监督信息的模糊化:
一种弱监督学习机制

王熙照
深圳大学计算机学院

Lecture 01

Introduction to Machine Learning

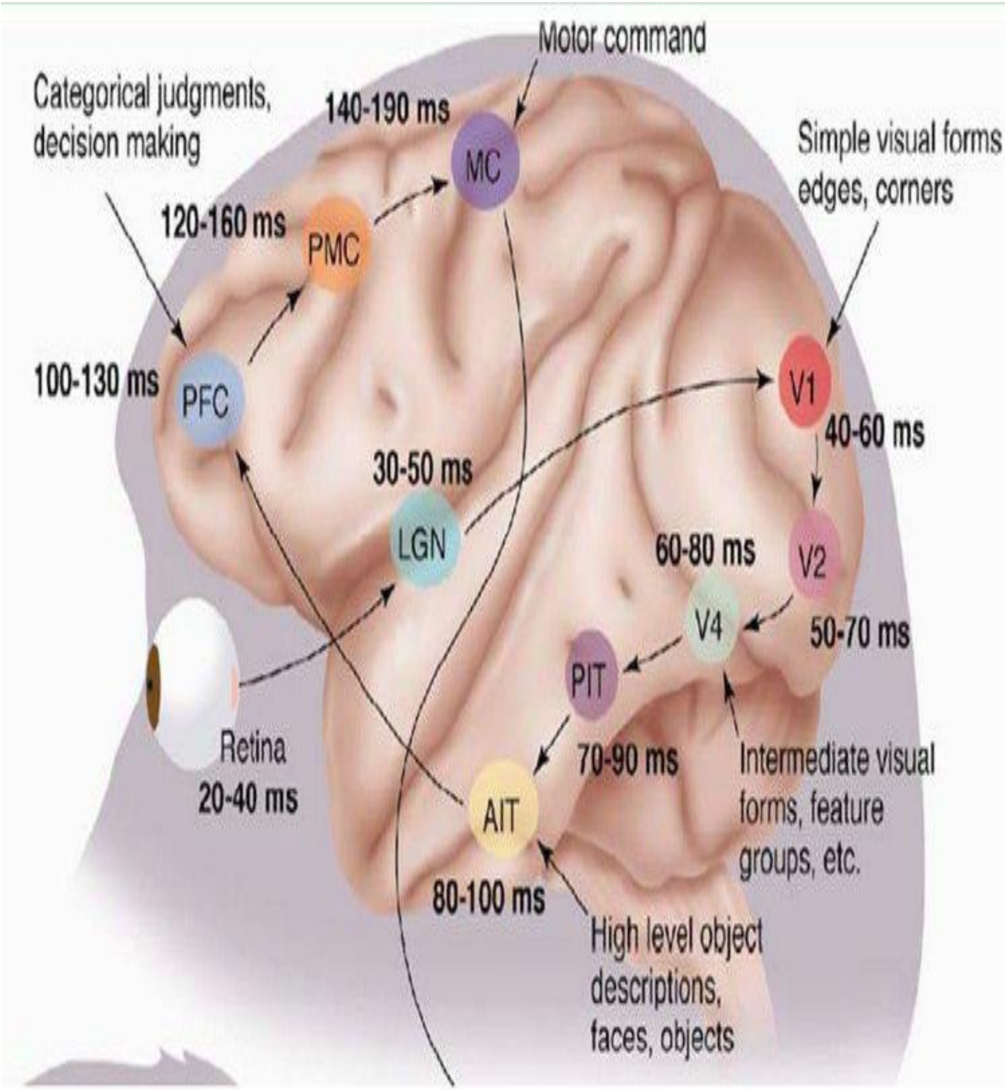
Xizhao WANG

Big Data Institute
College of Computer Science
Shenzhen University

March 2019

Lecture 01: Introduction to Machine Learning

1. AI brief history
2. What is ML?
3. AI and ML
4. ML categories
5. ML aims
6. References



Dartmouth Summer School

达特茅斯会议

- 1956 Summer School

J. McCarthy, M. Minsky, N. Lochester (IBM), C.E. Shannon

A. Samual (IBM), H.A. Simon (CMU), A. Newell (CMU),

T. More (Prinston), R. Solomonoff (MIT), O. Selfridge (MIT)

- Funded by the Rockefeller Foundation, $1,200 per person

洛克菲勒基金会资助，每人1200美元，报销往返车票

- Goal： Design a computer with real intelligence

目标：10个人2个月设计出具有真正智能的计算机

- Result： Laid a new science - artificial intelligence

结果：奠定了新的科学- 人工智能

Birth of AI
1956
Dartmouth Conference

1957
Invention of Perceptron Perceptual Neural Network

Trough
1970-1980
Computer performance is insufficient, data volume is seriously missing.

Expert system
1980
KB + Inference machine

1986
Invention and application of BP neural network algorithm

Statistical learning
1995
Representative technology is support vector machine: SVM

Deep learning
2006
Hintion proposed deep learning neural network, and achieved breakthrough progress

2013
Deep learning algorithm is successful in speech recognition and visual recognition rate, were more than 99% and 95%, entered the era of intelligent perception

2015
Big data + cloud computing is developing rapidly, AI is beginning to be fully applied.

New era
2016
Google AlphaGo wins World Cup champion Li Shishi, opening a new era

AI history

What is ML?

AI and ML

ML categories

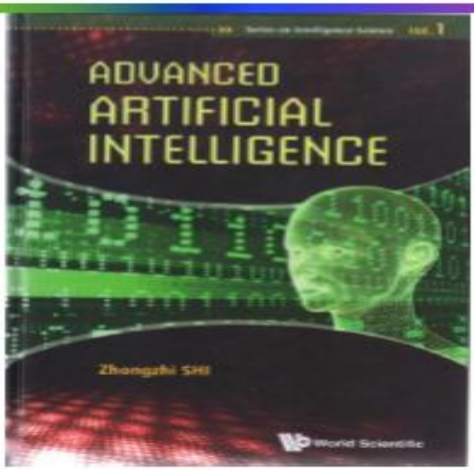
ML aims

References

人工智能

Using artificial methods and techniques to imitate, extend and expand human intelligence to realize machine intelligence.

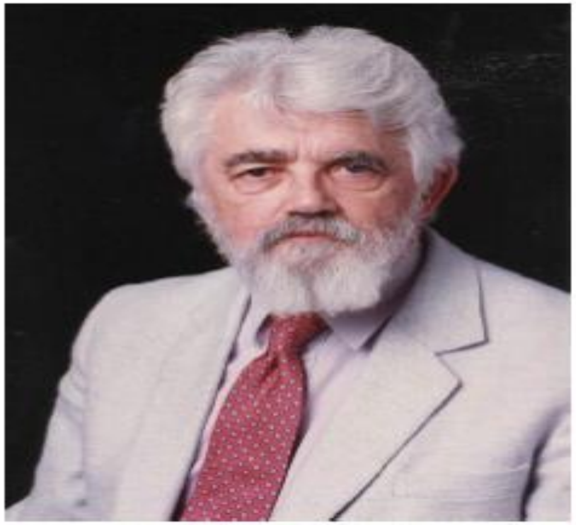
用人工的方法和技术，模仿、延伸和扩展人的 智能，实现机器智能。



The long-term goal of Artificial Intelligence is human-level Artificial Intelligence.

Cite from: John McCarthy. The Future of AI—A Manifesto. AI Magazine Volume 26 Number 4, 2005.

Intelligence Science Is The Road To Human-Level Artificial Intelligence



What is Machine Learning?

For certain types of tasks T and performance measure P, if a computer program on the T measures performance P with the experience E and self-improvement. So we call this computer program learning from experience E

—— Tom Mitchell

机器学习其实是一门多领域交叉学科，它涉及到计算机科学、概率统计、函数逼近论、最优化理论、控制论、决策论、算法复杂度理论、实验科学等多个学科。机器学习的具体定义也因此有许多不同的说法，分别以某个相关学科的视角切入。但总体上讲，其关注的核心问题是如何用计算的方法模拟类人的学习行为：从历史经验中获取规律（或模型），并将其应用到新的类似场景中。

—— Microsoft Research

机器学习是一门研究机器获取新知识与新技能，并识别现有知识的学问。从人工智能的角度，机器学习是指从经验中产生模型的一切方法论的总称。学习模型的构建是机器学习的核心研究内容。取决于已有知识表示形式、学习任务与学习环境，机器学习的研究内容十分广泛，涉及规则学习、类比学习、统计学习、强化学习、深度学习、大数据机器学习等多个方面。

—— Xiang and Liu

What is Machine Learning?

AI history

What is ML?

AI and ML

ML categories

ML aims

References

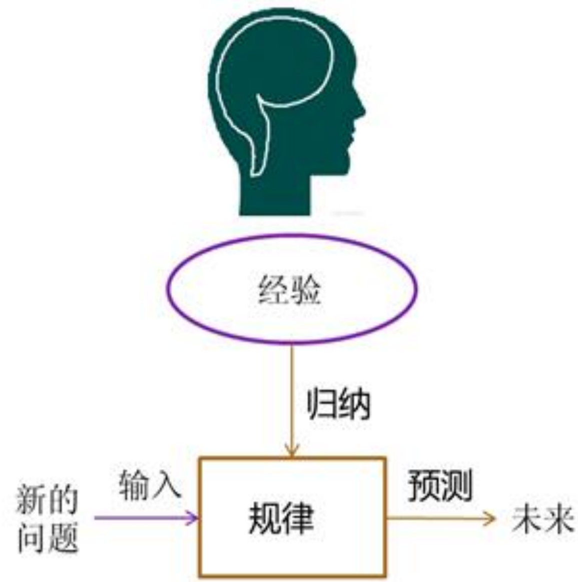
什么是机器学习？简单地说，计算机利用输入的大量样本数据，调整表示规律和分类通用数学模型的参数，然后以调好的模型作答。通常用线性函数的组合来表示数值规律和划分类别模式，实用中的线性函数参数是以万计到百亿计的数量。这样的数学模型虽然很简单，却因参数数量的巨大能够实现复杂的功能，足以涵盖各种预测和辨识情况。在数学上，这调整模型参数及应用模型的计算机制，都是精确有效的，但也因变量个数的巨大，难以分析归纳成像物理规律那样简单明晰的因果性机制，无法从人脑逻辑推演的角度来理解。

http://blog.sciencenet.cn/blog-826653-1029786.html

What is Machine Learning?

- AI history
- What is ML?
- AI and ML
- ML categories
- ML aims
- References

A computer discovers/extracts a model from existing data(experience), and then uses this model to complete a prediction task.



http://blog.sciencenet.cn/blog-826653-1029786.html

What is Machine Learning?

Machine Learning is a technique (skill) to study how do computers simulate/implement human's behavior of learning. It is to acquire new knowledge and then re-organize the existing knowledge structure in order to improve the its performance of problem-solving.

—— A summary

ZHOU Zhihua -—— A thinking on Machine Learning

What is the relationship between AI and ML?

AI history

What is ML?

AI and ML

ML categories

ML aims

References

Traditionally in text books, it is stated that:
Machine Learning is a key part of Artificial Intelligence.

Traditionally AI has 4 fundamental tasks:
Knowledge representation
Search
Learning
Reasoning

Another popular opinion is that:
AI = ML + (Big) Data

Learning is the essential way for human to get wisdom. Machine Learning is fundamental and indispensible for a computer to acquire intelligence. ML is a necessary part for any computers to intelligently solve problems.

------ Jiarong Hong

- AI history
- What is ML?
- AI and ML
- ML categories
- ML aims
- References

智能科学技术一级学科包括的2级学科

- Brain Cognition
- ML & PR
- NL processing & understanding
- Knowledge Engineering
- Robotics and Smart systems

AI history

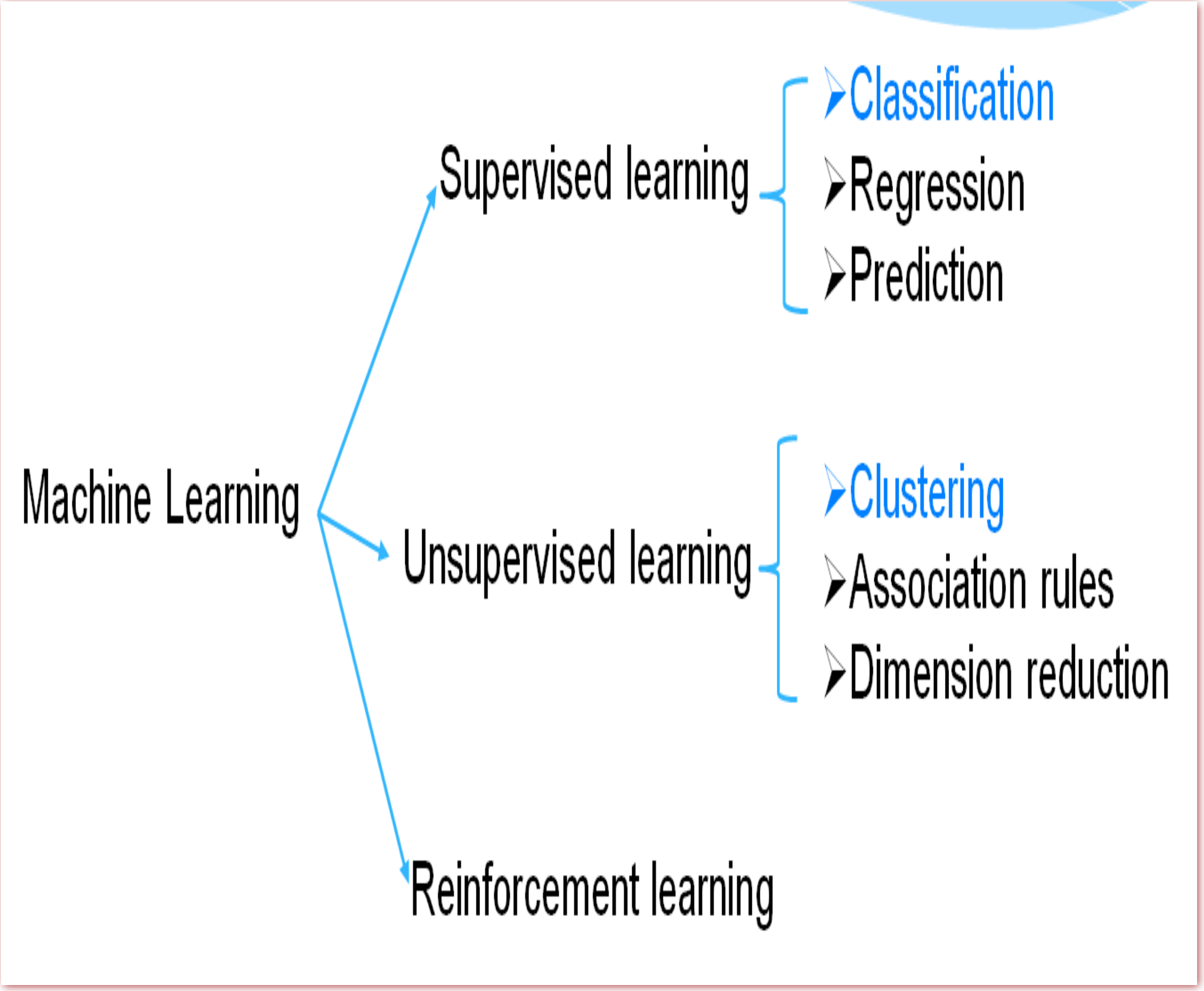
What is ML?

AI and ML

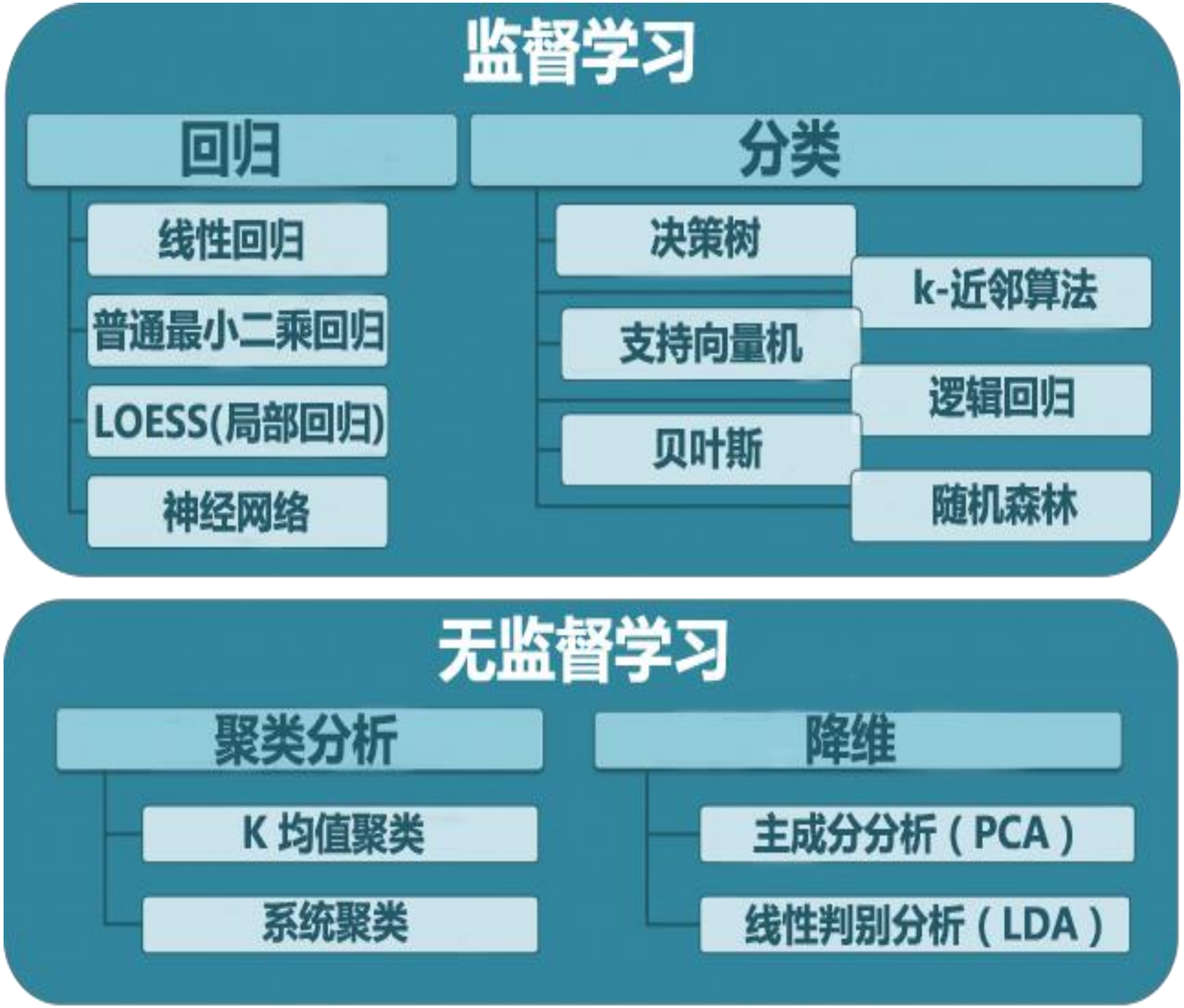
ML categories

ML aims

References



- AI history
- What is ML?
- AI and ML
- ML categories
- ML aims
- References



Supervised Learning: Prediction

AI history

What is ML?

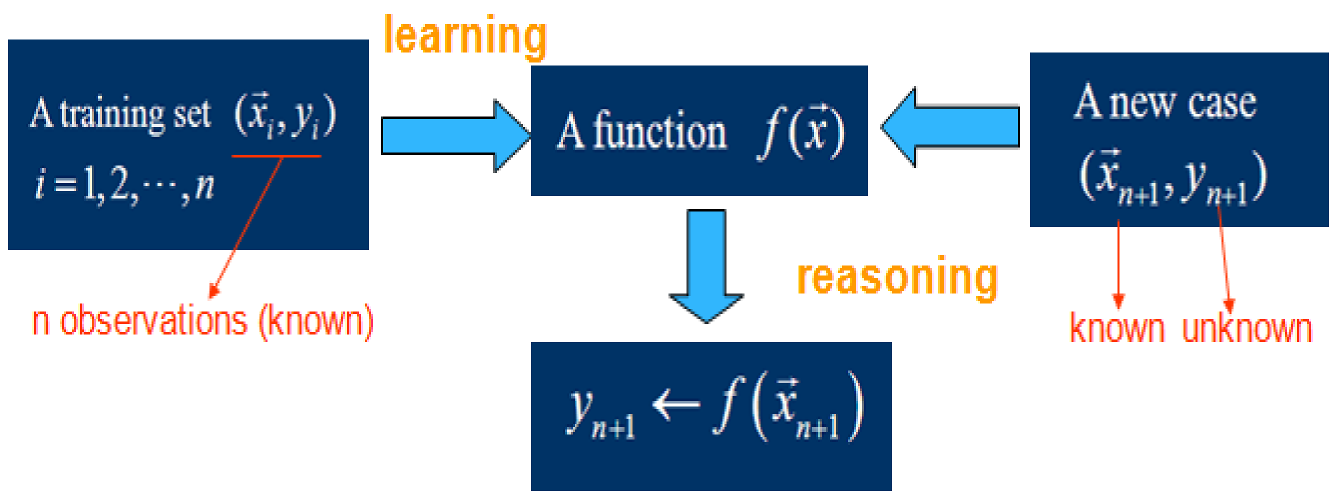
AI and ML

ML categories

ML aims

References

A general framework of supervised learning



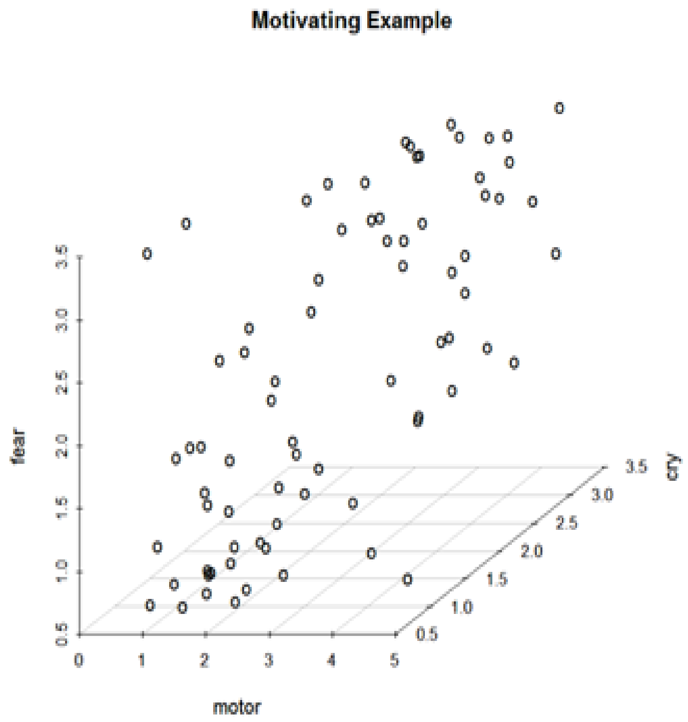
f(*x*) can have many forms, such as:

1. A regression function
2. A nearest Neighbor model
3. A set of rules
4. A neural network
5. A Bayesian network
6. etc

- y_i can be discrete or continuous

Unsupervised Learning: Clustering

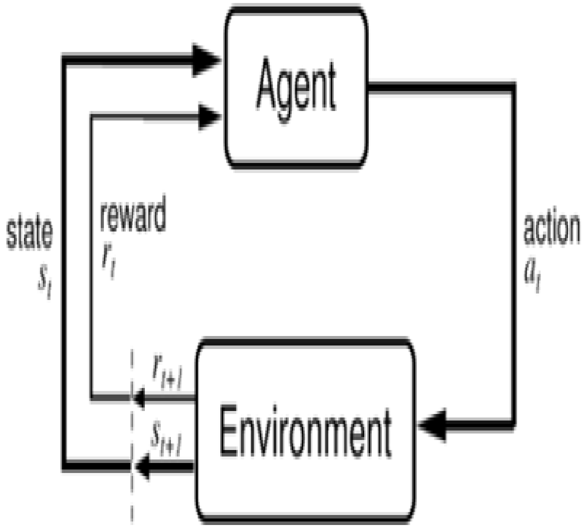
- Given a data set, can we find natural groupings or clusters in the data?
- How can we decide how many groups exist?
- Could there be subgroups within the groups?



Reinforcement Learning: learning from environment

"**Reinforcement learning** (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward"

— *Wikipedia*



The aim of machine learning is used to learn from existing data into knowledge to accurately predict unknown output as possible. Therefore, the accuracy of learning, known as ability to predict unknown output, or known as generalization. It has been a goal of machine learning all the time.

Improve the accuracy of prediction

Reducing the complexity of the search

Enhance understandability represented

The problems of machine learning are often attributed to the search problem, which is a very large search space to search in order to determine the best fit observed data and prior hypothesis of the learner. Therefore, machine learning improves learning accuracy while it also pays attention to reduce the complexity of the search, in order to improve learning efficiency.

Improve the accuracy of prediction

Reducing the complexity of the search

Enhance understandability represented

The knowledge that the system learns should be understandable.

- Cases:
- Rule 1: If a + b > c, then Joe Smith to play.
- Rule 2: If the weather is good, then Joe Smith to play.
- Obviously, Rule 1 is a poor understanding of the rules, and rule 2 is a strong the rules of understanding.

Improve the accuracy of prediction

Reducing the complexity of the search

Enhance understandability represented

Lecture 01: Introduction to Machine Learning

AI history

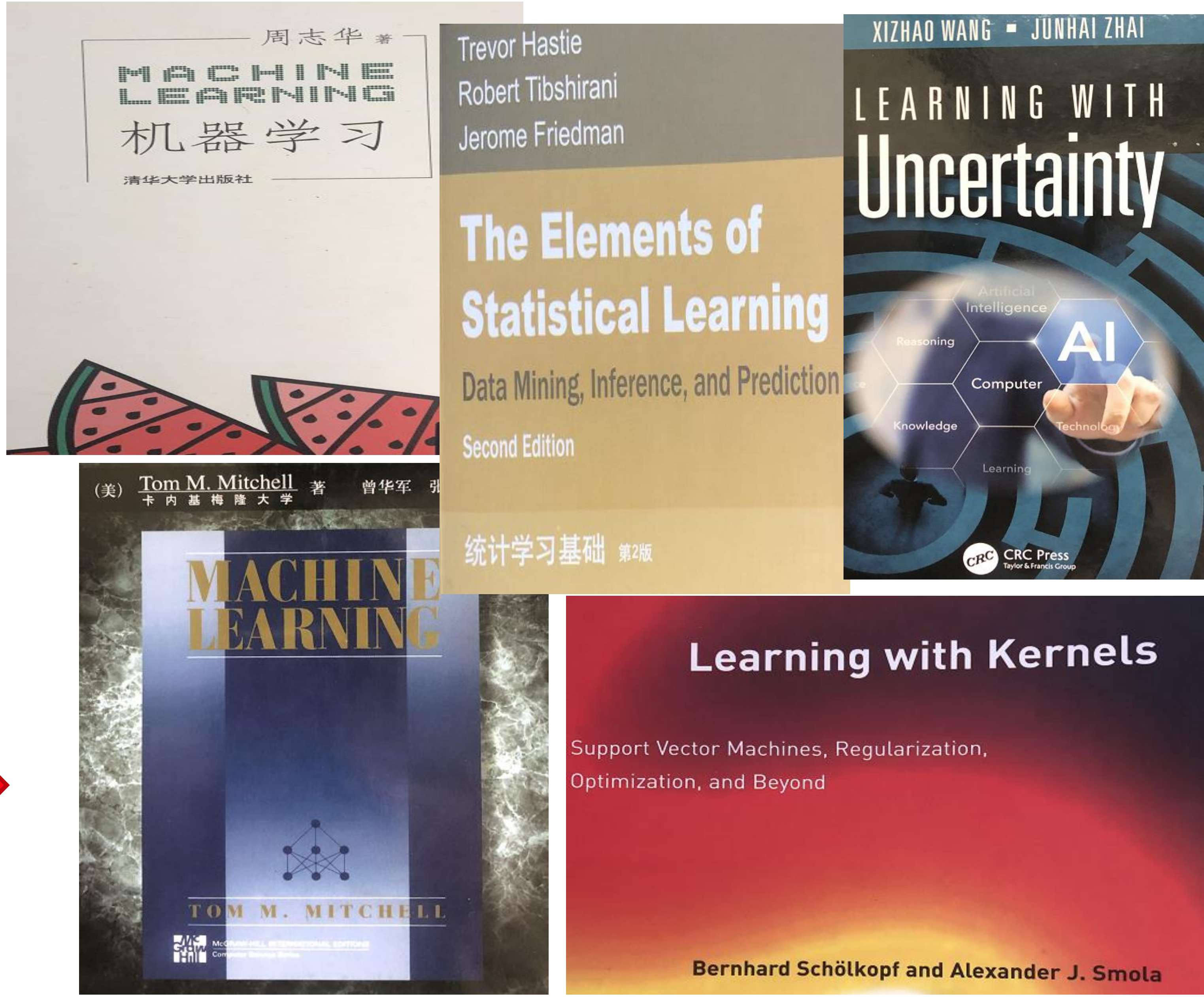
What is ML?

AI and ML

ML categories

ML aims

References



Lecture 01: Introduction to Machine Learning

AI history

What is ML?

AI and ML

ML categories

ML aims

References

No.1 《机器学习简明教程》作者：Vishal Maini

Machine Learning for Humans

链接：https://medium.mybridge.co/machine-learning-top-10-articles-for-the-past-month-v-sep-2017-c68f4b0b5e72

这篇文章用平实的语言阐释了什么是机器学习、机器学习的主要内容等，使用了少量的数学公式、代码和实例，内容涉及监督学习、无监督学习、神经网络和深度学习、强化学习等，同时列举了一些优秀的资源，附目录如下：

Lecture 02

Introduction to learning with weak supervision

Xizhao WANG

Big Data Institute
College of Computer Science
Shenzhen University

March 2019

Lecture 02: Introduction to learning with weak supervision

Big Data

Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning with Weak Supervision

Big Data

Essential Goal

Turn data into information and knowledge, so as to support sound decision making

Key Techniques

Cloud Computing → Managing Data

Crowdsourcing → Collecting Data

Machine Learning → Analyzing Data

Lecture 02: Introduction to learning with weak supervision

Big Data

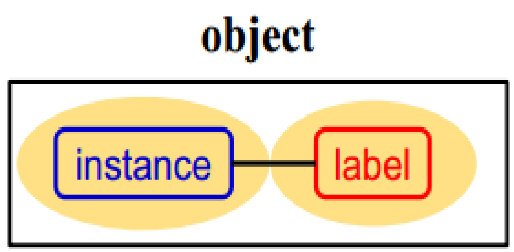
Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning with Weak Supervision

Traditional Supervised Learning

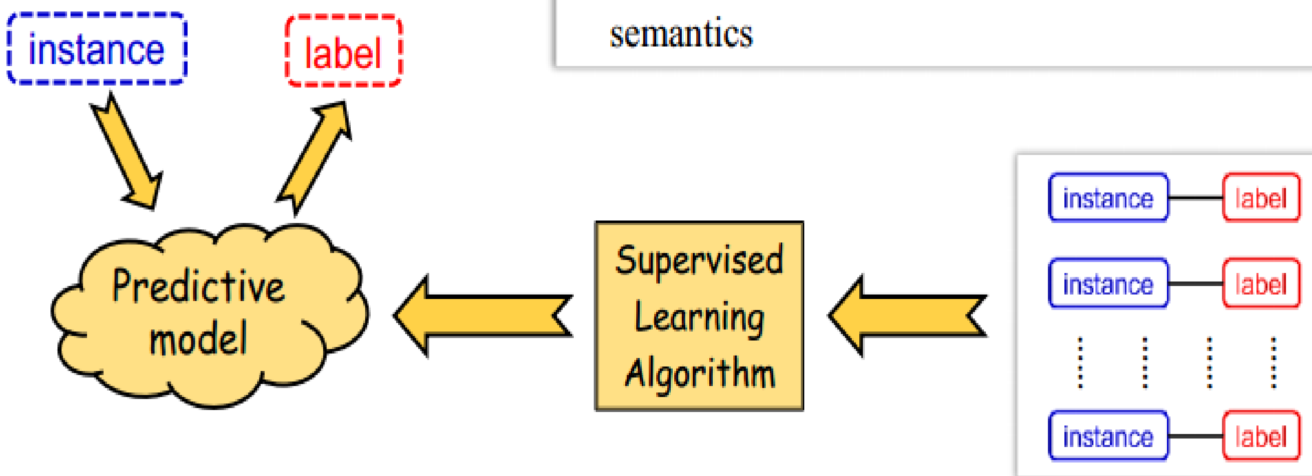


Input Space

represented by a single instance (feature vector) characterizing its properties

Output Space

associated with a single label characterizing its semantics



Lecture 02: Introduction to learning with weak supervision

Big Data

Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning with Weak Supervision

Basic Assumption: Strong Supervision

label ⟷ supervision information

Key factor for successful learning

(encoding *semantics* and *regularities* for the learning problem)

Strong supervision assumption

- Sufficient labeling
  abundant labeled training data are available
- Explicit labeling
  object labeling is unique and unambiguous

Big Data

Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning with Weak Supervision

But, Supervision Is Usually Weak

Difficult to have!

Constrained by:

- Limited resources
- Physical environment
- Problem properties
- ......

Strong supervision (sufficient & explicit) → Strong generalization ability

In practice, we usually have to learn with weak supervision

Lecture 02: Introduction to learning with weak supervision

Big Data

Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning with Weak Supervision

Learning with Weak Supervision

✓ Insufficient labeling

Labeled Data + Unlabeled Data

✓ Non-Unique labeling

Multi-Label Data (labeling with multiple valid labels)

✓ Ambiguous labeling

Partial-Label Data (labeling with multiple candidate labels)

Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

MLL

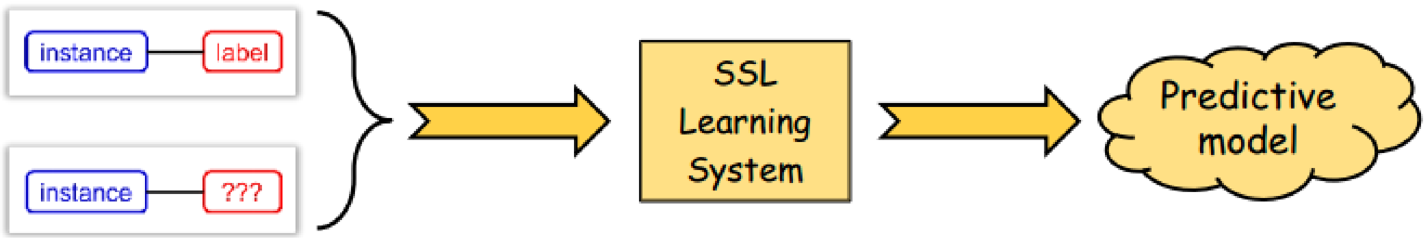
Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Semi-Supervised Learning (SSL)



Major paradigm in exploiting unlabeled data to improve generalization performance, without human interventions

- Generative methods [Miller & Uyar, NIPS'97] [Nigam et al., MLJ00]
- S3VMs [Joachims, ICML'99] [Chapelle & Zien, AIStats'05] [Grandvalet & Bengio, NIPS'05]
- Graph-based methods [Zhu et al., ICML'03] [Zhou et al, NIPS'04] [Belkin et al., JMLR06]
- Disagreement-based methods [Blum & Mitchell, COLT'98] [Zhou & Li, KAIS10]

SSL

Multi-Label Objects

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Multi-Label Objects


Russia to Ensure High Level of Security at 2018 FIFA World Cup

SPORT 21:53 17.04.2017 [illegible]

Topic: Russia hosts FIFA World Cup 2018 ( ) 1 206 3 5

Russia will provide a proper level of security on the streets and in public places during the 2018 FIFA World Cup in the country, Vladimir Markin, the head of the Russian Football Union's security committee, said Monday.

MOSCOW (Sputnik) — Earlier in the day, Russian President Vladimir Putin, amid preparations for the international football event, [illegible] that will toughen administrative responsibility for football fans and introduce penalties for gross violations of rules of conduct during sporting events.

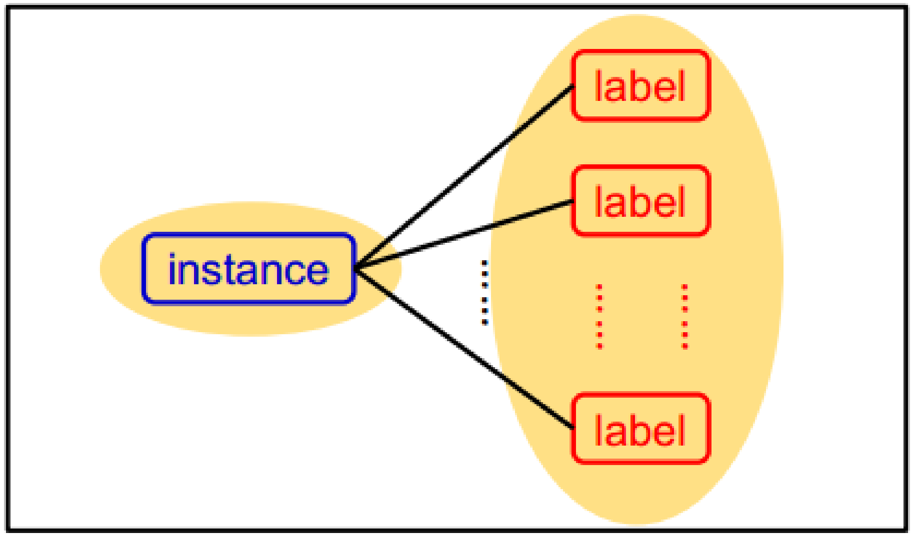
> *"The security on the streets and in other public places of the cities where the matches will be held is going to be provided on a proper level. Not only the law enforcement, but also private security services, volunteers and responsible citizens will make every effort for it," Markin told R-Sport.*

- Sports
- Europe
- Economics
- Travel
- Government
- ……

Multiple labels

Multi-Label Learning (MLL)

object

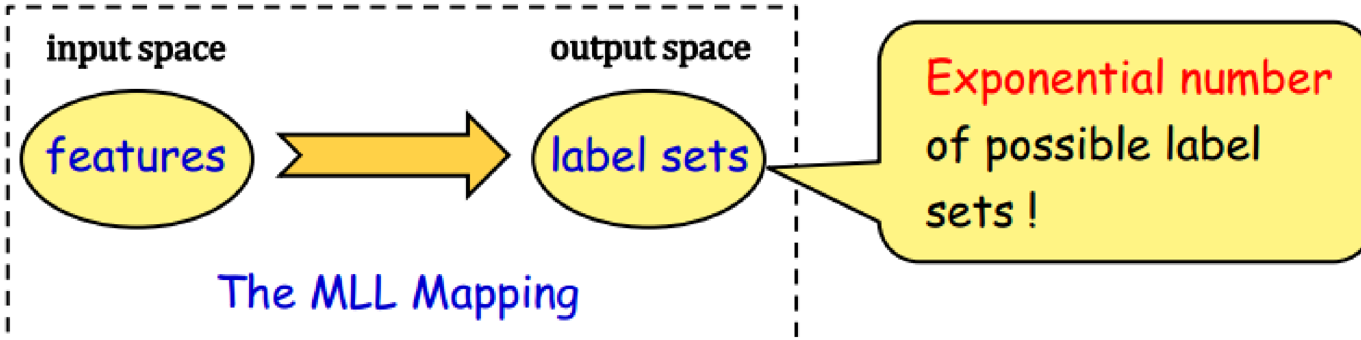


Multi-Label Learning (MLL)

Lecture 02: Introduction to learning with weak supervision

- SSL
- Multi-Label Objects
- MLL
- Major Challenge of MLL
- Partial Label
- PLL
- Other Scenarios Widely Exist

Major Challenge of MLL



q=5 ➔ 32 label sets

q=10 ➔ ~1k label sets

q=20 ➔ ~1M label sets

……

Supervision Info.

- Individually strong
- But, globally weak !

SSL

Multi-Label Objects

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Partial Label

Appreciator A ----------> Picasso style ×

Appreciator B ----------> Monet style ×

Appreciator C ----------> van Gogh style √

Widely exist in real-world applications

- Computer vision [Cour et al., JMLR11] [Tang & Zhang, AAAI'17]
- Image classification [Zeng et al., CVPR'13] [Chen et al., CVPR'13]
- Learning from crowds [Raykar et al., JMLR10] [Yu & Zhang, MLJ17]
- Ecoinformatics [Liu & Dietterich, NIPS'12] [Zhang & Yu, IJCAI'15]
- ……

SSL

Multi-Label Objects

MLL

Major Challenge of MLL

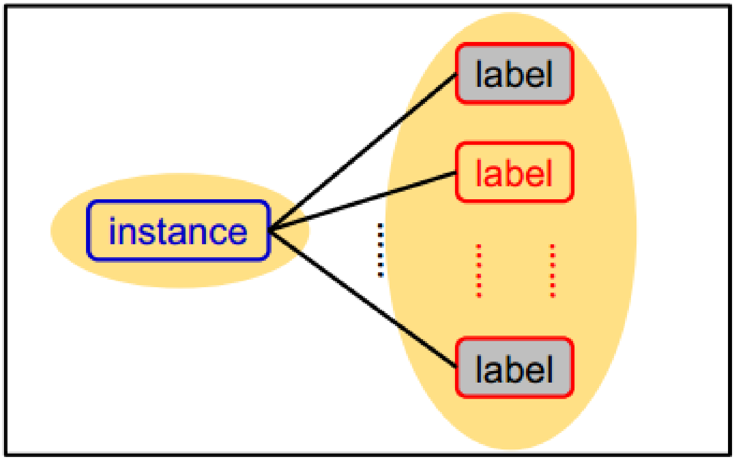
Partial Label

PLL

Other Scenarios Widely Exist

Partial-Label Learning (PLL)

object



- Each object is associated with multiple candidate labels
- Only one of the candidate label is the unknown ground-truth label

Partial-Label Learning (PLL)

Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

MLL

Major Challenge of MLL

Partial Label

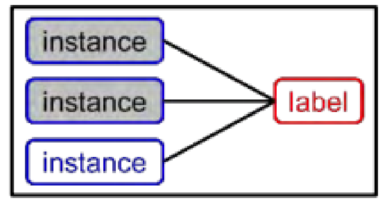
PLL

Other Scenarios Widely Exist

Other Scenarios Widely Exist

multi-instance learning

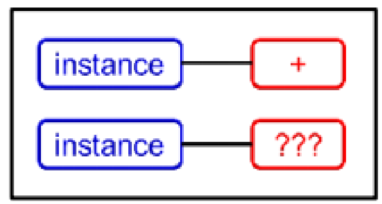
[Dietterich et al., AIJ97] [Foulds & Frank, KER10] [Amores, AIJ13]



ambiguous labeling

PU learning

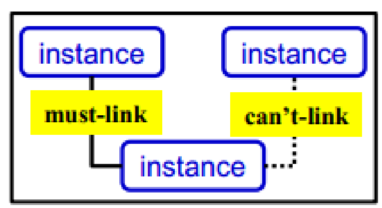
[Liu et al., ICML'02] [Liu et al., ICDM'03] [Li et al., ACL'10]



insufficient labeling

learning with constraints

[Wagstaff et al., ICML'01] [Basu et al., CRCBook08]



non-unique labeling

....

SSL

Multi-Label Objects

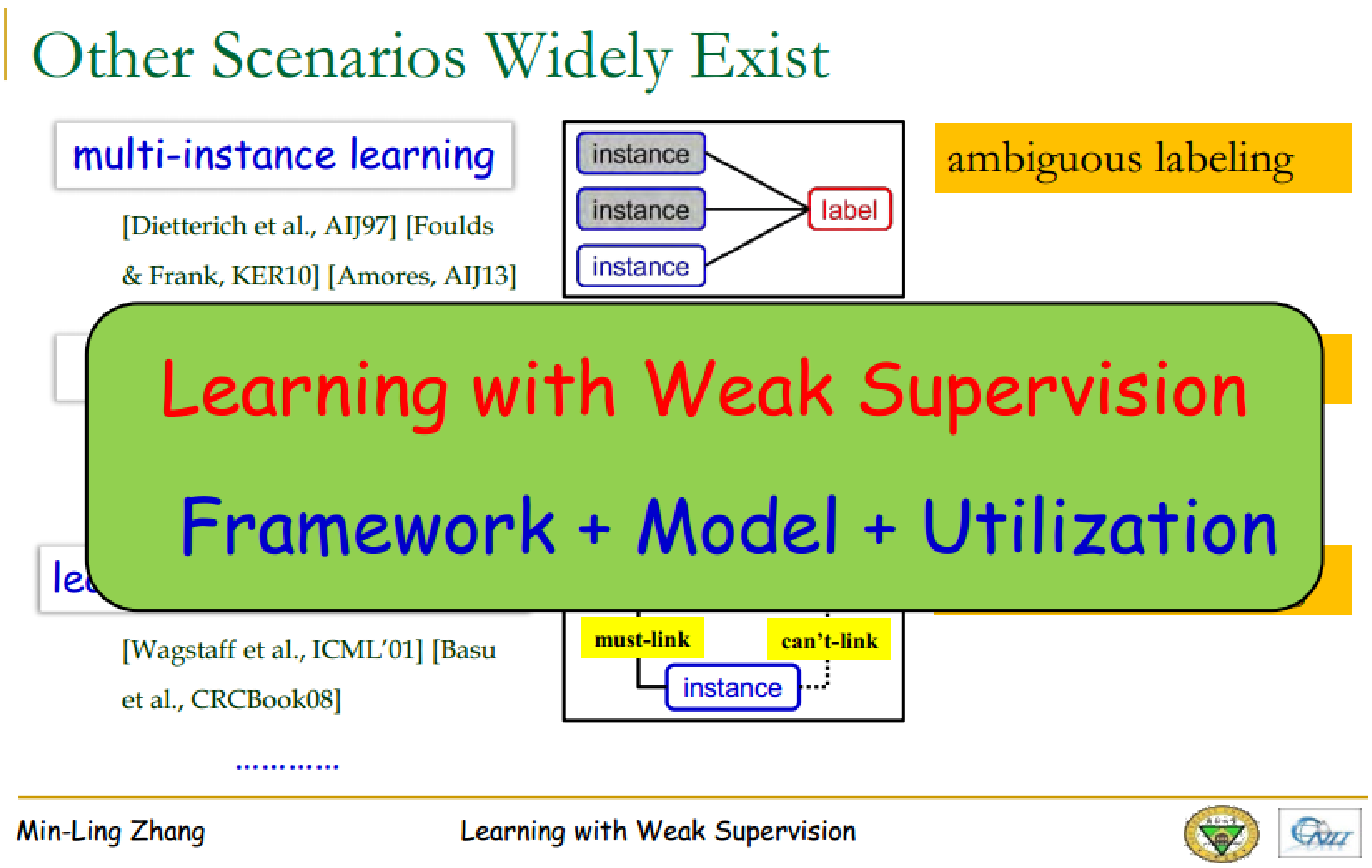
MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist



Lecture 03

Learning from mislabeled training data through ambiguous learning

Xizhao WANG

Big Data Institute
College of Computer Science
Shenzhen University

March 2019

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

Most existing studies assume that the training data are **perfect**, **sufficient** and **cost free**. However, in the real applications, these assumptions might be **false**. For example:

- The number of training examples might be insufficient;
- Obtaining the labels of training examples is expensive, and
- Only positive and unlabeled examples are available.

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

The performance of classification algorithms can be affected by **noisy training samples**.

Two types of noisy training samples:

- **Noisy features**: means that the values of the features of some training examples are incorrect.
- **Noisy labels**: means that some of the training examples are mislabeled.

Introduction

The proposed Approach

Experimental result

Summary

The **mislabeled data** can dramatically degrade the performance of the classifier. How can deal with mislabeled examples?

- **Algorithm** level approach: modifies the existing algorithm to make it robust against mislabeled data during model training, i.e., KNN and Edited Nearest Neighbor (ENN).
- **Data** level approach: directly handles the training samples, i.e., Majority Filtering (MF) and Consensus Filtering (CF).

In order to minimize the downside of Mislabeled training instances:

- **Noise tolerance**: tries to control the negative effect of noisy instances without removing them, and
- **Noise filtering**: tries to improve the quality of training data by **identifying** and **eliminating** the noisy instances prior to applying the learning algorithm.

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

The noise filtering mainly include:

- **Distance-based algorithms** usually adopt the idea of **k-nearest neighbors** and believe that the **nearby samples tend to have the same label**, and
- **Ensemble learning based algorithms** employ multiple classifiers to detect the noises.

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

Algorithm: Majority Filtering (MF)

Input: E (training set)

Parameter: n (number of subsets), y (number of learning algorithms),

$A_1, A_2, \ldots, A_y$(y kinds of learning algorithms)

Output: A (detected noisy subset of E)

(1) form n disjoint almost equally sized subset of E_i, where $\bigcup_i E_i = E$

(2) $A \leftarrow \emptyset$

(3) **for** i=1, . . . , n **do**

(4) form $E_t \leftarrow E \backslash E_i$

(5) **for** j=1,. . . y **do**

(6) induce H_j based on examples in E_t and A_j

(7) **end for**

(8) **for** every $e \in E_i$ **do**

(9) $ErrorCounter \leftarrow 0$

(10) **for** j=1,. . . ,y **do**

(11) **if** $H_j$incorrectly classifies e

(12) **then** $ErrorCounter \leftarrow ErrorCounter + 1$

(13) **end for**

(14) **if** $ErrorCounter > \frac{y}{2}$, **then** $A \leftarrow A \cup \{e\}$

(15) **end for**

(16) **end for**

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

Training sample	Given label	SVM	KNN	NB	$P(c_1)$	$P(c_2)$	Noise?
1	1	1	1	1	1	0	No
2	2	1	2	2	0.34	0.66	No
3	2	2	2	2	0	1	No
4	1	1	2	2	0.34	0.66	Yes
5	1	2	1	1	0.66	0.34	No
6	2	1	2	2	0.34	0.66	No
7	1 (MisL)	2	2	2	0	1	Yes
8	2	2	2	2	0	1	No
9	2 (MisL)	1	1	1	1	0	Yes
10	1	1	2	1	0.66	0.34	No

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

This work proposed an approach which **learns from mislabeled training data through ambiguous learning** (LeMAL).

In ambiguous learning, each training example is assigned with **a set of candidate labels**, among which only one is valid.

Experimental result

Summary

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Formally, let $X = R^d$ be the d-dimensional input space and $Y = y_1, y_2, ..., y_q$ be the output space including q classes. An ambiguous label training set is defined as follows:

$$D = \{(x_i, S_i, P_i)|1 < i \leq m\} \quad (1)$$

where $x_i \in X$ is a d-dimensional feature vector; $S_i \in Y$ is the set of candidate labels; P_i is the probabilities of each candidate labels.

Experimental result

Summary

Introduction

The proposed Approach

Experimental result

Summary

In k-NN classification, the label λ_{x0} assigned to a query sample x_0 is given by the label that is most frequent among the k-nearest neighbors of x_0, which can be found by using the distance function.

$$\lambda_{x0} = \underset{\lambda \in Y}{\operatorname{argmax}} \sum_{i=i}^{k} P_i \coprod(\lambda \in S_i) \qquad (2)$$

Where x_i is the i-th nearest neighbor; λ_{xi} is the label of x_i, and $\coprod()$ is the standard true, false $\rightarrow$ 0, 1 mapping.

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

Training sample	Given label	SVM	KNN	NB	$P(c_1)$	$P(c_2)$
1	1	1	1	1	1	0
2	2	1	2	2	0.34	0.66
3	2	2	2	2	0	1
4	1	1	2	2	0.34	0.66
5	1	2	1	1	0.66	0.34
6	2	1	2	2	0.34	0.66
7	1	2	2	2	0	1
8	2	2	2	2	0	1
9	2	1	1	1	1	0
10	1	1	2	1	0.66	0.34

Assume for a given test sample x_t, training samples 1, 4 and 8 are 3 nearest neighbors. Therefore:

$V(c_1) = 1 + 0.34 + 0 = 1.34$

$V(c_2) = 0 + 0.66 + 1 = 1.64$ (predicts x_t as class 2)

Introduction

The proposed Approach

Experimental result

Summary

Classification accuracy comparison on Wilttrain data set

Method	Dataset5							
	Noise ratio							
	10	15	20	25	30	35	40	Ave
A_CLEAN	0.920	0.926	0.906	0.916	0.913	0.916	0.913	0.915
A_NOISE	0.910	0.896	0.893	0.840	0.873	0.730	0.726	0.838
AENN	0.903	0.866	**0.896**	**0.880**	0.886	0.786	0.706	0.846(2)
MF	0.890	0.856	0.893	0.826	0.893	0.786	**0.763**	0.843(1)
CF	0.890	0.876	0.883	0.850	0.900	0.773	**0.763**	0.847(1)
MFMF	0.893	**0.883**	**0.896**	0.846	0.890	0.793	**0.763**	0.852(3)
CFMF	0.893	**0.883**	0.876	0.873	**0.903**	0.776	0.763	0.852(2)
LeMAL1	**0.903**	0.853	0.820	0.696	0.796	0.630	0.596	0.756(1)
LeMAL2	0.890	0.866	0.890	0.876	**0.903**	**0.810**	**0.763**	0.856(3)

where
MFMF and CFMF are multiple-voting based filtering methods.
ENN is a kNN-based noise filtering algorithm.

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Inaccurate supervision are less reliable in high-dimensional feature space because the identification of neighborhoods is usually less reliable when data are sparse.

Summary

Lecture 03:
Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

- Different level of noise, i.e., 10%, 15%,...,40%, were injected to the labels of training samples.
- Soft and probabilistic strategy is used to label training samples.
- The KNN classification algorithm is used to predict the labels of the test samples.

The End of Lecture

Thank you for your attention！