Topic Recommender

Parameter

Topic cutoff = 20

Num of neighbours = 25

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n25/")
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 0.409
##  2     2 0.554
##  3     3 0.632
##  4     4 0.682
##  5     5 0.728
##  6     6 0.754
##  7     7 0.778
##  8     8 0.803
##  9     9 0.818
## 10    10 0.828
## 11    11 0.842
## 12    12 0.851
## 13    13 0.858
## 14    14 0.863
## 15    15 0.872
```

```
## 16    16 0.879
## 17    17 0.883
## 18    18 0.886
## 19    19 0.889
## 20    20 0.892
```

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

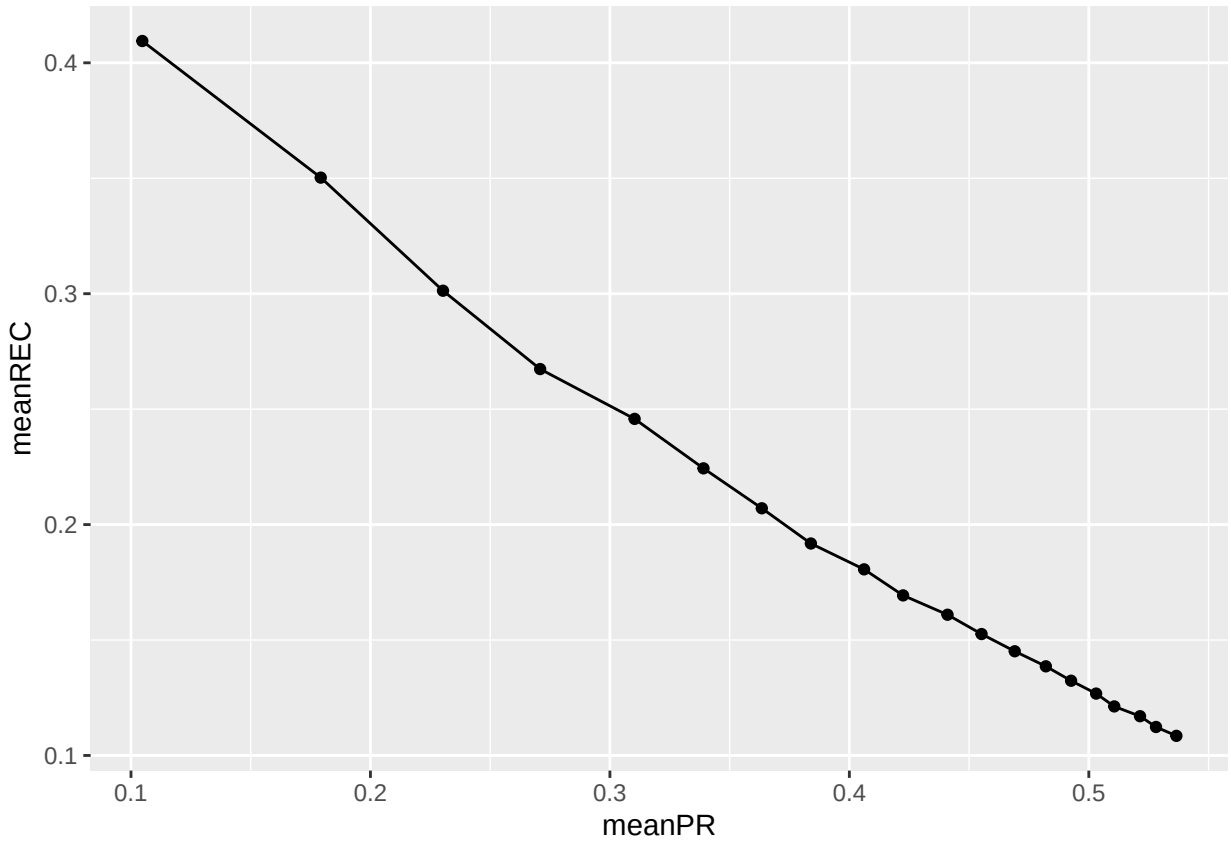
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4089  0.7478  0.8347  0.7851  0.8736  0.8922
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_n25/")
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  0.105   0.409
##  2     2  0.179   0.350
##  3     3  0.230   0.301
##  4     4  0.271   0.267
##  5     5  0.310   0.246
##  6     6  0.339   0.224
##  7     7  0.363   0.207
##  8     8  0.384   0.192
##  9     9  0.406   0.181
## 10    10  0.422   0.169
## 11    11  0.441   0.161
## 12    12  0.455   0.153
## 13    13  0.469   0.145
## 14    14  0.482   0.139
## 15    15  0.493   0.132
## 16    16  0.503   0.127
## 17    17  0.511   0.121
## 18    18  0.521   0.117
## 19    19  0.528   0.112
## 20    20  0.537   0.109
```

```
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.
##  0.1047  0.3319  0.4317  0.3975  0.4952  0.5366
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1085  0.1310  0.1651  0.1931  0.2297  0.4094
```