**中山大学研究生课程教学大纲**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 中文名称 | 数据挖掘 | | | 课程编号 | | DCS5225 | |
| 英文名称 | Data Mining | | | | | | |
| 总学时 | 36学时 其中实验课：0 学时 | | | | | 学分 | 2 |
| 开课院系 | 计算机学院 | 课程负责人 | 苏勤亮 | 课程性质 | | □必修 □选修 | |
| 授课团队 | 苏勤亮 | | | | | | |
| 课程类别 | □公共课 □专业基础课 □专业方向课  □专业技术实践课 □选修课 | | | | | | |
| 授课方式 | 面授课程 | | 授课语言 | | 中文 | | |
| 考核方式 | 课堂 | | | | | | |
| 先修课程要求 | 概率统计、矩阵分析 | | | | | | |
| 教学目标（100字以内）  本课程拟从大规模数据分析的角度，讲授前沿的机器学习与数据挖掘相关的概念、算法、系统等。要求学生掌握基本的机器学习与数据挖掘概念、算法，并能实现相关的系统，以及更进一步地能够做具有一定创新性的前沿学术研究。 | | | | | | | |
| 课程简介（教学内容及基本要求）  第一周——深度学习基础知识介绍  介绍深度学习的基础知识，包括几种广泛使用的神经网络结构，神经网络常见的优化算法，典型的应用场景等。  第二周——图像表示学习及应用  介绍图像数据的低维度表征获取方法，从最经典的PCA方法，到ImageNet预训练网络提取方法，到近年来获得巨大成功的对比学习方法，MoCo方法，MAE方法等。  第三周——大规模图像检索  介绍大规模图像检索的应用及原理，讨论具体的检索方法和所需的技术。  第四周——自然语言处理基础  介绍自然语言处理的基础知识，包括传统的TFIDF特征，现代的Word2Vec词向量训练原理，现代基于深度模型CNN和RNN的自然语言处理方法，已经基于预训练语言模型BERT的自然语言理解方法和技术。  第五周——深度生成模型原理及应用  介绍深度生成模型的基本原理和基础理论，包括变分自编码器VAE、生成对抗网络GAN、基于能量的模型、基于流的模型等，以及这些模型的具体应用。  第六周——异常检测  介绍异常检测任务和异常检测的方法，包括基于重构的方法、基于概率密度估计的方法、基于一分类器的方法，以及这些模型在深度模型方面的拓展。  第七周——推荐系统  介绍推荐系统的基本原理及常见的推荐系统方法，包括基于内容的推荐算法、协同滤波算法、基于矩阵分解的算法等。  第八周——大语言模型原理及应用  介绍ChatGPT等大语言模型的基本原理及其应用。介绍我国在大语言模型方向在国际竞争中大态势，鼓励同学投诉前沿技术研究。  第九周——生成式人工智能介绍  介绍生成式人工智能的最新进展，包括文生图、可控生成、视频生成等。介绍我国在生成式人工智能方向在国际竞争中大态势，鼓励同学投诉前沿技术研究。 | | | | | | | |
| 教材及主要参考书目、文献与资料 | Mining of Massive Datasets，2th Edition, 2014 （作者：Leskovec, Rajaraman， 出版社：Ullman. Cambridge University Press） | | | | | | |

注：每门课程都应填写此表。