

产品维度上话题差异的推断

实验数据：iphone5s，华为荣耀 3C 的产品评论，在产品维度上评价两者的 topic 分布。

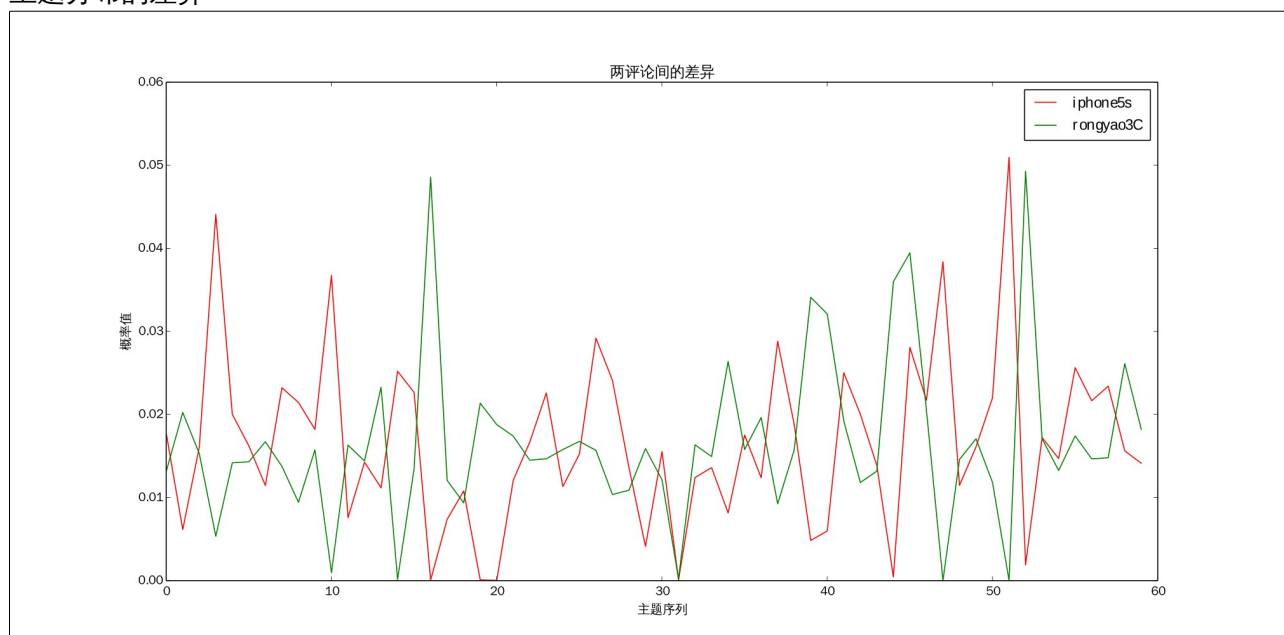
数据预处理：分词，去停用词，分句。

存在问题：取出高频形容词，去除的阈值设定，此次实验没去除（“不错，好，很好”）

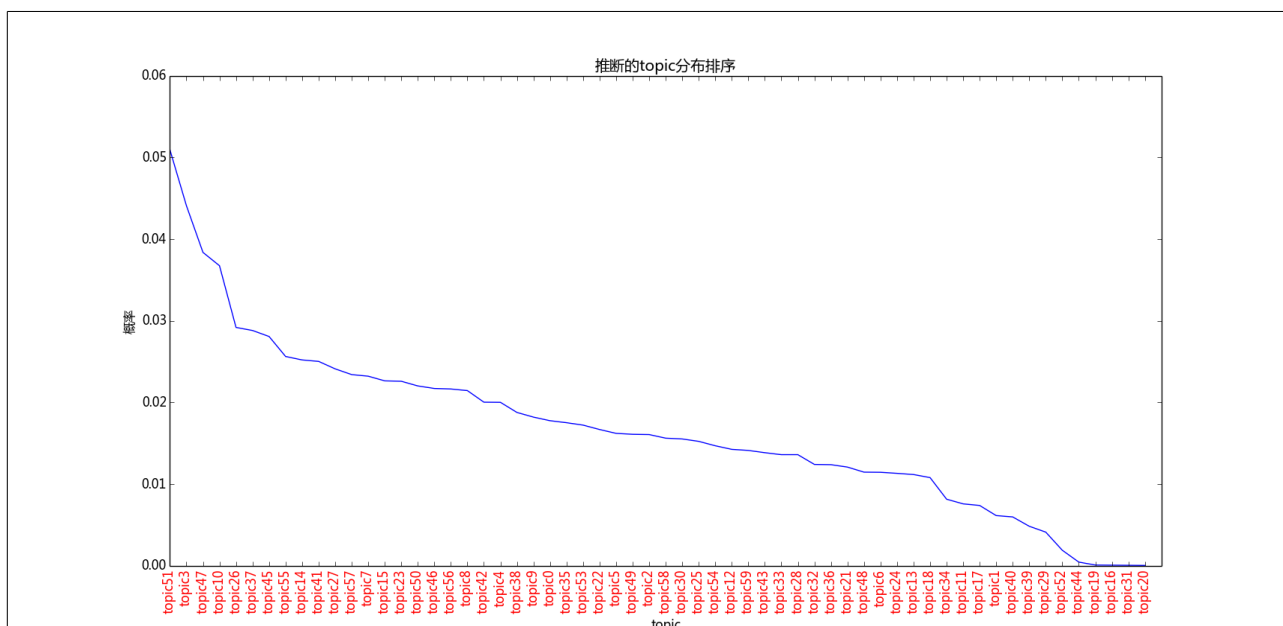
lda 模型：困惑度曲线，设定 $t=60$ 。

lda 推断：分别对 iphone5s 和 rongyao3c 的评论进行主题分布的推断。

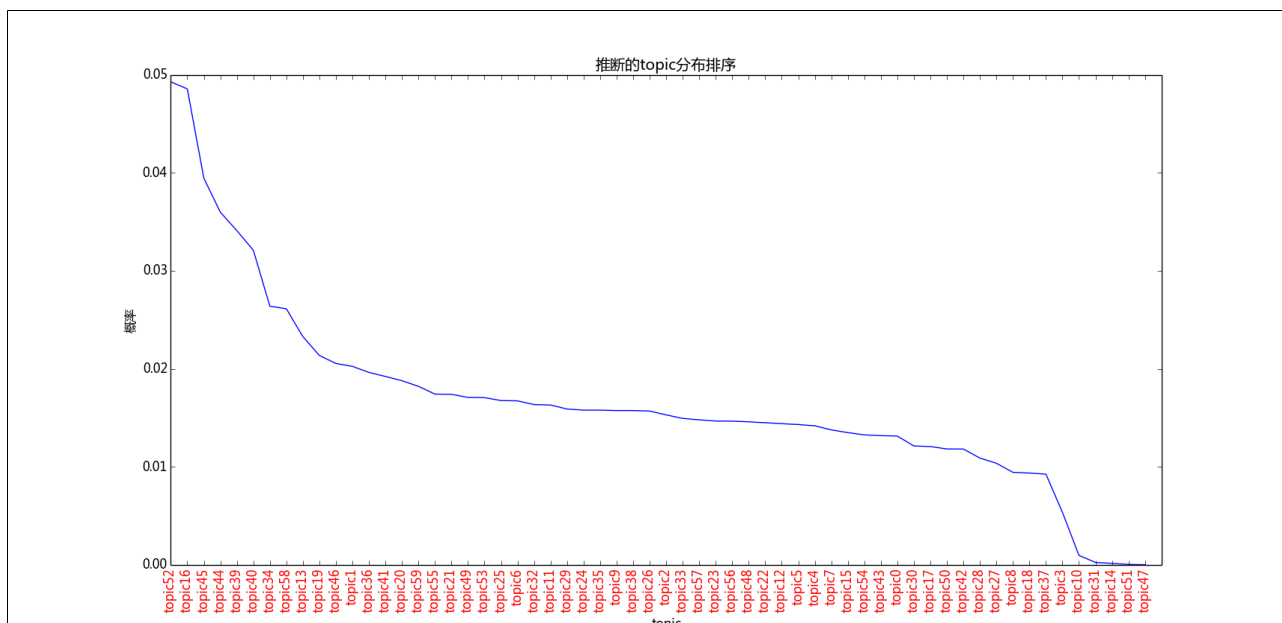
主题分布的差异



1,iphone5s



2,rongyao3c



推断结果：

iphone5s 的评论分布中，概率较高的前十个 topic 为：

topic51 手机，4G，TD，金色，发票，保障，LTE，WCDMA，SCDMA

topic3 正品，行货，机子，激活，全新，原装，国行，未激活

topic47 土豪，金，5S，5s，iPhone，质量，s，iphone

topic10 苹果，产品，确实，一如既往，手机，三星，没得说，不用说，没的说

topic26 收到，货，第二天，下单，下午，点，到货，号，晚上

topic37 说，不，客服，售后，打电话，检测，换货，申请，找，退货

topic45 不错，挺好用，价格便宜，真心，反正，触摸，听说，卖，真的说

topic55 太，不，差，服务，京东，评，真心，过程，习惯，适合，电量

topic14 京东，支持，一如既往，信任，华为，商城，服务，卖，自营

topic41 送货，速度，挺快，真，很快，超快，配送，屏，久，超赞

rongyao3c 的评论分布中，概率较高的前十个 topic 为：

topic52 京东，天，炸，屌，不错，发票，产品，价格便宜，cap，屌炸天，捆绑，玩游戏

topic16 华为，荣耀，3C，手机，1G，套装，白色，3G，版，ram，双卡双待

topic45 不错，挺好用，价格便宜，真心，反正，触摸，听说，卖，真的，说，未，很哈

topic44 支持，国产，国货，国产机，华为，精品，manhao，嘻嘻，国产品牌

topic39 性价比，高，分辨率，很高，超高，蛮高，删除，麻麻，整体，小米，超值

topic40 屏幕，清晰，效果，像素，照相，拍照，摄像头，够，很大，相机

topic34 京东，购物，首选，放心，够用，感谢，忙，爸爸，几天，再就是

topic58 买朋友，帮，同事，同学，介绍，年，代购，上网，代买，没得说，屏

topic13 感觉，总体，不错，整体，挺不错，算，国产手机，一般般

topic19 手机，价位，机子，款，配置，不错，算，这款，算是，适合，超值

两产品评论概率差距最大的前 10 个 topic:(绿色 iphone 高, 红色 rongyao3c 高)

topic51 手机, 4G, TD, 金色, 发票, 保障, LTE, WCDMA, SCDMA

topic16 华为, 荣耀, 3C, 手机, 1G, 套装, 白色, 3G, 版, ram, 双卡双待

topic52 京东, 天, 炸, 屌, 不错, 发票, 产品, 价格便宜, cap, 屌炸天, 捆绑, 玩游戏

topic3 正品, 行货, 机子, 激活, 全新, 原装, 国行, 未激活

topic47 土豪, 金, 5S, 5s, iPhone, 质量, s, iphone

topic10 苹果, 产品, 确实, 一如既往, 手机, 三星, 没得说, 不用说, 没的说

topic44 支持, 国产, 国货, 国产机, 华为, 精品, manhao, 嘻嘻, 国产品牌

topic39 性价比, 高, 分辨率, 很高, 超高, 蛮高, 删除, 麻麻, 整体, 小米, 超值

topic40 屏幕, 清晰, 效果, 像素, 照相, 拍照, 摄像头, 够, 很大, 相机

topic14 京东, 支持, 一如既往, 信任, 华为, 商城, 服务, 卖, 自营