综合主题模型LDA挖掘语意

[一．使用Scarpy框架爬取商品评论信息 1](#_Toc459633472)

[1.分析京东首页，获取商品ID 1](#_Toc459633473)

[2.通过ID，构造商品评论信息页面的URL 1](#_Toc459633474)

[3.通过ID，构造商品详细信息页面的URL 2](#_Toc459633475)

[4.通过ID，构造商品名称-ID映射页面的URL 2](#_Toc459633476)

[二．使用Python对获取的评论进行处理，进行初始的分析 2](#_Toc459633477)

[1.使用Python第三方扩展包，jieba分词，建立词向量模型 2](#_Toc459633478)

[2.使用IF\_IDF抽取关键字 2](#_Toc459633479)

[三．综合主题模型LDA挖掘语意 2](#_Toc459633480)

[1.得到形式化的输入格式，接下来选择对应的算法进行挖掘 2](#_Toc459633481)

[2.使用LSI主题模型，得到主题 3](#_Toc459633482)

# 一．使用Scarpy框架爬取商品评论信息

## 1.分析京东首页，获取商品ID

<https://github.com/kangsugo/sugo_coding/tree/master/diverdity_v0.2>

## 2.通过ID，构造商品评论信息页面的URL

<https://github.com/kangsugo/sugo_coding/tree/master/comment_statics>

## 3.通过ID，构造商品详细信息页面的URL

<https://github.com/kangsugo/sugo_coding/tree/master/goods_feature_v0.2>

## 4.通过ID，构造商品名称-ID映射页面的URL

<https://github.com/kangsugo/sugo_coding/tree/master/phone_id_name_db>

# 二．使用Python对获取的评论进行处理，进行初始的分析

## 1.使用Python第三方扩展包，jieba分词，建立词向量模型

## 2.使用IF\_IDF抽取关键字

<https://github.com/kangsugo/NLP_LDA>

# 三．综合主题模型LDA挖掘语意

## 1.得到形式化的输入格式，接下来选择对应的算法进行挖掘

<https://github.com/kangsugo/NLP_20116_8_15>

## 2.使用LSI主题模型，得到主题

<https://github.com/kangsugo/NLP_2016_8_16/tree/master/builded_models>