

Item-Based Collaborative Filtering

Yang LI

2017 年 11 月 16 日

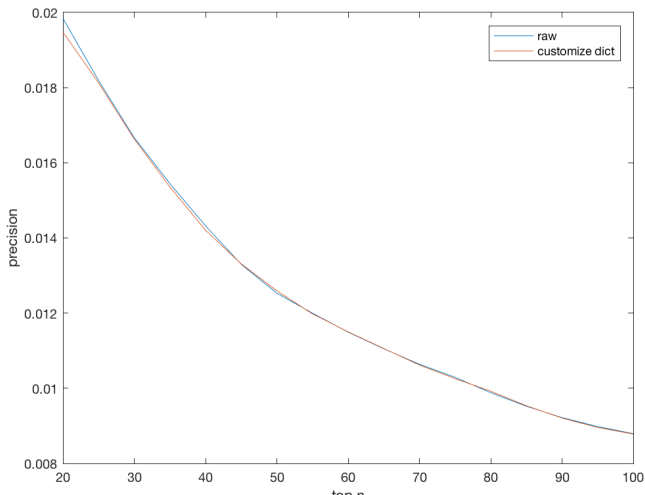
School of Software Engineering, Tongji University

Chinese Text Segmentation

分词的优化

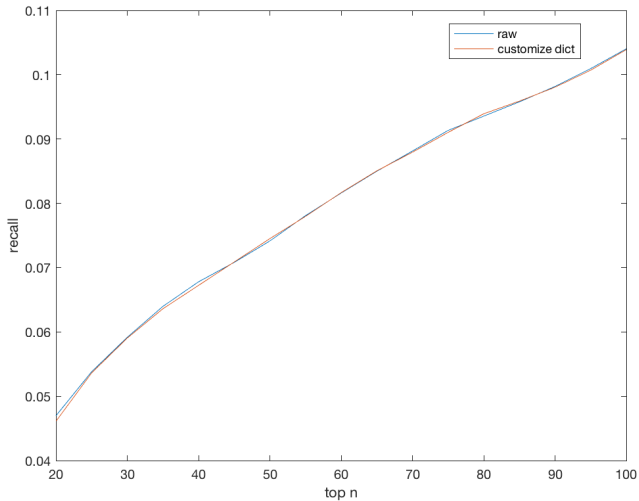
加入原数据的品牌名作为自定义词典, 提供给分词算法.

Precision:



分词的优化

Recall:

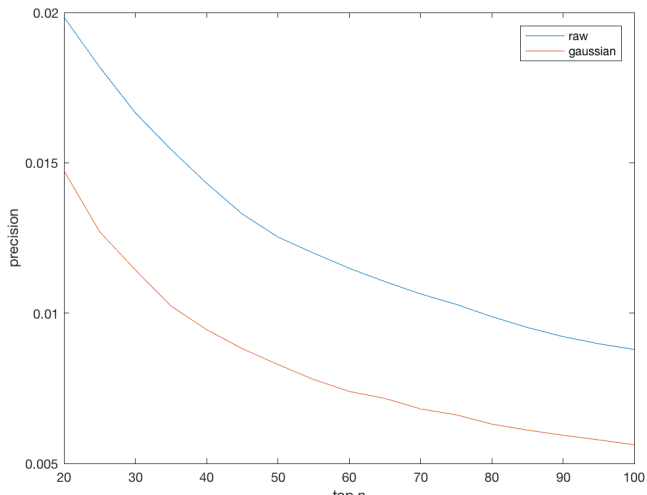


Item Cosine Similarity

商品距离的优化

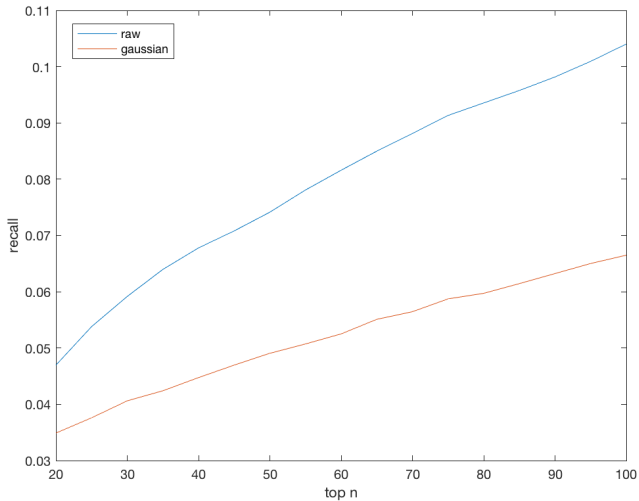
使用了高斯距离.

Precision:



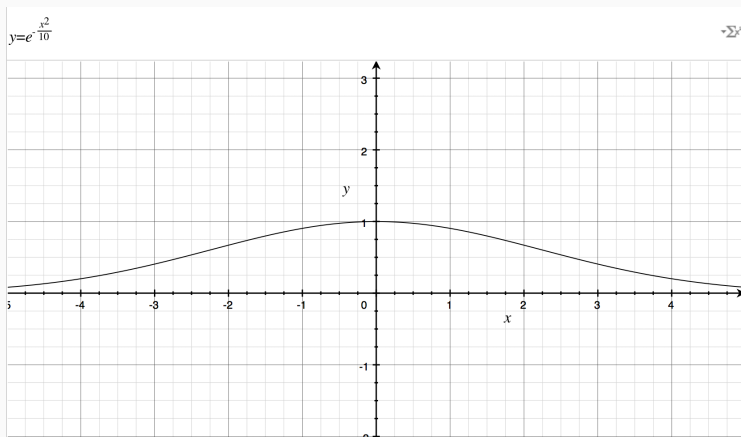
商品距离的优化

Recall:



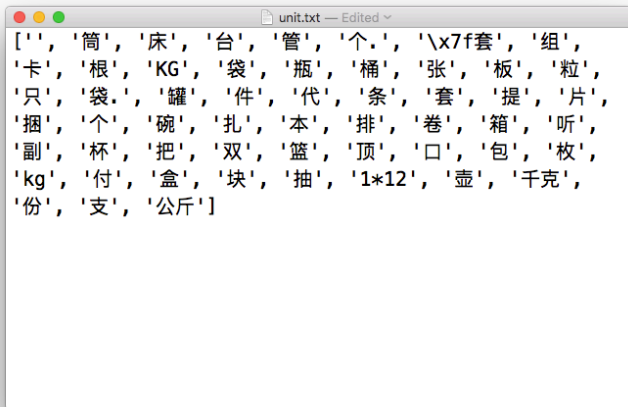
对多个 Gaussian 的 Sigma 进行测试:

prev: $\text{Sigma} = 10$



Other Attempts

分割量词



```
[', '筒', '床', '台', '管', '个.', '\x7f套', '组',  
'卡', '根', 'KG', '袋', '瓶', '桶', '张', '板', '粒',  
'只', '袋.', '罐', '件', '代', '条', '套', '提', '片',  
'捆', '个', '碗', '扎', '本', '排', '卷', '箱', '听',  
'副', '杯', '把', '双', '篮', '顶', '口', '包', '枚',  
'kg', '付', '盒', '块', '抽', '1*12', '壶', '千克',  
'份', '支', '公斤']
```

商品名的细分

