Information Retrieval (5LN712)
Information Extraction

Ali Basirat

Department of Computer and Information Science (IDA)
Linköping University
ali.basirat@liu.se

Department of Linguistics and Philology
Uppsala University
ali.basirat@lingfil.uu.se

May 6, 2021
1 Introduction

2 Named Entity Recognition

3 Relation Extraction

4 Summary
• To extract limited kinds of semantic content from text.
• To turn unstructured information in texts to structured data
- **Named entity recognition (NER):** to detect each mention of a named entity in a text and determine its type.
- **Relation extraction:** to determine the semantic relations between the entities (e.g., child-of, employment).
- **Event extraction:** to find the events in which the entities participate.
- **Temporal extraction:** to figure out when the event in a text happened.
- **Temporal normalization:** to normalize time expression onto specific calendar dates or times of day.
- **Template filling:** to find stereotypical events or situations that match a template and fill the slots of the template.
Table of contents

1 Introduction
2 Named Entity Recognition
3 Relation Extraction
4 Summary
Definition

Anything that can be referred to with a proper name is a **named entity**.

Named entities are often extended to **temporal expressions**, and **numerical expressions**

Example

A person, a location, an organization, a gene, a protein, prices, week days, commercial products
Generic named entity types:

- People (PER): people, characters
- Organization (ORG): companies, sport teams
- Location (LOC): regions, mountains, seas
- Geo-political entity (GPE): countries, states, provinces
- Facility (FAC): bridges, buildings, airports
- Vehicles (VEH): planes, trains, automobiles
Named entity recognition

NER consists of two major steps:

- To determine a span of text that contains a named entity
- To classify the type of the named entity
Difficulties:

- The span detection can be difficult because of the ambiguity of segmentation
- The ambiguity in the type classification

Example

- [PER Washington] was born into slavery on the farm of James Burroughs.
- [ORG Washington] went up 2 games to 1 in the four-game series.
- Blair arrived in [LOC Washington] for what may well be his last state visit.
- In June, [GPE Washington] passed a primary seatbelt law.
- The [VEH Washington] had proved to be a leaky ship, every passage I made...
**Named entity recognition**

NER as a sequence labeling problem

- To label the tokens in a text with tags that indicate the presence of a particular type of named entity
- **IOB** tagging scheme: to tag the **B**eginning, **I**nside, and **O**utside of a text span

<table>
<thead>
<tr>
<th>Words</th>
<th>IOB Label</th>
<th>IO Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Airlines</td>
<td>B-ORG</td>
<td>I-ORG</td>
</tr>
<tr>
<td>,</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>unit of AMR Corp.</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>, immediately matched the move</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>, spokesman Tim Wagner said</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>.</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Named entity recognition

NER evaluation:

- Precision: the ratio of the number of correctly labeled responses to the total labeled

- Recall: the ratio of the number of correctly labeled responses to the total that should have been labeled

- F1-measure: the harmonic mean of the precision and recall
Three approaches of NER:

- Feature-based NER
- Neural NER
- Rule-based NER
Feature-based NER:

- Train a sequence classifier (e.g., Maximum Entropy Markov Model (MEMM), or Conditional Random Field (CRF)) on an annotated corpus with predefined features
- This approach relies on word-shape features
- One can also consider the syntactic features of the words such as the part-of-speech tags, chunk tags, and morphological features
- It is also quite common to use a list of place names (gazetteer), a list of names (name-lists)
- The performance of this approach depends on the application, genre, media, and language.
Named entity recognition

Figure: Named entity recognition as sequence labeling.
Neural NER

- Word and character embeddings are fed to a left-to-right and a right-to-left LSTM
- The outputs of LSTMs are concatenated and fed to a CRF layer
Rule-based NER

1. First, use high-precision rules to tag unambiguous entity mentions.
2. Then, search for substring matches of the previously detected names.
3. Consult application-specific name lists to identify likely name entity mentions from the given domain.
4. Finally, apply probabilistic sequence labeling techniques that make use of the tags from previous stages as additional features.
Table of contents

1 Introduction

2 Named Entity Recognition

3 Relation Extraction

4 Summary
Finding the relations among the detected entities

Citing high fuel prices, [ORG United Airlines] said [TIME Friday] it has increased fares by [MONEY $6] per round trip on flights to some cities also served by lower-cost carriers. [ORG American Airlines], a unit of [ORG AMR Corp.], immediately matched the move, spokesman [PER Tim Wagner] said. [ORG United], a unit of [ORG UAL Corp.], said the increase took effect [TIME Thursday] and applies to most routes where it competes against discount carriers, such as [LOC Chicago] to [LOC Dallas] and [LOC Denver] to [LOC San Francisco].

We see binary relations such as:

- Tim Wagner is a spokesman for American Airlines
- United is a unit of UAL Corp
- American is a unit of AMR
Generic relations in news-style texts:

**Figure:** The 17 relations used in the ACE relation extraction task.
The sets of relations are domain specific

Example

- The Unified Medical Language System has 54 relations between medical-related entities
- Wikipedia offers a large supply of relations (see DBpedia and https://dl.acm.org/doi/10.1145/2629489)
- WordNet and other anthologies show hierarchical relations between words or concepts
- The TACRED dataset contains relations about people or organizations
- https://www.aclweb.org/anthology/S10-1006/ provides a standard benchmark to test semantic relation classification between nominals
- New York Times dataset
Relation extraction

Five approaches of relation extraction:

- Handwritten patterns
- Supervised
- Semi-supervised - bootstrapping
- Semi-supervised - distant supervision
- Unsupervised
Handwritten patterns:

- The relations are extracted based on predefined lexico-syntactic patterns

Example

A lexico-syntactic pattern like NP0 such as NP1, NP2..., (and|or)NPi, \(i \geq 1\) implies
\[ \forall NP_i, i \geq 1, \text{hyponym}(NP_i, NP0) \]
Hearst is a pioneer in using syntactic patterns to relation extraction
Later on, name-entity constraints are added to the Hearst’s approach
• The handwritten patterns can result in high precision outputs
• However, they often result in low recall
• Depending on the task and domain, it might be very hard to come up with a set of patterns that cover all relations
Supervised learning:

- A fixed set of relations and entities is chosen.
- A training corpus is hand-annotated with the relations and entities.
- The annotated texts are used to train classifiers to annotate an unseen test set.
Relation extraction

At test time:

- Find all pairs of named entities in a sentence
- Apply a filtering classifier to see if each pair of entities can be related to each other or not
- Pass the positive pairs to another classifier to determine the type of the relation

```
function FINDRELATIONS(words) returns relations

relations ← nil
entities ← FINDENTITIES(words)
forall entity pairs \( \langle e_1, e_2 \rangle \) in entities do
  if RELATED?(e_1, e_2)
    relations ← relations + ClassifyRelation(e_1, e_2)
```
Supervised approaches

- Require labeled data - which is expensive
- Two major approaches: feature-based and neural methods
- Can be very accurate if enough training data is provided
- Does not generalize well to unseen data
Feature-based supervised relation classifiers

- word features: embeddings, ngrams, etc.
- Named entity features: entity types, number of entities in between, etc.
- syntactic structure: dependency features, dependency path, constituents path
Neural supervised relation

- Neural networks are used for feature extractions
- An encoder provides a global representation of the sentence semantics
- Attention mechanism is used for context representation
- MLP is used for final relation extraction
Neural supervised relation classifiers

- Zhang et. al. (2017):
  - LSTM is used for global semantic representation of a sentence.
  - Positional information relative to subject and object are used to find attention weights
  - An MLP followed by a softmax layer predicts the relation holds between the subject and object

Figure: Zhang et. al. (2017)’s proposed position-aware neural sequence model.
Neural supervised relation classifiers

- Transformer-based models
  - A transformer like Bert encodes a sentence into a vector
  - Named-entities in the input are often marked
  - A linear layer is used for the relation classification
  - The input to these methods can be a long sequence of sentences.
Semi-supervised methods via bootstrapping

- Labelled data is expensive
- Bootstrapping labelled data from a small labelled data set
Semi-supervised methods via bootstrapping

- Starts with a set of seed tuples
- Find other mentions of the seed tuples in a corpus
- Learn new patterns through generalizing the contexts around the entities
- Find more tuples based on the extracted patterns
Semi-supervised methods via bootstrapping

```
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 tuples
    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
```

**Figure:** Bootstrapping from seed entity pairs to learn relations. Jurafsky et. al. (2019)
Semi-supervised methods via bootstrapping

- The new tuples should be evaluated through a confidence value
- Erroneous tuples lead to the introduction of new noisy tuples (semantic drift)
Semi-supervised methods via bootstrapping

• A confidence value should balance two factors:
  • hits: the overlap between the new tuples and the current tuples
  • finds: the total set of new tuples

• The confidence value for the newly proposed set of tuples \( p \) is:

\[
\text{Conf}(p) = \frac{|\text{hits}(p)|}{|\text{finds}(p)|} \log |\text{finds}(p)|
\]
Distant Supervision for RE

- Combines the advantages of bootstrapping with supervised learning
- A large database (e.g., DBPedia) is used to acquire huge number of seed examples
- Lots of noisy pattern features are created from these examples
- The data is combined in a supervised classifier
Distant Supervision for RE

\[
\text{function } \text{DISTANT SUPERVISION}(Database D, Text T) \text{ returns relation classifier } C
\]

\[
\text{foreach relation } R
\]

\[
\text{foreach tuple } (e1, e2) \text{ of entities with relation } R \text{ in } D
\]

\[
sentences \leftarrow \text{Sentences in } T \text{ that contain } e1 \text{ and } e2
\]

\[
f \leftarrow \text{Frequent features in } sentences
\]

\[
observations \leftarrow \text{observations + new training tuple } (e1, e2, f, R)
\]

\[
C \leftarrow \text{Train supervised classifier on } observations
\]

\[
\text{return } C
\]

**Figure:** Distant Supervision for RE (2019)
Distant Supervision for RE Shared advantages with other methods:

- Like **patterns-based** approaches, it learns patterns
- Like **supervised** methods, it uses a classifier
- Like **semi-supervised** methods, it relies on seed tuples
- Like **unsupervised** methods, it does not use labelled training corpus of text
Unsupervised Relation Extraction

- **Open IE**: no labelled data, no relation set.
- A set of candidate relations are extracted based on some *linguistic heuristics*
- The relation tuples with high *confidence values* are then selected
Unsupervised Relation Extraction

- Advantage: it handle a huge number of relations without having to specify them in advance
- Disadvantage: it needs to map large sets of strings into some canonical form for adding to databases or other knowledge sources.
Unsupervised Relation Extraction

Example

- **TEXTRUNNER** (Banko et al., 2007)
- **ReVerb** (Fader et al. 2011)
- **OLLIE** (Mausam et al. 2012)
- **ClausIE** (Del Corro and Gemulla, 2013)
- **Stanford OpenIE** (Angeli et al. 2015)
- **Neural OpenIE** (Cui et al. 2018)
- **OpenIE6** (Kolluru et al. 2020)
Unsupervised Relation Extraction

ReVerb

- **Input**: a POS-tagged and NP chunked sentence
- **Output**: a set of \((x, r, y)\) extraction triples
- **Algorithm**:
  - For each verb find the longest word sequence that satisfies syntactic and lexical constraints
  - For each relation phrase \(r\), find the nearest noun phrase \(x\) to the left of \(r\) such that \(x\) is not a relative pronoun, WH-adverb, or existential 'there'. Find the nearest noun phrase \(y\) to the right of \(r\)
  - Assign confidence \(c\) to the relation \(r = (x, r, y)\)
Unsupervised Relation Extraction
ReVerb (Cont.)

- syntactic constraints:
  - Purposes: 1) to eliminate incoherent extractions, and 2) to reduce uninformative extractions
  - A relation phrase must be a contiguous sequence of words that begins with a verb and end with a preposition may include nouns, adjectives, and adverbs in between
  - If multiple matches found in a sentence, then longest one is chosen
  - Example: Faust made a deal with the Devil

- lexical constraints:
  - The syntactic constraint may match very specific relation phrases with only few possible instances
  - Example: The Obama administration is offering only modest greenhouse gas reduction targets at the conference.
  - prune very rare long relation strings
Unsupervised Relation Extraction

Example

- Input sentence: United has a hub in Chicago, which is the headquarters of United Continental Holdings.
- Candidate relation phrases:
  - United <has a hub in> Chicago, which is the headquarters of United Continental Holdings.
  - United has a hub in Chicago, which <is the headquarters of> United Continental Holdings.
- Find the nearest noun phrase on the left and right:
  - United <has a hub in> Chicago, which is the headquarters of United Continental Holdings.
  - United has a hub in Chicago, which <is the headquarters of> United Continental Holdings.
Unsupervised Relation Extraction

ReVerb (Cont.)

- **Confidence function:**
  - Purposes: to trade recall for precision by tuning a confidence threshold
  - A logistic regression classifier assigns a confidence value to each relation tuple
  - The classifier is trained on 1000 manually annotated sentences

- **Annotation**
  - First, the relation tuples are extracted from the sentences
  - Then, the extracted relations are manually labelled
Evaluation of Relation Extraction

- **Supervised**: Precision, recall, and F-score for both labelled and unlabelled settings with regard to a **gold standard**
- **Semi-supervised**:
  - No gold standard (totally new relations are extracted)
  - No way to directly evaluate recall
  - Only the precision can be approximated on a random sample of outputs checked by a human
  - We are interested in the **set** of correct tuples rather than the **mentions** of relations

\[
\hat{p} = \frac{\text{# of correctly extracted relation tuples in the sample}}{\text{total # of extracted relation tuples in the sample}}
\]

- Precision at different levels of recall can be estimated based on the top \( k \) new relations with high probability or confidence value
<table>
<thead>
<tr>
<th></th>
<th>Table of contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Named Entity Recognition</td>
</tr>
<tr>
<td>3</td>
<td>Relation Extraction</td>
</tr>
<tr>
<td>4</td>
<td>Summary</td>
</tr>
</tbody>
</table>
• Information Extraction
• Named Entity Recognition
• Feature-based/ Neural/ Rule-based NER
• Relation Extraction
• Handwritten patterns/ Supervised/ Semi-supervised/ Unsupervised
• Semi-supervised via bootstrapping and distant supervision
• Evaluation