

课程大纲

课程编号: 081202M05004H **课时:** 40 **学分:** 2.00 **课程属性:** 专业普及课 **主讲教师:** 刘莹

课程名称: 数据挖掘21-22秋季

课程英文名称: Data Mining

教学目的、要求

本课程为计算机软件学科研究生的专业普及课程。主要介绍数据挖掘技术的起源、原理、主要算法、关键技术等。课程包含的主要议题包括：数据挖掘的重要性、特点、应用领域、数据仓库、数据预处理技术、关联规则、分类、预测、聚类、顺序模式、深度学习、大数据挖掘等。

本课程采用全英文教学，并将注重理论与实践相结合，使计算机专业研究生掌握数据挖掘的概念的同时，锻炼解决实际问题的能力，为将来的科研工作奠定基础。

预修课程

数据结构，算法，C编程，数据库，概率统计

教材

1、Data Mining: Concepts and Techniques Jianwei Han 2012-03-01 机械工业出版社

主要内容

第一章 Introduction 4学时 刘莹

第1节 Motivation

第2节 Major issues

第3节 Major applications

第4节 Data types

第5节 Characteristics

第二章 Data Warehouse 4学时 刘莹

第1节 Model

第2节 Architecture

第3节 Operations

第三章 Data Pre-processing 5学时 刘莹

第1节 Data cleaning

第2节 Data transformation

第3节 Data reduction

第四章 Association Rules Mining 5学时 刘莹

第1节 Apriori

第2节 Single-pass frequent itemset mining

第3节 FP-Growth

第4节 Multi-level & Multi-dimensional association rules mining

第五章 Classification 7学时 刘莹

第1节 Decision tree

第2节 Bayesian Classifier

第3节 Classification by backpropagation,

第4节 KNN classifier and others

第5节 Prediction models

第6节 Example

第六章 Clustering 5学时 刘莹

第1节 Partitioning methods

第2节 Hierarchical methods

第3节 Density-based methods

第4节 Grid-based methods

第5节 Outlier detection

第七章 Recommendation 2学时 刘莹

第1节 motivation

第2节 Content-based

第3节 Collaborative Filtering

第八章 Applications 3学时 刘莹

第1节 Credit scoring

第2节 oil exploration

第3节 Cosmological Simulation

第九章 Big data mining 5学时 刘莹

第1节 Big data

第2节 Deep Learning

第3节 Text mining

第4节 high performance mining

第5节 other topics

参考用书

课程教师信息

刘莹，女，博士，中国科学院大学计算机与控制学院，教授，博导，中国科学院大学数据挖掘与高性能计算实验室负责人。1999 年获北京大学学士学位，2005 年获美国西北大学博士学位。现任中国计算机学会高性能计算专业委员会委员。主要研究方向为数据挖掘、高性能计算、大数据、深度学习等。已在国际期刊和国际会议上发表学术论文80余篇。曾获北京市科学技术二等奖，获中国科学院朱李月华优秀教师奖，获美国英伟达公司（NVIDIA）中国“最佳教学中心”奖。