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
1. AI brief history
2. What is ML?
3. AI and ML
4. ML categories
5. ML aims
6. References
1956 Summer School
J. McCarthy, M. Minsky, N. Locheher (IBM), C. E. Shannon
A. Samual (IBM), H. A. Simon (CMU), A. Newell (CMU),
T. More (Prinstone), R. Solomonoff (MIT), O. Selfridge (MIT)

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

Goal: Design a computer with real intelligence

Result: Laid a new science - artificial intelligence
Lecture 01: Introduction to Machine Learning

AI history

What is ML?

AI and ML

ML categories

ML aims

References

AI history:

1956 Dartmouth Conference

1970-1980 Trough: KB + Inference machine

1980 Expert system

1986 Invention and application of BP neural network algorithm

1995 Statistical learning

1995 Representative technology: SVM

2006 Deep learning

2006 pioneering deep learning neural network, achieved breakthrough progress

2013 Deep learning algorithm is successful in speech recognition and visual recognition rate, 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

2016 New era

2016 Google AlphaGo wins World Cup champion Li Shishi, opening a new era
Using artificial methods and techniques to imitate, extend and expand human intelligence to realize machine intelligence.

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


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。

— Microsoft Research

Xiang and Liu
What is Machine Learning?

AI history

What is ML?

AI and ML

ML categories

ML aims

References

http://blog.scienccenet.cn/blog-826653-1029786.html
What is Machine Learning?

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

http://blog.scienecenet.cn/blog-826653-1029786.html
What is Machine Learning?

Machine Learning is a technique 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?

Traditional 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: $\text{AI} = \text{ML} + \text{(Big) Data}$

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

------ Jiarong Hong
Lecture 01: Introduction to Machine Learning

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

Brain Cognition
ML & PR
NL processing & understanding
Knowledge Engineering
Robotics and Smart systems
Lecture 01: Introduction to Machine Learning

AI history

What is ML?

AI and ML

ML categories

ML aims

References

Machine Learning

- Supervised learning
  - Classification
  - Regression
  - Prediction

- Unsupervised learning
  - Clustering
  - Association rules
  - Dimension reduction

- Reinforcement learning
Supervised Learning: Prediction

A general framework of supervised learning

- A training set \((\vec{x}_i, y_i)\)
  - \(i = 1, 2, \ldots, n\)
  - \(n\) observations (known)

- A function \(f(\vec{x})\)

- A new case \((\vec{x}_{n+1}, y_{n+1})\)
  - known
  - unknown

- \(y_{n+1} \leftarrow f(\vec{x}_{n+1})\)

\(f(x)\) 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

[Diagram showing the interaction between an agent and an environment, with state transitions, rewards, and actions labeled.]
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.
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.
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 understanding.
Lecture 01: Introduction to Machine Learning

- 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

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

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

Key Techniques
- Cloud Computing
- Crowdsourcing
- Machine Learning

Traditional Supervised Learning

Basic Assumption: Strong Supervision

Supervision Is Usually Weak

Learning 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

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

Predictive model

Supervised Learning Algorithm

Min-Ling Zhang

Learning 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**

---

### Basic Assumption: Strong Supervision

- **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
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**

---

**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

---

*Min-Ling Zhang*
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)

Min-Ling Zhang
**Semi-Supervised Learning (SSL)**

**Major Challenge of MLL**

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, KAI10]
Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

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

Sports
Europe
Economics
Travel
Government

Min-Ling Zhang

Learning with Weak Supervision
SSL

Multi-Label Learning (MLL)

Multi-Label Objects

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Multi-Label Learning (MLL)

Min-Ling Zhang

Learning with Weak Supervision
Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

Major Challenge of MLL

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

\( q = 5 \rightarrow 32 \text{ label sets} \)

\( q = 10 \rightarrow \sim 1k \text{ label sets} \)

\( q = 20 \rightarrow \sim 1M \text{ label sets} \)

\( \ldots \)

Supervision Info.

- Individually strong
- But, globally weak!

Exponential number of possible label sets!
Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

Partial Label

Appreciator A

Appreciator B

Appreciator C

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]

☐ ......
Lecture 02: Introduction to learning with weak supervision

SSL

Multi-Label Objects

Partial-Label Learning (PLL)

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

- 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

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Other Scenarios Widely Exist

Multi-instance learning

PU learning

learning with constraints

ambiguous labeling

insufficient labeling

non-unique labeling

Min-Ling Zhang

Learning with Weak Supervision

[Wagstaff et al., ICML’01] [Basu et al., CRCBook08]
SSL

Multi-Label Objects

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Learning with Weak Supervision

Framework + Model + Utilization

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

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

SSL

Multi-Label Objects

MLL

Major Challenge of MLL

Partial Label

PLL

Other Scenarios Widely Exist

Learning with Weak Supervision

Framework + Model + Utilization

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

[Wagstaff et al., ICML'01] [Basu et al., CRCBook08]
Lecture 03
Learning from mislabeled training data through ambiguous learning

Xizhao WANG
Big Data Institute
College of Computer Science
Shenzhen University

March 2019
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

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.
The proposed Approach

Experimental result

Summary

**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).
Lecture 03: Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

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.
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 \( i=1, \ldots, n \) do
4. form \( E_t \leftarrow E \setminus E_i \)
5. for \( j=1,\ldots,y \) 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. \( \text{ErrorCounter} \leftarrow 0 \)
10. for \( j=1,\ldots,y \) do
11. if \( H_j \) incorrectly classifies \( e \)
12. then \( \text{ErrorCounter} \leftarrow \text{ErrorCounter} + 1 \)
13. end for
14. if \( \text{ErrorCounter} > \frac{n}{2} \), then \( A \leftarrow A \cup \{e\} \)
15. end for
16. end for
**Introduction**

The proposed Approach

**Experimental result**

**Summary**

### Experimental result

<table>
<thead>
<tr>
<th>Training sample</th>
<th>Given label</th>
<th>SVM</th>
<th>KNN</th>
<th>NB</th>
<th>$P(c_1)$</th>
<th>$P(c_2)$</th>
<th>Noise?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.66</td>
<td>0.34</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>1 (MisL)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>2 (MisL)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.66</td>
<td>0.34</td>
<td>No</td>
</tr>
</tbody>
</table>
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.
Introduction

The proposed Approach

Experimental result

Summary

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

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

(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.
In k-NN classification, the label $\lambda_{x_0}$ 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.

$$\lambda_{x_0} = \arg\max_{\lambda \in Y} \sum_{i=1}^{k} P_i \prod (\lambda \in S_i)$$

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

The proposed Approach

Experimental result

Summary

Lecture 03:
Learning from mislabeled data through ambiguous learning

<table>
<thead>
<tr>
<th>Training sample</th>
<th>Given label</th>
<th>SVM</th>
<th>KNN</th>
<th>NB</th>
<th>$P(c_1)$</th>
<th>$P(c_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.66</td>
<td>0.34</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.34</td>
<td>0.66</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.66</td>
<td>0.34</td>
</tr>
</tbody>
</table>

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)
Lecture 03: Learning from mislabeled data through ambiguous learning

Introduction

The proposed Approach

Experimental result

Summary

where

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

- 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.

Experimental result

Summary
The End of Lecture

Thank you for your attention!