0.108
## 18    18 0.462   0.103
## 19    19 0.462   0.0977
## 20    20 0.462   0.0930
```

```
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.
## 0.09893 0.31138 0.40120 0.36324 0.44839 0.46236
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.09303 0.11797 0.15371 0.17907 0.21649 0.38611
```