Lecture 10 Information Extraction II: Relation Extraction and Sentiment Analysis

Relation Extraction

— Task of ***extracting semantic relationships from a text*** *. Ectracted relationships ususally occur* ***between two or more entities of a certain type*** *and* ***fall into a number of semantic categories***

- Why extract relation?

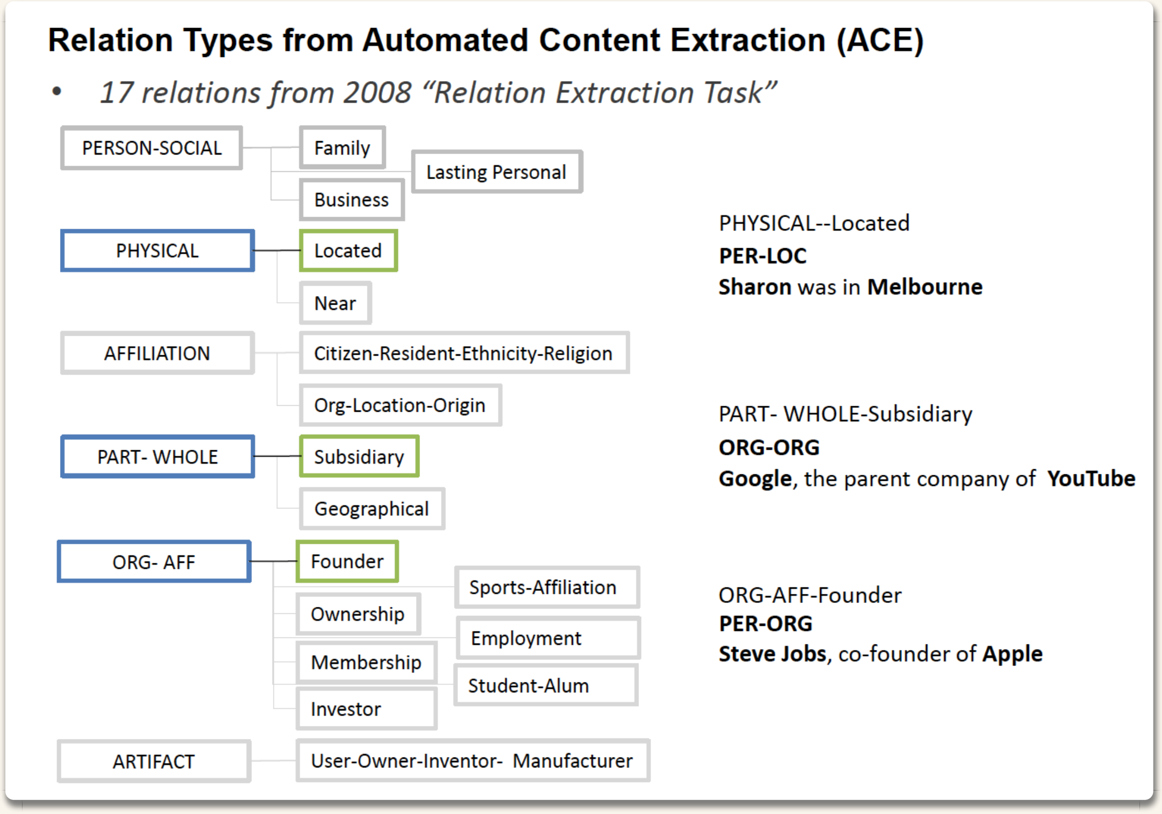
  — most applications require structured knowledge bases

  — most NLP application require word sense
  - QA

    —*Who is the granddaughter of which actor starred in the movie"E.T."?*

    *[granddaughter-of(*X*,* Y*)]* *[is-a[*Y*, actor]* *[acted-in(*X*, "E.T.")]*
  - Coversational Agent
  - Summarisation

- What relations should extract?

- 

- relations in Wikipedia: Database

  — RDF(REsource Description Framework) triples

  — Two serialisations:

  - turtle (til): *provides data in n-triple format(.) as a subset oft urtle serialization*
  - *Quad-turtle (tql): the quad turtle serialization (<graph/context>.) adds context information to every triple, containing the graph name and provenance information on each triple*
  - *relations in WordNet (missing for many words relations)*

    — WordNet(ontology) expresses relations between two words

    - *Hyponymy: San Francisco is an instance of a city*
    - *Antonymy: Acidic is the opposite of basic*
    - *Meronymy: An alternator is a part of a car*

Relation Extraction Approaches

(Hand-build) Pattern/Rule-based Approach

- *Methods*

— *Hearst(1992) proposed patterns for Hyponymy (is-a relation)*

— *Starting with Named Entity tags: usually relations hold between specific entities*

— *Rules and Named Entities*

- *Advantages*

  — *tends to have* *high-precision* *but* *low-recall*

  — *can be tailored to specific domains(domain-dependent)*

- *Disadvantages*

  — *Impossible to build pattern/rules for all possible relations*

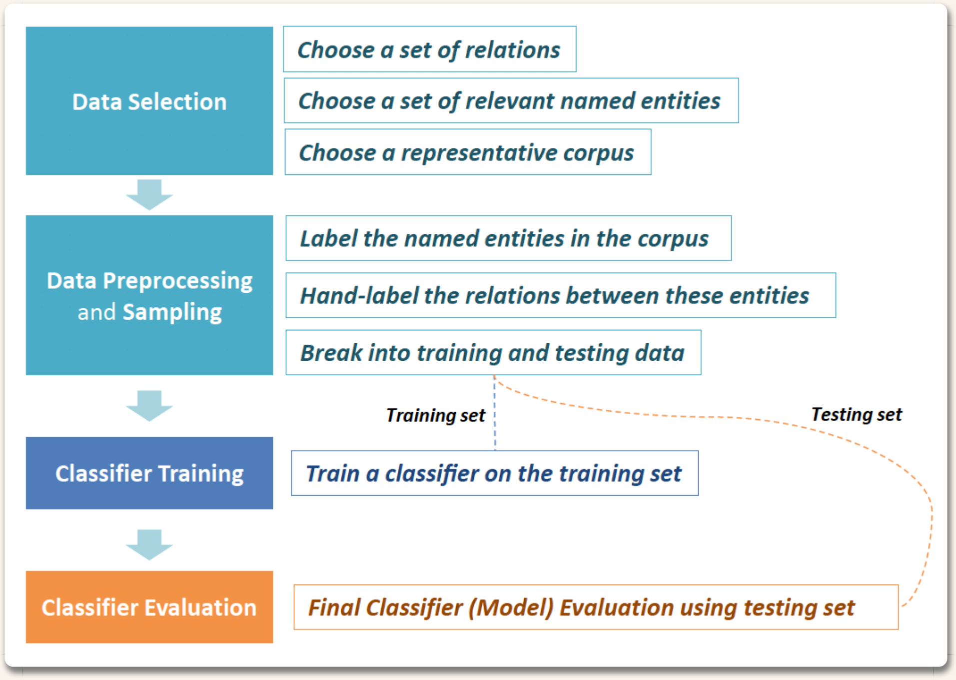
  — *difficult to generalize into new domain*

  — *produces low accuracy*

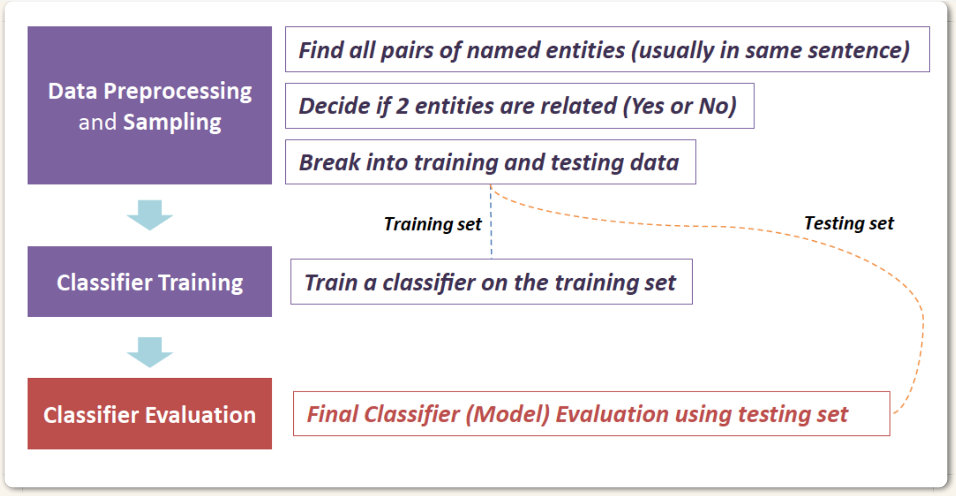
  - *Hearst(1992): 66% accuracy*
  - *Berland and Charniak: 55% accuracy*

Supervised Approach

- *How to build up?*
  - *Tranditional (classifier #2)*



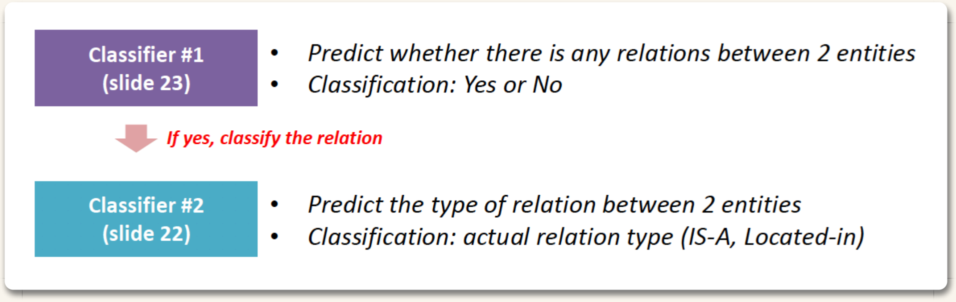
  - *Efficient (classifier #1)*



- *Apply both classifiers*

  — *Fast classification training* *by eliminating most pairs*

  — *canuse* *distinct feature-sets appropriate* *for each task*



- ***Feature exraction***

Mention1 *Mention2*
"Sydney University is an public research university in Australia."

- *Feature 1: word feature*

— Head words of mention 1 (M1) and mention 2(M2) and combination

e.g. University Australia University-Australia

— Bag of words and bigrams in M1 and M2

e.g. {Sydney, University,Australia, Sydney University}

— Words or bigrams in particular postitions left and right of M1/M2

e.g. M2: -1 in M1: +1 is

— Bag of words or bigrams between the two entities

e.g. {is, an, public, research, university, in}

- *Feature 2: NE type and mention level*

  — Named entity types

  e.g. M1: ORG. M2:LOC

  — Concatenation of the two named-entity types

  e.g. ORG-LOC

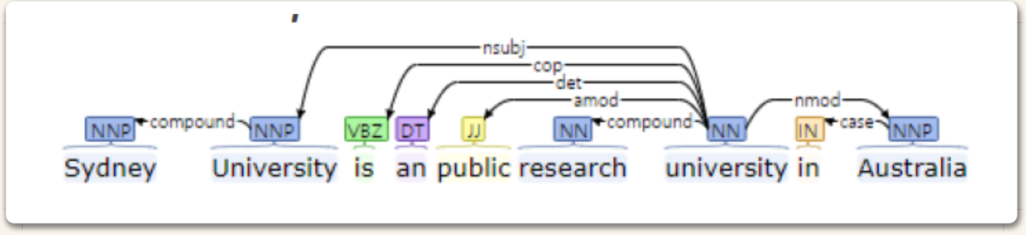
  — Entity leve of M1 and M2 (NAME, NOMINAL, PRONOUN)

  e.g. M1:NAME M2: NAME
- *Feature 3: NE based on parse feature*

  — Base on syntactic chunk sequence from one to the other

  NP. VP. NP. PP. NP

  — Dependency path: head and dependencies



- *Feature 4: Trigger or gazetteer level*

  — Trigger list of family: kinship terms

  e.g parent, wife, husband, gramdparent,etc. [form WordNet]

  — Gazetteer
  - *List of usefule geo or geopolitical words*
  - *country name list*
  - *other sub-entities: names of river, road etc*

  - *Person name*

Classifiers for supervised methods

Use any types of classifier that you would like to use:

- *Naive Bayes*
- *SVM*
- *Decision Tree*
- *Neural Network*

Evaluation: precinct, recall, or F-measure for each relation

- *Precision: # of correctly extracted relations/Total # of extracted relations*
- *Recall: # of correctly extracted relations / Total #of gold relations*
- *F-measure: 2PR/P+R*

Pros and Cons

- *Can get high accuracies with enough hand-labeled training data*

  — if test similar enough to training
- *Expensive: should label a large training set*
- *Cannot generalize well to different genres*

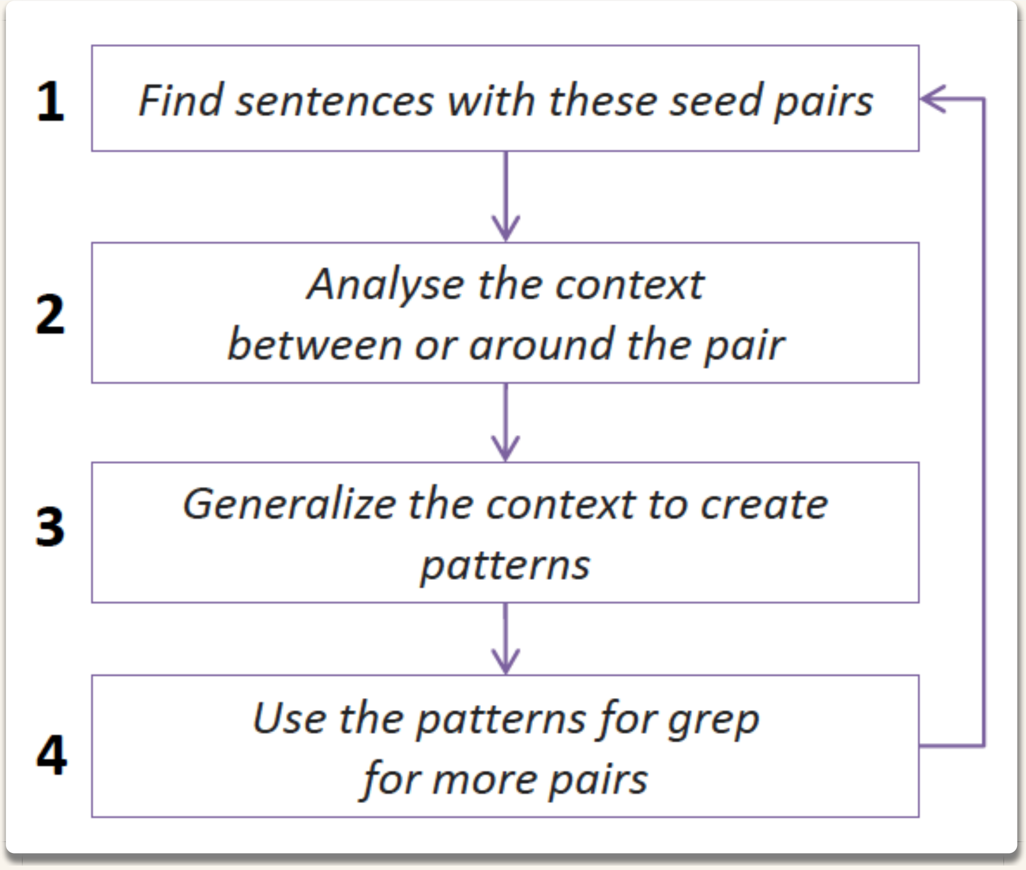
Unsupervised/semi-supervised approach

- *No well-structured training dataset*
- *Solution*
  - *build a few seed tuples*
  - *build a few high-precision patterns*
  - *Bootstrapping: use the seeds to directly learn to populate a relation*

- *Bootstrapping (Hearst 1992)*

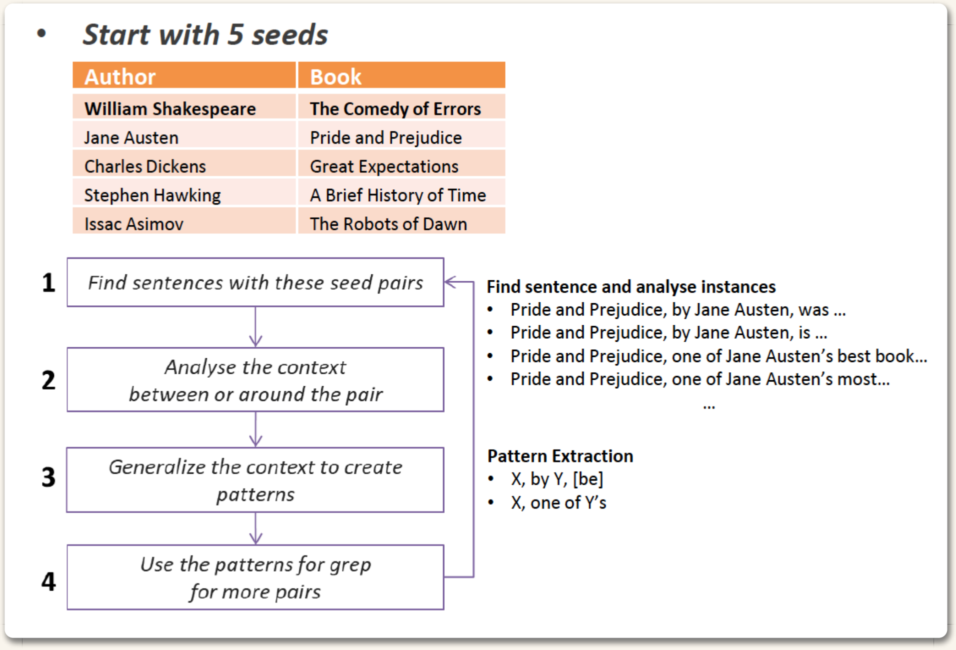
  — Setup: Gather a set of seed pairs that have relation

  — Process



— *Example*

1. *Dipre: RE using <author,book> pairs (Sergei 1998)*



2. *Snowball(2000)*

- *Use similar iterative algorithm*

Organization	Location of Headquarters
Microsoft	Redmond
Google	California
IBM	Armonk

- *Group instances w/similar prefix, middle, suffix, extract patterns*
  - *Require that X and Y be named entities*
  - ***Compute a confidence for each pattern***

0.69	ORG	{'s, in, headquarters}	LOC
0.85	LOC	{in, based}	ORG
…	…	…	…

- *Distant supervision*

  Distant Supervision = Bootstrapping + Supervised Learning

  - *Use a large database to get huge # of seed examples (instead of 5 seeds)*
  - *Create lots of features from all these examples*
  - *Combine in a supervised classifier*

Distant Supervision Paradigm

Aspect: supervised classification

- *Uses a classifier with lots of features*
- *Supervised by detailed hand-created knowledge*
- *Doesn't require iteratively expanding patterns*

Aspect: unsupervised classification

- *Uses very large amounts of unlabeled data*
- *Not sensitive to genre issues in training corpus*

  - *How to proces distant supervised learning*

Step	Process	Example
1	Go through each relation	Born-In
2	Go through each tuple in big database	<Vincent van Gogh, Netherlands> <Albert Einstein, Germany>
3	Find sentences in large corpus that have both entities	Albert Einstein, born 1879, Germany ... Gogh was born in Netherlands... Einstein was born in Ulm, Germany .... Gogh's birthplace in Germany ... .
4	Extract frequent features (parse, words, etc)	PER, born XXXX, LOC PER was born in LOC PER's birthplace in LOC
5	Train supervised classifier using thousands of patterns (positive and negative instances)	P(born-in | f1,f2,f3,...,fn)

- *Evaluation*

  — extact totally new relations from the web

  — No gold set of correct instances of relations

  - *Don't know which ones are correct: cannot compute precision*
  - *Don't know which ones were missed: cannot compute recall*

  — Solution: approximate precision

  - *Draw a random sample of relations from output, check precision manually*

    P= #of correctly extracted relations in the sample/Total # of extracted relations in the sample

  - *Can also compute precision at different levels of recall*

    Precision for top 1000 new relations, top 10,000 new relations, top 100,000

    (In eanch case taking a random sample of that set)
  - *Still no way to evaluate recal (without labelling whole entire relations)*

Sentiment Analysis

Sentiment analysis is the operation of understanding the intent or emotion behind a given piece of text. It is part of text classification but it is usefule for extracting structured information

- *Sentiment analysis = the detection of attitudes*

  Eduring, affectively coloured beliefs, dispositions towards objects/persons

- *Main Factors*
  - *Target Object: an entity that can be a product, person, event, organisation, or topic*
  - *Attribute: An object usually has two 2 types*

    — Componensts (e.g. touch screen, battery)

    — Properties (e.g. size, weight, colour, voice quality)

    — Explicit and implicit attributes:
    - *Explicit attributes: appearing in the attitude (e.g. "the battery like of this phone was not long")*
    - *Implicity attributes: not appearing in the attitude(e.g. "this phone is too expensive"— the property price)*
  - *Attitude Holder: the person or organisation that expresses the opinion (e.g. my mother was mad with me)*
  - *Type of attitude: positive, negative,or neutral or set of types (e.g happy)*
  - *Time: the time that expresses the opinion*

- *Why useful?*
  - *sentiment coud be considered a latent vairable in social behaviour.*
  - *Measuring and understanding this behaviour, could lead to better understanding of social phenomena*

— *Sentiment analysis often correlates well with real word observables*

- *Commercial: Brand Awareness*
- *Stock fluctuations and public opinoin*
- *Health related: vaccine sentiment vs coverage*
- *Public safety: situational awareness in mass emergencies via Twitter*

- *Build up sentiment lexicon*
  - *Bootstrap style: semi-supervised learning of lexicons*

    — *use a small amount of information*

    — *a few labeled examples*

    — *a few hand—build patterns*

    — *bootstrapping lexicon*

- *Feature vectors*

  — *word ngrams(up to 4),skip ngrams w/1 missing word*

  — *character ngrams up to 5*

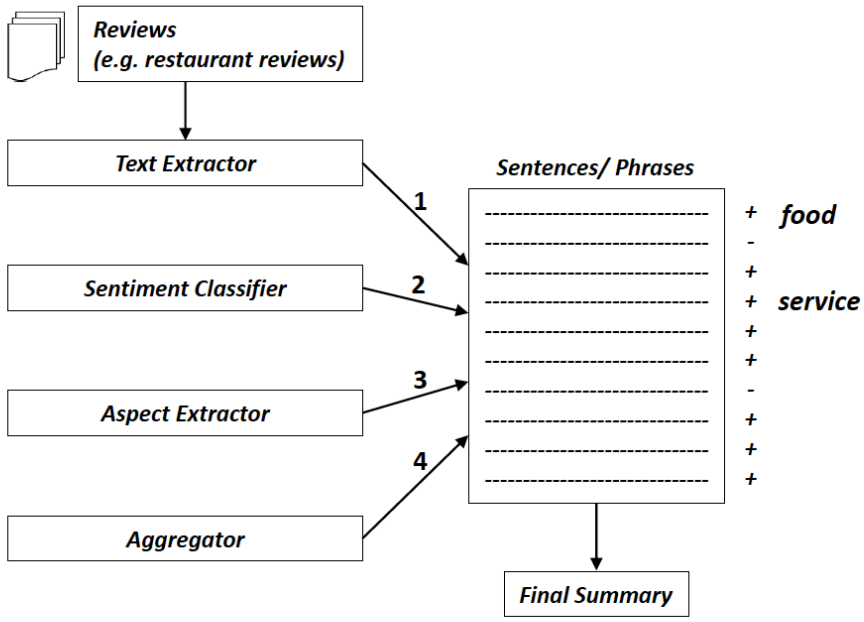
  — *all caps: number of words in capitals*

  — *Number of continuous punctuation marks, either exclamation ot question or mixed. Also whether last char contains one of these.*

  — *presence of emotions*

- *Finding sentiment of a sentence*
  - *finding aspects or attributes: target of sentiment*

Finding aspect/attribute/target of sentiment



Finding aspect/attribute/target of sentiment

Title: Sharp, Solid, but Harder to Hold than IPhone 7
- By Tristan on March 13, 2017

"my thoughts on the iPhone 7 are:
1) Retina display is awesome. Everything looks more defined and sharper. There is much color and clarity out there... or should I say, in those digital images and videos... needless to say, the camera as well captures great images.
......."

Attribute based Summary

- Attribute 1: display
  - Positive
    1. Retina display is awesome
    2. There is much color and clarity out there
    3. ...
- Attribute 2: camera
  - Positive
    1. the camera as well captures great images.
    2. .....

Attribute based Visualisation