IE through Distant Supervision.

Hypothesis:

If two entities maintain a certain relationship then any sentence containing these two entities is highly likely to express that relationship.

Key idea:

Use data to get lots of training examples.

Unlike
→ hand-crafting a few seed examples (bootstrapping)

Unlike
→ using hand labeled corpus (supervised setup).

Approach:

For each pair of entities obtained from a large database

- Grab sentences containing these entities from a corpus.
- Extract a lot of noisy features from these sentences.
  - lexical, syntactic ...
- Combine these to train a classifier.

Advantages over supervised frameworks.

— leverage existing reliable hand crafted knowledge.

— relation have canonical names.
{"located in", "situated in", "found in"}

— rich features.

Advantages on ~~unsupe~~ unsupervised methods.

— leverage unlimited amount of data

— many weak features.

— not sensitive to the training corpus → learn is domain independent.

Hypernyms for DB.

Construct a noisy training set containing occurrences from a corpus that correspond to a ~~hypernym~~ hyponym-hypernym pair.

from WordNet

DB
WordNet

Shakespear – author

"... consider authors like Shakespeare ..."
"Some authors (including Shakespeare) ..."
"Shakespeare was the author of several ..."
"Shakespeare, author of The Tempest ..."

Noisy examples:

"... authors in the Shakespeare festival ..."
"The author of Shakespeare in Love is ..."

Learning hypernym patterns.

① Take a corpus of sentences.
— "---- doubly heavy hydrogen atom called deuterium ----"

② Collect noun pairs:
(atom, deuterium).
Count the no. of such pairs in a large corpus (6M newswire sentences).

752311 pairs.

③ Taking pairs in each sentence adjudge if the pair is IS-A relationship as per the WordNet.

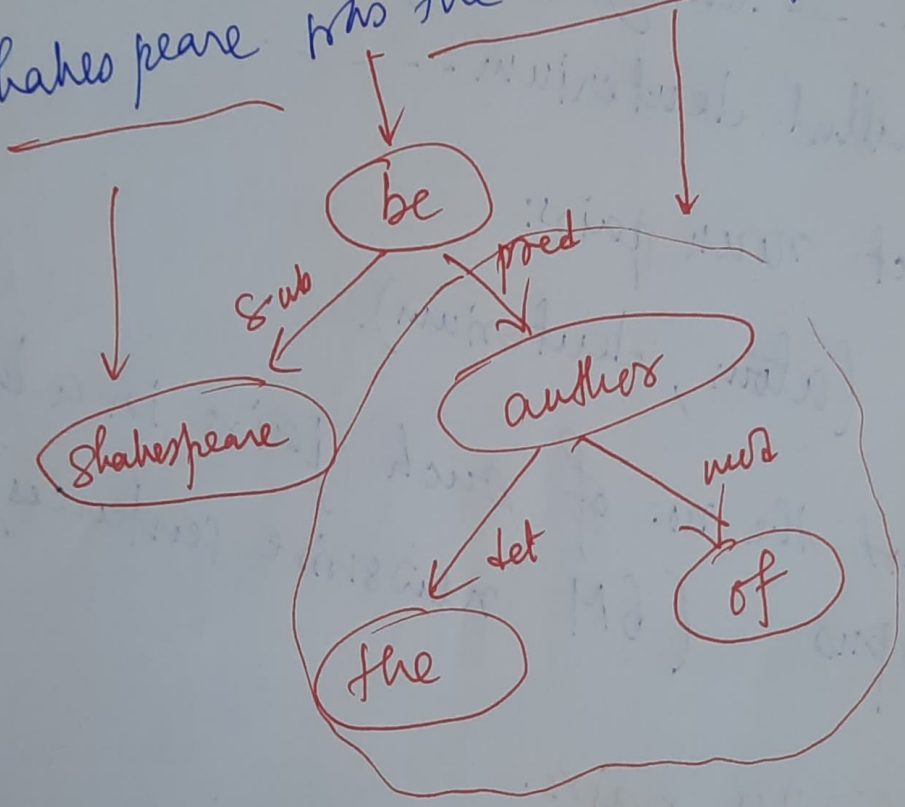
Yes — 14,387
No — 737924.

④ Parse the sentences — extract features — train classifier
↳ build patterns.

Syntactic dependency path between the noun pair in the sentence.

noun pair: (Shakespeare, author).

"Shakespeare was the author of several plays..."



Path:

N:sub:VBE, be, VBE:pred:N.

Noun subject of VBE(be) → Shakespeare

Noun predicate of the VBE(be) → author

N be N → NP be NP.

Satellite terms are added.

Such NP as NP.

X and/or other Y
Y such as X
such X as Y
Y including X
Y, especially X
Y like X
Y called X
X is a Y
X, a Y (appositive).

Coverage??

Mintz, Bills, Snow, Jurafsky (2009).

Training set

Freebase.

102 relations.

940,000 entities

1.8 million instances.

+

Running corpus-

Wikipedia.

Corpus text

Bill Gates founded Microsoft in 1975

Bill Gates, founder of Microsoft

Bill Gates attended Harvard from

Google was ~~found~~ founded by Larry Page

Free base:

Founder: (Bill Gates, Microsoft)

Founder: (Larry Page, Google)

CollegeAttended: (Bill Gates, Harvard).

Training data

(Bill Gates, Microsoft)

Label: Founder.

Feature: X founded Y.

Feature: X, founder of Y

Feature: Y founded by X.

(Larry Page, Google)

(Bill Gates, Harvard)

Label: CollegeAttended.

Feature: X attended Y.

Caveat of this process?

Corpus kept

Larry Page took a swipe at Microsoft. ... after Harvard invited Larry Page to .... Google is Bill Gates worst fear ....

Negative samples?

→ Sample 1% of unrelated pairs.

(Larry Page, Microsoft)

Label : NO-RELATION

Feature: (X took a swipe at Y.) → Vector.

→ lexical

→ Syntactic features.

POS patterns.

dependency patterns.

(Larry Page, Harvard)

Label : NO_RELATION

features : Y invited X

(Bill Gates, Google)

Label : NO_RELATION

feature: Y is X's worst fear

(POS, dependency, other lexical features.)

V_1 → FOUNDER

V_2 → COLLEGEATTENDED

V_3 — NO-RELATION

⋮