



人工智能

Lecture 1: 简介

赵培海

东华大学
计算机科学与技术学院

2023 年 9 月 6 日



内容组织

1. 课程简介
2. 人工智能简介
 - 人工智能历史
3. 形式化表示
 - 知识表示
 - 特征表示



内容组织

1. 课程简介
2. 人工智能简介
3. 形式化表示



关于我

赵培海

- 东华大学计算机科学与技术学院，讲师
- 办公地址：松江校区 1 号学院楼 245
- E-mail: peihaizhao@dhu.edu.cn

研究方向

- 人工智能，统计机器学习，随机过程，概率图模型
- 交易风控，行为建模，无人机



关于本课

授课及评定：

- 授课方式：

讲授：2 课时 \times 15

课外作业：1-2 次

- 成绩评定：

考勤 + 作业：40%

期末考试：60%

- 考试方式：

随堂、纸质开卷考试



关于本课

计算机学院 FTP: ftp://10.199.227.254/

- 课件及资料:

ftp://10.199.227.254//课件/赵培海/人工智能/

- 作业提交:

ftp://10.199.227.254//作业提交/赵培海/人工智能/



课程资源

参考教材：

- Artificial Intelligence: A Modern Approach, 3ed
- 人工智能及其应用 - 蔡自兴 (5ed) & 王万良 (3ed)
- Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 6ed

公开课：

- CMU: Artificial Intelligence: Representation and Problem Solving
- Stanford: Artificial Intelligence: Principles and Techniques
- MIT: Artificial Intelligence
- Coursera, Udacity, Edx, Mooc 学院, 等
- 期刊, 会议文章 (KDD, AAAI, ICML, ICDM, JMLR, ...)



课程资源

机器学习：

- Foundations of Machine Learning, Mehryar Mohri, 2ed, 2018.
- Machine Learning: A Bayesian and Optimization Perspective, Sergios Theodoridis, 2ed, 2020
- The Elements of Statistical Learning, Trevor Hastie, 2ed, 2009

动手实践：

- Kaggle
- 阿里天池
- UCI Machine Learning Repository
- Aminer



编程语言

成熟框架：

- TensorFlow, Keras
- PyTorch
- MXNet
- PaddlePaddle

基础语言：

- Python (scikit-learn, Anaconda)
- C++ / Java
- R / Julia
- Matlab



课程目标

- ① 了解人工智能基本内容
- ② 了解人工智能研究范畴
- ③ 了解人工智能的发展态势
- ④ 掌握人工智能基础模型
- ⑤ 运用人工智能方法解决实际问题
- ⑥ 为下一步研究打好基础



涉及内容

- ① 确定性推理（搜索）
- ② 不确定性推理（贝叶斯推理）
- ③ 机器学习
- ④ 深度学习
- ⑤ 群体智能
- ⑥ 强化学习
- ⑦ ...



内容组织

1. 课程简介
2. 人工智能简介
 - 人工智能历史
3. 形式化表示



人工智能发展

人工智能并不是新鲜事物

- ① 1997 年, DeepBlue vs Kasparov
- ② 2011 年, IBM Watson 机器人参加 Jeopardy
- ③ 2016 年, 李世乂 vs AlphaGo
- ④ 2017 年, 柯洁 vs AlphaGo
- ⑤ 2019 年, Openai Five vs OG
- ⑥ 2019 年, 微软超级麻将 AI Suphx



人工智能发展

- 为什么我们如此热衷于人机大战?



人工智能发展

- 为什么我们如此热衷于人机大战?
- 希望创造具有智能的机器



人工智能发展

- 为什么我们如此热衷于人机大战?
- 希望创造具有智能的机器
- 机器胜出是智能还是算力的碾压?



人工智能发展

- 为什么我们如此热衷于人机大战？
- 希望创造具有智能的机器
- 机器胜出是智能还是算力的碾压？
- 人工智能经历了什么？



人工智能

用人工的方式去实现智能



人工智能

用人工的方式去实现智能

什么是智能，如何定义它？



智能

什么是智能？

- 一种独立才能 or 一系列能力的总称？
- 后天学习的 or 先天存在的？
- 学习时发生了什么？
- 在生物体中，知识以何种形式表示？
- ...

难于实现普适的智能系统

图灵测试

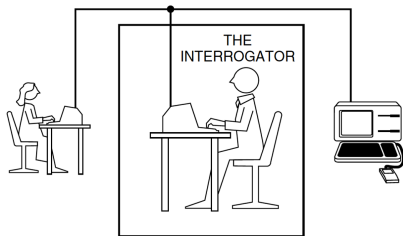
给出了一个客观的智能概念

- 根据对一系列特定问题的反应来确定是否是智能体的行为
- 为判断智能提供一个标准，避免有关智能“真正”特征的讨论

免于受到目前无法回答的问题的牵制

- 如，计算机是否真的意识到其动作？

完全图灵测试



图：图灵测试



要实现智能，不一定是人类智能
正若要飞行，不一定要长翅膀

从其它角度

表: 智能

像人一样思考的系统	理性地思考的系统
要使计算机能思考.....有头脑的机器 (Haugeland, 1985) [使之自动化] 与人类的思维相关的活动, 诸如决策、问题求解、学习等活动 (Bellman, 1978)	通过对计算模型的使用来进行心智能力的研究 (Charniak & McDermott, 1985) 对使得知觉、推理和行动成为可能的计算的研究 (Winston, 1992)
像人一样行动的系统	理性地行动的系统
创造机器来执行人需要智能才能完成的功能 (Kurzweil, 1990) 研究如何让计算机能够做到那些目前人比计算机做得更好的事情 (Rich & Knight, 1991)	计算智能是对设计智能化智能体的研究 (Poole et al., 1998) AI 关心的是人工制品中的智能行为 (Nilsson, 1998)

从其它角度

表: 智能

像人一样思考的系统	理性地思考的系统
要使计算机能思考.....有头脑的机器 (Haugeland, 1985) [使之自动化] 与人类的思维相关的活动, 诸如决策、问题求解、学习等活动 (Bellman, 1978)	通过对计算模型的使用来进行心智能力的研究 (Charniak & McDermott, 1985) 对使得知觉、推理和行动成为可能的计算的研究 (Winston, 1992)
像人一样行动的系统	理性地行动的系统
创造机器来执行人需要智能才能完成的功能 (Kurzweil, 1990) 研究如何让计算机能够做到那些目前人比计算机做得更好的事情 (Rich & Knight, 1991)	计算智能是对设计智能化智能体的研究 (Poole et al., 1998) AI 关心的是人工制品中的智能行为 (Nilsson, 1998)

借助于计算机, AI 为智能理论探索提供了方法与验证环境.



计算 → 智能

↓ 智能是关于：思维感知行动
智人、哲学



计算 → 智能

- ↓ 智能是关于：思维感知行动
智人、哲学
- ↓ 将其形式化：建立目标模型
工程、计算机



计算 → 智能

- ↓ 智能是关于：思维感知行动
智人、哲学
- ↓ 将其形式化：建立目标模型
工程、计算机
- ↓ 形式化要求：合适的表示
数学、表示方法



计算 → 智能

- ↓ 智能是关于：思维感知行动
智人、哲学
- ↓ 将其形式化：建立目标模型
工程、计算机
- ↓ 形式化要求：合适的表示
数学、表示方法
- ↓ 外界的限制：在约束条件下
约束优化、规划、运筹学



计算 → 智能

- ↓ 智能是关于：思维感知行动
智人、哲学
- ↓ 将其形式化：建立目标模型
工程、计算机
- ↓ 形式化要求：合适的表示
数学、表示方法
- ↓ 外界的限制：在约束条件下
约束优化、规划、运筹学
- ↓ 计算机实现：编写算法
程序设计、数据结构、算法、分布式、数据库



AI 使得一些课程变得 “有用” 了!



AI 使得一些课程变得 “有用” 了！
大家都在说人工智能，但所在的角度可能大相径庭



哲学

哲学贡献：

- ① 形式化规则能用来抽取合理的结论吗？
- ② 精神的意识是如何从物质的大脑产生出来的？
- ③ 知识是从哪里来的？
- ④ 知识是如何导致行动的？



数学

数学贡献：

- ① 如何抽取形式化规则来得到合理结论？
- ② 什么可以被计算？
- ③ 如何用不确定的知识进行推理？
- ④ 形式化工具：逻辑、概率、线性代数等等

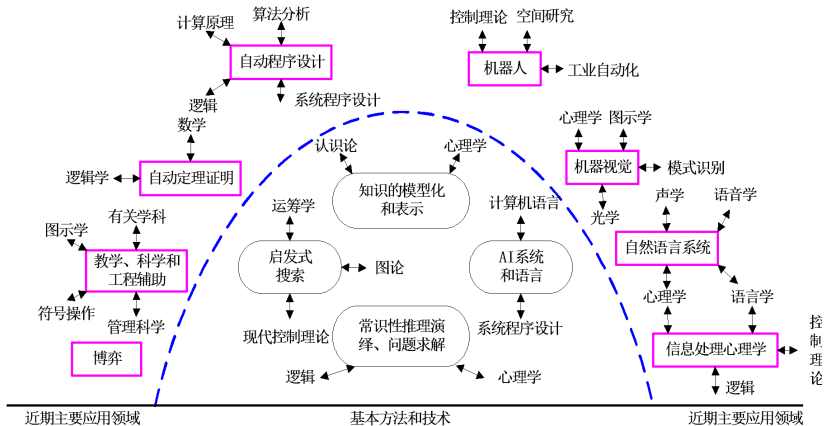


经济学

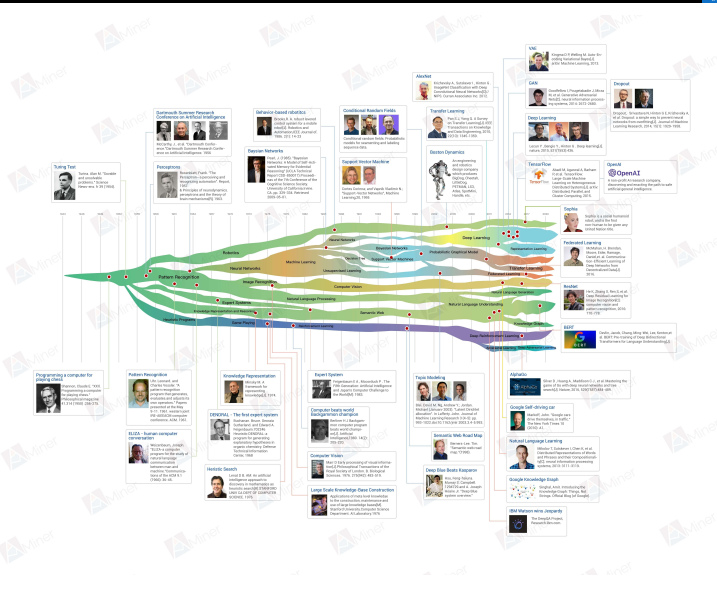
经济学贡献：

- ① 如何决策以获得最大收益？
- ② 在他人不合作的情况下如何做到这点？
- ③ 在收益遥遥无期的情况下如何做到这点？

多学科交叉知识结构



图：多学科交叉知识结构

图: 人工智能方法及领域发展¹



形式化方法发展

形式化大致的发展阶段：

- ① 符号/逻辑推理
- ② 确定性推理
- ③ 不确定性推理、统计推理
- ④ 群体智能
- ⑤ 机器学习、随机过程
- ⑥ 深度学习



人工智能应用

- 风控系统
- 推荐系统
- 自然语言理解
- 自动推理和定理证明
- 图像/视频处理
- 机器人
- 自动控制
- ...



内容组织

1. 课程简介
2. 人工智能简介
3. 形式化表示
 - 知识表示
 - 特征表示



表示

表示很重要!!!:

- 表示决定了后续的推理分析过程;
- 好的知识表示, 提高系统推理能力;
- 好的特征表示, 优化模型效率;



表示

表示很重要!!!:

- 表示决定了后续的推理分析过程;
- 好的知识表示, 提高系统推理能力;
- 好的特征表示, 优化模型效率;

在实际应用中不要无脑深度学习, 很多问题通过良好表示, 简单模型就可达到效果;



表示系统

表示系统：所有表示模式的功能都是要捕捉（或抽取）**问题域中的本质特征**并使这一信息能够被问题求解过程所访问。**抽象（abstraction）**是处理复杂性的一种关键工具，也是保证最终程序计算高效性的一个重要因素。

- 表现力（特征抽取结果）和效率（特征抽取算法的计算复杂度）是评价知识表示语言的主要尺度
- 有时，为了提高效率必须牺牲表现力。但这是以不限制捕捉关键的问题求解知识的表示能力为前提的
- 如，浮点数表示
- 如，罗马数字、阿拉伯数字表示



什么是知识

知识：

- 在长期的生活及社会实践中、在科学研究及实验中积累起来的对客观世界的认识与经验；
- 把有关信息关联在一起所形成的信息结构；
- 反映了客观世界中事物之间的关系，不同事物或者相同事物间的不同关系形成了不同的知识；

什么是知识

知识就是力量：

- 知识和创新是推动人类发展的动力；
- 人类个体比其他动物没有多大优势，掌握了符号语言，人类社会的结构发生了突变，有了一个连接在一起的集体大脑；



图：猩球崛起



什么是知识

知识的种类：

- ① 事实性知识：采用直接表示的形式
如：凡是猴子都有尾巴
- ② 过程性知识：描述做某件事的过程
如：电视使用说明书
- ③ 类比性知识：既不给出外延，也不给出内涵，只给出他与其他事物的某些相似之处
如：比喻、谜语
- ④ ...



知识的表示

知识的形式化表示方法：

- ① 命题/谓词表示
- ② 产生式
- ③ 状态空间表示
- ④ 语义网络表示
- ⑤ 框架表示
- ⑥ 其它, 如 Petri Net 表示



命题表示

A proposition is the non-linguistic bearer of truth or falsity which makes any sentence that expresses it either true or false.

–Wikipedia



命题表示

A proposition is the non-linguistic bearer of truth or falsity which makes any sentence that expresses it either true or false.

–Wikipedia

命题表示：

- 明确定义的形式语义 (formal semantics)
- 可靠 (sound) 和完备 (complete) 的推理规则

命题演算：

- 为命题语句赋真值被称为解释 (Interpretation)
- 1956 年，逻辑理论家 (Logic Theorist)

命题逻辑

命题的简单推理: 交换律, 结合律, 分配律, 真值表

P	Q	$\neg P$	$\neg P \vee Q$	$P \Rightarrow Q$	$(\neg P \vee Q) = (P \Rightarrow Q)$
T	T	F	T	T	T
T	F	F	F	F	T
F	T	T	T	T	T
F	F	T	T	T	T

Basic and derived argument forms [edit]

Name	Sequent	Description
Modus Ponens	$((p \rightarrow q) \wedge p) \vdash q$	If p then q ; p ; therefore q
Modus Tollens	$((p \rightarrow q) \wedge \neg q) \vdash \neg p$	If p then q ; not q ; therefore not p
Hypothetical Syllogism	$((p \rightarrow q) \wedge (q \rightarrow r)) \vdash (p \rightarrow r)$	If p then q ; if q then r ; therefore, if p then r
Disjunctive Syllogism	$((p \vee q) \wedge \neg p) \vdash q$	Either p or q , or both; not p ; therefore, q
Constructive Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (p \vee r)) \vdash (q \vee s)$	If p then q ; and if r then s ; but p or r ; therefore q or s
Destructive Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (\neg q \vee \neg s)) \vdash (\neg p \vee \neg r)$	If p then q ; and if r then s ; but not q or not s ; therefore not p or not r
Bidirectional Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (p \vee \neg s)) \vdash (q \vee \neg r)$	If p then q ; and if r then s ; but p or not s ; therefore q or not r
Simplification	$(p \wedge q) \vdash p$	p and q are true; therefore p is true
Conjunction	$p, q \vdash (p \wedge q)$	p and q are true separately; therefore they are true conjointly
Addition	$p \vdash (p \vee q)$	p is true; therefore the disjunction (p or q) is true
Composition	$((p \rightarrow q) \wedge (p \rightarrow r)) \vdash (p \rightarrow (q \wedge r))$	If p then q ; and if p then r ; therefore if p is true then q and r are true

命题逻辑

命题的简单推理: 交换律, 结合律, 分配律, 真值表

P	Q	$\neg P$	$\neg P \vee Q$	$P \Rightarrow Q$	$(\neg P \vee Q) = (P \Rightarrow Q)$
T	T	F	T	T	T
T	F	F	F	F	T
F	T	T	T	T	T
F	F	T	T	T	T

Basic and derived argument forms [edit]

Name	Sequent	Description
Modus Ponens	$((p \rightarrow q) \wedge p) \vdash q$	If p then q ; p ; therefore q
Modus Tollens	$((p \rightarrow q) \wedge \neg q) \vdash \neg p$	If p then q ; not q ; therefore not p
Hypothetical Syllogism	$((p \rightarrow q) \wedge (q \rightarrow r)) \vdash (p \rightarrow r)$	If p then q ; if q then r ; therefore, if p then r
Disjunctive Syllogism	$((p \vee q) \wedge \neg p) \vdash q$	Either p or q , or both; not p ; therefore, q
Constructive Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (p \vee r)) \vdash (q \vee s)$	If p then q ; and if r then s ; but p or r ; therefore q or s
Destructive Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (\neg q \vee \neg s)) \vdash (\neg p \vee \neg r)$	If p then q ; and if r then s ; but not q or not s ; therefore not p or not r
Bidirectional Dilemma	$((p \rightarrow q) \wedge (r \rightarrow s) \wedge (p \vee \neg s)) \vdash (q \vee \neg r)$	If p then q ; and if r then s ; but p or not s ; therefore q or not r
Simplification	$(p \wedge q) \vdash p$	p and q are true; therefore p is true
Conjunction	$p, q \vdash (p \wedge q)$	p and q are true separately; therefore they are true conjointly
Addition	$p \vdash (p \vee q)$	p is true; therefore the disjunction (p or q) is true
Composition	$((p \rightarrow q) \wedge (p \rightarrow r)) \vdash (p \rightarrow (q \wedge r))$	If p then q ; and if p then r ; therefore if p is true then q and r are true

命题演算无法分析断言的各个部分;

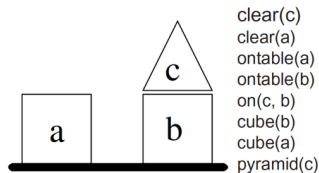
谓词表示

谓词演算：

- 允许表达式中含有变量
- 允许使用对象的函数
- 量词: \forall, \exists
- 谓词语义

likes(george,kate)
likes(george,susie)
likes(george,sarah,tuesday)
friends(bill,george)
helps(bill,george)

likes(X,george)
likes(X,X)
friends(bill,richard)
friends(father_of(david),father_of(andrew))
helps(richard,bill)





产生式表示

确定性规则产生式

- 基本形式: IF P THEN Q
- 或者: $P \rightarrow Q$
- 如:
Rule: IF 动物会飞 AND 会下蛋 THEN 该动物是鸟

不确定性规则产生式

- 基本形式: IF P THEN Q (置信度)
- 或者: $P \rightarrow Q$ (置信度)
- 如:
Rule: IF 发烧 THEN 感冒 (0.6)

状态空间表示

一个状态用一个节点表示, 如: 拼图游戏 (8-Puzzle)

1	2	3
8	6	
7	5	4

图: 一个状态节点

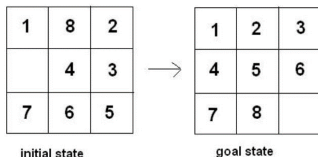


图: 初始与目标状态节点

语义网络表示

语义网络:

- 表示形式 (采用网络形式表示人类的知识)
- 语义网络一般由一些最基本的语义单元组成. 可用如下三元组来表示: (结点 1, 弧, 结点 2) 每一个要表达的事实用一个节点表示, 事实之间的关系用有向弧表示
- 当把多个语义基元用相应的语义联系关联在一起的时候, 就形成了一个语义网络

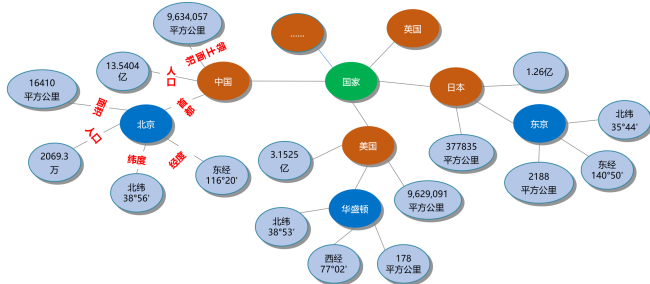
例如 “香蕉是黄色的”



语义网络表示

知识图谱：

- 旨在描述真实世界中存在的各种实体或概念及其关系，一般用三元组表示
- 可被看作是一张巨大的图，节点表示实体或概念，边则由属性或关系构成



语义网络表示

Google

北京

✕



[全部](#) [图片](#) [新闻](#) [视频](#) [更多](#)
[设置](#) [工具](#)

找到约 1,270,000,000 条结果 (用时 0.58 秒)

zh.wikipedia.org › zh-hans › 北京市

北京市- 维基百科，自由的百科全书

北京市，通称**北京**（汉语拼音：Běijīng；邮政式拼音：Peking），简称“京”，是中华人民共和国的首都及直辖市，是中国的政治、文化、科技和国际交往中心，是 ...

- 市长: 陈吉宁

最大区县: 密云区

市树: 国槐、银杏

- 密度: /km²

东城区 · 通州区 · 中华人民共和国省级行政区土地 ... · 北京市交通

baike.baidu.com › item › 北京

北京（中华人民共和国首都）_百度百科

北京市，简称“京”，古称燕京、北平，是中华人民共和国的首都、省级行政区、直辖市、国家中心城市、超大城市，国务院批复确定的中国政治中心、文化中心、 ...

面积: 16410.54 km²

地理位置: 中国华北地区，华北平原北部

人口数量: 2153.6万人（2019年常住人口）

地区生产总值: 35371.3亿元（2019年）

故言 · 北京大学 · 朱刚演唱歌曲 · 燕京

www.beijing.gov.cn

首都之窗_北京市人民政府门户网站

首都之窗是北京市国家机关在互联网上统一建立的网站群，包括北京市人民政府门户网站、市级各部门、各区政府和各级国家机关网站。首都之窗网站由北京市人民 ...

www.bj.xinhuanet.com

新华网北京频道_北京新闻_首都



北京

中国的首都

北京市，通称北京，简称「京」，是中华人民共和国的首都及直辖市，是中国的政治、文化、科技和国际交往中心，是世界人口第三多的城市和人口最多的首都，具有重要的国际影响力。北京位于华北平原的西北边缘，背靠燕山，有永定河流经老城西南，毗邻天津市、河北省，为京津冀城市群的重要组成部分。 [维基百科](#)

面积: 6,490 m²

天气: 26°C, 风向东南, 风速 6 公里/时, 湿度 49%

人口: 2154 万 (2018 年)

当地时间: 星期一 下午 1:34

市树: 国槐、银杏

市花: 月季、菊花

人口 (2017) : 2170.7 万

图: Google 搜索引擎

知识 → 特征

有时一些知识或概念无法形式化描述, 就需要再往下一层, 寻找表示方法, 如:

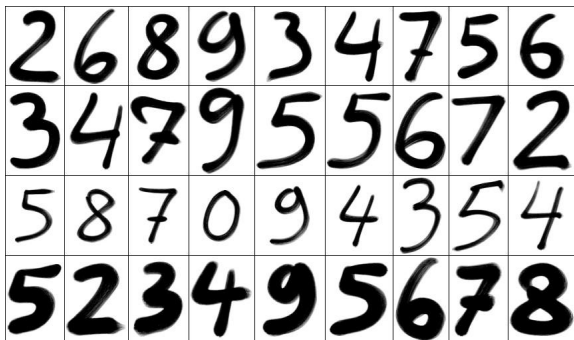


图: 手写数字识别

知识 → 特征

有时一些知识或概念无法形式化描述, 就需要再往下一层, 寻找表示方法, 如:

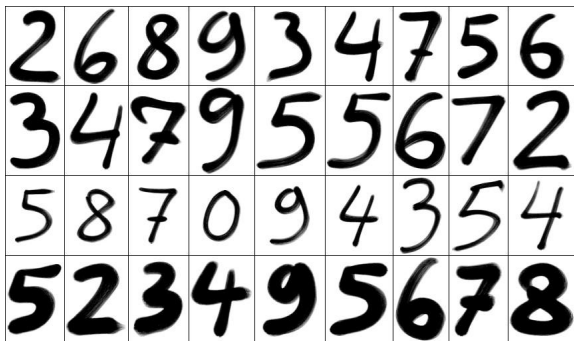


图: 手写数字识别

特征提取与表示, 是机器学习模型的基石.



特征表示

西瓜特征提取与表示

回头看第一段话, 我们会发现这里涉及很多基于经验做出的预判. 例如, 为什么看到微湿路面、感到和风、看到晚霞, 就认为明天是好天呢? 这是因为在我们的生活经验中已经遇见过很多类似情况, 头一天观察到上述特征后, 第二天天气通常会很好. 为什么色泽青绿、根蒂蜷缩、敲声浊响, 就能判断出是正熟的好瓜? 因为我们吃过、看过很多西瓜, 所以基于色泽、根蒂、敲声这几个特征我们就可以做出相当好的判断. 类似的, 我们从以往的学习经验知道, 下足了工夫、弄清了概念、做好了作业, 自然会取得好成绩. 可以看出, 我们能做出有效的预判, 是因为我们已经积累了许多经验, 而通过对经验的利用, 就能对新情况做出有效的决策.

图: 机器学习, 周志华



特征 → 知识

从特征提取知识的过程就是一个学习的过程, 用计算机通过批量数据来实现这一过程, 就是机器学习.



特征 → 知识

从特征提取知识的过程就是一个学习的过程, 用计算机通过批量数据来实现这一过程, 就是机器学习.

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .

– Tom M. Mitchell



特征表示

样本空间与特征空间



特征表示

样本空间与特征空间

One-Hot 表示

- 离散表示, 局部表示



特征表示

样本空间与特征空间

One-Hot 表示

- 离散表示, 局部表示

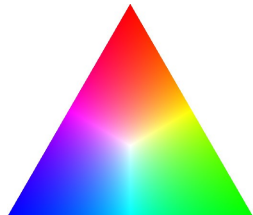
分布式 (distributed) 表示

- 压缩、低维、稠密向量
- 用 $O(N)$ 个参数表示 $O(2^k)$ 区间

分布式表示

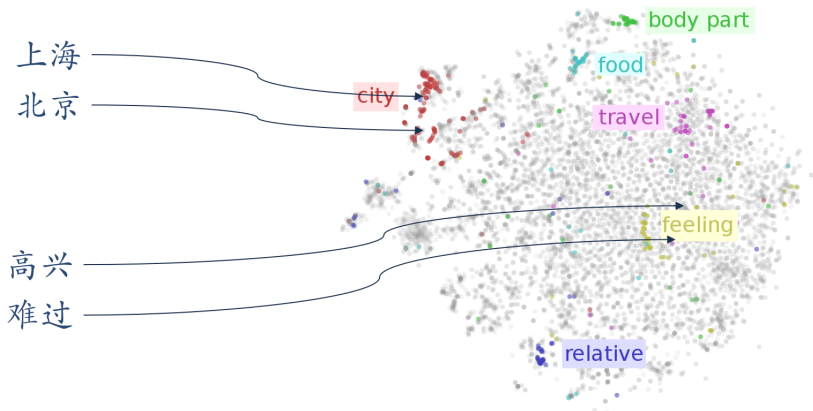
分布式表示的实例: 颜色

颜色	局部表示	分布式表示
琥珀色	$[1, 0, 0, 0]^T$	$[1.00, 0.75, 0.00]^T$
天蓝色	$[0, 1, 0, 0]^T$	$[0.00, 0.5, 1.00]^T$
中国红	$[0, 0, 1, 0]^T$	$[0.67, 0.22, 0.12]^T$
咖啡色	$[0, 0, 0, 1]^T$	$[0.44, 0.31, 0.22]^T$



分布式表示

嵌入 (Embedding) : 词嵌入, 节点嵌入, 图嵌入等





特征表示的研究

特征抽取：

- 特征抽取与特征提取



特征表示的研究

特征抽取：

- 特征抽取与特征提取

传统特征抽取方式：

- 线性投影 (子空间)：PCA, LDA (Linear Discriminant Analysis)
- 非线性嵌入：LLE, Isomap, 谱方法
- 自编码器



特征表示的研究

表示学习：

- 通过学习产生分布式特征向量
- 优点：不用手动构建特征向量
- 难点：没有明确定义什么是“好”



特征表示的研究

表示学习：

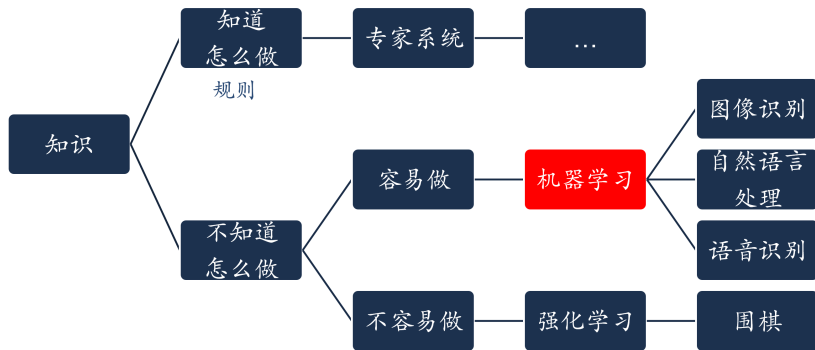
- 通过学习产生分布式特征向量
- 优点：不用手动构建特征向量
- 难点：没有明确定义什么是“好”

特征抽取 vs. 表示学习：

- 特征抽取：基于任务或先验对去除无用特征
- 表示学习：通过深度模型学习高层语义特征



人工智能系统的设计





Question & Answer