Topic Recommender

Parameter

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

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n20/")
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 0.401
##  2     2 0.548
##  3     3 0.633
##  4     4 0.685
##  5     5 0.727
##  6     6 0.753
##  7     7 0.778
##  8     8 0.796
##  9     9 0.81
## 10    10 0.822
## 11    11 0.832
## 12    12 0.84
## 13    13 0.847
## 14    14 0.856
## 15    15 0.861
## 16    16 0.865
## 17    17 0.872
## 18    18 0.881
## 19    19 0.884
## 20    20 0.889
```

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

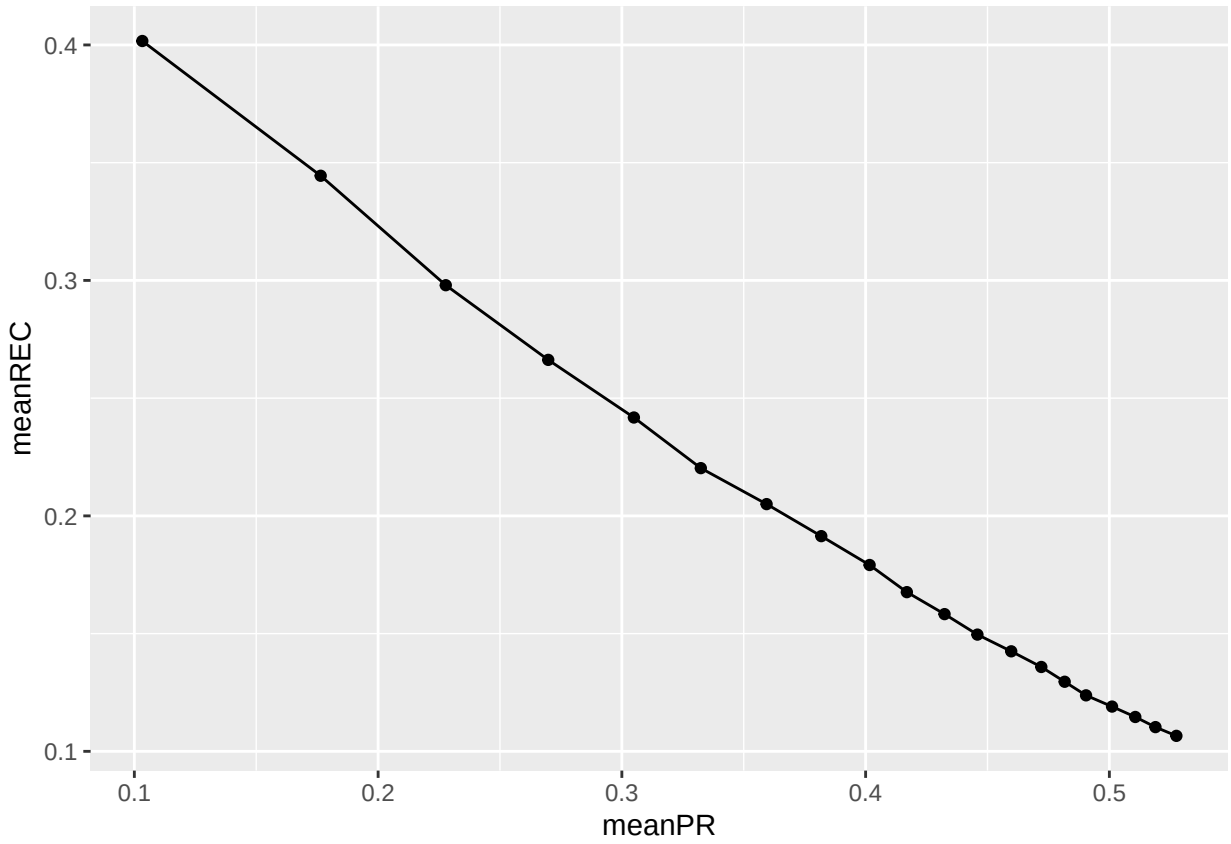
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4011  0.7464  0.8267  0.7789  0.8617  0.8894
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n20/")
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  0.103   0.402
##  2     2  0.176   0.344
##  3     3  0.228   0.298
##  4     4  0.270   0.266
##  5     5  0.305   0.242
##  6     6  0.332   0.220
##  7     7  0.359   0.205
##  8     8  0.382   0.191
##  9     9  0.402   0.179
## 10    10  0.417   0.168
## 11    11  0.432   0.158
## 12    12  0.446   0.150
## 13    13  0.460   0.142
## 14    14  0.472   0.136
## 15    15  0.482   0.130
## 16    16  0.490   0.124
## 17    17  0.501   0.119
## 18    18  0.511   0.115
## 19    19  0.519   0.110
## 20    20  0.527   0.107
```

```
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.
##  0.1033  0.3255  0.4247  0.3907  0.4839  0.5275
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1066  0.1281  0.1629  0.1903  0.2257  0.4017
```