Relation Extraction

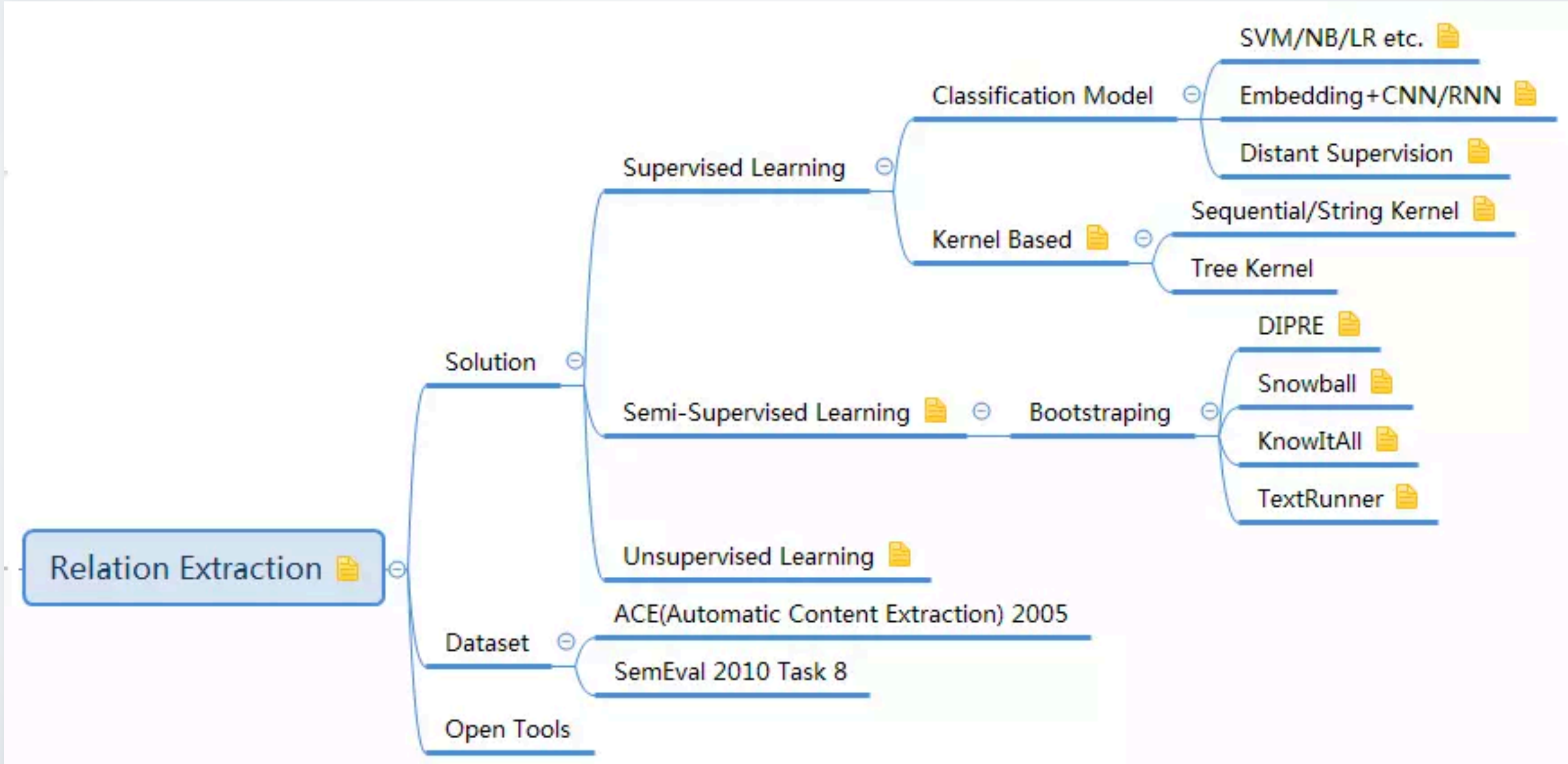
A Survey

Yifei Li

Relation Extraction

- Occurrences of entities in a sentence are often linked through well-defined relations; e.g., occurrences of person and organization in a sentence may be linked through relations such as employed at.
- The task of **Relation Extraction (RE)** is to identify such relations automatically.

Relation Extraction



Relation Extraction: Challenges

- There is a variety of possible relations, which vary from domain-to-domain.
- Non-binary relations
- Supervised techniques: lack of sufficient training data
- Ambiguity
- Language-dependent

Datasets

- ACE: Automatic Content Extraction
    - Entity Detection and Tracking (EDT): detecting mentions of these NEs, and identifying their co-references
    - Relation Detection and Characterization (RDC): detecting relations between entities identified by the EDT task

Global level vs Mention level

- Global level RE
  - To produce a list of entity pairs for which a certain semantic relation exists
  - Generally takes a large text corpus as input and produces such a list as output
- Mention level RE
  - Takes as input an entity pair as well as the sentence which contains it
  - Identifies whether a certain relation exists for that entity pair
  - Also called: Relation Detection and Characterization (RDC).

Outline

- Supervised techniques
  - Feature-based methods
  - Kernel-based methods (unfinished)
- Approaches for joint extraction of entities and relations
- Semi-supervised techniques
- Unsupervised techniques
- Open Information Extraction (Open IE)
- Distant supervision based techniques
- Recently advanced techniques

Supervised Approaches

- Required Labeled Data: each pair of entity, labelled with pre-defined relation types (including: NONE)
- Can be considered as a multi-class classification problem
- Two kinds of methods:
  - Feature-based methods
  - Kernel-based methods

Feature-based Methods

- For each relation instance (pair of entity mentions) in the labelled data, a set of features is generated and a classifier is then trained to classify any new relation instance.
- Kambhatla et al. trained a maximum entropy classifier with 49 classes
  - ACE 2003 has 24 relation subtypes, each has 2 classes considering order of relation arguments
  - a NONE class where two mentions are not related

Feature Types	Example
Words: Words of both the mentions and all the words in between	M11_leaders, M21_Venice; B1_of, B2_Italy, B3_'s, B4_left-wing, B5_government, B6_were, B7_in
Entity Types: Entity types of both the mentions	E1_PERSON, E2_GPE
Mention Level: Mention types (NAME, NOMINAL or PRONOUN) of both the mentions	M1_NOMINAL, M2_NAME
Overlap: #words separating the two mentions, #other mentions in between, flags indicating whether the two mentions are in the same NP, VP or PP	7_Words_Apart, 2_Mentions_In_Between (Italy & government), Not_Same_NP, Not_Same_VP, Not_Same_PP
Dependency: Words, POS and chunk labels of words on which the mentions are dependent in the dependency tree, #links traversed in dependency tree to go from one mentions to another	M1W_were, M1P_VBD, M1C_VP, M2W_in, M2P_IN, M2C_PP, DepLinks_3
Parse Tree: Path of non-terminals connecting the two mentions in the parse tree, and the path annotated with head words	PERSON-NP-S-VP-PP-GPE, PERSON-NP:leaders-S -VP:were-PP:in-GPE

Table 2: Various feature types with examples described by Kambhatla [57]

Improvement

- Features:
  - Word based features
  - Base phrase chunking based features
  - Features based on semantic resources
- SVM Classifier: one vs others (because SVM is a binary classifier)
- Some interesting features blablabla... (没仔细看 -_-||)

Summary for Feature-based Methods

- Given relation instance (pair of entity mentions), do feature engineering and classify it into a relation type
- Major problem: *Class Imbalance* --- the number of **negative** instances (entity pairs with no meaningful relation) outnumber the number of **positive** instances (entity pairs with one of the pre-defined relation type), which results in a higher precision but lower recall because there are lots of NONE class.
- Most of the efforts are spent in designing the "right" set of features, which requires careful analysis of contribution of each feature and knowledge of underlying linguistic phenomena.

Kernel-based Methods

- Sequence Kernel
- Syntactic Tree Kernel
- Dependency Tree Kernel
- Dependency Graph Path Kernel
- Composite Kernels
- 这部分由于缺乏kernel function和kernel methods的背景知识，加之时间比较紧张，感觉理解起来比较困难，因此先跳过，以后有时间再补。如果老师/学长学姐能稍微指点一下，起个头，那就更好。

Jointly extract entities and relations

- Most of RE techniques: knowledge about boundaries and types of entity mentions are known before hand.
- If such knowledge is not available, entity extraction techniques needs to be used first.
- Once entity mentions and their entity types are identified, then RE techniques can be applied.
- This "pipeline" propagates errors from the first phase to the second phase. To avoid this, a line of research were done to extract entities and relations jointly.

Integer Linear Programming based Approach

- First learns independent local classifiers for entity extraction and RE
- Given a sentence, produce a global decision that domain-specific or task-specific constraints are satisfied.
- To solve the decision consistent with constraints, an Integer Linear Programming (ILP) approach is proposed. It minimizes the cost function including two parts:
  - Assignment cost: the most probable prediction of a local classifier
  - Constraint cost: to impose cost for breaking constraints between connected entities and relations
- Performance is much better than the simple "pipeline" method

Graphical Models based Approach

- The dependencies between entities and relations are encoded through a Bayesian belief network which is a bipartite, directed acyclic graph
- Entities are represented as nodes in one layer, whereas relations in another layer
- Each relation instance node Rij two incoming edges from its argument entity instance nodes Ei,Ej
- Given the feature vector X which characterizes the sentence, the local entity and relation classifiers are used to compute Pr(Ei|X) and Pr(Rij|X), respectively
- The constraints are encoded through the conditional probabilities Pr(Rij|Ei,Ej),
- Goal: $$arg \max_{e_i, r_{jk}} Pr(E_1, E_2, \cdots E_n, R_{12}, R_{21}, \cdots R_{n(n-1)}) \quad (3)$$

Card-Pyramid Parsing

- Encodes mutual dependencies among the entities and relations in a graph structure
- Aim to jointly label the nodes in the card-pyramid graph
- The authors propose a parsing algorithm analogous to the bottom-up CYK parsing algorithm for Context Free Grammar (CFG) parsing
  - Entity Productions of the form EntityT ype → Entity, e.g. PER→leaders. A local entity classifier is trained to compute the probability that entity in the RHS being of the type given in the LHS of the production.
  - Relation Productions of the form RelationType → EntityType1 EntityType2, e.g. PHYS→PER GPE. A local relation classifier is trained to predict the probability that the relation type in the LHS holds between the two entities in the RHS of the production.

Card-Pyramid Parsing

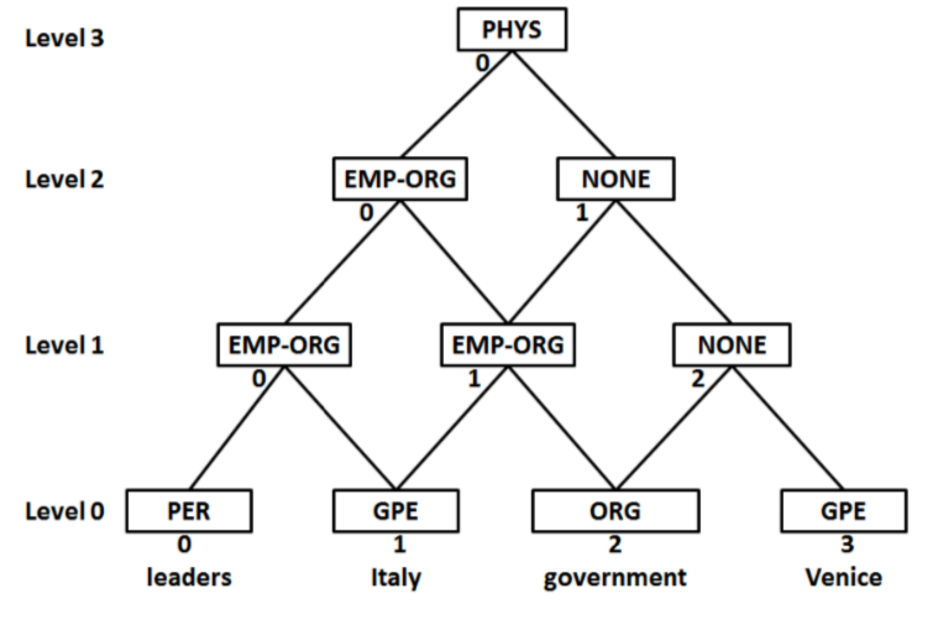


Figure 5: Card-pyramid graph for our example sentence

Semi-supervised techniques

- Advantages:
  - Reduce the manual efforts required to create labelled data
  - Exploit the unlabeled data which is generally easily available without investing much efforts
- Bootstrapping Approaches
- Active Learning
- Label Propagation Method
- Other Methods

Bootstrapping Approaches

- DIPRE (Dual Iterative Pattern Rela- tion Expansion): Pattern Relation Duality
  - Given a good set of patterns, a good set of tuples (entity pairs following a certain relation type) can be found.
  - Given a good set of tuples, a good set of patterns can be learned

Input: Seed set S of tuples, i.e. entity pairs known to be related with certain relation type R
Output: Set S grown over multiple iterations
1. Find all *occurrences* of the tuples from the seed set S on the Web
2. Learn *patterns* from these *occurrences*
3. Search the web using these *patterns* and find new tuples and add to the set S
4. Go to step 1 and iterate till there are no new tuples to be added

Table 6: Overview of DIPRE [12] algorithm

Bootstrapping Approaches - 2

- Snowball: an improvement on two points below:
    - Pattern representation and generation
    - Evaluation of patterns and tuples

```
sub GenerateTuples(Patterns)
  foreach text_segment in corpus
(1)  {<o, ℓ>, <l_s, t_1, m_s, t_2, r_s>} =
        = CreateOccurrence(text_segment);
     T_C = <o, ℓ>;
     Sim_Best = 0;
     foreach p in Patterns
(2)    sim = Match(<l_s, t_1, m_s, t_2, r_s>, p);
       if (sim ≥ τ_sim)
(3)      UpdatePatternSelectivity(p, T_C);
         if(sim ≥ Sim_Best)
           Sim_Best = sim;
           P_Best = p;
     if(Sim_Best ≥ τ_sim)
       CandidateTuples[T_C].Patterns[P_Best] =
         = Sim_Best;
  return CandidateTuples;
```

Figure 4: Algorithm for extracting new tuples using a set of patterns.

Bootstrapping Approaches - 3

- BootProject: SVM, Co-training algorithm
- Bootstrapping approaches like DIPRE and SnowBall:
  - mostly apply relation patterns when both the entities are present as ***name*** mentions
  - NOT good at extracting general relations like EMP-ORG relation in ACE 2004 dataset
  - Depends on the choice of initial ***seed*** examples

Active Learning

- To be fulfilled...

Label Propagation Method

- The label information for any node is propagated to nearby nodes through weighted edges iteratively and finally the labels of unlabeled examples are inferred when the propagation process is converged
- Each entity pair in the dataset is considered as a node in a graph, with a feature vector

Other Methods

- Multi-task transfer learning to solve a weakly-supervised RE problem
- Condition: only a few seed instances of the relation type of interest are available but a large amount of labelled instances of other relation types is also available
- Idea: different relation types can share certain common structures (e.g. EMP-ORG : employees of TCS; GPE-AFF : residents of India).
- The proposed framework uses a multi-task transfer learning method along with human guidance in the form of entity type constraints. The commonality among different relation types is modelled through a shared weight vector, enabling the knowledge learned from other relation types to be transferred to the target relation type.

Unsupervised techniques

- Advantage: do not require any labelled data
- Methods:
  - Clustering based approaches
  - Other approaches

Clustering-based approaches

- The earliest completely unsupervised RE: only require a NER tagger.
- The approach can be described in following steps:

  1. The named entities in the text corpora are tagged

  2. Co-occurring named entity pairs are formed and their contexts are recorded

  3. Context similarities among the pairs identified in the step 2, are computed

  4. Using the similarity values computed in previous step, the pairs are clustered

  5. As each of these clusters represent one relation, a label is automatically assigned to each cluster describing the relation type represented by it

Non Clustering-based approaches

- URES: Unsupervised RE System
- Input: the definitions of the relation types of interest
  - A relation type is defined as a small set of keywords, including relation type and entity types of its arguments. e.g. for the relation type *Acquisition*, the keywords can be *acquired, acquisition*
- The direct successor of KnowItAll system, which extracts entities
- URES: extract relations

Unsupervised paraphrase acquisition for RE

- *Paraphrases:* The text expressions that convey roughly the same meaning.
- The approach begins with one text expression (and corresponding syntactic structure like dependencies structure) representing the target relation and finds its paraphrases using an unsupervised paraphrase acquisition approach.
- e.g. **X interact with Y**, paraphrase acquisition algorithm would produce new expressions - **X bind to Y**, **X activate Y**, **X stimulate Y**, **interaction between X and Y**, etc.

Open Information Extraction (Open IE)

- Traditional RE: focuses on precise, pre-specified set of relations
  - require human involvement to design extraction rules and creating labelled training data
  - hard to change into different domains
- Open IE
  - Automatically discover possible relations in the text corpus without any human involvement
  - No additional efforts are required to switch to a different domain
  - TextRunner system

Self-supervised Learner

- Automatically labels a set of extracted entity tuples as positive or negative with some heuristic rules
    - Here, positive class indicates that the corresponding tuple represents some valid relation
- After automatic labelling, each tuple is mapped to a feature vector representation and a Naive Bayes classifier is trained

Single Pass Extractor

- Traverse the entire corpus and obtain POS and NP (base noun phrases) information for all sentences
- Each pair of NPs (E1 and E2) becomes a candidate tuple
- Heuristically decide whether to include words between E1 and E2 in R
- Candidate tuples are presented to Naive Bayes classifier, "positive"s are extracted and stored

Redundancy-based Assessor

- TextRunner automatically merges tuples with same entities & relations
- Number of distinct sentences containing the tuple is recorded, and count a probability of correctness of each tuple

Improvements to TextRunner

- Use O-CRF (self supervised Conditional Random Field sequence classifier) instead of Naive Bayes classifier observed better performance
- Use Wikipedia infoboxes to more accurately generate training data for the Self-supervised Learner module
- Bootstrapping methods like **Snowball**:
    - Significantly reduce the number of initial training examples
    - Do not perform OpenIE (?)
    - StatSnowball: can perform open IE along with traditional RE

Limitations of TextRunner

- Incoherent Extractions: No meaningful interpretation of extracted re- lation phrases can be made
  - This is caused by a word-by-word decision making about whether to include a word in a relation phrase.
- Uninformative Extractions: These extractions omit critical information
  - Generally caused by improper handling Light Verb Constructions (LVCs) (复合短语).
  - e.g. "John made a promise to Alice", TextRunner: (John, made, a promise), Correct: (John, made a promise to, Alice)
- Overly-specific Extractions: extract very specific relations which are not useful
  - e.g. The Obama administration, [is offering only modest greenhouse gas reduction targets at], the conference

ReVerb: an advanced OpenIE system

- This system improves over TextRunner by overcoming following limitations of TextRunner:
  - Syntactic Constraint: constraint relation phrases to match several POS tag patterns. This avoids both "Incoherent" problem and "Uninformative" problem since LVCs are also captured.

$$\boxed{\begin{gathered} V|VP|VW^*P \\ V = verb\ particle?\ adv? \\ W = (noun|adj|adv|pron|det) \\ P = (prep|particle|inf.\ marker) \end{gathered}}$$

Table 7: Syntactic Constraint

Summary for OpenIE

Recently, Open IE has been an active area of research within RE systems.

- Major advantage
  - Unsupervised nature
  - Scalability to the Web scale
- Limitation
  - Same semantic relation may be represented by multiple relation phrases

Distant Supervision

- Idea: use a large semantic database for automatically obtaining relation type labels
- Labels may be noisy, but the huge amount of training data is expected to offset this noise.

Distant Supervision - 2

- Advantages of both supervised and unsupervised:
  - It combines thousands of features using a probabilistic classifier as in the case of supervised paradigm.
  - It extracts a large number of relations from large corpora of any domain as in the case of unsupervised paradigm.
- Shortcomings:
  - Failed to model overlapping relations, i.e. the same pair of entities with multiple valid relations.
  - e.g. *FoundedBy* **(Steve Jobs, Apple)** and *CEO* **(Steve Jobs, Apple)**
  - Some solutions: Multi-instance Multi-label learning based approach (**MIML-RE**)

MIML-RE

- A novel graphical model is used to represent "multiple instances" as well as "multiple labels" of an entity pair
- Mention level relation classifiers
- Entity pair level classifiers
- Outperforms many traditional distant supervision approaches
- Major advantage: the entity pair level classifiers
- Details are to be continued...

Distant Supervision - 3

- Recently: many new approaches are actively being proposed to overcome specific problems
  - FALSE negative instances: due to the incompleteness of semantic database
  - MIML-RE's data likelihood: is a non-convex formulation
- Ontological-Smoothing (本体平滑)
- Details need to read the cited papers listed by the survey's author.

Recent Advances in Relation Extraction

- Universal Schemas
- n-ary Relation Extraction
- Cross-sentence Relation Extraction
- Convolutional Deep Neural Network
- Cross-lingual Annotation Projection
- Domain Adaptation

Reference

- [1] Pawar, Sachin, Girish K. Palshikar, and Pushpak Bhattacharyya. "Relation extraction: A survey." *arXiv preprint arXiv:1712.05191* (2017).
- [2] Kumar, Shantanu. "A survey of deep learning methods for relation extraction." *arXiv preprint arXiv:1705.03645* (2017).
- [3] Slide: A-survey-on-Relation-Extraction-Slides. CMU.
- [4] Slide: cse517wi13-RelationExtraction. Stanford.

Thanks

Yifei Li