

# Topic Recommender

## Parameter

Topic cutoff = 20 Num of neighbours = 5 Sim function = topic-based ## Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_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  0.352
## 2     2  0.475
## 3     3  0.558
## 4     4   0.6
## 5     5  0.631
## 6     6  0.661
## 7     7  0.679
## 8     8  0.695
## 9     9  0.705
## 10    10  0.706
## 11    11  0.708
## 12    12  0.705
## 13    13  0.688
## 14    14  0.654
## 15    15  0.609
## 16    16  0.562
## 17    17  0.509
## 18    18  0.456
## 19    19  0.406
## 20    20  0.357
```

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

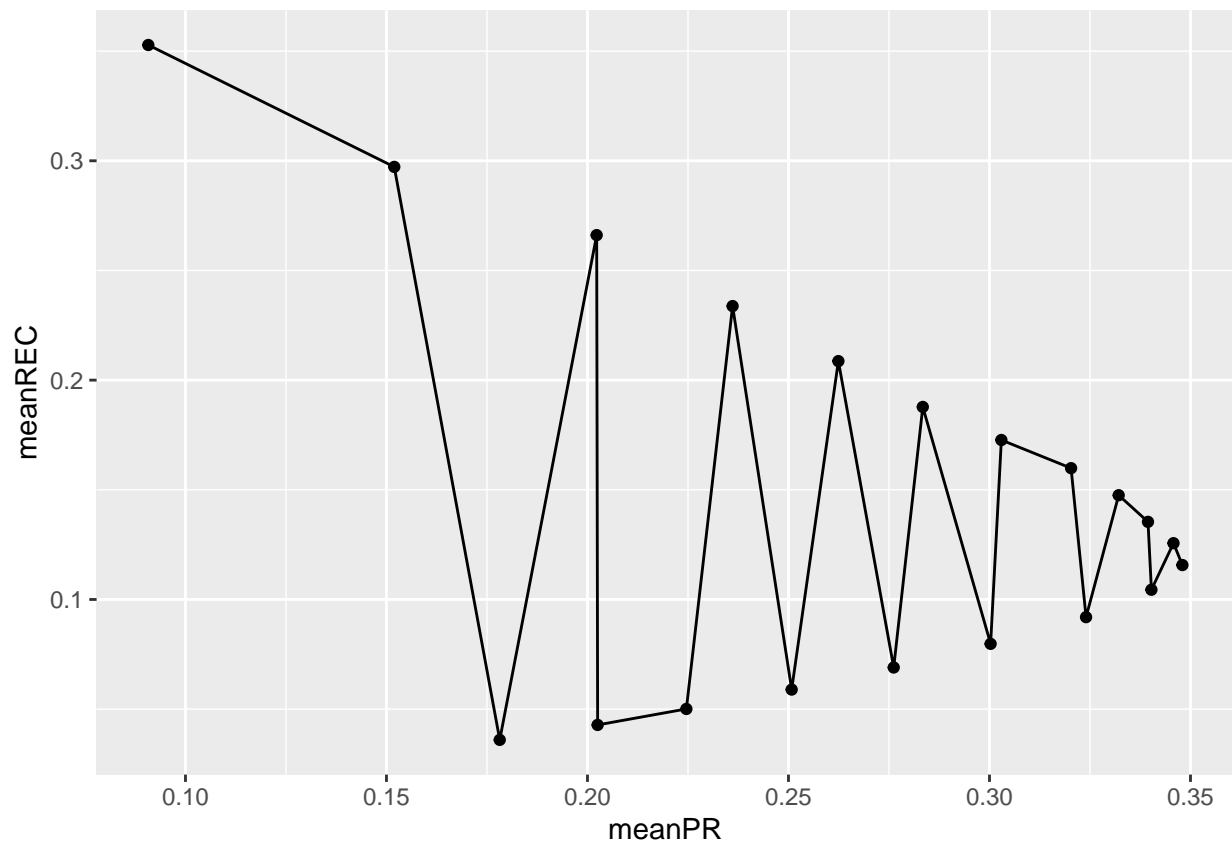
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3522  0.5004   0.6203  0.5858  0.6900  0.7078
```

## Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k20_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$V1) %>% summarise(meanPR = mean(V2), meanREC = mean(V3)) %>% rename(k = pr_tot$V1)
pr_tot
```

```
## # A tibble: 20 x 3
##       k meanPR meanREC
##   <int> <dbl> <dbl>
## 1     1 0.0908 0.353
## 2     2 0.152 0.297
## 3     3 0.202 0.266
## 4     4 0.236 0.234
## 5     5 0.262 0.209
## 6     6 0.283 0.188
## 7     7 0.303 0.173
## 8     8 0.320 0.160
## 9     9 0.332 0.148
## 10    10 0.339 0.135
## 11    11 0.346 0.126
## 12    12 0.348 0.116
## 13    13 0.340 0.104
## 14    14 0.324 0.0919
## 15    15 0.300 0.0798
## 16    16 0.276 0.0690
## 17    17 0.251 0.0589
## 18    18 0.225 0.0501
## 19    19 0.203 0.0428
## 20    20 0.178 0.0360
```

```
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.
## 0.09076 0.21908 0.27979 0.26561 0.32605 0.34798
```

### Recall summary

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

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03600 0.07708 0.13052 0.14681 0.19300 0.35278
```