Relation Recognition

Fang Li

Dept. of Computer Science & Engineering

Contents

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

1.Knowledge Engineering Approach

2.Machine Learning Approach

--Supervised learning

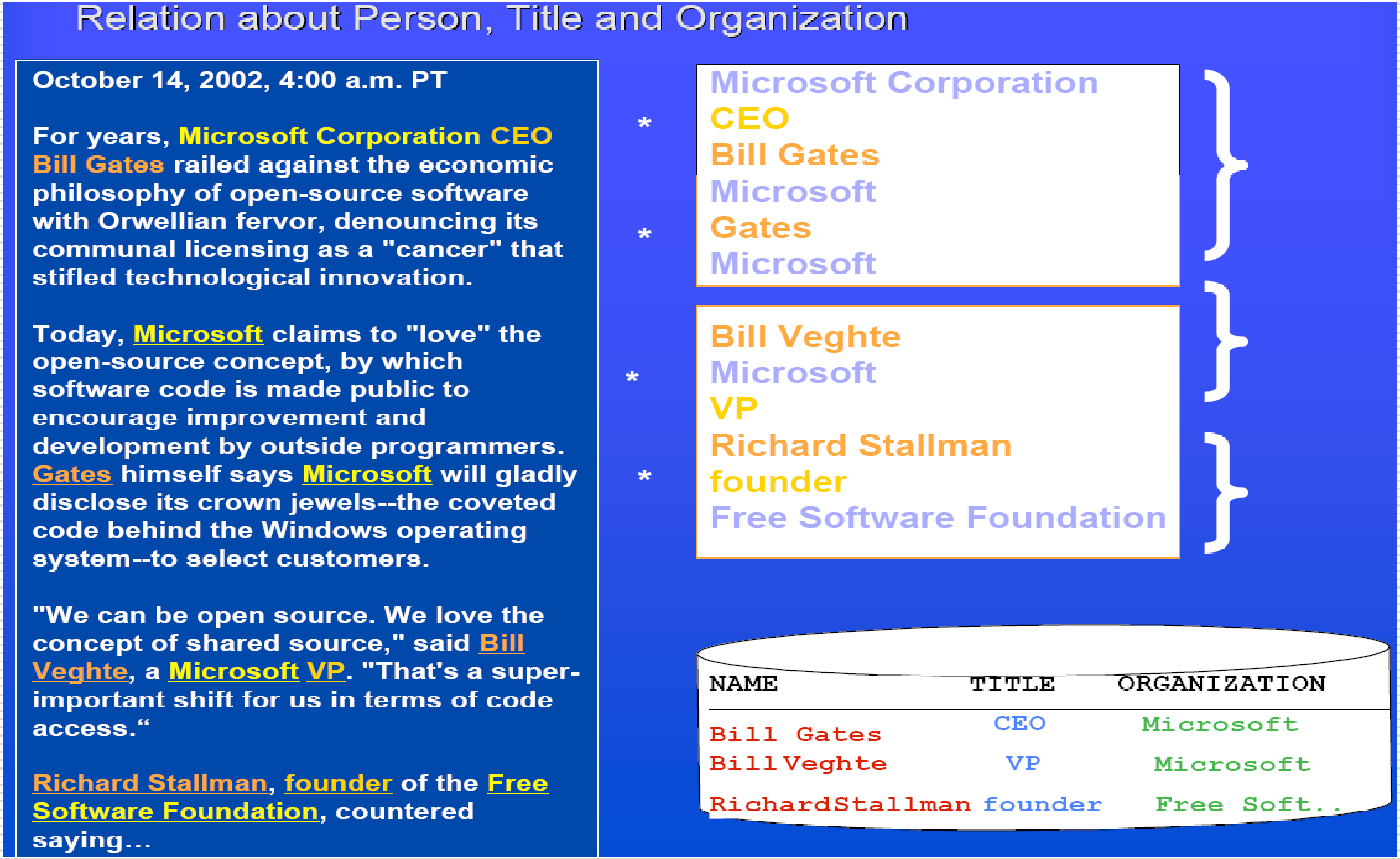
--Semi-supervised learning

-- Distant supervised learning

-- Deep Learning

3. Knowledge graph

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*….

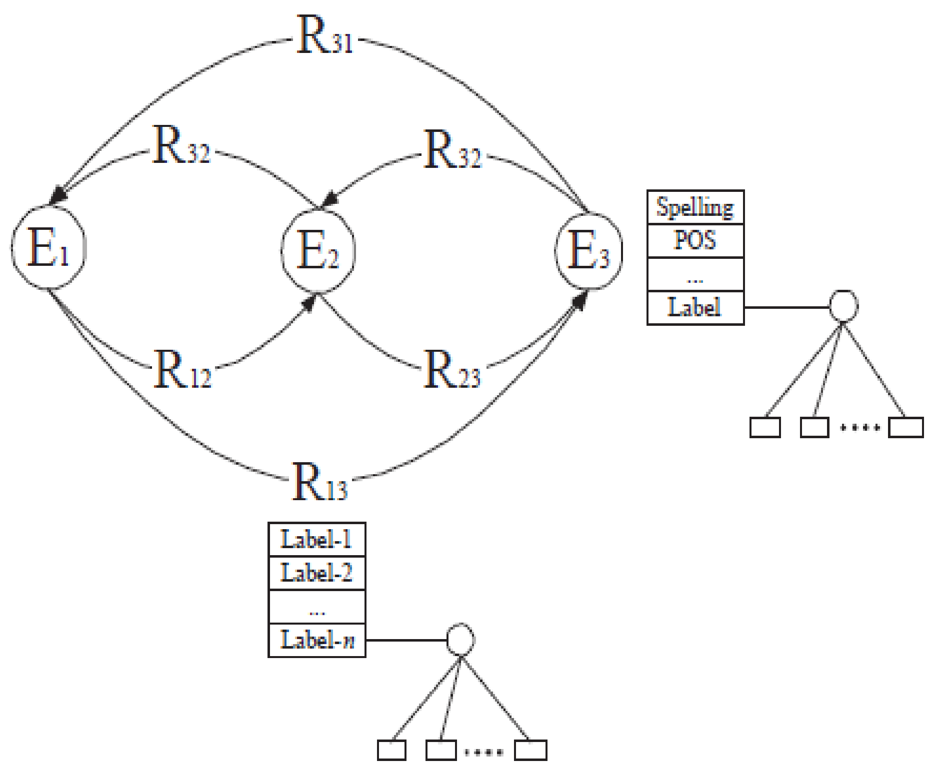
Explicit and Implicit Relations (cont.)

- 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: Was Tony Blair a British Prime Minister?

A conceptual view of entities and relations

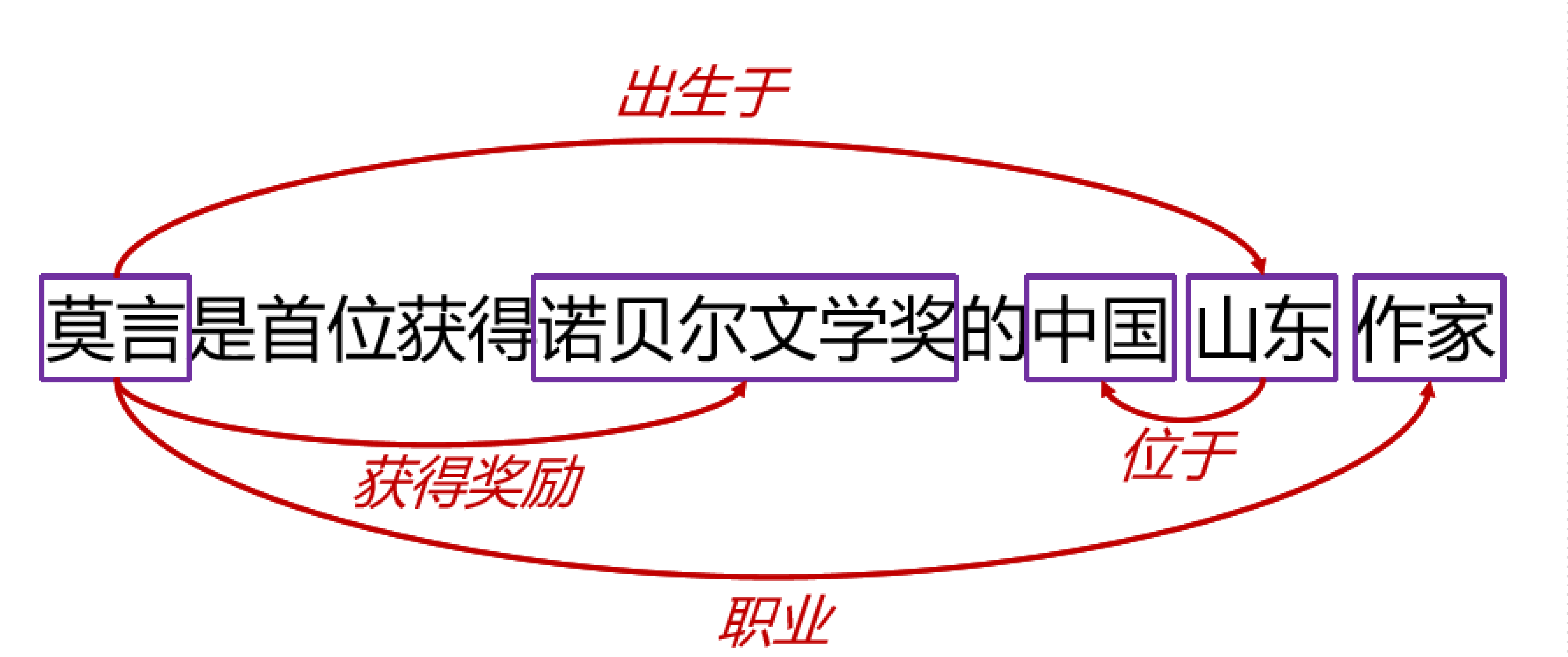


A conceptual view of entities and relations

- 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.

Preprocessing 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

Preprocessing 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
5. Deep learning method
6. …

Hearst’s Patterns

Some patterns extracted from the sentence or between the two entities: X is an instance of Y

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 Inc.Smith
- **Bag of words and bigrams in M1 and M2**

{Acme, Inc, hired, 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**

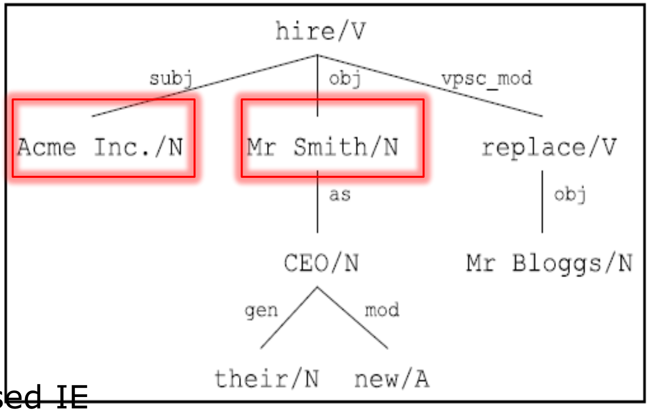
  VP

- **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.

- **Entity-based features**

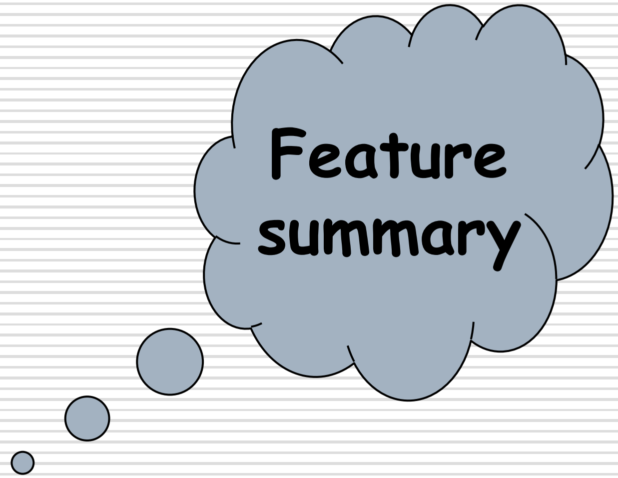
  M1 type: ORG
  M1 head: Inc
  M2 type: PERS
  M2 head: Smith

- **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



- ***American Airlines**, a unit of AMR, immediately matched the move, spokesman **Tim Wagner** sa*

From Dan Jurafsky

为什么需要两个实体的Entity head特征呢？

Entity-based features		
	Entity$_1$ type	ORG
	Entity$_1$ head	*airlines*
	Entity$_2$ type	PERS
	Entity$_2$ head	*Wagner*
	Concatenated types	ORGPERS
Word-based features		
	Between-entity bag of words	{ *a, unit, of, AMR, Inc., immediately, matched, the, mo* *spokesman* }
	Word(s) before Entity$_1$	NONE
	Word(s) after Entity$_2$	*said*
Syntactic features		
	Constituent path	$NP \uparrow NP \uparrow S \uparrow S \downarrow NP$
	Base syntactic chunk path	$NP \rightarrow NP \rightarrow PP \rightarrow NP \rightarrow VP \rightarrow NP \rightarrow NP$
	Typed-dependency path	$Airlines \leftarrow_{subj} matched \leftarrow_{comp} said \rightarrow_{subj} Wagner$

Features → Feature representation of the model

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

- Feature between Mention1 and mention2: word sequences or number of words between them.
- Morphologic feature of mention 1（形态特征）：

  ACME INC, A.C.M.E Inc, Acme Inc, acme inc
- Combination feature （顺序关系）：

(company name, person name), (person name, company name)

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

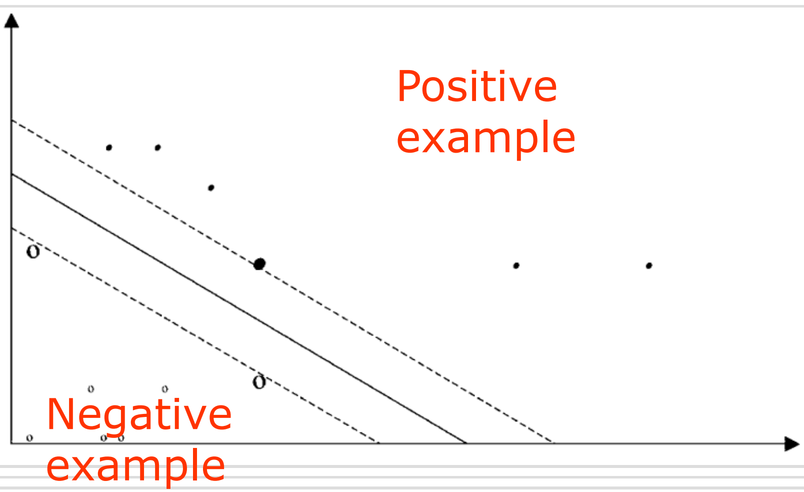
- Choose models:

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.*



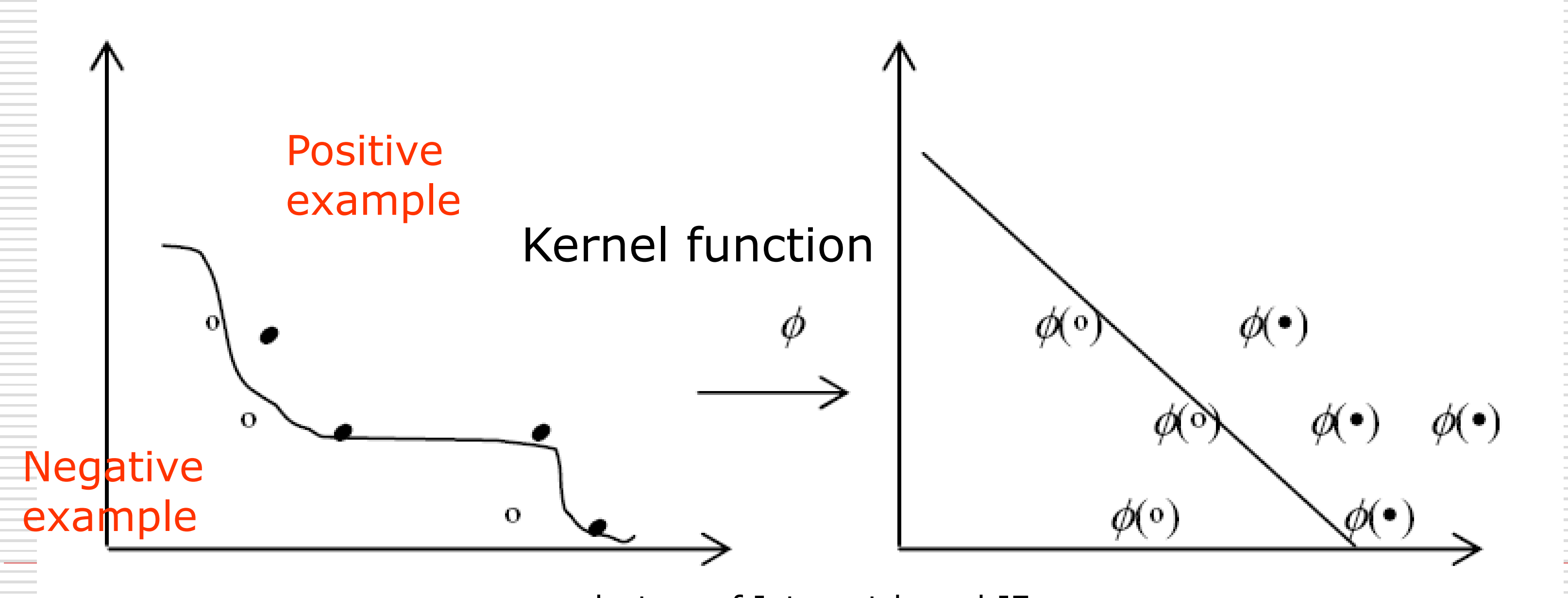
Given the set S of n training examples:

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

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.

Comparison of Classification Models

- Test corpus: Reuters-21578 Data Set
- 21578 documents
- 118 categories
  - An article can be in more than one category
  - Learn 118 binary category distinctions
- Common categories (#train, #test)

  - Earn (2877, 1087)
  - Acquisitions (1650, 179)
  - Money-fx (538, 179)
  - Grain (433, 149)
  - Crude (389, 189)
  - Trade (369,119)
  - Interest (347, 131)
  - Ship (197, 89)
  - Wheat (212, 71)
  - Corn (182, 56)

(a)

	NB	Rocchio	kNN		SVM
micro-avg-L (90 classes)	80	85	86		89
macro-avg (90 classes)	47	59	60		60

(b)

	NB	Rocchio	kNN	trees	SVM
earn	96	93	97	98	98
acq	88	65	92	90	94
money-fx	57	47	78	66	75
grain	79	68	82	85	95
crude	80	70	86	85	89
trade	64	65	77	73	76
interest	65	63	74	67	78
ship	85	49	79	74	86
wheat	70	69	77	93	92
corn	65	48	78	92	90
micro-avg (top 10)	82	65	82	88	92
micro-avg-D (118 classes)	75	62	n/a	n/a	87

Evaluation measure: F_1

Semi-supervised method

Relation Bootstrapping

- Gather a set of seeds
- Iterate:

1. Find sentence with these seeds
2. Look at the context between or around the seeds to define a pattern
3. Use the pattern for more examples

Bootstrapping from seed entity pairs to learn relations

function BOOTSTRAP(*Relation R*) **returns** *new relation tuples*

tuples ← Gather a set of seed tuples that have relation *R*
iterate
sentences ← find sentences that contain entities in *seeds*
patterns ← generalize the context between and around entities in *sentences*
newpairs ← use *patterns* to grep for more tuples
newpairs ← *newpairs* with high confidence
tuples ← *tuples* + *newpairs*
return tuples

Confidence Value for Bootstrpping

- Given a document collection ***D***, a current set of tuples ***T***, and a proposed pattern ***P,*** two factors need to be considered:
  - Hits: the set of tuples in ***T*** that ***p*** matches while looking in D.
  - Finds: The total set of tuples that ***p*** finds in ***D***

$$Conf_{RlogF}(p) = \frac{hits_p}{finds_p} \times log(finds_p)$$

Conf(CD)=2/4 X log (4)=30%

A corpus D:

ABCDEF
BDCDFE
CDEFG
HHECDE

Tuple set T:

BCDE
ECDE
AUD

Dipre: Extract <author,book> pairs

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

The Comedy of Errors, by William Shakespeare, was

The Comedy of Errors, by William Shakespeare, is

The Comedy of Errors, one of William Shakespeare's earliest attempts

The Comedy of Errors, one of William Shakespeare's most

- Extract patterns (group by middle, take longest common prefix/suffix)

`?x , by ?y ,`　　　　`?x , one of ?y 's`

- Now iterate, finding new seeds that match the pattern

Example: Extract Person name and position title

- Search Engine Keywords: Wang Ning + vice Mayor

王宁出任北京市副市长 优秀区县委书记首成副部级 中国网络电视台 1天前
王宁出任北京市副市长 新浪新闻 2天前
王宁被任命为北京市副市长 为土生土长北京人 腾讯新闻 2天前
王宁当选北京市副市长 张延昆不再担任(图|简历)_网易新闻中心

- **Patterns:**

[person][was assigned | was selected | was appointed as][position]

- **New examples:**

陆志鹏同志当选南通市委书记--组织人事-人民网
易炼红当选长沙市委书记_网易新闻
曹炯芳当选湘潭市委书记_凤凰资讯

上海市发改委主任俞北华被任命为市政府副秘书长(图) 网易新闻
方洪添被任命为广东省食品药品监督管理局副局长 南方网
2015新增院士公布 浙大四位教授当选

Summarization

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

回家作业课程网站下载，两篇文章选一篇阅读，完成截止日10月28日24点以前

- Mintz,Bills,Snow,Jurafsky.Distant supervision for relation extraction without labeled data. ACL 2009
- TransE

Distant supervision method

"Distant supervision for relation extraction without labeled data"

- What means "**distant supervision**"?
- What are the **advantages** of the method?
- What are the **disadvantages** of the method?

lecture of Internet-based IE Technology