Topic Recommender

Parameter

Topic cutoff = 20 Num of neighbours = 15 Sim function = topic-based ## Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n15/")
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 0.391
##  2     2 0.534
##  3     3 0.621
##  4     4 0.672
##  5     5 0.716
##  6     6 0.74
##  7     7 0.763
##  8     8 0.783
##  9     9 0.796
## 10    10 0.808
## 11    11 0.819
## 12    12 0.832
## 13    13 0.839
## 14    14 0.846
## 15    15 0.854
## 16    16 0.862
## 17    17 0.869
## 18    18 0.873
## 19    19 0.877
## 20    20 0.878
```

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

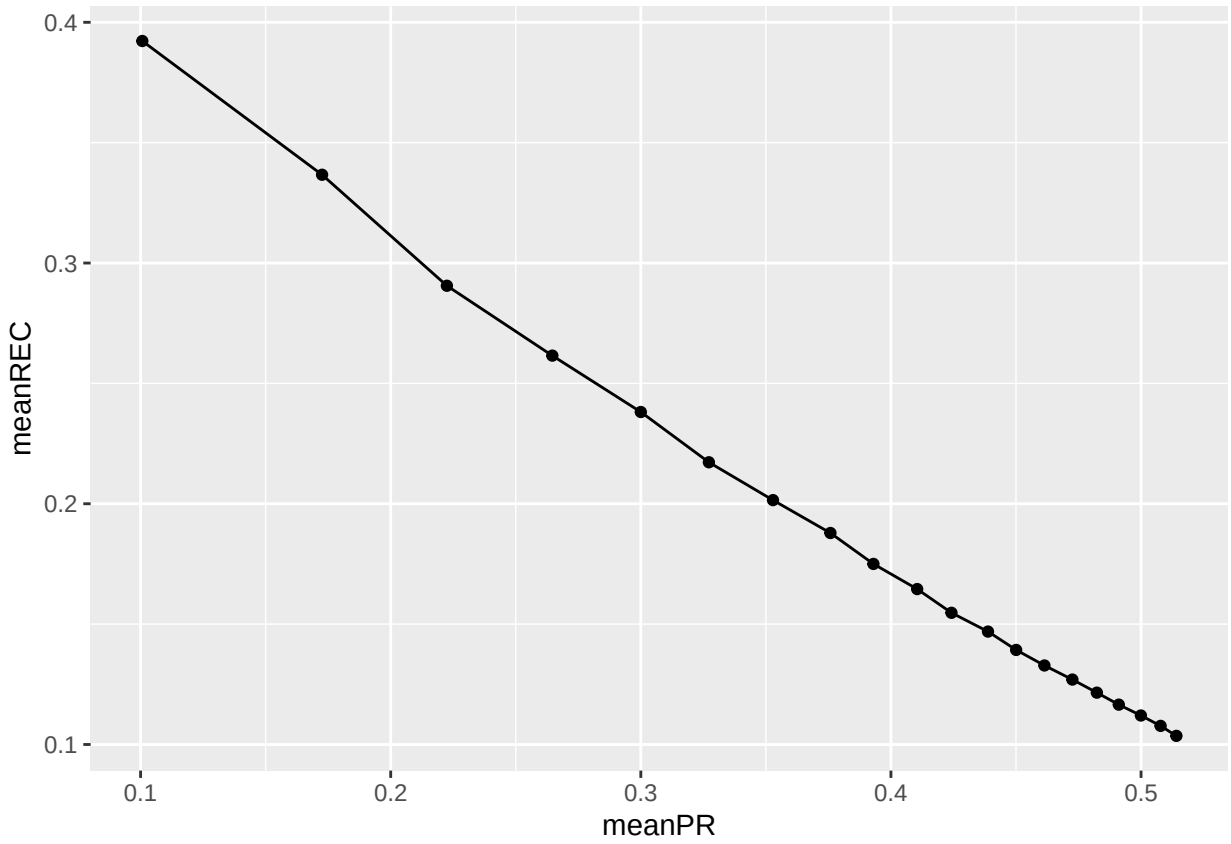
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3911  0.7340  0.8139  0.7687  0.8558  0.8783
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n15/")
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  0.101   0.392
##  2     2  0.173   0.337
##  3     3  0.222   0.291
##  4     4  0.265   0.262
##  5     5  0.300   0.238
##  6     6  0.327   0.217
##  7     7  0.353   0.202
##  8     8  0.376   0.188
##  9     9  0.393   0.175
## 10    10  0.410   0.165
## 11    11  0.424   0.155
## 12    12  0.439   0.147
## 13    13  0.450   0.139
## 14    14  0.461   0.133
## 15    15  0.473   0.127
## 16    16  0.482   0.122
## 17    17  0.491   0.117
## 18    18  0.500   0.112
## 19    19  0.508   0.108
## 20    20  0.514   0.104
```

```
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.
##  0.1007  0.3204  0.4174  0.3831  0.4750  0.5141
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1036  0.1256  0.1596  0.1864  0.2224  0.3922
```