# Relation Recognition

**Fang Li**

**Dept. of Computer Science & Engineering**

*Some slides are From Dan Jurafsky NLP in stanford university.

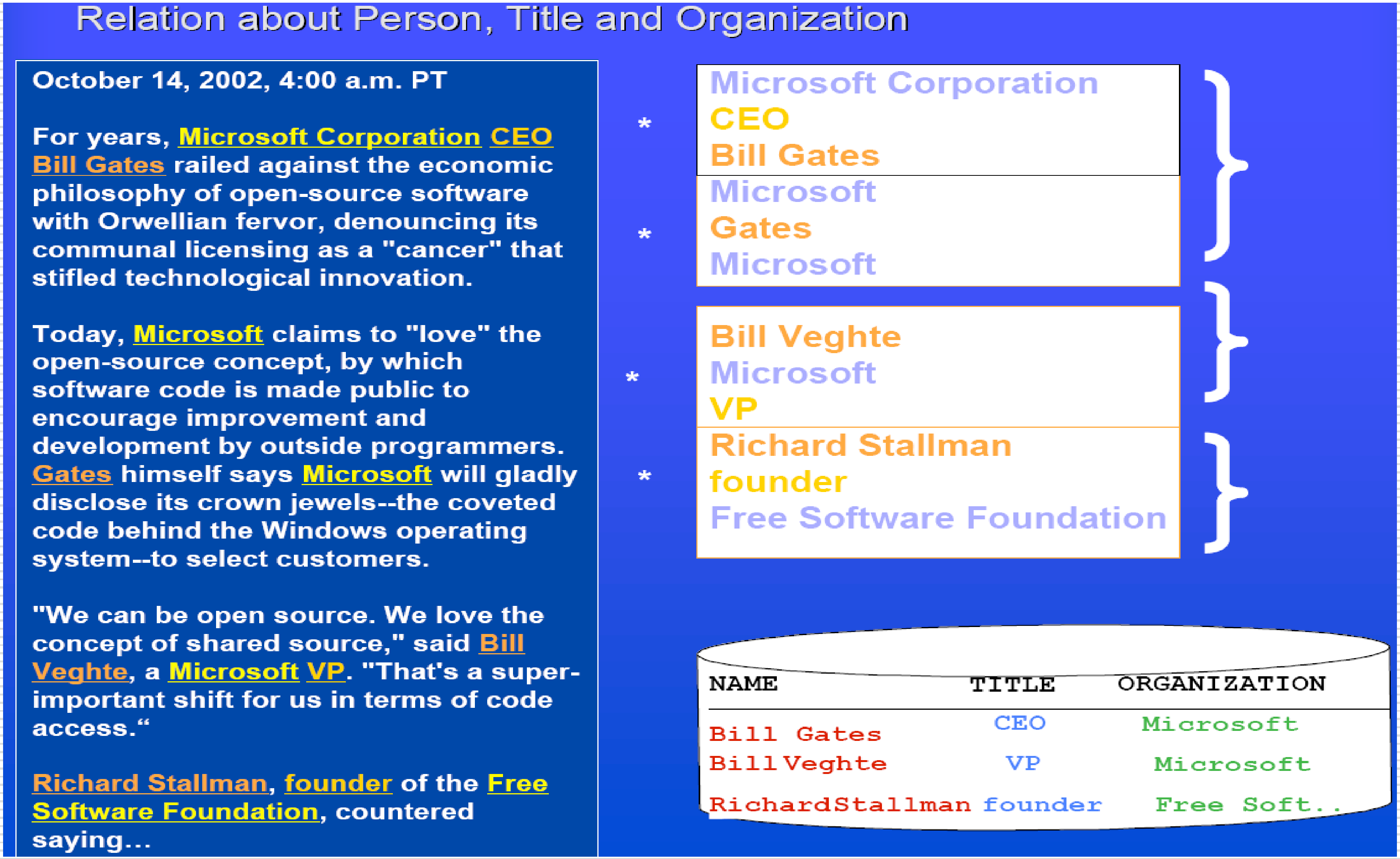
# Contents

- **Relation Representation**
- **Relation Identification**

1. **Knowledge Engineering Approach**
2. **Machine Learning Approach**

   **--Supervised learning**

# Relation Example:



# Relation with/without Time

- Relations may be timeless or bound to time intervals. For example, father(x,y) vs. boss(x,y)
- Time type is divided into temporal points and intervals

# Relation and Event

- Events: a particular kind of simple or complex relation among entities involving a change in relation state at the end of a time interval.
- Eg: Company-founding

**Company**: IBM
**Location**: New York
**Date**: June 16,1911
**Original-Name**: Coputer-Tabulating-Recording Co.

**Founding-year** (IBM, 1911)
**Founding-location** (IBM, New York)

# Relation examples

- Physical--Located PER---GPE
  He was in Tennessee.
- Part--Whole-Subsidiary ORG---ORG
  XYZ, the parent company of ABC.
- Person--Social--Family PER---PER
  John's wife Yoko!
- Org--AFF-Founder PER---ORG
  Steve Jobs, co-founder of Apple.

# Explicit and Implicit Relations

Explicit relations are expressed by certain surface linguistic forms:

- Prepositional phrase: *The CEO of Microsoft…*
- Prenominal modification: *the American envoy…*
- Possessive: *Microsoft's chief scientist…*
- Nominalizations: *Anan's visit to Baghdad..*
- Apposition: *Tony Blair, Britain's prime minister….*

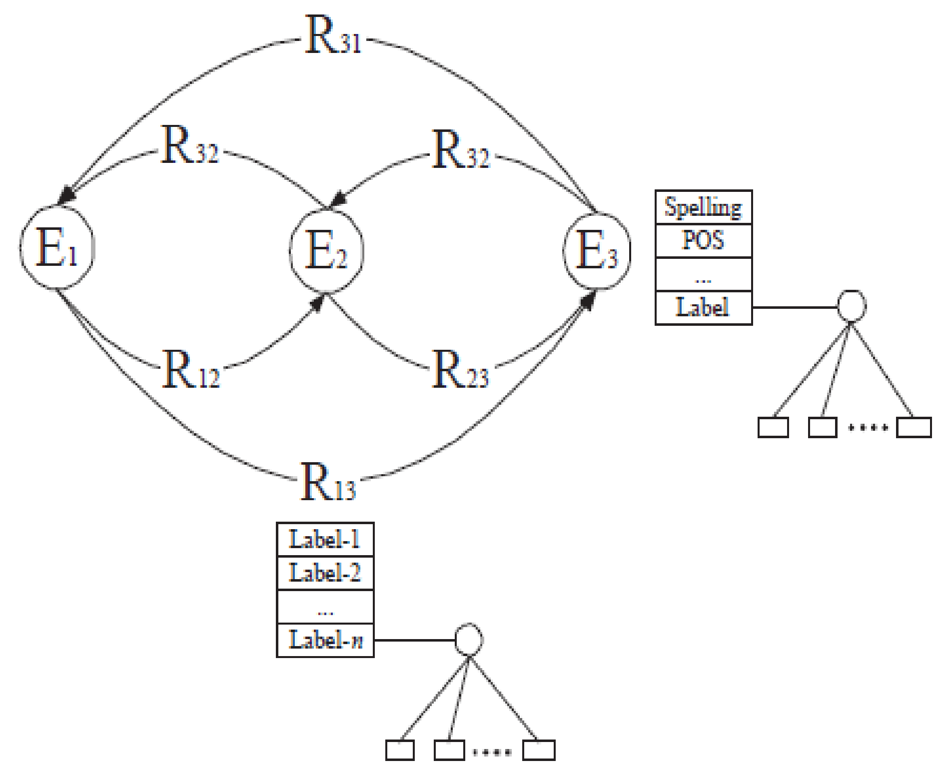
# Explicit and Implicit Relations (cont.)

- A relation is **implicit** in the text if the text provides convincing evidence that the relation actually holds.
- Example:

**Prime Minister** Tony Blair attempted to convince **the British** Parliament of the necessity of intervening in Iraq.

*Question: Is Tony Blair a British Prime Minister?*

# A conceptual view of entities and relations

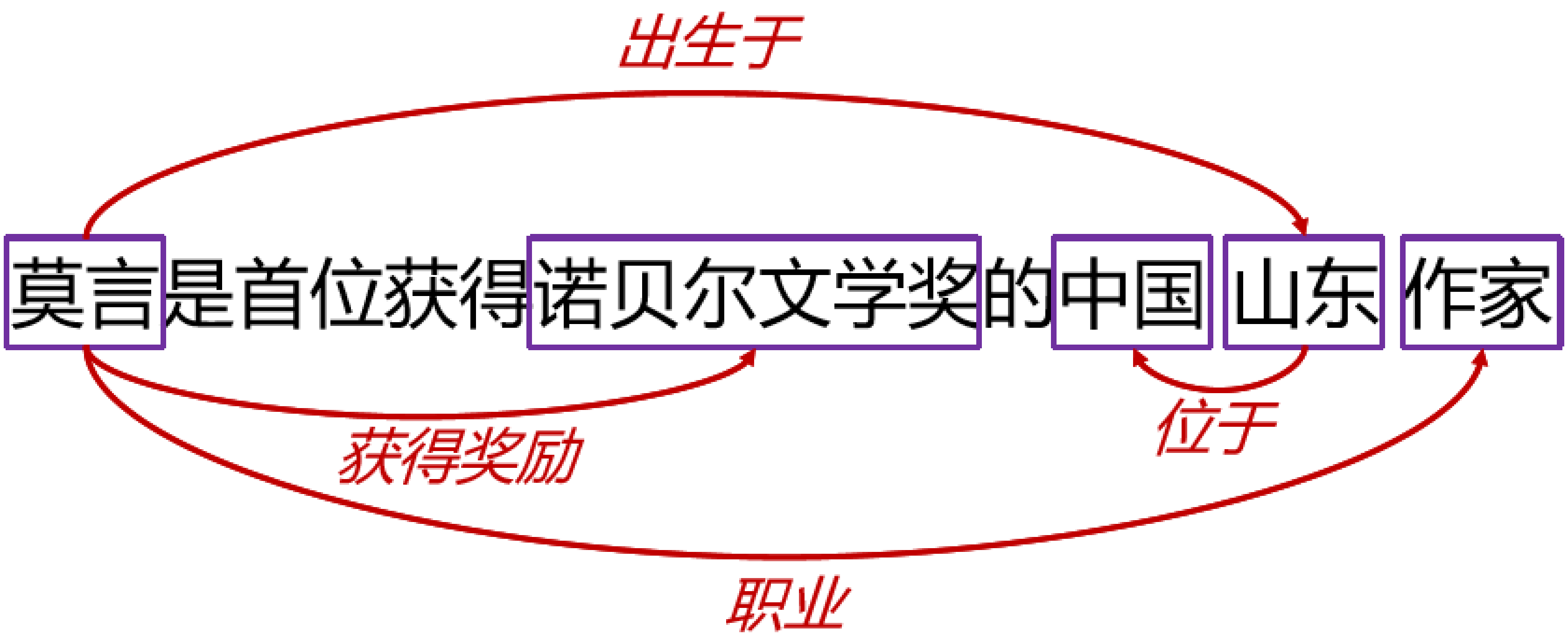



A conceptual view of entities and relations

- E's are the entities found in a sentence.
- R's are the relations between any two entities.
- mutually dependent

# Example:

**Mo Yan** is the first Chinese **writer** to win the **Nobel Prize in Literature,** who was born in **Shan Dong** province.



# Three Cases of Binary Relation Extractions $R(E_1,E_2)$

- For a given fixed pair of entities $(E_1,E_2)$, to find out the type of relationship (R) that exists between the pair.
- Given relationship R and an entity name $E_1$, to extract the entities ($E_2$ ) with which $E_1$ has relationship R.
- Given a fixed relationship type R, to find all the entity pairs $(E_1,E_2)$.

# Relation Extraction

- A harder task than entity extraction
- Relation extraction requires a skillful combination of local and nonlocal noisy clues from diverse syntactic and semantic structures in a sentence.

# Steps for relation extraction

E.g. **Haifa located 53 miles from Tel Aviv will host ICML in 2010. → located**

1) Named entity identification:

<LOC>Haifa</LOC> located 53 miles from <LOC>Tel Aviv</LOC> will host ICML in 2010.

2) POS tagging:

Haifa/NNP located/VBN 53/CD miles/NNS from/IN Tel/NNP Aviv/NNP will/MD host/VB ICML/NNP in/IN 2010/CD

# Steps of Relation Extraction (cont.)

## 3)Syntactic Parse Tree

Haifa located 53 miles from Tel Aviv will host ICML in 2010

```
(ROOT
  (S
    (NP
      (NP (NNP Haifa))
      (VP (VBN located)
        (PP
          (NP (CD 53) (NNS miles))
          (IN from)
          (NP (NNP Tel) (NNP Aviv)))))
    (VP (MD will)
      (VP (VB host)
        (NP
          (NP (NNP ICML))
          (PP (IN in)
            (NP (CD 2010))))))))
```

Parse tree of a sentence.

## 4) dependency Graph


Dependency parse of a sentence.

# Methods of Relation Recognition

1. Pattern-based methods:
   - hand made patterns
   - learning based on seeded pattern.
2. Supervised method
3. Semi-supervised method
4. Distant-supervised method

# Pattern-based methods

Some patterns extracted from the sentence or between the two entities:

| Hearst pattern | Example occurrences |
|---|---|
| X and other Y | ...temples, treasuries, and other important civic buildings. |
| X or other Y | Bruises, wounds, broken bones or other injuries... |
| Y such as X | The bow lute, such as the Bambara ndang... |
| Such Y as X | ...such authors as Herrick, Goldsmith, and Shakespeare. |
| Y including X | ...common-law countries, including Canada and England... |
| Y , especially X | European countries, especially France, England, and Spain... |

# Learning Patterns -- based on seeds

- *<Mark Twain, Elmira> Seed tuple*

"MarkTwain is buried in Elmira, NY."

X is buried in Y --pattern1 induced

"The grave of Mark Twain is in Elmira"

The grave of X is in Y –pattern2 induced

"Elmira is Mark Twain's final resting place"

Y is X's final resting place. –pattern3 induced

→ ***Use those patterns to grep for new facts.***

# Problems with patterns

- Hand-built

Plus: High-precision, tailored to specific domains

Minus: low-recall, huge labor

- Learning based on seeds

Plus: high-recall, less human labor

Minus: noise, low-precision

# Supervised machine learning methods (overview)

1. Choose a set of relations to extract
2. Find and label data
   - ✓ Label the named entities and the relations between these entities.
   - ✓ Break into training, development and test sets
3. Train a classifier on the training set
4. Find all pairs of named entities (usually in the same sentence)
5. Use the classifier to identify the relation

# For example: to identify the *employee relation* (Org, Per)

Input:

*Acme Inc. hired Mr Smith as their new CEO, replacing Mr Bloggs.*

**Org**: Acme Inc.
**Per**: Smith and Bloggs

(Acme Inc., Smith) → Model → true
(Acme Inc., Bloggs) → Model → false

# Train the Model

- Extract features:

1. Features similar to those used in the entity recognition: capitalized, suffix and so on.
2. Conjunctions of the features from the two entities: spouse_of needs person type of both entities.

- Choose models: many models.

# Word Features

Acme Inc (mention 1). hired Mr Smith (mention 2) as their new CEO, replacing Mr Bloggs.

- **Headwords of M1 and M2, and combination**

  Inc. Smith

- **Bag of words and bigrams in M1 and M2**

{Acme, Inc, Mr., Smith, Acme Inc, Mr. Smith}

- **Words or bigrams in particular positions left and right of M1 and M2**

M1: +1 hired M2: +1 as, -1 hired

- **Bag of words or bigrams between M1 and M2**

{hired}

# Named entity type and mention level Features

Acme Inc (mention 1). hired Mr Smith (mention 2) as their new CEO, replacing Mr Bloggs.

- **Named-entity types (ORG, PER, etc)**

M1: ORG M2: person

- **Concatenation of the two entity type**

ORG-PERSON

- **Entity level of M1 and M2 {name, nominal, pronoun}**

M1: name

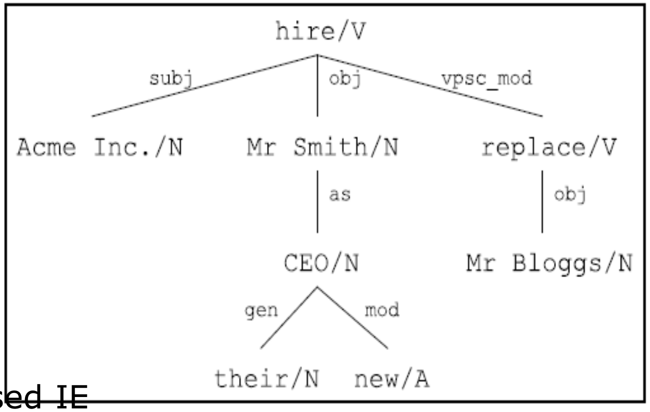
M2: name

# Parse Features

Acme Inc (mention 1). hired Mr Smith (mention 2) as their new CEO, replacing Mr Bloggs.

- **Base syntactic chunk sequence from one to the other**
  VP
- **Constituent path through the tree from one to the other**
  NP ↑ VP ↓ NP
- **Dependency path**

Acme Inc. hired Mr Smith



# Other Features: Gazeteers and trigger words

- Personal relative trigger list from Wordnet: parent, wife, husband,...
- Country name list
- Wikipedial

**Acme Inc** (mention 1). hired **Mr Smith** (mention 2) as their new CEO, replacing Mr Bloggs.

Feature summary

- **Entity-based features**

  M1 type: ORG
  M1 head: Inc
  M2 type: PERS
  M2 head: Smith

- **Word-based features**

  Between entity bag of words: {hired}
  Words before M1: none
  Words after M2: as

- **Syntactic features**

  Constituent path: NP VP NP
  Basic syntactic chunk path: VP
  Typed-dependency path:
  Acme Inc.<- subj hired→ obj Mr.Smith

# Classifiers for supervised methods (ref. chapter 5 of textbook)

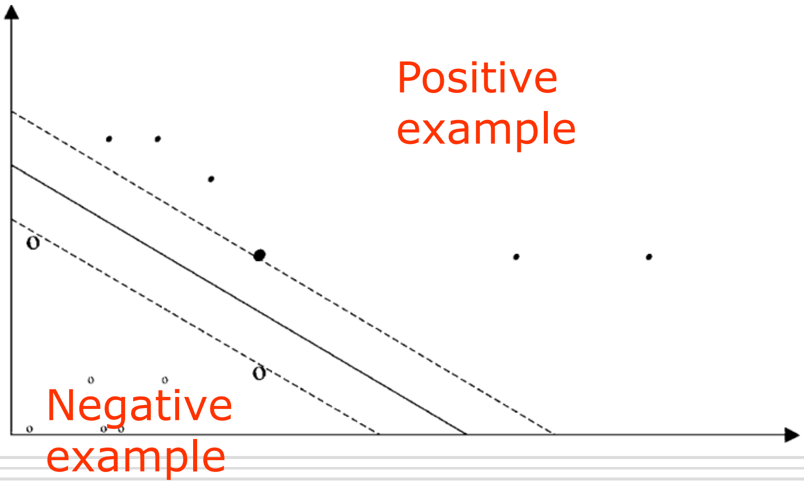
- Choose models:

1. MaxEnt(maximum entropy model)
2. NB(Naïve Bayes)
3. SVM(support vector machines)
4. …

- Train it on the **training set**, turn on the **development set**, test on the **test set**.

# Relationship Extraction using Support Vector Machine (SVM)

- ***Support vector machine (SVM)*** *is recognized as one of the best classification algorithm over various applications and domains.*
- *SVM is a method that finds a function that discriminates between two classes.*



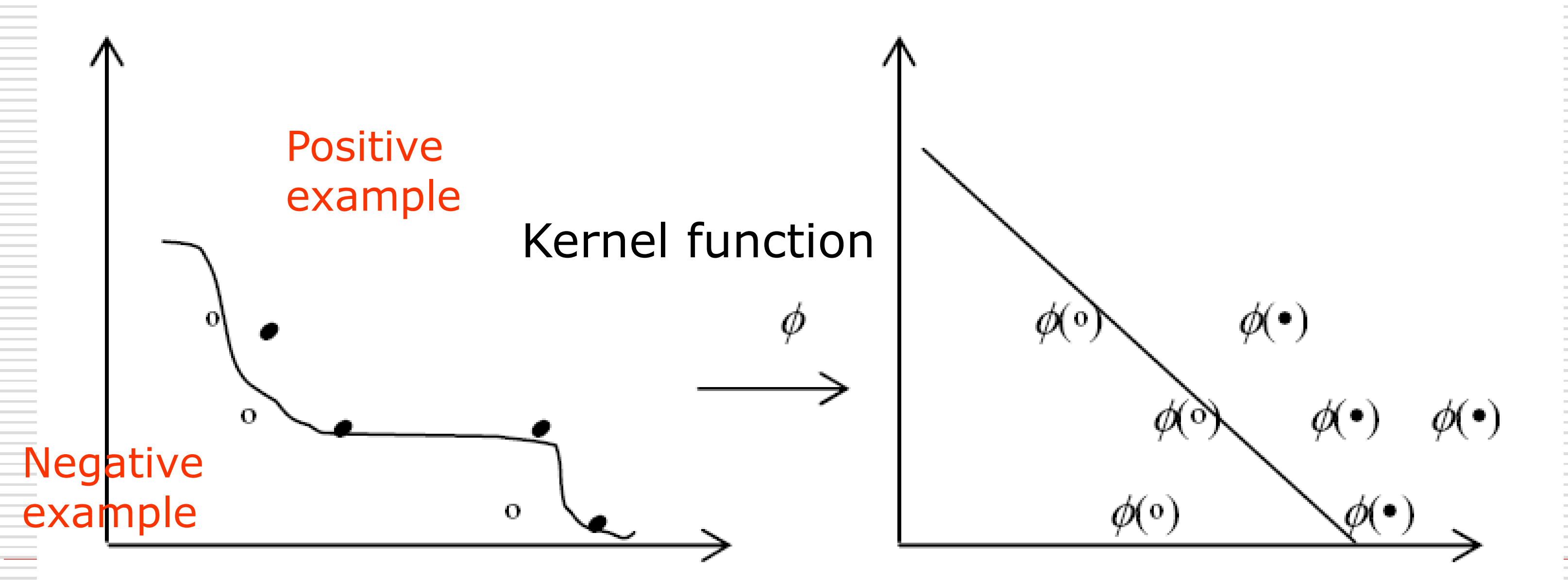

Given the set $S$ of $n$ training examples:

$$S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

where $x_i \in \Re^p$ ($p$-dimensional space) and $y_i \in \{-1, +1\}$ indicating that $x_i$ is respectively a negative or a positive example.

# Support Vector Machine (SVM)

- When classifying natural language data, it is not always possible to linearly separate the data → map them into a feature space where they are linearly separable.



# SVMLight: an open software

- Install an SVM package such as SVMlight (http://svmlight.joachims.org/)
- Transfer your training data format in order to be matched.
- Use training command for SVMlight.

SVM Ref:

http://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-the-linearly-separable-case-1.html#svm-sv-classifier

# A Guide to SVM

- Transform data to the format of an SVM package.
- Conduct simple scaling on the data.
- Choose a kernel for SVM.
- Use cross-validation to the best parameter.
- Train the whole training set.
- Test

# Data Preprocessing

- SVM requires that each data instance is represented as a vector of real numbers.
- Use m numbers to represent a m-category attribute. For example a three-category attribute such as (red, green, blue) can be represented as (0,0,1), (0,1,0), and (1,0,0).

# Scaling

- Some attribute may be a value, such as the length of a sentence.
- Scaling before using SVM → [0,1] or [-1,1], for example, [-10,10] to [-1,1]
- How ?

  X= (x-min)/(max-min)

**Using the same scaling factors for training and test sets, obtain better result.**

# Choose a kernel

- Linear kernel when the number of features is very large.
- RBF kernel can handle nonlinear problem.

# Cross-validation & grid-search

- In v-fold cross-validation, first divide the training set into v subsets of equal size. Sequentially one subset is tested using the classier trained on the remaining v-1 subsets.
- Each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified.
- Grid-search parameter using cross-validation.

# Problems of Supervised methods

- High precision with enough hand-labeled training data.
- Labeling is expensive.
- Supervised models can not generalize well to different genres.

# Summarization

- *What is relations recognition? Three cases*
- *How to identify relations?*
- *Pattern-based methods*
- *Supervised methods*

# References

- Text book chapter 5 Supervised Classification
- Sunita Sarawagi. Information Extraction Foundations and Trends in Databases vol.1,No.3 2007  261-377.
- Jun Zhu,et al. StatSnowball:a statistical apprpach to extracting entity relationships  In Proceedings of WWW 2009, Madrid.
- Mintz,Bills,Snow,Jurafsky.Distant supervision for relation extraction without labeled data. ACL 2009
- Standford Book about IR: http://www-nlp.stanford.edu/IR-book/html/htmledition/contents-1.html

# About the Project

- Task: Employment relation extraction
- Training corpus:本报北京１２月３０日讯新华社记者胡晓梦、本报记者吴亚明报道：新年将至，国务院侨务办公室主任郭东坡今天通过新闻媒介，向海外同胞和国内归侨、侨眷、侨务工作者发表新年贺词。

(胡晓梦,新华社)

(吴亚明,新民晚报)

(郭东坡,国务院侨务办)

# About the Project (cont.)

- Methods:
  - Pattern-based
  - Supervised method or semi-supervised or unsupervised methods

*Training corpus are put online*.

- Evaluation:

Use test corpus with human annotated results to evaluate your algorithm.