

MathorCup 全球大学生数学建模挑战赛

暨 CAA 世界大学生数学建模竞赛

C 题 垃圾观点文档识别

目前商务网站或博客论坛允许用户发表针对产品或话题的一些评论看法，难免会存在一些虚假的或是与产品及话题无关的评论信息，这极大地误导了商家、读者以及观点挖掘系统，造成了资源的浪费。因此，垃圾观点文档的识别具有重要研究价值和实用意义。

本任务是对给定的语料集合中，要求参赛系统识别出文档是否为垃圾观点文档。要求完成以下问题：

(1) 针对下面介绍的情形，请建立合理的数学模型进行识别，并给出你的算法流程。并通过程序验证，给出你的正确识别率。

(2) 请在网络上收集一个更大的关于某件产品的评价集合，建立合理的数学模型和算法进行识别，并给出你的结论。

(3) 对一般的产品评价集合，讨论并建立更一般的模型，并谈谈你的该类识别问题的看法。

附件 1

下面四个句子,其中前面三个不是垃圾观点文档，第四个则是垃圾观点文档。

<Doc1>买的 6plus, 轻薄分辨率高, 电池耐用, 屏幕够大, 散热比之前的有所改进</Doc1>

<Doc2>看到个视频, iphone6 轻轻一掰就弯了</Doc2>

<Doc3>我还是喜欢保时捷, 外观、内饰都大气奢华 </Doc3>

<Doc4>总是有人都没用过买不起, 就在哪里瞎说说 </Doc4>

背景知识

对于垃圾评论（或者称之为无用评论），我们主要考虑如下几种情形：

1. 确实是评论，但只对品牌和制造商，甚至是站点评论，而没有针对当前产品本身进行评论，因此这种评论是无意义的，比如下面例子。

苹果 iPhone 6（全网通）点评

[づ我的世界、你都是无](#) 01 月 15 日 17:58

苹果手机有什么牛逼的嘛！当年的诺基亚比它牛逼到哪儿去了！现在诺基亚还不是被收购了！别有事没事装装逼，买个苹果手机穷得瑟！

[倚栏笑看花满楼_13](#) 01 月 15 日 15:01

苹果就是一部普通手机，有什么好抢的，都 tm 脑残粉

[广东省网友\(119.142.*.*\)](#) 01 月 09 日 19:51

乔布斯死了，苹果也就不怎么样了

[陕西省网友\(117.37.*.*\)](#) 12 月 24 日 13:54

用苹果的大多是脑残

[qq_1t1939y13430](#) 12 月 24 日 11:50

苹果很快流落到诺基亚的地步。。】

[为你而平凡](#) 12 月 20 日 21:04

苹果手机就是烂货，一点不好用，坚决不买。中国人应该买国产手机

[明明哥。](#) 12月03日 12:30

我 宁可把钱 砸地下城中 绝不会 买 苹果手机 支持 3星手机

2. 确实是对产品进行了评论，但是评错了产品，例子如下所示。

苹果 iPhone 6（全网通）点评

[Tonyboy^o^](#) 01月17日 22:50

买 SONYZ3 也不错啊 很漂亮 旗舰机

[北京市网友\(60.206.*.*\)](#) 01月11日 21:53

我的诺基亚 1209 ,我用了 7 年了,还是很好用, 对了,我还去马桶了捞过它一次,真耐用呀

[安徽省网友\(120.210.*.*\)](#) 12月30日 12:40

苹果 4 掉进厕所，捞起来，在给它洗个澡，晒干了，依旧好用，

[空杯心情](#) 12月27日 10:12

我现在还用诺基亚 1100 感觉挺好，公交车上，大街上，回头率也高。也不咋耗流量。可以。

[北京市网友\(123.126.*.*\)](#) 12月24日 11:10

4S 的核心创新是什么啊

3. 广告评论，如下所示。

苹果 iPhone 6（全网通）点评

[找 Q122929079 买](#) 01月16日 21:13

苹果 iPhone 6 Plus 超爱大屏幕 3500 元拿现货哦 QQ 热购 1229 再 29079

[QQ 热购 390565208](#) 01月13日 21:39

苹果 iPhone 6 Plus 超爱大屏幕 3300 元拿现货哦

[淘宝《果真好数码》](#) 11月24日 20:37

买个好手机，当然要买个好壳，买好壳，就来淘宝《果真好数码》

4. 无关的文本，大致包括个人的消费体经历、人身攻击和其它无关文本三类，例子如下。

4.1 个人的消费体经历

苹果 iPhone 6（全网通）点评

[北京市网友\(221.176.*.*\)](#) 01月07日 12:18

苹果 6 手机

[Yolanda 仰望星空](#) 01月06日 22:03

再烂都永远有人疯抢，漂洋过海甚至就为一部手机，实在不懂。

[湖南省网友\(222.240.*.*\)](#) 01月05日 18:12

就是买个装逼 就是任性

[波多也解衣](#) 12月28日 12:42

国产就是党产 党一家老小移民美国 却不停的煽动中国人支持中国专制统治？

[波多也解衣](#) 12月28日 12:39

马劣畜生政权

[浙江省网友\(123.97.*.*\)](#) 12月27日 10:24

什么炫富不炫富都是自己思想上有问题

[茧棋子](#) 12月10日 12:46

一帮什么也不懂的人就知道 YY，人家买飞机的都不任性，不知道买个苹果为什么任性。。。。攀比而已。。。幼稚

[『不再问→』](#)10月14日 17:25

评论纷纭，笑死个人了，唉

4.2 人身攻击

[倚栏笑看花满楼_13](#)01月15日 15:01

苹果就是一部普通手机，有什么好抢的，都 tm 脑残粉

[波多也解衣](#)12月28日 12:42

国产就是党产 党一家老小移民美国 却不停的煽动中国人支持中国专制统治？

[波多也解衣](#)12月28日 12:39

马劣畜生政权

[陕西省网友\(117.37.*.*\)](#)12月24日 13:54

用苹果的大多是脑残

[山东省网友\(60.213.*.*\)](#)12月20日 13:14

一个成本在 1000 快而已，一群 SB 还 JB 买，你们跟开日本车的有什么区别，还成天说人汉奸，SB 在这里叨叨

4.3 其他无关文本

[星空_219](#)12月06日 23:09

转给我呗？？？

[广东省网友\(223.73.*.*\)](#)12月06日 00:09

信号好不好

[浙江省网友\(39.190.*.*\)](#)12月03日 03:19

你好

[望断的秋水](#)11月01日 09:58

凑个热闹！

[ben**dan](#)10月14日 14:37

现在售价到底是好多啊？