Topic Recommender

Parameter

Topic cutoff = 10

Num of neighbours = 15

Sim function = topic-based

Success Rate

```
setwd("/Users/juri/Desktop/entangled/ResultsIN5/")
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 0.136
##  2     2 0.217
##  3     3 0.301
##  4     4 0.389
##  5     5 0.476
##  6     6 0.601
##  7     7 0.668
##  8     8 0.704
##  9     9 0.734
## 10    10 0.754
## 11    11 0.773
## 12    12 0.788
## 13    13 0.803
## 14    14 0.808
## 15    15 0.822
```

```
## 16    16 0.829
## 17    17 0.834
## 18    18 0.84
## 19    19 0.847
## 20    20 0.855
```

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1361  0.5694  0.7636  0.6591  0.8239  0.8550
```

Precision and Recall

```
setwd("/Users/juri/Desktop/entangled/ResultsIN5/")
pr_r1 <-  read.csv("PRC_Round1", sep="\t", header = F)
colnames(pr_r1)
```

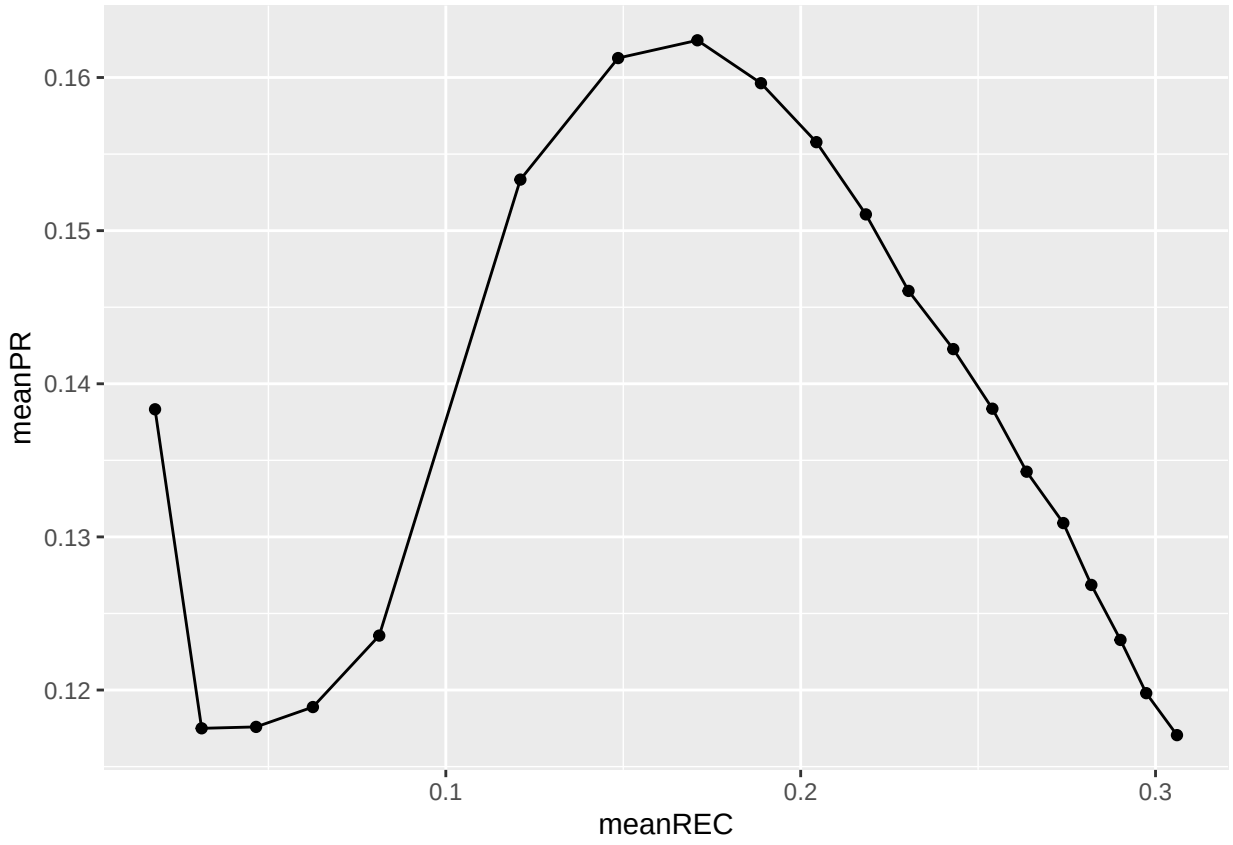
```
## [1] "V1" "V2" "V3"
```

```
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_r1 %>% merge(pr_r2, "V1") %>% merge(pr_r3, "V1")
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(meanREC = mean(V2), meanPR = mean(V3)) %>% renam
pr_tot
```

```
## # A tibble: 20 x 3
##        k meanREC meanPR
##    <int>   <dbl>  <dbl>
##  1     1  0.0181  0.138
##  2     2  0.0312  0.118
##  3     3  0.0465  0.118
##  4     4  0.0625  0.119
##  5     5  0.0812  0.124
##  6     6  0.121   0.153
##  7     7  0.149   0.161
##  8     8  0.171   0.162
##  9     9  0.189   0.160
## 10    10  0.204   0.156
## 11    11  0.218   0.151
## 12    12  0.230   0.146
## 13    13  0.243   0.142
## 14    14  0.254   0.138
## 15    15  0.264   0.134
## 16    16  0.274   0.131
## 17    17  0.282   0.127
## 18    18  0.290   0.123
```

```
## 19    19  0.297   0.120
## 20    20  0.306   0.117
```

```
pr_tot %>% ggplot(aes(x=meanREC, y=meanPR)) + geom_line() + geom_point()
```



Precision summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1171  0.1224  0.1363  0.1369  0.1516  0.1624
```

Recall summary

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.01805 0.11106 0.21140 0.18661 0.26626 0.30605
```