Topic Recommender

Parameter

Topic cutoff = 15

Num of neighbours = 20

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_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 0.391
##  2     2 0.528
##  3     3 0.617
##  4     4 0.675
##  5     5 0.713
##  6     6 0.738
##  7     7 0.763
##  8     8 0.783
##  9     9 0.798
## 10    10 0.806
## 11    11 0.817
## 12    12 0.829
## 13    13 0.837
## 14    14 0.846
## 15    15 0.852
```

```
## 16    16 0.858
## 17    17 0.860
## 18    18 0.864
## 19    19 0.874
## 20    20 0.877
```

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

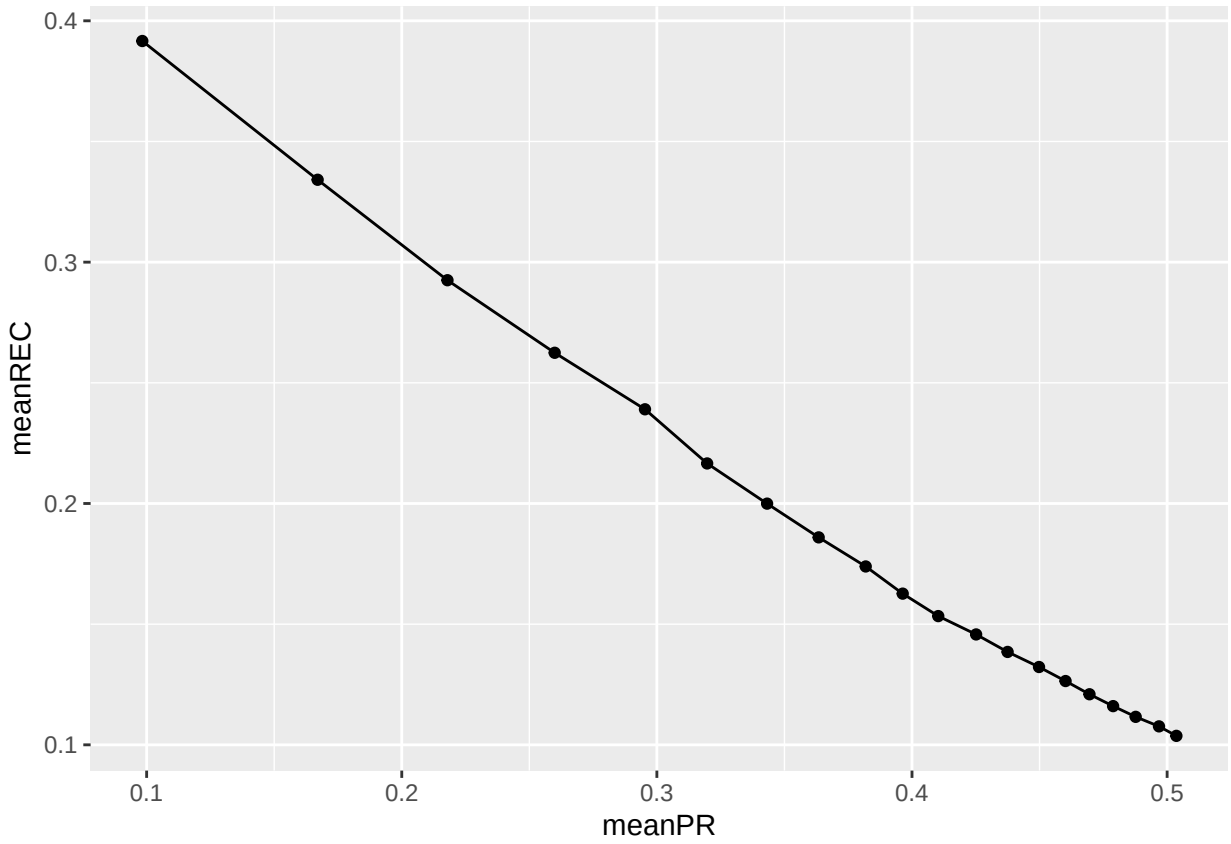
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3907  0.7317  0.8115  0.7663  0.8537  0.8771
```

Precision and Recall

```
setwd("/Users/juri/Desktop/RFiles/Results_k15_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 0.0983   0.392
##  2     2 0.167    0.334
##  3     3 0.218    0.293
##  4     4 0.260    0.262
##  5     5 0.295    0.239
##  6     6 0.320    0.217
##  7     7 0.343    0.200
##  8     8 0.363    0.186
##  9     9 0.382    0.174
## 10    10 0.396    0.163
## 11    11 0.410    0.153
## 12    12 0.425    0.146
## 13    13 0.437    0.138
## 14    14 0.450    0.132
## 15    15 0.460    0.126
## 16    16 0.470    0.121
## 17    17 0.479    0.116
## 18    18 0.488    0.112
## 19    19 0.497    0.108
## 20    20 0.504    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.
##  0.0983  0.3136  0.4033  0.3731  0.4625  0.5036
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1037  0.1251  0.1580  0.1858  0.2222  0.3916
```