目录(8页)：

1. Lda – intro
2. Topic Modeling
   1. Plsa
   2. Lda – implement（写一下如何实现的具体细节好了）
      1. Preprocessing

Stop words

Tokens, esp. special token” 0s\_token0”

* + 1. Gibbs sampler
    2. Gibbs Lda 时间复杂度: O(iter\_num\*corpus\_word\_count\*K topics)
    3. Gibbs LDA 空间复杂度: O(K topics \* vocabulary size)

1. Document classification
   1. Data set 数据如何抽取获得，有哪些变量，我们需要选取的变量
   2. Choose Topics numbers：依据 perplexity, 结论
   3. Application on Comment text
      1. Topics: 展示主要的topics，简要分析
      2. Auto – tagging – 用theta 和 phi
      3. Sentiment Binary Classification：logistics: lda feature vs. word feature
      4. Sentiment 3-class classification: logistics ovr, lda vs. word
   4. Application on Comment summary
      1. (隐藏在过渡中)为何想到用summary来做？因为前面的结果不竟如人意
      2. Topics：展示10个topic 里面有什么词
      3. Binary classification: logistics
         1. Word2vec
         2. Lda
         3. 比较
   5. 分析：
      1. ~~Gibbs Lda 时间复杂度: O(iter\_num\*corpus\_word\_count\*K topics)(转移到上面了)~~
      2. 为何用lda分类的结果不好: 数据集随机抽取1500，导致部分商品可能只有一条评论，score的分布和原始分布不同。
2. Collaborative filtering
   1. Data set:
   2. 这部分南哥自己发挥吧hhh
3. Reference
4. Appendix
   1. Results/or 数学推导？
   2. Program + readme