Introduction of Structured Learning

Hung-yi Lee
Structured Learning

• We need a more powerful function $f$
  • Input and output are both objects with structures
  • Object: sequence, list, tree, bounding box ...

$f : X \rightarrow Y$

$X$ is the space of one kind of object
$Y$ is the space of another kind of object
Example Application

• **Speech recognition**
  • $X$: Speech signal (sequence) $\rightarrow Y$: text (sequence)

• **Translation**
  • $X$: Mandarin sentence (sequence) $\rightarrow Y$: English sentence (sequence)

• **Syntactic Paring**
  • $X$: sentence $\rightarrow Y$: parsing tree (tree structure)

• **Object Detection**
  • $X$: Image $\rightarrow Y$: bounding box

• **Summarization**
  • $X$: long document $\rightarrow Y$: summary (short paragraph)

• **Retrieval**
  • $X$: keyword $\rightarrow Y$: search result (a list of webpage)
Unified Framework

Step 1: Training

• Find a function $F$

$$F : X \times Y \rightarrow \mathbb{R}$$

• $F(x,y)$: evaluate how compatible the objects $x$ and $y$ is

Step 2: Inference (Testing)

• Given an object $x$

$$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

$f : X \rightarrow Y \quad \Rightarrow \quad f(x) = \tilde{y} = \arg \max_{y \in Y} F(x, y)$

Unified Framework – Object Detection

• Task description
  • Using a bounding box to highlight the position of a certain object in an image
  • E.g. A detector of Haruhi

\[
X : \text{Image} \quad \rightarrow \quad Y : \text{Bounding Box}
\]

Haruhi
(the girl with yellow ribbon)
Unified Framework – Object Detection

Step 1: Training

- Find a function $F$
  \[ F : X \times Y \rightarrow R \]
- $F(x,y)$: evaluate how compatible the objects $x$ and $y$ is

$x$: Image

$y$: Bounding Box

$F(x,y) \rightarrow F($

the correctness of taking range of $y$ in $x$ as “Haruhi”
Unified Framework – Object Detection

Step 1: Training
- Find a function $F$
  $$F : X \times Y \rightarrow R$$
- $F(x,y)$: evaluate how compatible the objects $x$ and $y$ is

Step 2: Inference (Testing)
- Given an object $x$
  $$\tilde{y} = \arg \max_{y \in Y} F(x, y)$$

input $x =$

Enumerate all possible bounding box $y$
Unified Framework - Summarization

• Task description
  • Given a long document
  • Select a set of sentences from the document, and cascade the sentences to form a short paragraph

\[ X = \{s_1, s_2, s_3, \ldots s_i, \ldots\} \]

\[ Y = \{s_1, s_3, s_5\} \]

\( s_i \): the \( i \)th sentence
Unified Framework
- Summarization

Step 1: Training

Step 2: Inference

F(x, y)

x
y

d_1

F(x, y)

x
y

d'_1
{s_1, s_3, s_5}

d'_2
{s_2, s_4, s_6}

d'_3
{s_3, s_6, s_9}
Unified Framework - Retrieval

• Task description
  • User input a keyword Q
  • System returns a *list* of web pages

```
X
“Obama” (keyword)
```

```
Y
```

```
d10011
```
```
d98776
```
```
A list of web pages (Search Result)
```
Unified Framework - Retrieval

Step 1: Training

Step 2: Inference

\[ F(x,y) \]

- \( x = \) Obama, \( y = d_666 \ldots \)
- \( x = \) Haruhi, \( y = d_203 d_330 \ldots \)
- \( x = \) Haruhi, \( y = d_{103} d_{304} \ldots \)
- \( x = \) Haruhi, \( y = d_{103} d_{305} \ldots \)
**Unified Framework**

**Step 1: Training**
- Find a function $F$
  
  $F : X \times Y \rightarrow \mathbb{R}$

  - $F(x,y)$: evaluate how compatible the objects $x$ and $y$ is

**Step 2: Inference**
- Given an object $x$

  $\tilde{y} = \arg \max_{y \in Y} P(y | x)$

  $\quad = \arg \max_{y \in Y} \frac{P(x, y)}{P(x)}$

  $\quad = \arg \max_{y \in Y} P(x, y)$

**Statistics**

**Step 1: Training**
- Estimate the probability $P(x,y)$
  
  $P : X \times Y \rightarrow [0,1]$
**Statistics**

**Unified Framework**

\[ F(x, y) = P(x, y) \]

**Step 1: Training**

- Estimate the probability \( P(x, y) \)

\[ P : X \times Y \rightarrow [0, 1] \]

**Step 2: Inference**

- Given an object \( x \)

\[ \tilde{y} = \arg \max_{y \in Y} P(y \mid x) \]

\[ = \arg \max_{y \in Y} \frac{P(x, y)}{P(x)} \]

\[ = \arg \max_{y \in Y} P(x, y) \]

**Drawback for probability**

- Probability cannot explain everything
- 0-1 constraint is not necessary

**Strength for probability**

- Meaningful
Link to DNN?

**Step 1: Training**

\[ F: X \times Y \rightarrow R \]

\[ F(x, y) = -\|y - N(x)\|^2 \]

- **N(x)**
- **DNN**
- **x**
- **y**

**Step 2: Inference**

\[ \tilde{y} = \arg \max_{y \in Y} F(x, y) \]

In handwriting digit classification, there are only 10 possible \( y \).

\[ y = [1 \ 0 \ 0 \ 0 \ldots] \]
\[ y = [0 \ 1 \ 0 \ 0 \ldots] \]
\[ y = [0 \ 0 \ 1 \ 0 \ldots] \]
\[ \vdots \]

Find max

The same as what we have learned.
Unified Framework

- Solve any tasks by two steps
  - Easier than putting an elephant into a refrigerator

Really? No, we have to answer three problems.
Problem 1

• **Evaluation**: What does \( F(x,y) \) look like?
  
  • How \( F(x,y) \) compute the “compatibility” of objects \( x \) and \( y \)

  \[
  \text{Object Detection:} \quad F(x= \quad , \quad y= \quad ) \\
  \text{Summarization:} \quad F(x= \quad , \quad y= \quad ) \\
  \text{(a long document) (a short paragraph)}
  \]

  \[
  \text{Retrieval:} \quad F(x= \text{“Obama”} \quad , \quad y= \quad ) \\
  \text{(keyword) (Search Result)}
  \]
Problem 2

- **Inference**: How to solve the “arg max” problem

\[ y = \arg \max_{y \in Y} F(x, y) \]

The space \( Y \) can be extremely large!

**Object Detection**: \( Y=\)All possible bounding box (maybe tractable)

**Summarization**: \( Y=\)All combination of sentence set in a document ...

**Retrieval**: \( Y=\)All possible webpage ranking ....
Problem 3

- **Training**: Given training data, how to find $F(x,y)$

**Principle**

Training data: $\{(x^1, \hat{y}^1), (x^2, \hat{y}^2), \ldots, (x^r, \hat{y}^r), \ldots\}$

We should find $F(x,y)$ such that ......
Three Problems

Problem 1: Evaluation

• What does $F(x, y)$ look like?

Problem 2: Inference

• How to solve the “arg max” problem

$$y = \arg \max_{y \in Y} F(x, y)$$

Problem 3: Training

• Given training data, how to find $F(x, y)$
Have you heard the three problems elsewhere?

### Three Problems

**Problem 1: Evaluation**
- What does $F(x,y)$ look like?

**Problem 2: Inference**
- How to solve the "arg max" problem:
  $$y = \arg \max_y P(y|x)$$

**Problem 3: Training**
- Given training data, how to train?

---

### Hidden Markov Model

- **Three Basic Problems for HMMs**
  Given an observation sequence $O=(o_1,o_2,\ldots,o_T)$, and an HMM $\lambda=(A,B,\pi)$
  
  - **Problem 1:**
    How to efficiently compute $P(O|\lambda)$?  
    $\Rightarrow$ Evaluation problem
  
  - **Problem 2:**
    How to choose an optimal state sequence $q=(q_1,q_2,\ldots,q_T)$? 
    $\Rightarrow$ Decoding Problem
  
  - **Problem 3:**
    Given some observations $O$ for the HMM $\lambda$, how to adjust the model parameter $\lambda=(A,B,\pi)$ to maximize $P(O|\lambda)$? 
    $\Rightarrow$ Learning / Training Problem
Preview

• **Viterbi Algorithm**
  • 數位語音處理:
  • 演算法
  • 數位通信相關課程