《数据挖掘方法与应用》练习与拓展

参考答案（要点）

第 1 章 数据挖掘概述

1.什么是数据挖掘？请结合实例加以说明。

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据中，提取隐含在其中的、有效的、新颖的、潜在有用的并且最终可理解的模式的非平凡过程。实例（略）。

2.检索近几年数据挖掘国际学术会议的入选论文，分析数据挖掘研究现状及热点问题。

https://www.kdd.org/

https://www.aminer.cn/research_report/5f23ba8921d8d82f52e5aebf

3. 查找相关资料，分析第四代数据挖掘系统的特点。

第四代数据挖掘系统旨在挖掘嵌入式系统、移动系统及各种普适计算设备产生的各种类型数据。其与移动设备及各种计算设备结合，集成了数据库管理系统、预言模型系统及移动系统等，支持多个算法。

物联网的不断发展，云计算、雾计算技术的广泛应用，将会进一步推动第四代数据挖掘系统的研究与发展。

4. 什么是云计算？

云计算不仅是技术，更是一种全新的商业服务模式。

云计算服务，以云资源为实现基础，以云计算技术为实现保障，以低成本、按需付费的形式，向用户提供软(硬)件基础设施、计算平台和软件服务，使用户在无基础投入的前提下直接实现数据的存储、管理和分析，也可利用提供的云服务平台创建和开发应用程序，或直接使用云服务平台提供的各类服务软件。

5. 什么是大数据？大数据是否等于大数据分析？

大数据是超出了典型(传统、常用)硬件环境和软件工具收集、存储、管理和分析能力的数据集。

大数据 “4V” 特征：数据量大(Volume)、产生和处理的速度快(Velocity)、形式多样(Variety)、价值密度低但价值量大(Value)。

大数据不等于大数据分析，大数据包括了大数据的生产、采集、传输、存储、分析及应用。大数据分析只是其中的一个重要环节。

6. 如何理解大数据被认为是下一个社会发展阶段的石油和金矿。

各个国家把大数据当作一种全新的社会资源，并把大数据产业的发展提升到国家战略发展的高度。类比于石油资源，从石油的勘探、开采、运输、提炼到石油产品的生产与销售等多个环节形成了石油产业，对于大数据的生产、采集、传输、存储、分析及应用则形成了大数据产业。

物理的网络（物体的流动）、交易网络、社交网络产生的大数据的流动、融合、分析及应用，会产生无限的可能，从而改变我们的生产方式与生活方式。

7. 什么是 Web 数据挖掘？

Web 数据挖掘是数据挖掘技术在 Web 上的应用，是从 Web 的网页内容、超链接结构和用户使用日志中获取有用知识的过程，包括 Web 内容挖掘、Web 结构挖掘和 Web 使用挖掘。

8. 什么是文本数据挖掘？

文本数据挖掘是从半结构化或非结构化文本中获取用户有用信息的过程。这一过程主要包括文本数据的获取、文本预处理、挖掘分析和结果可视化四个步骤。

9. 分析说明 Fayyad 过程模型。

Fayyad 等将知识发现过程定义为：从数据中鉴别出有效模式的非平凡过程，该模式是新颖的、可能有用的和最终可理解的。Fayyad 过程模型包括数据选择(data selection)、数据预处理(data preprocessing)、数据转换(data transformation)、数据挖掘(data mining)、模式解释与评价(pattern interpretation)。

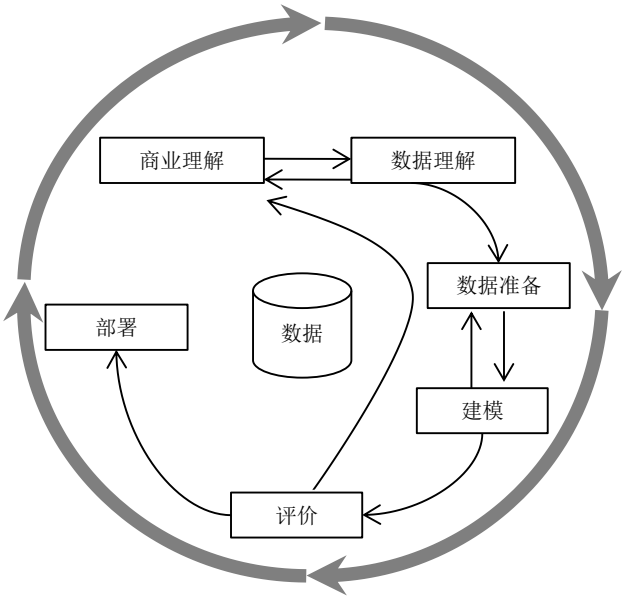
这个过程包括了数据挖掘过程中必要的各个处理阶段，并且也形成了一个可以根据各个处理阶段的结果来决定是否返回以前的阶段进行再处理的闭环过程。但是，Fayyad 过程模型从数据入手，到知识结束，过多地偏重于从技术的角度来理解数据挖掘过程，缺乏应用层面的体现，如商业问题的分析、知识的使用等。

10. 分析说明 CRISP-DM 过程模型。

CRISP-DM（cross-industry standard process for data mining，即跨行业数据挖掘过程标准）强调，数据挖掘不单是数据的组织或者呈现，也不仅是数据分析和统计建模，而是一个从理解业务需求、寻求解决方案到接受实践检验的完整过程。

CRISP-DM 过程模型包括商业理解(business understanding)、数据理解(data understanding)、数据准备(data preparation)、建模(modeling)、评价(evaluation)

和部署(deployment)六个阶段。如下图所示：



外圈形象地表达了数据挖掘过程的循环特性。一个数据挖掘项目并不是一次部署完就结束，在挖掘的过程中或部署过程中获得的经验可能会触发新的商业问题。后续的挖掘过程将从前一次的经验中受益。内部的箭头表示阶段之间最重要和最频繁发生的关联关系。阶段间的顺序不是严格不变的，可以根据具体的任务需要进行来回选择。(每一阶段的具体内容请参考教材)

11. 结合 CRISP-DM 过程模型，自选一个感兴趣的商业问题，以小组为单位，制订一份数据挖掘项目计划。(略)

12. 数据挖掘的功能有哪些？

常见的数据挖掘功能可以概括为六个方面：数据描述、聚类、偏差检测(孤立点检测)、关联分析、预测和分类。

13. 结合数据挖掘使用技术，分析其与相关学科之间的关系。(略)

14. 什么是机器学习？按学习方式不同，机器学习可以分成哪几种？分别具有什么特点？

机器学习是指计算机利用各种学习算法，从输入的数据中学习，识别复杂的模式，从而作出智能的决断。

机器学习的基础是数据，核心是各种学习算法，只有通过这些算法，机器才能识别分析这些数据，获得知识，从而不断提升自身性能。机器学习的算法很多，根据学习方式不同，可以分为有监督学习(supervised learning)、无监督学习(unsupervised learning)、半监督学习(semi-supervised learning)和强化学习(reinforcement learning)。

用于有监督学习训练的数据集包含输入(特征)和输出(目标)，也称为有标记的

数据集。

用于无监督学习训练的数据集只包含输入(特征)，而没有输出(目标)，也称为无标记数据集。

有两个数据集用于半监督学习，一个为有标记的数据集，一个为无标记的数据集，通常无标记数据集的数据量要远远大于有标记数据集的数据量。

强化学习是智能体(agent)在尝试的过程中学习在特定的环境下选择哪种行动可以得到最大的回报。

15. 为什么说对于更大的数据库或是要满足更多在线处理任务的数据库，直接从日常数据库中读取数据用于数据挖掘是不可行的？

第一，联机事务处理系统强调的是数据处理性能和系统的安全与可靠性，并不关心数据查询的方便与快捷，直接从日常数据库中读取数据用于数据挖掘会给日常 DBMS 带来很大负担，影响其运行性能；

第二，用于数据挖掘的数据通常来源于不同的事务数据库，不同事务数据库数据的模式是针对具体事务处理而设计的，可能存在不一致等问题，不适合直接用于数据挖掘；

第三，联机事务处理(OLTP)系统可能缺少数据挖掘需要的大量历史数据。

16. 查阅相关资料，说明什么是模式识别。（略）

17. 数据挖掘可视化包含哪些方面？

数据可视化、挖掘过程可视化、挖掘结果可视化。

18. 查阅相关资料，说明什么是分布式计算，什么是并行计算，两者有什么关系。（略）

19. 结合教材中提到的数据挖掘应用领域，请举例说明，并谈谈你的理解。（略）

20. 除了教材中提到的数据挖掘应用领域，请思考还有哪些应用领域，并举例说明。（略）

第 2 章 数据挖掘工具

1. 下载并安装 Weka 软件，了解探索者界面、实验者界面、知识流界面、工作台界面和简单命令行界面及相应的功能。

探索者界面是 Weka 系统提供的最容易使用的图形用户界面(GUI)，用户通过选择菜单就可以调用 Weka 的所有功能。

实验者界面允许使用多种算法对多个数据集进行操作，突破了时间的限制，

包含了一些分布式计算的功能。

知识流界面允许用户从设计面板中选择数据源、预处理工具、学习算法、评估方法和可视化等 Weka 组件，放置在布局区域，并将它们连接起来形成“知识流”，进行数据处理和分析。

从 Weka 3.8.0 版本开始，在原有探索者、实验者、知识流和简单命令行四个界面的基础上，新增了工作台界面。工作台界面集成了原有的四个界面，方便操作。

简单命令行界面是为不具有命令行界面的操作系统提供的，通过该界面，用户可以直接执行 Weka 命令。

2. 下载并安装 IBM SPSS Modeler 软件，了解数据流构建区、节点区、流管理区和项目管理区的相应功能。

数据流构建区是数据挖掘分析人员的主要工作区域。分析人员可以根据分析需要，从节点区选择功能节点，添加到数据流构建区，构建分析所需的数据流。

节点区包含了分析需要的所有节点，按功能分成：收藏夹、源、记录选项、字段选项、图形、建模、输出、导出、Python、Spark 和 Text Analytics。

流管理区包含“流”“输出”和“模型”选项卡。

项目管理区为用户提供管理数据挖掘项目过程的相关文件，包含 CRISP-DM 和“类”选项卡。

3. 下载并安装 R 和 RStudio，了解从数据加载、预处理、模型建立到模型评估各个环节 R 可提供的相应软件包。（略）

4. 下载并安装 Python，了解 Python 可用于数据分析的相关库。（略，参考表 2.1）

5. 下载并安装 Anaconda，了解相应的功能。（略）

第 3 章 数据与数据平台

1. 从数据形态看，数据可以分为哪些类型，各有什么特征？

数据按存在形态可以分为结构化数据、半结构化数据和非结构化数据三种。

结构化数据指可以按特定的数据结构来表示的数据，主要指存储在关系型数据库或数据仓库中的数据。这类数据由二维表结构来表示，以行为单位，一行数据表示一个实体的信息，每一行数据的属性(列)是相同的，每一列数据不可以再细分且数据类型相同。

半结构化数据介于完全结构化数据和完全无结构化数据之间。它具有结构化

的特点，包含某些标记，可以用来分隔语义元素以及对记录和字段进行分层，所以不能简单地将它组织成一个文件按照非结构化数据处理；但因为结构会变化，所以也不能使用关系型数据库或其他数据库二维表的形式来表示。

非结构化数据是指结构不规则或不完整，没有预定义的数据模型，不方便用二维逻辑表结构来表示的数据，包括文本文档、图像、音频、视频等。

2. 从数据所处的环境看，数据可以分为哪些类型？

根据数据所处的环境，可以把数据分为三种类型：生产数据、原始数据和分析数据。生产数据存在于生产环境，是生产应用系统实时运行所产生的数据；分析数据是对原始数据经过 ETL 过程等优化后存放于数据分析环境的数据；原始数据既不属于生产环境，也不属于分析环境，是一种从生产环境解耦的过渡形态的数据。

3. 分析关系型数据库的特点、优点及其局限性。

请参考教材 3.2.1。

4. 目前流行的 NoSQL 数据库有哪几种模式？各有什么特点？

从数据存储类型看，目前流行的 NoSQL 数据库主要为键值存储、文档存储、列族存储和图存储四种模式。特点请参考教材 3.3 相应章节。

5. 数据仓库有哪些主要特征？与传统数据库有什么区别？

根据数据仓库概念的含义，数据仓库拥有以下四个特点。

(1) 数据仓库的数据是面向主题的；

(2) 数据仓库的数据是集成的；

(3) 数据仓库的数据是随时间变化的；

(4) 数据仓库的数据是非易失的。

数据仓库与传统数据库的区别（略）。

6. 分析星形模型、雪花模型和星座模型的异同。

同：都符合数据仓库的四大特点，都包含事实表和维表。

异：星形模型通常包括一个大的高度规范化的事实表和多个允许非规范化的维表。以事实表为核心，维度表只与事实表关联，维度表之间没有联系。

雪花模型是对星形模型的扩展，每一个维表都可以连接多个详细类别表，使星形模型中非规范化的维表进一步规范化。

一般一个星形模型或一个雪花模型对应一个主题，它们都有多个维表，但是只有一个事实表。在一个多主题的数据仓库中，存在多个事实表共享某一个或多个维表的情况，就形成星座模型。

7. 了解大数据平台的核心组件及其功能。（略）

8. 了解主流的分布式计算框架。（略）

第 4 章 数据预处理

1. 对于格式化数据而言，原始数据一般会存在哪些问题？

对于格式化数据而言，原始数据一般存在以下问题：

数据缺失；数据异常；数据重复；数据不一致；数据高维性及数据不平衡等。

2. 数据预处理的主要任务可以概括为哪几个方面，每个方面主要解决什么问题？

数据预处理的主要任务可以概括为四个方面：数据清洗；数据集成；数据变换及数据归约。

数据清洗主要处理的是每个数据源中的数据缺失、数据异常和数据重复的问题；数据集成主要处理的是多数据源集成时数据不一致和数据冗余的问题；数据变换主要是把数据变换成适合挖掘的形式；数据归约即数据缩减，对于格式化数据而言，主要指数据行(记录)的缩减、列(属性)的缩减及数值的归约。

3. 对于缺失数据，常用的处理方法有哪些？请举例说明。

常用的缺失数据处理方法可归纳为以下几个方面：

直接删除；使用一个固定的值代替缺失值；使用属性平均值代替缺失值；使用同一类别的均值代替缺失值；使用成数推导值代替缺失值；使用最可能的值代替缺失值。举例（略）。

4. 如何处理异常数据？

首先使用一定的方法（如对于结构化数据，常用的方法有：可视化方法、置信区间检验方法、箱型图分析法、基于距离的方法和基于聚类的方法等）找出异常数据；然后再判断找出的异常是错误导致的还是事物、现象的异常但真实地发展变化，若是错误导致的，则可将孤立点视为噪声或异常而丢弃，或者运用数据平滑技术按数据分布特征修匀数据，或者寻找不受异常点影响的健壮性建模方法。若为后一类情况，则可以寻找原因，进一步挖掘出有意义的信息。

5. 数据集成主要考虑哪些问题？

数据集成主要考虑模式匹配及数值一致化和删除冗余数据两个方面问题。

模式匹配及数值一致化主要解决的是不同数据源中行和列的识别与匹配及一致化属性值的计算方法、计量单位、空间范围和时间范围等方面。

冗余是指存在重复的信息。最明显的冗余是数据中存在两个或多个重复的记录，或者是同一个属性多次出现，或某个属性和其他属性具有明显的相关性。这

类冗余较为容易发现，可以直接删除。而有些冗余比较隐蔽，我们可以使用相关分析加以判别。

6. 什么是分箱？分箱常用的方法有哪些？

分箱是一种将连续型变量转换成序数变量或者类别变量的技术。分箱可分为无指导的简单分箱和有指导的信息分箱。无指导的简单分箱常用的有等宽分箱法和等深分箱法等。有指导的信息分箱常用的如最小熵分箱法等。

7. 什么是不平衡数据？对于不平衡数据，常用的处理方法是什么？

数据不平衡指的是原始数据中不同类别的样本量差异非常大，主要出现在与分类相关的挖掘任务中。对于不平衡数据，常用的处理方法有：通过过采样或欠采样解决不平衡；通过集成方法解决不平衡；通过调整模型类别权重解决不平衡，这种方法不需要对样本本身做处理，只需要在计算和建模过程中，针对不同类别调整其权重进行平衡化处理。

8. 什么是数据归约？数据归约要注意哪些问题？

数据归约即数据缩减，对于格式化数据而言，主要指数据行(记录)的缩减、列(属性)的缩减及数值的归约。数据归约要注意以下两个问题：用于数据归约的时间应当小于在归约后的数据上挖掘节省的时间；归约后的数据集可以获得与原数据集相同或几乎相同的分析结果。

9. 属性归约常用的方法有哪些？

首先可以对属性进行预处理，若存在数值型属性为常量或差异较小，属性值为空值，属性值呈现稀疏性，属性为单调类别变量这四类情况，可以考虑去掉该属性，以实现属性归约的目的；

然后可以采用属性子集选择、主成分分析、聚类分析等方法进行属性选择，以实现属性归约的目的。

10. 结合教材 5-8 章相关案例，调整案例中的参数，练习使用 IBM SPSS Modeler 或 R 语言实现数据预处理。（略）

第 5 章 关联分析

1. 关联规则挖掘问题分哪两个步骤？

关联规则挖掘的第一步，就是根据最小支持度阈值，找出事务数据集中所有的频繁项集；第二步就是根据最小置信度阈值，产生强关联规则。

2. 简述 Apriori 算法原理，并分析其优缺点。

参考教材 5.2。

3. 查阅相关资料，了解关联规则挖掘的其他算法(如 FP-tree)，并与 Apriori 算法比较，分析各自的优缺点。（略）

4. 结合教材案例，调整案例中的参数，练习使用 IBM SPSS Modeler 实现关联分析。（略）

5. 结合教材案例，调整案例中的参数，练习使用 R 语言实现关联分析。（略）

第 6 章 决策树

1. 分析基于有监督学习方法训练模型前需要把数据集划分为哪几个部分？常用的划分方法有哪些？

在有监督的学习中，我们通常至少把数据集划分为训练集和测试集两部分。当模型有超参数存在时，我们需要把数据集划分为训练集、测试集和验证集三部分。

以划分为训练集和测试集两部分为例，常用的划分方法：

留出法：直接把数据集划分成两个互斥的集合，一个为训练集，另一个为测试集。

交叉验证法：先将数据集划分为 k 个大小相似的互斥子集，然后每次用 k-1 个子集的并集作为训练集，余下的那个子集作为测试集，这样可进行 k 次训练和测试。

留一法：为交叉验证法的一个特例，每次只留出一个样本作为测试样本，数据集中的其余样本都作为训练样本，如数据集共有 m 个样本，则可进行 m 次训练和测试。

自助法：如数据集共有 m 个样本，以自助重抽样的方法从数据集中抽取 m 次，得到一个新样本作为训练集，以数据集中没有被抽中的其余的样本作为测试集。

2. 决策树方法的核心问题是什么？如何实现？

（1）决策树的生长问题：第一，如何从众多的输入变量中选择一个当前最佳的分组变量；第二，如何从选定的分组变量中找到最佳的分割点用于分枝。不同的解决方法形成了不同的决策树算法，如 C5.0 等。

（2）决策树的剪枝问题：主要分为预修剪和后修剪两种。预修剪技术主要用于限制树的完全生长，常用的方法有：第一，节点中的样本为同类或空时，停止生长；第二，指定节点样本量的最小值，当节点中样本量少于最小值时，停止生长；第三，指定树的深度，当树的深度达到指定值时，停止生长。后修剪技术是

等待决策树充分生长后再根据一定的标准进行修剪，剪去不具有一般代表性的树枝。采用后剪枝技术的决策树算法，其剪枝的标准也不尽相同，最基本的标准，如可以使用错误率等。

3. 结合 ID3 算法的基本原理，分析其优缺点。

ID3 算法以信息增益为度量标准，用于决策树节点的属性选择，每次优先选取对分类提供信息量最多的属性，以构造一棵熵值下降最快的决策树，到叶子节点处的熵值为零，即每个叶子节点对应的实例集中的实例都属于同一类。

ID3 算法只能处理离散型属性；ID3 算法通过信息增益的方式来选择最优划分属性，会使其倾向于选择取值类别多的属性；ID3 算法不能处理带有缺失值的数据集；ID3 算法没有考虑剪枝。

4. 分析 C5.0 算法基于 ID3 算法改进了哪些方面，并说明每一方面是如何改进的。

C5.0 算法基于 ID3 算法所作的改进：集成了对连续属性的离散化，可以处理连续型变量；使用信息增益率来选择最优划分属性；增加了对缺失值的处理策略，遇到属性值有缺失，会将带有缺失值的样本临时剔除，并进行权重调整处理；使用悲观估计法对树进行剪枝，还可以结合损失代价。

5 结合教材并查阅相关资料，分析集成学习常用的集成方法有哪些。

常用的集成方法有：bagging；boosting；stacking 等。

6. 查阅相关资料，学习 CART(classification and regression tree)算法基本原理，并与 C5.0 算法进行比较，分析其差异性。（略）

7. 查阅相关资料，了解随机森林方法的基本原理。（略）

8. 结合教材案例，调整案例中的参数，练习使用 IBM SPSS Modeler 实现决策树分析。（略）

9. 结合教材案例，调整案例中的参数，练习使用 R 语言实现决策树分析。（略）

10. 查阅相关资料，说明分类模型常用的评价指标有哪些。（略）

第 7 章 贝叶斯分类

1. 阐述朴素贝叶斯分类原理。

朴素贝叶斯分类假设输入变量之间条件独立，即对已知类别，假设所有属性相互独立，也即每个属性独立地对分类结果发生影响。

给定一个未知类别的样本 X，朴素贝叶斯分类将 X 划分到具有最大后验概率

的类C_i中，也就是把X预测为C_i类，当且仅当$p(C_i \mid X) > p(C_j \mid X)$，$1 \leqslant j \leqslant m$，$i \neq j$。

根据贝叶斯定理可得

$$p(C_i \mid X) = \frac{p(X \mid C_i)p(C_i)}{p(X)}$$

由于$p(X)$对所有的类为常数，只需求出分子部分$p(X \mid C_i)p(C_i)$最大的值即可。$p(X \mid C_i)$是一个联合条件概率，X_k之间互相条件独立，所以有

$$p(X \mid C_i)=p(X_1,\ X_2,\ \cdots,\ X_n \mid C_i) = \prod_{k=1}^{n} p(X_k \mid C_i)$$

对于某个新样本x，其描述属性为$\{x_1,\ x_2,\ \cdots,\ x_n\}$，则可通过式(7-4)预测其所属类别为

$$c' = \arg\max_{c_i} \{p(C_i)\prod_{k=1}^{n} p(x_k \mid C_i)\}$$

$p(C_i)$是类先验概率，$p(x_k \mid C_i)$是类条件概率。

2. 阐述 TAN 贝叶斯网络构建过程。

(1) 计算所有输入变量对X_i和X_j的条件互信息，

$$I(X_i;\ X_j \mid Y) = \sum_{x_i, x_j, y} P(x_i,\ x_j,\ y)\log_2 \frac{P(x_i,\ x_j \mid y)}{P(x_i \mid y)P(x_j \mid y)}$$

(2) 依次找到与变量X_i具有最大条件互信息的变量X_j，并以无向边连接节点X_i和节点X_j，得到最大权重跨度树。

(3) 将无向边转为有向边。任选一个输入变量节点作为根节点，所有无向边改为有向边，方向朝外。

(4) 输出变量节点作为父节点与所有输入变量节点相连。

3. 当我们使用搜索引擎输入 “julh” 时，系统会提示 “您是不是要找：july”，结合书中提供的资料，说明系统为什么提示的是 “july” 这个单词，而不是其他？（略）

4. 结合教材案例，调整案例中的参数，练习使用 IBM SPSS Modeler 实现贝叶斯分析。（略）

5. 查阅相关资料，学习 IBM SPSS Modeler 中 “贝叶斯网络” 节点 “模型” 选项卡下马尔可夫覆盖网络的基本原理，说明其与朴素贝叶斯网络和 TAN 贝叶斯网络的差异。（略）

6. 结合教材案例，调整案例中的参数，练习使用 R 语言实现贝叶斯分析。

（略）

7. 结合 7.5 节内容，总结英文文本预处理流程。

英文文本数据可能存在的问题：单词大小写不一致；包含数字；包含标点符号及连词、介词等停用词；多余的空格等。

R 中 tm 添加包提供了文本挖掘的综合处理功能，包括创建语料库、语料库数据预处理和建立文档-词条矩阵等。

(1) 使用 Corpus()函数创建语料库；

(2) 使用 tm_map()函数实现数据转换：主要包括大小写的转换（可使用 tolower 参数），去除所有的数字（可使用 removeNumbers 参数），去除数据中的停用词（可结合使用 stopwords()函数），去除标点符号（可使用 removePunctuation 参数），去除额外的空格（可使用 stripWhitespace 参数）等；

(3) 使用 DocumentTermMatrix()建立行为文档编号，列为词条的文档-词条矩阵，也可以使用 TermDocumentMatrix()建立行为词条，列为文档编号的文档-词条矩阵。

8. 查阅相关资料，了解中文文本预处理流程，说明与英文文本预处理流程的差异。（略）

第 8 章 神经网络

1. 简述神经元的特点。

人工神经元模拟生物神经元的工作方式，首先从各输入端接受输入信息；然后根据连接权值,汇总所有输入信息 ;最后对汇总信息使用激活函数 *f* 进行变换，将其结果映射到一定的取值范围内。

2. 常用的激活函数有哪些？各有什么特点？

常用的激活函数包括线性函数、[0，1]阶跃函数、(0，1)型 Sigmoid 函数和 ReLU 函数等。

特点（略）。

3. 神经网络的拓扑结构包含哪些元素？从信息传播的方向分，可以分为哪些类型？

神经网络的拓扑结构：网络中层的数目、网络中每一层内的节点数以及网络

中的信息是否允许向后传播。

从信息传播的方向分，可以分为前馈神经网络和反馈神经网络。

4. 前馈神经网络训练时，常用的迭代结束条件有哪些？

常用的迭代结束条件：预测误差小于设定的阈值；或前一周期所有的权值调整量都小于设定的阈值；或前一周期正确分类的样本百分比达到设定的阈值；或训练周期数超过设定的阈值等。

5. 以二层前馈神经网络为例，简述 BP 算法的学习过程。

(1) 选择一组训练样本集，每个样本由输入信息和期望的输出结果两部分组成；

(2) 随机初始化隐藏层和输出层中每条有向加权边的权值及偏置；

(3) 从训练样本集中取一个样本，把输入信息输入到神经网络输入层中；

(4) 分别计算经神经元处理后隐藏层及输出层节点的输出；

(5) 计算网络的实际输出和期望输出的误差；

(6) 从输出层反向计算到隐藏层，并按照某种能使误差向减小方向发展的原则（梯度下降法），调整网络中各神经元的连接权值及偏置；

(7) 对训练样本集中的每一个样本重复 (3) - (6) 步骤，直到对整个训练样本集的误差达到要求为止。

也可从工作信号的正向传播和误差信号的反向传播两个方面来分析。

6. 什么是梯度下降法？一般梯度下降法会出现什么问题？查阅相关资料，了解现有的解决方法。(略)

7. 简述卷积神经网络的结构及每个组成部分的作用。

除了输入层，典型的卷积神经网络通常还包括若干个卷积层、激活层、池化层和全连接层(隐藏层和输出层)。其中，池化层不是必需的，有时候会被省略。

卷积层是卷积神经网络的核心。卷积层通过卷积运算，基于“局部感知”和“参数共享”实现降维处理和提取特征的目的。

激活层的作用类似于 BP 神经网络中神经元使用激活函数的作用，其将前一卷积层中的输出，通过非线性的激活函数转换，用以模拟任意函数，从而增强网络的表征能力。

池化就是把小区域的特征通过整合得到新特征的过程。池化层实际上在卷积层的基础上又进行了一次特征提取，最直接的结果是降低了下一层待处理的数据量。

在 CNN 的前面几层是卷积、激活和池化(可以被省略)的多轮交替转换，这些层中的数据通常是多维的。而全连接层就是传统的多层感知机，它的拓扑结构就是一个简单的 $n\times1$ 的模式。

8. 解释卷积核、步长及填充，举例说明卷积运算的实现过程。（略）

9. 结合教材案例，调整案例中的参数，练习使用 IBM SPSS Modeler 实现神经网络分析。（略）

10. 结合教材案例，调整案例中的参数，练习使用 R 语言实现神经网络分析。（略）