



小组项目题目列表 (Topic List)

Topic 1-8 助教: 卢昊桢

Topic 01 (week 4, 图灵机) 介绍图灵机的变种及其一致性

Description “可计算理论 (Computability Theory)” 研究的是现代计算机可以求解的问题是什么样的问题。科学家们在可计算理论中提出了很多种计算模型用于刻画计算机求解问题的能力。尽管科学家们提出了很多的计算模型，然而神奇的是，这些模型的计算能力都和图灵机的计算能力等价 (Church-Turing Thesis)。我们在课堂上已经讲了单带图灵机 (Single-Tape Turing Machine)，希望同学能够了解图灵机的更多变种，如单带图灵机 (Single-Tape Turing Machine) vs 多带图灵机 (Multi-Tape Turing Machine)、单向无限图灵机 (Unidirectional Tape Turing Machine) vs 多向无限图灵机 (Bidirectional Tape Turing Machine)、 $\{0,1\}$ 语言系统图灵机 vs 任意语言系统图灵机。并且了解这些图灵机变种的一致性，即，这些图灵机变种均可以用最简单的图灵机进行表示。希望同学对图灵机的变种及其一致性进行简要介绍。

Topic 02 (week 4, 图灵机) 简要介绍上下文无关文法 (context-free grammar) 和下推自动机 (push-down automata)

Description 在编译原理中，“文法”描述的是高级语言程序的构成规则。乔姆斯基(Chomsky)把文法分成四种类型：0 型、1 型、2 型和 3 型。其中，“0 型文法”又称“短语结构文法”，而图灵机正是 0 型文法的识别器；“2 型文法”又称“上下文无关文法”，它可以描述现今程序设计语言的绝大多数语法结构，而下推自动机正是 2 型文法的识别器。希望同学能够了解乔姆斯基层次结构 (chomsky hierarchy)，并对上下文无关文法 (context-free grammar) 和下推自动机 (push-down automata) 进行简要介绍。

Topic 03 (week 4, 信息论) 简介信道复用技术、有线宽带接入技术 (主要是 ADSL 和 FTTx) 和速率转换关系 (Mbps, MB/s)

Description 复用是通信技术中的基本概念，在计算机网络的信道广泛地使用了各种复用技术。此外，宽带接入技术是目前通信技术中最为活跃的领域之一。希望同学能够了解信道复用技术及有线宽带接入技术 (主要是 ADSL 和 FTTx)，对这两项技术进行简要介绍。

Topic 04 (week 4, 信息论) 介绍 5G 网络的概念、关键技术及其应用

Description 第五代移动通信网络 (5G 网络) 是最新一代蜂窝移动通信技术。希望同学能够了解 5G 网络的基本概念、关键技术、及其应用。对 5G 网络的基本概念、关键技术、及其应用进行简要介绍。

Topic 05 (week 5, 计算机体系结构) 介绍计算机的层次存储结构及其分层价值，分析生活中常见的设备都使用了何种存储部件及原因，并对磁盘阵列技术 (RAID) 进行简要介绍

Description “层次存储”和“通过冗余提高可靠性”是计算机体系结构中八大伟大思想中的两大思想。希望同学能够了解这两大思想。由于课上会对层次存储结构进行简要介绍，所以这里主要是希望同学分析常见的设备使用了何种存储部件及为什么要选用该存储部件，并对磁盘阵列技术 (RAID) 进行介绍。



Topic 06 (week 5, 计算机体系结构) 简介 CPU 的流水线机制 (如 MIPS 的五级流水线机制), 并举例说明“流水线”思想在其他方面的应用

Description “通过流水线提高性能”是计算机体系结构中的又一伟大思想。希望同学能够了解计算机时钟频率、时钟周期、指令周期的概念,了解流水线机制(了解计算机是如何将一条硬件指令的执行过程划分为若干个阶段来进行的)。希望同学对流水线机制及其在其他领域方面的应用进行简要介绍。

Topic 07 (week 5, 数据库) 介绍关系型数据库 (E-R model) 和非关系型数据库 (NoSQL) 的相关概念

Description 按数据模型对数据库进行分类,数据库可分为关系型数据库以及非关系型数据库。希望同学能够了解关系型数据库和非关系型数据库的相关概念,并对他们作简要介绍(比如介绍关系型数据库中的 E-R model,以及非关系型数据库中的基于对象的数据库等)。

Topic 08 (week 5, 数据库) 介绍分布式数据库的相关概念及其关键技术

Description 分布式数据库是用计算机网络将物理上分散的多个数据库单元连接起来组成的一个逻辑上统一的数据库。随着的互联网高速发展,构架于单机系统的数据库已无法满足越来越高的并发请求和越来越大的数据存储需求,因此,分布式数据库被愈加广泛的采用。希望同学能够了解分布式数据库的相关概念及其关键技术,对它们作简要介绍。

Topic 9-16 助教: 陈纪坤

Topic 09 (week 6, 操作系统) 介绍操作系统的概念、结构和功能, 介绍主流操作系统

Description 操作系统是作为计算机硬件和计算机用户之间的中介的程序。操作系统庞大且复杂,因此被构造成功能明确的几个部分。此外,操作系统也不局限于以 Windows, Linux 为代表的桌面操作系统,还包括手持系统,乃至云计算操作系统等很多种类。希望同学们介绍操作系统的概念,操作系统的结构和对应的功能,并对主流的计算机、手机操作系统进行介绍和比较。

Topic 10 (week 6, 操作系统) 介绍进程的概念、生命周期和调度方式

Description 操作系统管理着计算机的资源和运行在计算机系统里的各种应用程序。现代计算机系统允许将多个程序调入内存并发执行,这要求对各种程序提供更严格的控制和更好的划分,这些需求产生了进程的概念,即执行中的程序。进程如何创建,何时终止,如何进行调度,是本课题所要研究的问题。希望同学们对进程的概念、生命周期和调度方式进行介绍。

Topic 11 (week 6, 软件工程) 介绍一个软件项目的完整工业化流程, 包括需求、设计、开发、测试等

Description 现代社会离不开软件。不只是娱乐业,工业制造和分销、金融系统都已经完全计算机化了,因此软件工程对于社会的运转是必不可少的。软件是抽象的,很容易变得极为复杂,对其理解和修改会很困难。因此,使用系统的,有组织的方法开发软件是非常重要的。希望同学们介绍软件过程和软件过程模型的概念,介绍三个一般的软件过程模型:瀑布模型、增量式开发、面向复用的软件工程,以及何时使用它们。在此基础上,请梳理一下需求工程、软件开发、测试和进化中所涉及的基本活动过程。



Topic 12 (week 6, 软件工程) 简单介绍软件工程中的设计模式

Description 软件设计模式 (Design pattern), 是一套被反复使用、多数人知晓的、经过分类编目的、代码设计经验的总结。使用设计模式是为了保证代码的可读性、可靠性和程序的重用性。设计模式分为三大类: 创建型、结构型、行为型。希望同学们介绍设计模式的定义和构成要素、设计模式的分类依据, 并结合自己学习过的编程内容, 选择三种有代表性的设计模式进行介绍。

Topic 13 (week 7, 人工智能) 介绍人工智能的概念、发展和应用领域, 简单讲解如何使用感知机完成分类任务

Description “人工智能”以理解和模拟人类智能, 研究其行为规律为目的, 在计算技术日益提高的过程中受到了人们的格外重视和期待。它包含知识工程、机器学习、模式识别、自然语言处理、机器人学等诸多内容。希望同学们介绍人工智能的概念和发展历程; 简单举例介绍人工智能的应用领域 (如计算机视觉、博弈); 最后, 介绍使用感知机模型进行二分类的数学原理 (关键词: 优化目标, 激活函数, 梯度下降)。

Topic 14 (week 7, 人工智能) 介绍自然语言处理技术的概念和分类, 选择其中一个典型任务 (如文本分类), 介绍经典算法

Description 自然语言处理 (NLP, Natural Language Processing), 是一门融语言学、计算机科学、数学于一体的科学, 是人工智能研究的重要内容。其终极目标是使计算机能理解并生成人类语言。自然语言处理的典型任务有文本分类、语义表示、句法分析、机器翻译等。希望同学们介绍自然语言处理技术的概念和分类, 并选择一个典型任务, 介绍其经典算法。

Topic 15 (week 7, 计算机网络) 介绍计算机网络的分层体系结构, 描述各模块 (网络协议、设备等) 在其中扮演的角色

Description 计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备, 通过通信线路连接起来, 在网络操作系统, 网络管理软件及网络通信协议的管理和协调下, 实现资源共享和信息传递的计算机系统。因特网的协议栈由五个层次组成, 这是计算机网络模型中最重要的概念。希望同学们结合分层体系结构, 介绍计算机网络的重要组成部分: 网络协议、端系统、传输介质、应用软件的概念和作用。

Topic 16 (week 7, 计算机网络) 介绍 IP 协议的概念和运作方式, 路由器工作原理, 讨论 IP 安全性问题

Description 因特网与我们的生活密不可分, 其中有许多重要的概念, 我们可能在正式学习计算机网络课程以前就有所接触。例如, 在网络的体系结构中, 网络层负责主机到主机的通信服务, 我们熟知的 IP 协议、路由器就在其中扮演重要的角色。希望同学们以路由器、IP 地址和协议为着眼点, 介绍这几个常见概念, 以及它们是如何实现转发和路由选择这两大重要功能的。在此基础上, 请同学们简要讨论 IP 协议可能遇到的安全性问题。

Topic 17-24 助教: 卢广霖

Topic 17 (week 8, 机器学习) 介绍有监督学习的概念、特点、及几种经典算法

Description 有监督学习是从标签化训练数据集中推断出函数的机器学习任务。其可以由训练资料中学到或建立一个模型, 并依此模型推测新的实例。训练资料是由输入和预期输出所



组成。函数的输出可以是一个连续的值（称为回归分析），或是预测一个分类标签（称作分类）。希望同学们在介绍有监督学习的概念，特点的基础上，简要介绍有监督学习的经典算法，分别是决策树(decision trees)，线性回归(linear regression)，神经网络(Neural Networks)。

Topic 18 (week 8, 机器学习) 介绍无监督学习的概念、特点、及几种经典算法

Description 现实生活中常常会有这样的问题：缺乏足够的先验知识，因此难以人工标注类别或进行人工类别标注的成本太高。很自然地，我们希望计算机能代我们完成这些工作，或至少提供一些帮助。根据类别未知(没有被标记)的训练样本解决模式识别中的各种问题，称之为无监督学习。希望同学们在了解无监督学习概念，特点的基础上，简要介绍无监督学习的经典算法，即主成分分析方法(PCA)，K均值聚类算法(K-Means)。

Topic 19 (week 8, 物联网) 浅谈 5G 与物联网的关系，以及未来场景中的 5G 物联网用例

Description 5G，即第五代移动通信技术，是最新一代蜂窝移动通信技术。5G 的性能目标是高数据速率、减少延迟、节省能源、降低成本、提高系统容量和大规模设备连接。物联网又称 IoT 技术，是互联网、传统电信网等的信息承载体，让所有能行使独立功能的普通物体实现互联互通的网络。5G 与物联网都是下一代影响人们生活的技术，希望同学们查阅资料，从智能交通，自动驾驶，智能电网方面浅谈 5G 与物联网对人们生活的影响。

Topic 20 (week 8, 物联网) 浅谈人工智能和物联网的关系，以及人工智能在物联网中的实际应用

Description 自 2017 年开始，“AIoT”一词便开始频频刷屏，成为物联网的行业热词。“AIoT”即“AI+IoT”，指的是人工智能技术与物联网在实际应用中的落地融合。当前，已经有越来越多的人将 AI 与 IoT 结合到一起来看，AIoT 作为各大传统行业智能化升级的最佳通道，已经成为物联网发展的必然趋势。希望同学们浅谈人工智能和物联网的关系，以及人工智能在物联网中的实际应用，比如智能家居等。

Topic 21 (week 9, 计算机视觉) 简要介绍如何识别手写数字

Description 数字识别(Digit Recognition)，是计算机从纸质文档，照片，或其他来源接收和理解并识别可读的数字的能力。根据数字来源的产生方式的不同，目前数字识别问题可以区分为手写体数字识别，印刷体数字识别，光学数字识别，自然场景下的数字识别等，具有很大的实际应用价值。请同学们借鉴课堂中如何识别猫和狗的例子，简要介绍如何识别手写数字。

Topic 22 (week 9, 计算机视觉) 简要介绍如何实现人脸识别

Description 人脸识别，是基于人的脸部特征信息进行身份识别的一种生物识别技术。用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸进行脸部识别的一系列相关技术，通常也叫做人像识别、面部识别。请同学们查阅相关资料，简要介绍如何实现人脸识别，即人脸识别的一般步骤，如特征点提取等。

Topic 23 (week 9, 信息安全) 浅谈阿里巴巴的 OceanBase 是如何实现数据安全的

Description OceanBase 是阿里巴巴集团研发的数据库软件，用于淘宝网和诸多阿里集团的云端服务、部分政府机构、银行，擅长于海量资料处理，目前速度为全球排名第一。数据库



安全，是指以保护数据库系统、数据库服务器和数据库中的数据、应用、存储，以及相关网络连接为目的，是防止数据库系统及其数据遭到泄露、篡改或破坏的安全技术。请从保密性、完整性和可用性三个方面浅谈 OceanBase 是如何保障数据安全的。

Topic 24 (week 9, 信息安全) 介绍黑客攻击的一般流程及其技术和方法，简述我们该如何保护自身信息安全

Description 黑客攻击已经成为近年来经常出现的问题。黑客利用计算机系统、网络协议及数据库等方面的漏洞和缺陷，采用后门程序、信息炸弹、拒绝服务、网络监听、密码破解等手段侵入计算机系统，盗窃系统保密信息，进行信息破坏或占用系统资源。请同学们查阅资料，简述黑客攻击的一般流程，并介绍几种黑客攻击方法，如木马攻击，网络蠕虫，SQL 注入攻击。

Topic 25-35 助教：陈鸿滨

Topic 25 (week 10, 语音识别) 介绍 GMM-HMM 声学模型

Description 语音识别系统主要由四部分组成：信号处理和特征提取、声学模型、语言模型和解码搜索部分。其中，声学模型是很重要的一环，它将声学和发音学的知识进行整合，以特征提取部分生成的特征为输入，并为可变长特征序列生成声学模型分数。GMM-HMM 声学模型是传统的经典模型，在 DNN-HMM 流行起来前，构成了现代语音识别系统的基础。希望本组同学简要介绍混合高斯模型（GMM）和隐马尔可夫模型（HMM）的概念和原理，以及如何将这两者有机结合起来组成声学模型（包括如何对语音特征建模、GMM 和 HMM 在声学模型中分别起什么作用等）。

Topic 26 (week 10, 语音识别) 介绍 DNN-HMM 声学模型

Description 语音识别系统主要由四部分组成：信号处理和特征提取、声学模型、语言模型和解码搜索部分。其中，声学模型是很重要的一环，它将声学和发音学的知识进行整合，以特征提取部分生成的特征为输入，并为可变长特征序列生成声学模型分数。GMM-HMM 声学模型是传统的经典模型，是早期现代语音识别系统的基础。近年来，深度学习的兴起也给语音识别技术带来新的发展，比如，深度神经网络（DNN）被应用到声学模型中，代替了 GMM 的位置，与 HMM 一起组成 DNN-HMM 声学模型。希望本组同学能够简要介绍深度神经网络（DNN）的概念及大概的运作原理，隐马尔可夫模型（HMM）的概念，以及 DNN 如何与 HMM 融合（包括如何对语音特征建模、GMM 和 HMM 在声学模型中分别起什么作用等），同时，要与传统的 GMM-HMM 模型简单对比，指出 DNN-HMM 相比 GMM-HMM 的优点。

Topic 27 (week 10, 隐私保护) 介绍大数据隐私保护技术

Description 现在是大数据的时代，大数据分析给我们带来方便的同时，也使得隐私泄露问题日趋严重。大数据具有规模大、来源多、动态更新等特点，传统的隐私保护技术不再适用。希望本组同学从大数据生命周期的发布、存储、分析和使用的这 4 个阶段出发，简要介绍各个阶段中的隐私泄露问题以及对应的隐私保护关键技术。

Topic 28 (week 10, 隐私保护) 介绍差分隐私保护原理及其应用

Description 对于研究机构、信息咨询组织以及政府决策部门来说，数据是非常重要的基础资源。这种需求极大地促进了数据的发布、共享与分析。然而，数据集里通常包含着许多个



人的隐私信息，如医疗诊断结果、个人消费习惯以及其它能够体现个人特征的数据，这些信息会随着数据集的发布和共享而被泄露。为保护个人数据隐私，k-匿名化等一系列匿名化方法被提出，而差分隐私保护具有k-匿名化没有的优点，是一种优秀的数据隐私保护方法。希望本组同学介绍差分隐私的概念，差分隐私保护的原理和该技术的一些应用。

Topic 29 (week 11, 数据可视化) 介绍六种高级数据可视化图表

Description 数据可视化活跃在科学研究、商业分析等领域，降低数据分析的门槛，帮助人们快速方便地理解数据，使用图表进行数据可视化是常见且重要的一环，常见的图表包括折线图、条形图、散点图、饼图等等。除了这些常见图表外，还有一些不常见但在某些场景下十分有表现力的图表，比如箱线图、树状图、热力图等。希望本组同学介绍箱线图、小提琴图、树状图、热力图、甘特图、词云图这六种图表，每种图表需介绍相关概念、适用场景、使用样例等。

Topic 30 (week 11, 数据可视化) 介绍数据可视化工具 Echarts 的特性及如何使用

Description 数据可视化在科学研究、商业分析等方面都有广泛的应用，因此也诞生了许多的可视化工具。我们最熟悉可能是简单易上手的 excel，它是软件类的工具；另外有一类数据可视化库，通过简单的代码来绘制图表，其中，Echarts 是一个开源免费的纯 javascript 的数据可视化库，其具有图表种类多、主题可选、支持动态可视化效果等优点，并且 Echarts 的官网上有详尽的教程文档及使用样例。希望本组同学学习并上手 Echarts，简要介绍 Echarts 的特性及基本使用，基本使用的具体内容包括但不限于安装运行、Echarts 中的基本名词和概念、图表样式简介、如何管理数据、如何加入交互组件等，可以结合具体实例进行介绍。

Topic 31 (week 11, 区块链) 介绍区块链中的共识算法

Description 区块链架构是一种分布式的架构。利用区块链构造基于互联网的去中心化账本，需要解决的首要问题是如何实现不同账本节点上的账本数据的一致性和正确性。这就需要借鉴已有的在分布式系统中实现状态共识的算法，确定网络中选择记账节点的机制，以及如何保障账本数据在全网中形成正确、一致的共识。基于在 20 世纪 80 年代出现的分布式系统共识算法，人们发展出了许多区块链共识算法。希望本组同学简要介绍 PoW 机制、PoS 机制、DPoS 机制的概念和算法的简单流程，并比较它们的异同、优缺点。

Topic 32 (week 11, 区块链) 浅谈区块链应用前景

Description 我国《“十三五”国家信息化规划》中把区块链作为一项重点前沿技术，明确提出需加强区块链等新技术的创新、试验和应用，以实现抢占新一代信息技术主导权。目前，我国区块链技术持续创新，区块链产业初步形成，开始在供应链金融、征信、产品溯源、版权交易、数字身份、电子证据等领域快速应用。希望本组同学在学习了区块链相关知识、对区块链的原理有大体了解后，结合自己对区块链的认识和相关资料，介绍区块链如何在金融、征信、版权保护等方面进行应用，谈谈区块链的应用前景。

Topic 33 (week 12, 推荐系统) 介绍协同过滤推荐

Description 大多数互联网用户都会遇上推荐系统，比如在浏览网页的时候，会被针对性投放广告；购物的时候会被推荐相关商品；还有诸如短视频推荐、新闻推荐等，推荐系统近年来得到了广泛的应用。推荐系统会根据用户的社交关系、浏览记录等来进行推荐，其中涉及到多种多样的推荐算法，比如协同过滤推荐、基于内容的推荐、基于知识的推荐以及混合推



荐方法。希望本组同学介绍协同过滤推荐的概念、算法实际应用案例等，包括基于用户的协同过滤算法，基于物品的协同过滤算法，矩阵因子分解，关联规则挖掘等。

Topic 34 (week 12, 推荐系统) 介绍基于内容的推荐和基于知识的推荐

Description 大多数互联网用户都会遇上推荐系统，比如在浏览网页的时候，会被针对性投放广告；购物的时候会被推荐相关商品；还有诸如短视频推荐、新闻推荐等，推荐系统近年来得到了广泛的应用。推荐系统会根据用户的社交关系、浏览记录等来进行推荐，其中涉及到多种多样的推荐算法，比如协同过滤推荐、基于内容的推荐、基于知识的推荐以及混合推荐方法。希望本组同学介绍基于内容推荐的概念、算法、实际应用案例等，包括内容表示和相似度、基于内容相似度检索等；介绍基于知识推荐的概念、算法、实际应用案例等，包括知识表示和推理、与基于约束的推荐系统交互、与基于实例的推荐系统交互等。

Topic 35 (week 12, 推荐系统) 介绍分析今日头条推荐系统原理

Description 2018 年，今日头条向社会公布了其推荐算法原理。今日头条是一个新闻资讯软件，其针对性地向用户推荐用户喜爱的新闻资讯，据今日头条官方，他们的其他 app 如抖音、火山视频等也使用了同样的推荐系统框架。头条系软件拥有庞大的用户群体，学习他们的推荐系统原理有助于我们了解推荐系统在工业应用中可能会遇到的问题以及解决方案。今日头条较为详细地分享了他们庞大的推荐系统，希望本组同学根据自己对推荐系统的理解，分析并介绍今日头条推荐系统的架构，特点，以及他们遇到的挑战和对应的解决方案。