第12章 电商产品评论数据情感分析

教案

**课程名称：**Python数据分析与挖掘实战

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**112学时（其中理论49学时，实验63学时）

**总学分：**7.0学分

**本章学时**：14学时

# 材料清单

* 1. 《Python数据分析与挖掘实战》教材。
  2. 配套PPT。
  3. 引导性提问。
  4. 探究性问题。
  5. 拓展性问题。

# 教学目标与基本要求

### 教学目标

主要介绍通过Python语言进行项目数据爬取，对文本数据进行预处理，分词，去除停用词等操作，在知网情感词表上进行优化，进行基于词表的情感分析，最后使用LDA主题模型对正负面评论进行主题分析。

### 基本要求

1. 熟悉电商产品评论数据实现情感分析的步骤与流程。
2. 了解如何使用Python语言对互联网信息进行爬取。
3. 掌握文本分析的预处理方法。
4. 对预处理后的评论数据进行情感分析。
5. 使用LDA模型对正、负面评论数据进行主题分析。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

1. 哪里有电商产品评论的数据？
2. 情感分析的方法有哪些？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. LDA模型得到的主题是否还能优化？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 除了文中介绍的获取评论数据的方法外，还有哪些Python库获取评论数据？
2. 主题分析除了LDA模型外，还有哪些？

# 主要知识点、重点与难点

### 主要知识点

1. 了解电商企业现状。
2. 熟悉电商评论数据情感分析的步骤与基本流程。
3. 掌握获取评论数据的方法。
4. 对电商产品评论数据进行预处理，包括行分词处理，去除停用词，查看分词效果。
5. 基于情感词表进行情感词匹配，并对情感词的倾向进行修正。
6. 对情感分析结果进行检验。
7. 了解主题模型，以及LDA模型原理与参数估计方法。
8. 掌握寻找最优主题数的方法。
9. 建立相应的LDA模型。
10. 分析LDA模型的结果。

### 重点

1. 对电商产品评论数据进行预处理。
2. 了解主题模型，以及LDA模型原理与参数估计方法。
3. 掌握寻找最优主题数的方法。
4. 建立相应的LDA模型。
5. 分析LDA模型的结果。

### 难点

1. 掌握寻找最优主题数的方法。
2. 分析LDA模型的结果。

# 教学过程设计

### 理论教学过程

1. 了解电商企业现状。
2. 了解网络上发布内容的技术和Web文档中提取信息的技术，以获取网络数据。
3. 对电商产品评论数据进行预处理，包括行分词处理，去除停用词，查看分词效果。
4. 基于情感词表进行情感词匹配，并对情感词的倾向进行修正。
5. 对情感分析结果进行检验。
6. 了解主题模型，以及LDA模型原理与参数估计方法。
7. 掌握寻找最优主题数的方法。
8. 建立相应的LDA模型。
9. 进行LDA主题分析。

### 实验教学过程

1. 去除评论数据的数字、字母。
2. 对评论数据进行去重。
3. 对评论数据进行分词处理。
4. 根据停用词库去除评论文本中的停用词。
5. 绘制词云图，查看分词效果。
6. 将情感词表与分词结果进行匹配。
7. 对情感值的方向进行修正，并计算情感分析的准确率。
8. 分别对正面评论和负面评论绘制词云，查看情感分析效果。
9. 对正面情感词与负面情感词构建语料库，并建立文档-词条矩阵。
10. 使用LDA主题模型，找出不同主题数下的主题词，寻找最优主题数。
11. 输入正面情感与负面情感评论求解LDA模型，并分析结果。

# 教材与参考资料

### 教材

张良均，谭立云，刘名军，江建明．《Python数据分析与挖掘实战（第2版）》[M]．北京：机械工业出版社．2019．