Learning Semantic Relations from Text

Preslav Nakov[1], Diarmuid Ó Séaghdha[2], Vivi Nastase[3], Stan Szpakowicz[4]

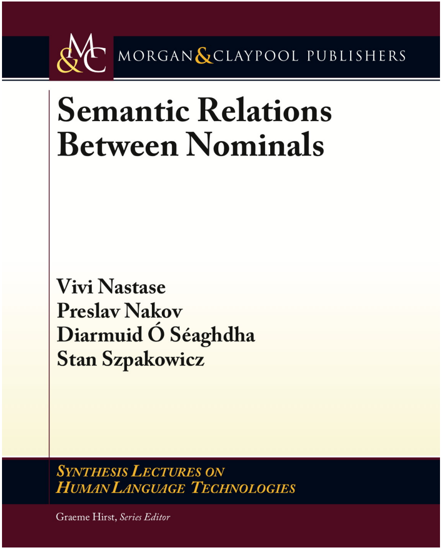
[1] Qatar Computing Research Institute, HBKU

[2] Vocal IQ

[3] Fondazione Bruno Kessler

[4] University of Ottawa

EMNLP 2015
CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING
LISBON



Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. Embeddings
7. Wrap-up

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. Embeddings
7. Wrap-up

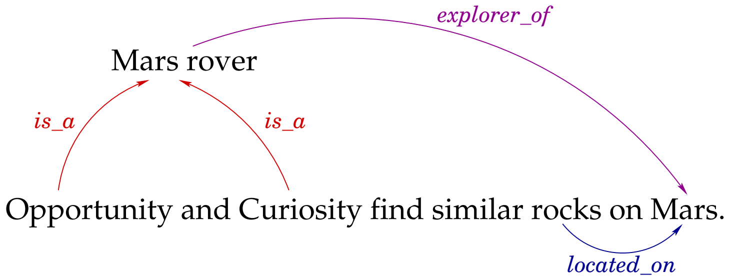
Motivation

The connection is indispensable to the expression of thought. Without the connection, we would not be able to express any continuous thought, and we could only list a succession of images and ideas isolated from each other and without any link between them.
[Tesnière, 1959]

What Is It All About?

Opportunity and Curiosity find similar rocks on Mars.

What Is It All About?



What Is It All About? (1)

Semantic relations

- matter a lot
- connect up entities in a text
- together with entities make up a good chunk of the meaning of that text
- are not terribly hard to recognize

What Is It All About? (2)

Semantic relations between nominals

- matter even more in practice
- are the target for knowledge acquisition
- are key to reaching the meaning of a text
- their recognition is fairly feasible

Historical Overview (1)

Capturing and describing world knowledge

- Artistotle's *Organon*
  - includes a treatise on *Categories*
    - objects in the natural world are put into categories called τὰ λεγόμενα (*ta legomena*, things which are said)
    - organization based on the class inclusion relation
- then, for 20 centuries:
  - other philosophers
  - some botanists, zoologists
- in the 1970s: realization that a robust Artificial Intelligence (AI) system needs the same kind of knowledge
  - capture and represent knowledge: machine-friendly
  - intersection with language: inevitable

Historical Overview (2)

Indian linguistic tradition

- Pāṇini's Aṣṭādhyāyī
  - rules describing the process of generating a Sanskrit sentence from a semantic representation
  - semantics is conceptualized in terms of *kārakas*, semantic relations between events and participants, similar to semantic roles
  - covers noun-noun compounds comprehensively from the perspective of word formation, but not semantics
  - later, commentators such as Kātyāyana and Patañjali: compounding is only supported by the presence of a semantic relation between entities

Historical Overview (3)

Ferdinand de Saussure

- *Course in General Linguistics* [de Saussure, 1959]
  - taught 1906-1911; published in 1916

Historical Overview (4)

Ferdinand de Saussure

- *Course in General Linguistics*: two types of relations which "correspond to two different forms of mental activity, both indispensable to the workings of language"
  - syntagmatic relations
    - hold in context
  - associative (paradigmatic) relations
    - come from accumulated experience
- BUT no explicit list of relations was proposed

Historical Overview (5)

Ferdinand de Saussure

- *Syntagmatic* relations hold between two or more terms in a sequence *in praesentia*, in a particular context: “words as used in discourse, strung together one after the other, enter into relations based on the linear character of languages – words must be arranged consecutively in a spoken sequence. Combinations based on sequentiality may be called syntagmas.”

Historical Overview (6)

Ferdinand de Saussure

- *Associative (paradigmatic)* relations come from accumulated experience and hold *in absentia*: "Outside the context of discourse, words having something in common are associated together in the memory. [. . .] All these words have something or other linking them. This kind of connection is not based on linear sequence. It is a connection in the brain. Such connections are part of that accumulated store which is the form the language takes in an individual's brain."

Historical Overview (7)

Syntagmatic vs. paradigmatic) relations

- Harris [1987]: frequently occurring instances of syntagmatic relations may become part of our memory, thus becoming paradigmatic
- Gardin [1965]: instances of paradigmatic relations are derived from accumulated syntagmatic data
- This reflects current thinking on relation extraction from open texts.

Historical Overview (8)

Predicate logic [Frege, 1879]

- inherently relational formalism
- *e.g.*, the sentence "Google buys YouTube." is represented as

$$\textbf{\textit{buy}}(\texttt{Google}, \texttt{YouTube})$$

Historical Overview (9)

Neo-Davidsonian logic representation

- additional variables represent the event or relation
- it can thus be explicitly modified and subject to quantification

$\exists e$ ***InstanceOfBuying***$(e) \wedge$ ***agent***$(e,$ `Google`$) \wedge$ ***patient***$(e,$ `YouTube`$)$

or perhaps

$\exists e$ ***InstanceOf***$(e, Buying) \wedge$ ***agent***$(e,$ `Google`$) \wedge$ ***patient***$(e,$ `YouTube`$)$

- existential graphs [Peirce, 1909]

Historical Overview (10)

The dual nature of semantic relations

- in logic: predicates
  - used in AI to support knowledge-based agents and inference
- in graphs: arcs connecting concepts
  - used in NLP to represent factual knowledge
  - thus, mostly binary relations
    - in ontologies
    - as the target in IE
    - ...

Historical Overview (11)

The rise of reasoning systems

- (McCarthy, 1958): logic-based reasoning, no language
- early NLP systems with semantic knowledge
  - (Winograd, 1972): interactive English dialogue system
  - (Charniak, 1972): understanding children's stories
  - conceptual shift from the "shallow" architecture of primitive conversation systems such as ELIZA [Weizenbaum, 1966]
- large-scale hand-crafted ontologies
  - Cyc
  - OpenMind Common Sense
  - MindPixel
  - FreeBase – truly large-scale

Historical Overview (12)

At the cross-roads between knowledge and language

- Spärck-Jones [1964]: lexical relations found in a dictionary can be learned automatically from text
- Quillian [1962]: *semantic network*
  - a graph in which meaning is modelled by labelled associations between words
    - vertices are concepts onto which words in a text are mapped
    - connections – relations between such concepts
- WordNet [Fellbaum, 1998]
  - 155,000 words (nouns, verbs, adjectives, adverbs)
  - a dozen semantic relations, *e.g.,* synonymy, antonymy, hypernymy, meronymy

Historical Overview (13)

Automating knowledge acquisition

- learning ontological relations
  - ***is-a*** [Hearst, 1992]
  - ***part-of*** [Berland & Charniak, 1999]
- bootstrapping [Patwardhan & Riloff, 2007; Ravichandran & Hovy, 2002]
- open relation extraction
  - no pre-specified list/type of relations
  - learn patterns about how relations are expressed, *e.g.,*
    - POS [Fader et al., 2011]
    - paths in a syntactic tree [Ciaramita et al., 2005]
    - sequences of high-frequency words [Davidov & Rappoport, 2008]
  - hard to map to “canonical” relations

Why Should We Care about Semantic Relations?

Relation learning/extraction can help

- building knowledge repositories
- text analysis
- NLP applications
    - Information Extraction
    - Information Retrieval
    - Text Summarization
    - Machine Translation
    - Question Answering
    - Paraphrasing
    - Recognizing Textual Entailment
    - Thesaurus Construction
    - Semantic Network Construction
    - Word-Sense Disambiguation
    - Language Modelling

Example Application: Information Retrieval

[Cafarella et al., 2006]

- list all *X* such that *X* **causes** *cancer*
- list all *X* such that *X* **is part of** an *automobile engine*
- list all *X* such that *X* **is material for** making a *submarine's hull*
- list all *X* such that *X* **is a type of** *transportation*
- list all *X* such that *X* **is produced from** *cork trees*

Example Application: Statistical Machine Translation

[Nakov, 2008]

- if the SMT system knows that
  - *oil price hikes* is translated to Spanish as
  - *alzas en los precios del petróleo*
    - note: this is hard to get word-for-word!
- if we further interpret/paraphrase *oil price hikes* as
  - *hikes in oil prices*
  - *hikes in the prices of oil*
  - ...
- then we can use the same fluent Spanish translation for the paraphrases

Outline

1 Introduction

2 Semantic Relations

3 Features

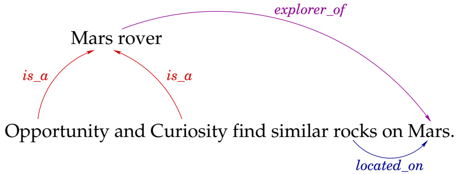
4 Supervised Methods

5 Unsupervised Methods

6 Embeddings

7 Wrap-up

Two Perspectives on Semantic Relations



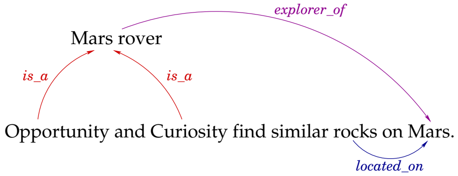
Relations between concepts

. . . arise from, and capture, knowledge about the world

Relations between nominals

. . . arise from, and capture, particular events/situations expressed in texts

Two Perspectives on Semantic Relations



Relations between concepts

... arise from, and capture, knowledge about the world

... can be found in texts!

Relations between nominals

... arise from, and capture, particular events/situations expressed in texts

... can be found using information from knowledge bases

[Casagrande & Hale, 1967]

Asked speakers of an exotic language to give definitions for a given list of words, then extracted 13 relations from these definitions.

Relation	**Example**
attributive	toad - small
function	ear - hearing
operational	shirt - wear
exemplification	circular - wheel
synonymy	thousand - ten hundred
provenience	milk - cow
circularity	X is defined as X
contingency	lightning - rain
spatial	tongue - mouth
comparison	wolf - coyote
class inclusion	bee - insect
antonymy	low - high
grading	Monday - Sunday

[Chaffin & Hermann, 1984]

Asked humans to group instances of 31 semantic relations.
Found five coarser classes.

Relation	Example
constrasts	night - day
similars	car - auto
class inclusion	vehicle - car
part-whole	airplane - wing
case relations – ***agent***, ***instrument***	

Semantic Relations in Noun Compounds (1)

Noun compounds (NCs)

- Definition: sequences of two or more nouns that function as a single noun, *e.g.,*
  - *silkworm*
  - *olive oil*
  - *healthcare reform*
  - *plastic water bottle*
  - *colon cancer tumor suppressor protein*

Semantic Relations in Noun Compounds (2)

Properties of noun compounds

- Encode implicit relations: hard to interpret
  - *taxi driver* is ‘a driver who drives a taxi’
  - *embassy driver* is ‘a driver who is employed by/drives for an embassy’
  - *embassy building* is ‘a building which houses, or belongs to, an embassy’
- Abundant: cannot be ignored
  - cover 4% of the tokens in the Reuters corpus
- Highly productive: cannot be listed in a dictionary
  - 60% of the NCs in BNC occur just once

Semantic Relations in Noun Compounds (3)

Noun compounds as a microcosm: representation issues reflect those for general semantic relations

- voluminous literature on their semantics

`www.cl.cam.ac.uk/~do242/Resources/compound_bibliography.html`

- two complementary perspectives
  - linguistic: find the most comprehensive explanatory representation
  - NLP: select the most useful representation for a particular application
    - computationally tractable
    - giving informative output to downstream systems

Semantic Relations in Noun Compounds (4)

Do the relations in noun compounds come from a small closed inventory?

In other words, is there a (reasonably) small set of relations which could cover completely what occurs in texts in the vicinity of (simple) noun phrases?

- affirmative: most linguists
  - early descriptive work [Grimm, 1826; Jespersen, 1942; Noreen, 1904]
  - generative linguistics [Levi, 1978; Li, 1971; Warren, 1978]
- negative: some linguists *e.g.*, [Downing, 1977]

[Warren, 1978] (1)

Relations arising from a comprehensive study of the Brown corpus:

- a four-level **hierarchy** of relations
- six major semantic relations

Relation	**Example**
Possession	family estate
Location	water polo
Purpose	water bucket
Activity-Actor	crime syndicate
Resemblance	cherry bomb
Constitute	clay bird

[Warren, 1978] (2)

A four-level **hierarchy** of relations

L1: ***Constitute***

- L2: ***Source-Result***
- L2: ***Result-Source***
- L2: ***Copula***
  - L3: ***Adjective-Like_Modifier***
  - L3: ***Subsumptive***
  - L3: ***Attributive***
    - L4: ***Animate_Head*** (*e.g., girl friend*)
    - L4: ***Inanimate_Head*** (*e.g., house boat*)

[Levi, 1978] (1)

Relations (Recoverable Deletable Predicates) which underlie all compositional non-nominalized compounds in English

RDP	Example	Role	Traditional name
$\texttt{CAUSE}_1$	tear gas	object	causative
$\texttt{CAUSE}_2$	drug deaths	subject	causative
$\texttt{HAVE}_1$	apple cake	object	possessive/dative
$\texttt{HAVE}_2$	lemon peel	subject	possessive/dative
$\texttt{MAKE}_1$	silkworm	object	productive/composit.
$\texttt{MAKE}_2$	snowball	subject	productive/composit.
USE	steam iron	object	instrumental
BE	soldier ant	object	essive/appositional
IN	field mouse	object	locative
FOR	horse doctor	object	purposive/benefactive
FROM	olive oil	object	source/ablative
ABOUT	price war	object	topic

[Levi, 1978] (2)

Nominalizations

	Subjective	**Objective**	**Multi-modifier**
Act	*parental refusal*	*dream analysis*	*city land acquisition*
Product	*clerical errors*	*musical critique*	*student course ratings*
Agent	—	*city planner*	—
Patient	*student inventions*	—	—

Problem: spurious ambiguity

- *horse doctor* is ***for*** (RDP)
- *horse healer* is ***agent*** (nominalization)

[Vanderwende, 1994]

Relation	Question	Example
Subject	*Who/what?*	*press report*
Object	*Whom/what?*	*accident report*
Locative	*Where?*	*field mouse*
Time	*When?*	*night attack*
Possessive	*Whose?*	*family estate*
Whole-Part	*What is it part of?*	*duck foot*
Part-Whole	*What are its parts?*	*daisy chain*
Equative	*What kind of?*	*flounder fish*
Instrument	*How?*	*paraffin cooker*
Purpose	*What for?*	*bird sanctuary*
Material	*Made of what?*	*alligator shoe*
Causes	*What does it cause?*	*disease germ*
Caused-by	*What causes it?*	*drug death*

Desiderata for Building a Relation Inventory

1. the inventory should have **good coverage**
2. relations should be **disjoint**, and should each describe a **coherent concept**
3. the **class distribution** should not be overly skewed or sparse
4. the concepts underlying the relations should generalize to other **linguistic phenomena**
5. the guidelines should make the **annotation process** as **simple** as possible
6. the categories should provide **useful semantic information**

[adapted from (Ó Séaghdha, 2007)]

[Ó Séaghdha, 2007]

- ***BE*** (`identity`, `substance-form`, `similarity`)
- ***HAVE*** (`possession`, `condition-experiencer`, `property-object`, `part-whole`, `group-member`)
- ***IN*** (`spatially located object`, `spatially located event`, `temporarily located object`, `temporarily located event`)
- ***ACTOR*** (`participant-event`, `participant-participant`)
- ***INST*** (`participant-event`, `participant-participant`)
- ***ABOUT*** (`topic-object`, `topic-collection`, `focus-mental activity`, `commodity-charge`)

e.g., tax law is `topic-object`, *crime investigation* is `focus-mental activity`, and they both are also ***ABOUT***.

[Barker & Szpakowicz, 1998]

An inventory of 20 semantic relations.

Relation	Example	Relation	Example
Agent	*student protest*	***Possessor***	*company car*
Beneficiary	*student price*	***Product***	*automobile factory*
Cause	*exam anxiety*	***Property***	*blue car*
Container	*printer tray*	***Purpose***	*concert hall*
Content	*paper tray*	***Result***	*cold virus*
Destination	*game bus*	***Source***	*north wind*
Equative	*player coach*	***Time***	*morning class*
Instrument	*laser printer*	***Topic***	*safety standard*
Located	*home town*		
Location	*lab printer*		
Material	*water vapor*		
Object	*horse doctor*		

[Nastase & Szpakowicz, 2003]

A two-level hierarchy of 31 semantic relations

Causal (4 relations)
cause: flu virus,
effect: exam anxiety, . . .

Participant (12 relations)
Agent: student protest,
Instrument: laser printer, . . .

Quality (8 relations)
Manner: stylish writing,
Measure: expensive book, . . .

Spatial (4 relations)
Direction: outgoing mail,
Location: home town, . . .

Temporal (3 relations)
Frequency: daily experience,
Time_at: morning exercise, . . .

[Girju, 2005]

A list of 21 noun compound semantic relations: a subset of the 35 general semantic relations of Moldovan & al. (2004).

Relation	Example	Relation	Example
Possession	*family estate*	***Manner***	*style performance*
Attribute-Holder	*quality sound*	***Means***	*bus service*
Agent	*crew investigation*	***Experiencer***	*disease victim*
Temporal	*night flight*	***Recipient***	*worker fatalities*
Depiction-Depicted	*image team*	***Measure***	*session day*
Part-Whole	*girl mouth*	***Theme***	*car salesman*
Is-a	*Dallas city*	***Result***	*combustion gas*
Cause	*malaria mosquito*		
Make/Produce	*shoe factory*		
Instrument	*pump drainage*		
Location/Space	*Texas university*		
Purpose	*migraine drug*		
Source	*olive oil*		
Topic	*art museum*		

[Tratz & Hovy, 2010]

- Tratz and Hovy [2010]
  - new inventory
  - 43 relations in 10 categories
  - developed through an iterative crowd-sourcing
  - maximize agreement between annotators
- Analysis: all previous inventories have commonalities
  - *e.g.*, have categories for locative, possessive, purpose, etc.
  - cover essentially the same semantic space
    - BUT differ in the exact way of partitioning that space

[Rosario, 2001]: Biomedical Relations (1)

18 biomedical noun compound relations (initially 38).

Relation	Example
Subtype	*headaches migraine*
Activity/Physical_process	*virus reproduction*
Produce_genetically	*polyomavirus genome*
Cause	*heat shock*
Characteristic	*drug toxicity*
Defect	*hormone deficiency*
Person_Afflicted	*AIDS patient*
Attribute_of_Clinical_Study	*headache parameter*
Procedure	*genotype diagnosis*
Frequency/time_of	*influenza season*
Measure_of	*relief rate*
Instrument	*laser irradiation*
...	...

[Rosario, 2001]: Biomedical Relations (2)

18 biomedical noun compound relations (initially 38).

Relation	Example
...	...
Object	*bowel transplantation*
Purpose	*headache drugs*
Topic	*headache questionnaire*
Location	*brain artery*
Material	*aloe gel*
Defect_in_location	*lung abscess*

The Opposite View: No Small Set of Semantic Relations

Much opposition to the previous work

- (Zimmer, 1971): so much variety of relations that it is simpler to categorize the semantic relations that CANNOT be encoded in compounds
- (Downing, 1977)
  - *plate length* ("what your hair is when it drags in your food")
  - "The existence of numerous novel compounds like these guarantees the futility of any attempt to enumerate an absolute and finite class of compounding relationships."

Noun Compounds: Using Lexical Paraphrases (1)

Lexical items instead of abstract relations

The hidden relation in a noun compound can be made explicit in a paraphrase.

- *e.g., weather report*
  - abstract
    - ***topic***
  - lexical
    - *report about the weather*
    - *report forecasting the weather*

Noun Compounds: Using Lexical Paraphrases (2)

Using prepositions: the idea

- (Lauer, 1995) used just eight prepositions
  - of, for, in, at, on, from, with, about
    - *olive oil* is "oil from olives"
    - *night flight* is "flight at night"
    - *odor spray* is "spray for odors"
  - easy to extract from text or the Web [Lapata & Keller, 2004]
- (Srikumar & Roth, 2013) 32 relations / 34 prepositions
  - *good at boxing* → ***activity***
  - *opened by Annie* → ***agent***
  - *travel by road* → ***journey***
  - ...

Noun Compounds: Using Lexical Paraphrases (3)

Using prepositions: the issues

- prepositions are polysemous, *e.g.*, different ***of***
  - *school of music*
  - *theory of computation*
  - *bell of (the) church*
- unnecessary distinctions, *e.g.*, ***in*** vs. ***on*** vs. ***at***
  - *prayer in (the) morning*
  - *prayer at night*
  - *prayer on (a) feast day*
- some compounds cannot be paraphrased with prepositions
  - *woman driver*
- strange paraphrases
  - *honey bee* – is it "bee for honey"?

Noun Compounds: Using Lexical Paraphrases (4)

Using paraphrasing verbs

- (Nakov, 2008): a relation is represented as a distribution over verbs and prepositions which occur in texts
  - *e.g., olive oil* is “oil that is extracted from olives” or “oil that is squeezed from olives”
  - rich representation, close to what Downing [1977] demanded
  - allows comparisons, *e.g., olive oil* vs. *sea salt*
    - similar: both match the paraphrase “N1 is extracted from N2”
    - different: salt is not *squeezed* from the sea

Noun Compounds: Using Lexical Paraphrases (5)

Abstract Relations vs. Prepositions vs. Verbs

- Abstract relations [Nastase & Szpakowicz, 2003; Kim & Baldwin, 2005; Girju, 2007; Ó Séaghdha & Copestake, 2007]
  - malaria mosquito: ***Cause***
  - olive oil: ***Source***
- Prepositions [Lauer, 1995]
  - malaria mosquito: ***with***
  - olive oil: ***from***
- Verbs [Finin, 1980; Vanderwende, 1994; Kim & Baldwin 2006; Butnariu & Veale 2008; Nakov & Hearst 2008]
  - malaria mosquito: ***carries, spreads, causes, transmits, brings, has***
  - olive oil: ***comes from, is made from, is derived from***

Noun Compounds: Using Lexical Paraphrases (6)

Note 1 on paraphrasing verbs

- Can paraphrase a noun compound
  - chocolate bar: ***be made of, contain, be composed of, taste like***
- Can also express an abstract relation
  - ***MAKE***$_2$: ***be made of, be composed of, consist of, be manufactured from***
- ... but can also be NC-specific
  - orange juice: ***be squeezed from***
  - bacon pizza: ***be topped with***
  - chocolate bar: ***taste like***

Noun Compounds: Using Lexical Paraphrases (7)

Note 2 on paraphrasing verbs

- Single verb
  - malaria mosquito: ***cause***
  - olive oil: ***be extracted from***
- Multiple verbs
  - malaria mosquito: ***cause, carry, spread, transmit, bring, ...***
  - olive oil: ***be extracted from, come from, be made from, ...***
- Distribution over verbs (SemEval-2010 Task 9)
  - malaria mosquito: ***carry (23), spread (16), cause (12), transmit (9), bring (7), be infected with (3), infect with (3), give (2), ...***
  - olive oil: ***come from (33), be made from (27), be derived from (10), be made of (7), be pressed from (6), be extracted from (5), ...***

Noun Compounds: Using Lexical Paraphrases (8)

Free paraphrases at SemEval-2013 Task 4 [Hendrickx & al., 2013]

- *e.g.*, for *onion tears*
  - tears from onions
  - tears due to cutting onion
  - tears induced when cutting onions
  - tears that onions induce
  - tears that come from chopping onions
  - tears that sometimes flow when onions are chopped
  - tears that raw onions give you
  - ...

Relations between Concepts: Semantic Relations in Ontologies

The easy ones:

- ***is-a***
- ***part-of***

The backbone of any ontology.

Relations between Concepts: Semantic Relations in Ontologies

The easy ones?

- ***is-a***
  - CHOCOLATE ***is-a*** FOOD – class inclusion
  - TOBLERONE ***is-a*** CHOCOLATE – class membership

  and also [Wierzbicka, 1984]

  - CHICKEN ***is-a*** BIRD – taxonomic (is-a-kind-of)
  - ADORNMENT ***is-a*** DECORATION – functional (is-used-as-a-kind-of)
  - ...
- ***part-of***

Relations between Concepts: Semantic Relations in Ontologies

The easy ones?

- ***is-a***
- ***part-of*** [Winston & al., 1987]

Relation	**Example**
component-integral object	pedal - bike
member-collection	ship - fleet
portion-mass	slice - pie
stuff-object	steel - car
feature-activity	paying - shopping
place-area	Everglades - Florida

Relations between Concepts: Semantic Relations in Ontologies

The easy ones?

- ***is-a***
- ***part-of*** [Winston & al., 1987]
  - motivation: lack of transitivity
    1. Simpson's arm is part of Simpson('s body).
    2. Simpson is part of the Philosophy Department.
    3. *Simpson's arm is part of the Philosophy Department.
  - ***component-object*** is incompatible with ***member-collection***

Relations in WordNet

Relation	Example
Synonym	day (Sense 2) / time
Antonym	day (Sense 4) / night
Hypernym	berry (Sense 2) / fruit
Hyponym	fruit (Sense 1) / berry
Member-of holonym	Germany / NATO
Has-member meronym	Germany / Sorbian
Part-of holonym	Germany / Europe
Has-part meronym	Germany / Mannheim
Substance-of holonym	wood (Sense 1) / lumber
Has-substance meronym	lumber (Sense 1) / wood
Domain - TOPIC	line (Sense 7) / military
Domain - USAGE	line (Sense 21) / channel
Domain member - TOPIC	ship / porthole
Attribute	speed (Sense 2) / fast
Derived form	speed (Sense 2) / quick
Derived form	speed (Sense 2) / accelerate

Conclusions

- No consensus on a comprehensive list of relations fit for all purposes and all domains.
- Some shared properties of relations, and of relation schemata.

Properties of Relations (1)

Useful distinctions

- Ontological vs. Idiosyncratic
- Binary vs. *n*-ary
- Targeted vs. Emergent
- First-order vs. Higher-order
- General vs. Domain-specific

Properties of Relations (2)

Ontological vs. Idiosyncratic

- Ontological
  - come up practically the same in numerous contexts
    - *e.g.*, ***is-a***(*apple*, *fruit*)
  - can be extracted with both supervised and unsupervised methods
- Idiosyncratic
  - highly sensitive to the context
    - *e.g.*, ***Content-Container***(*apple*, *basket*)
  - best extracted with supervised methods

Note: Parallel to paradigmatic vs. syntagmatic relations in the *Course in General Linguistics* [de Saussure, 1959].

Properties of Relations (3)

Binary vs. *n*-ary

- Binary
  - most relations
  - our focus here
- *n*-ary
  - good for verbs that can take multiple arguments, *e.g., sell*
  - can be represented as *frames*
    - *e.g.,* a *selling* event can invoke a frame covering relations between a *buyer*, a *seller*, an *object_bought* and *price_paid*

Properties of Relations (4)

Targeted vs. Emergent

- Targeted
  - coming from a fixed inventory
  - *e.g.*, {***Cause***, ***Source***, ***Target***, ***Time***, ***Location***}
- Emergent
  - not fixed in advance
  - can be extracted using patterns over parts-of-speech
  - *e.g.*, `(V | V (N | Adj | Adv | Pron | Det)* PP)` can extract ***invented***, ***is located in*** or ***made a deal with***
  - could also use clustering to group similar relations
    - but then naming the clusters is hard

Properties of Relations (5)

First-order vs. Higher-order

- First-order
  - *e.g.*, ***is-a***(*apple*, *fruit*)
  - most relations
- Higher-order
  - *e.g.*, ***believes***(*John*, ***is-a***(*apple*, *fruit*))
  - can be expressed as conceptual graphs [Sowa, 1984]
  - important in *semantic parsing* [Liang & al., 2011; Lu & al., 2008]
  - also in *biomedical event extraction* [Kim & al., 2009]
  - *e.g.*, "*In this study we hypothesized that the phosphorylation of TRAF2 inhibits binding to the CD40 cytoplasmic domain.*"
    - E1: phosphorylation(Theme:TRAF2),
    - E2: binding(Theme1:TRAF2, Theme2:CD40, Site:cytoplasmic domain),
    - E3: negative_regulation(Theme:E2, Cause:E1).

Properties of Relations (6)

General vs. Domain-specific

- General
  - likely to be useful in processing all kinds of text or in representing knowledge in any domain
  - *e.g.*, ***location***, ***possession***, ***causation***, ***is-a***, or ***part-of***
- Domain-specific
  - only relevant to a specific text genre or to a narrow domain
  - *e.g.*, ***inhibits***, ***activates***, ***phosphorylates*** for gene/protein events

Properties of Relation Schemata (1)

Useful distinctions

- Coarse-grained vs. Fine-grained
- Flat vs. Hierarchical
- Closed vs. Open

Properties of Relation Schemata (2)

Coarse-grained vs. Fine-grained

- Coarse-grained
  - *e.g.,* 5 relations
- Fine-grained
  - *e.g.,* 30 relations
- Infinite, in the extreme
  - every interaction between entities is a distinct relation with unique properties
  - not very practical as there is no generalization
  - however, a distribution over paraphrases is useful

Properties of Relation Schemata (3)

Flat vs. Hierarchical

- Flat
  - most inventories
- Hierarchical
  - *e.g.,* Nastase & Szpakowicz's [2003] schema has 5 top-level and 30 second-level relations
  - *e.g.,* Warren's [1978] schema has four levels: *e.g.,* ***Possessor-Legal Belonging*** is a subrelation of ***Possessor-Belonging***, which is a subrelation of ***Whole-Part*** under the top-level relation ***Possession***

Properties of Relation Schemata (4)

Closed vs. Open

- Closed
  - most inventories
- Open
  - used for Web

Reflects the distinction between targeted and emergent relations.

The Focus of this Tutorial

- Our focus
  - relations between entities mentioned in the same sentence
  - expressed linguistically as *nominals*
- Terminology
  - Relation *type*
    - *e.g.*, hyponymy, meronymy, ***container***, ***product***, ***location***
  - Relation *instance*
    - *e.g.*, "*chocolate contains caffeine*"

Nominal (1)

The standard definition

- a phrase that behaves syntactically like a noun or a noun phrase [Quirk & al., 1985]

Nominal (2)

Our narrower definition

- a *common noun* (`chocolate`, `food`)
- a *proper noun* (`Godiva`, `Belgium`)
- a *multi-word proper name* (`United Nations`)
- a *deverbal noun* (`cultivation`, `roasting`)
- a *deadjectival noun* ([the] `rich`)
- a *base noun phrase* built of a head noun with optional premodifiers (`processed food`, `delicious milk chocolate`)
- (recursively) a sequence of nominals (`cacao tree`, `cacao tree growing conditions`)

Some Clues for Extracting Semantic Relations (1)

Explicit clue

- A phrase linking the entity mentions in a sentence
  - *e.g.*, "Chocolate ***is a*** raw or processed food produced from the seed of the tropical Theobroma cacao tree."
  - issue 1: ambiguity
    - `in` may indicate a temporal relation (chocolate ***in*** the 20th century)
    - but also a spatial relation (chocolate ***in*** Belgium)
  - issue 2: over-specification
    - the relation between chocolate and cultures in "Chocolate ***was prized as a health food and a divine gift by*** the Mayan and Aztec cultures."

Some Clues for Extracting Semantic Relations (2)

Implicit clue

- The relation can be implicit
  - *e.g.*, in noun compounds
    - clues come from knowledge about the entities
    - *e.g.*, cacao tree: CACAO are SEEDS produced by a TREE

Some Clues for Extracting Semantic Relations (3)

Implicit clue

When an entity is an occurrence (event, activity, state) expressed by a deverbal noun such as `cultivation`

- The relation mirrors that between the underlying verb and its arguments
  - *e.g.*, in "the ancient Mayans cultivated chocolate", chocolate is the ***theme***
    - thus, a ***theme*** relation in *chocolate cultivation*
- We do not treat nominalizations separately: typically, they can be also analyzed as normal nominals
  - but they are treated differently
    - in some linguistic theories [Levi, 1978]
    - in some computational linguistics work [Lapata, 2002]

Our Assumptions

- Entities are given
  - no entity identification
  - no entity disambiguation
- Entities in the same sentence, no coreference, no ellipsis

  Angela Merkel's spokesman has insisted that the German chancellor's first meeting with François Hollande, France's president-elect, will be a "getting to know you" exercise, and not "decision making" [meeting].

- Not of direct interest: existing ontologies, knowledge bases and other repositories
  - though useful as seed examples or training data

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. Embeddings
7. Wrap-up

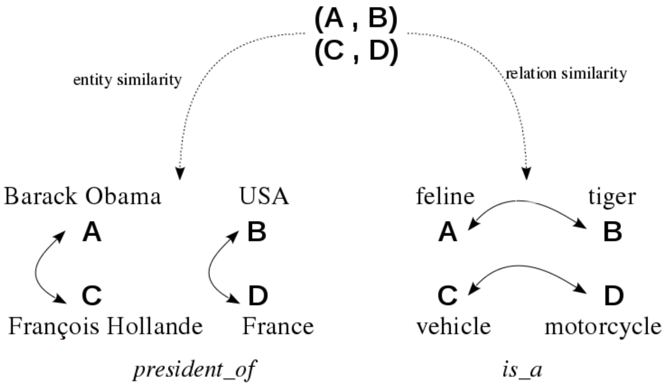
Learning Relations

Methods of Learning Semantic Relations

- Supervised
  - PROs: perform better
  - CONs: require labeled data and feature representation
- Unsupervised
  - PROs: scalable, suitable for open information extraction
  - CONs: perform less well

Learning Relations: Features

- **Purpose: map a pair of terms to a vector**
- **Entity features** and **relational features** [Turney, 2006]



Features

Entity features

...capture some representation of the meaning of an entity – the arguments of a relation

Relational features

...directly characterize the relation – the interaction between its arguments

Entity Features (1)

Basic entity features

- The string value of the argument (possibly lemmatized or stemmed)
- Examples:
  - string value
  - individual words/stems/lemmata

PROs: often informative enough for good relation assignment
CONs: too sparse

Entity Features (2)

Background entity features

- Syntactic information (*e.g.*, grammatical role) or semantic information (*e.g.*, semantic class)
- Can use task-specific inventories, *e.g.*,
  - ACE entity types
  - WordNet features

PROs: solve the data sparseness problem
CONs: manual resources required

Entity Features (3)

Background entity features

- clusters as semantic class information
  - Brown clusters [Brown et al., 1992]
  - Clustering By Committee [Pantel & Lin, 2002]
  - Latent Dirichlet Allocation [Blei et al., 2003]

Entity Features (4)

Background entity features

- Direct representation of co-occurrences in feature space
  - coordination (and/or) [Ó Séaghdha & Copestake, 2008], *e.g., dog and cat*
  - distributional representation
  - relational-semantic representation
- Word embeddings [Nguyen & Grishman, 2014; Hashimoto et al., 2015]

Entity Features (5)

Background entity features

- Distributional representation

Word	Syntactic relation	Co-occurring words
paper-n	coordination	pen-n:69, pencil-n:51, paper-n:32, glass-n:22, ink-n:20, …
	subject_of	say-v:86, make-v:39, propose-v:39, describe-v:31, set-v:30, …
	object_of	publish-v:147, read-v:129, use-v:78, write-v:62, take-43, …
	modified_by_adj	white-j:923; local-j:159, green-j:63, brown-j:56, non-stick-j:71…
	modified_by_n	consultation-n:117, government-n:94, discussion-n:84, tissue-n:71, blotting-n:59, …
	modifies_n	bag-n:150, money-n:44, cup-n:37, mill-n:36, work-n:34, …
	pp_with	number-n:6, address-n:3, title-n:2, note-n:2, word-n:2, …
	pp_on	reform-n:13, future-n:13, policy-n:10, environment-n:9, subject-n:8, …
	…	

Entity Features (6)

Background entity features

- Distributional representation for the noun *paper*
  - what a paper can do: *propose, say*
  - what one can do with a paper: *read, publish*
  - typical adjectival modifiers: *white, recycled*
  - noun modifiers: *toilet, consultation*
  - nouns connected via prepositions: *on environment, for meeting, with a title*

- PROs: captures word meaning by aggregating all interactions (found in a large collection of texts)
- CONs: lumps together different senses
  - *ink* refers to the medium for writing
  - *propose* refers to writing/publication/document

Entity Features (7)

Background entity features

- Relational-semantic representation:
  it uses related concepts from a semantic network or a formal ontology

PROs: based on word senses, not on words
CONs: word-sense disambiguation required

Entity Features (8)

Background entity features

- Determining the semantic class of relation arguments
  - Clustering
  - The descent of hierarchy
  - Iterative semantic specialization
  - Semantic scattering

Entity Features (9)

Background entity features

- The descent of hierarchy [Rosario & Hearst, 2002]:
  the same relation is assumed for all compounds from the same hierarchies
  - *e.g.*, the first noun denotes a *Body Region*, the second noun denotes a *Cardiovascular System*:
    limb vein, scalp arteries, finger capillary, forearm microcirculation
  - generalization at levels 1-3 in the MeSH hierarchy
  - generalization done manually
  - 90% accuracy

Entity Features (10)

Background entity features

- Iterative Semantic Specialization [Girju & al., 2003]
  - fully automated
  - applied to *Part-Whole*
  - given positive and negative examples
    1. generalize up in WordNet from each example
    2. specialize so that there are no ambiguities
    3. produce rules
- Semantic Scattering [Moldovan & al., 2004]
  - learns a boundary (a cut)

Relational Features (1)

Relational features

- characterize the relation directly
  (as opposed to characterizing each argument in isolation)

Relational Features (2)

Basic relational features

- model the context
  - words between the two arguments
  - words from a fixed window on either side of the arguments
  - a dependency path linking the arguments
  - an entire dependency graph
  - the smallest dominant subtree

Relational Features (3)

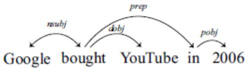
Basic relational features: examples

Bag of words {*2006, bought, Google, in, YouTube*}

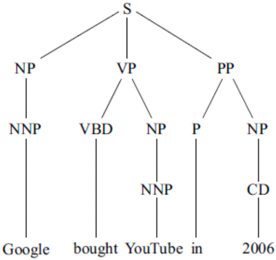
Word sequence (*Google, bought, YouTube, in, 2006*)

Dependency path Google $\xleftarrow{nsubj}$ bought $\xrightarrow{dobj}$ YouTube

Dependency graph



Constituent tree



Relational Features (4)

Background relational features

- encode knowledge about how entities typically interact in texts beyond the immediate context, *e.g.,*
  - paraphrases which characterize a relation
  - patterns with place-holders
  - clustering to find similar contexts

Relational Features (5)

Background relational features

- characterizing noun compounds using paraphrases
  - Nakov & Hearst [2007] extract from the Web verbs, prepositions and coordinators connecting the arguments
    - `"X that * Y"`
    - `"Y that * X"`
    - `"X * Y"`
    - `"Y * X"`
  - Butnariu & Veale [2008] use the Google Web 1T n-grams

Relational Features (6)

Background relational features

- [Nakov & Hearst, 2007]: example for *committee member*

Freq.	Pattern	POS	Direction
2205	of	P	$2 \rightarrow 1$
1923	be	V	$1 \rightarrow 2$
771	include	V	$1 \rightarrow 2$
382	serve on	V	$2 \rightarrow 1$
189	chair	V	$2 \rightarrow 1$
189	have	V	$1 \rightarrow 2$
169	consist of	V	$1 \rightarrow 2$
148	comprise	V	$1 \rightarrow 2$
106	sit on	V	$2 \rightarrow 1$
81	be chaired by	V	$1 \rightarrow 2$
78	appoint	V	$1 \rightarrow 2$
77	on	P	$2 \rightarrow 1$
66	and	C	$1 \rightarrow 2$
66	be elected	V	$1 \rightarrow 2$
58	replace	V	$1 \rightarrow 2$
48	lead	V	$2 \rightarrow 1$
47	be intended for	V	$1 \rightarrow 2$
45	join	V	$2 \rightarrow 1$
...	...	...	...

Relational Features (7)

Background relational features

- using features with placeholders: Turney [2006] mines from the Web patterns like
  - "`Y * causes X`" for **Cause** (*e.g., cold virus*)
  - "`Y in * early X`" for **Temporal** (*e.g., morning frost*).

Relational Features (8)

Background relational features

- can be distributional
  - Turney & Littman [2005] characterize the relation between two words as a vector with coordinates corresponding to the Web frequencies of 128 fixed phrases like "`X for Y`" and "`Y for X`" (`for` is one of a fixed set of 64 joining terms: `such as`, `not the`, `is *`, etc. etc.)
    - can be used directly, or
    - in singular value decomposition [Turney, 2006]

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. Embeddings
7. Wrap-up

Supervised Methods

Supervised relation extraction: setup

- Task: given a piece of text, find instances of semantic relations
- Subtasks
  - argument identification (often ignored)
  - relation classification (core subtask)
- Needed
  - an inventory of possible semantic relations
  - annotated positive/negative examples: for training, tuning and evaluation

Data

Annotated data for learning semantic relations

- small-scale / large-scale
- general-purpose / domain-specific
- arguments marked / not marked
- additional information about the arguments (*e.g.*, senses) / no additional information

Data: MUC and ACE

Relation Type	Subtypes
Physical	***Located***
	Near
Part-Whole	***Geographical***
	Subsidiary
Personal-Social	***Business***
	Family
	Lasting-Personal
Organization-Affiliation	***Employment***
	Ownership
	Founder
	Student-Alum
	Sports-Affiliation
	Investor-Shareholder
	Membership
Agent-Artifact	***User-Owner-Inventor-Manufacturer***
General Affiliation	***Citizen-Resident-Religion-Ethnicity***
	Organization-Location-Origin

Data: MUC and ACE

Relation Type	Subtypes
Physical	***Located***
	Near
Part-Whole	***Geographical***
	Subsidiary
Personal-Social	***Business***
	Family
	Lasting-Personal
Organization-Affiliation	***Employment***
	Ownership
	Founder
	Student-Alum
	Sports-Affiliation
	Investor-Shareholder
	Membership
Agent-Artifact	***User-Owner-Inventor-Manufacturer***
General Affiliation	***Citizen-Resident-Religion-Ethnicity***
	Organization-Location-Origin

The arguments of relations are tagged for type!

Employment(Person, Organization):
<PER>He</PER> had previously worked at <ORG>NBC Entertainment</ORG>.

Near(Person, Facility):
<PER>Muslim youths</PER> recently staged a half dozen rallies in front of <FAC>the embassy</FAC>.

Citizen-Resident-Religion-Ethnicity(Person, Geo-political entity):
Some <GPE>Missouri</GPE> <PER>voters</PER>...

Data: SemEval

- a small number of relations
- annotated entities
- additional entity information (WordNet senses)
- sentential context + mining patterns

SemEval-2007 Task 4 (1)

Semantic relations between nominals: inventory

Relation	Training positive	Training size	Test positive	Test size
Cause-Effect laugh [*Cause*] wrinkles [*Effect*]	52.1%	140	51.3%	80
Instrument-Agency laser [*Instrument*] printer [*Agency*]	50.7%	140	48.7%	78
Product-Producer honey [*Product*] bee [*Producer*]	60.7%	140	66.7%	93
Origin-Entity message [*Entity*] from outer-space [*Origin*]	38.6%	140	44.4%	81
Theme-Tool news [*Theme*] conference [*Tool*]	41.4%	140	40.8%	71
Part-Whole the door [*Part*] of the car [*Whole*]	46.4%	140	36.1%	72
Content-Container the apples [*Content*] in the basket [*Container*]	46.4%	140	51.4%	74
Average	**48.0%**	**140**	**48.5%**	**78**

SemEval-2007 Task 4 (2)

Semantic relations between nominals: examples

"Among the contents of the <e1>vessel</e1> were a set of carpenter's <e2>tools</e2>, several large storage jars, ceramic utensils, ropes and remnants of food, as well as a heavy load of ballast stones."

WordNet(e1) = "vessel%1:06:00::",
WordNet(e2) = "tool%1:06:00::",
Content-Container(e2, e1) = "true",
Query = "contents of the * were a"

"<e1>Batteries</e1> stored in <e2>contact</e2> with one another can generate heat and hydrogen gas."

WordNet(e1) = "battery%1:06:00::",
WordNet(e2) = "contact%1:26:00::",
Content-Container(e1, e2) = "false,"
Query = "batteries stored in"

SemEval-2010 Task 8 (1)

Multi-way semantic relations between nominals: inventory

Relation	Training positive	Training size	Test positive	Test size
Cause-Effect radiation [*Cause*] cancer [*Effect*]	12.5%	1003	12.1%	328
Instrument-Agency phone [*Instrument*] operator[*Agency*]	6.3%	504	5.7%	156
Product-Producer suits [*Product*] factory [*Producer*]	9.0%	717	8.5%	231
Content-Container wine [*Content*] is in the bottle [*Container*]	6.8%	540	7.1%	192
Entity-Origin letters [*Entity*] from the city [*Origin*]	9.0%	716	9.5 %	258
Entity-Destination boy [*Entity*] went to bed [*Destination*]	10.6%	845	10.8%	292
Component-Whole kitchen [*Component*] apartment [*Whole*]	11.8%	941	11.5 %	312
Member-Collection tree [*Member*] forest [*Collection*]	8.6%	690	8.6%	233
Message-Topic lecture [*Message*] on semantics [*Topic*]	7.9%	634	9.6 %	261
Other people filled with joy	17.6%	1410	16.7%	454
Total		**8000**		**2717**

SemEval-2010 Task 8 (2)

Multi-way semantic relations between nominals: examples

The <e1>collision</e1> resulted in two more <e2>crashes</e2> in the intersection, including a Central Concrete truck that was about to turn left onto College Ave.
Relation = Cause-Effect(e1, e2)

Entity-Origin(e_1, e_2)
He removed the <e1>apples</e1> from the <e2>basket</e2> and put them on the table.

Content-Container(e_1, e_2)
When I entered the room, the <e1>apples</e1> were put in the <e2>basket</e>.

Entity-Destination(e_1, e_2)
Then, the <e1>apples</e1> were put in the <e_2 >basket</e_2 > once again.

Algorithms for Relation Learning (1)

Pretty much any machine learning algorithm can work, but some are better for relation learning.

Classification with kernels is appropriate because relational features (in particular) may have complex structures.

Neural networks are appropriate for capturing complex interactions and compositionality

Sequential labelling methods are appropriate because the arguments of a relation have variable span.

Algorithms for Relation Learning (2)

Classification with kernels: overview

- idea: the similarity of two instances can be computed in a high-dimensional feature space without the need to enumerate the dimensions of that space (*e.g.,* using dynamic programming)
- convolution kernels: easy to combine features, *e.g.,* entity and relational
- kernelizable classifiers: SVM, logistic regression, kNN, Naïve Bayes

Algorithms for Relation Learning (3)

Kernels for linguistic structures

- string sequencies [Cancedda & al., 2003]
- dependency paths [Bunescu & Mooney, 2005]
- shallow parse trees [Zelenko & al., 2003]
- constituent parse trees [Collins & Duffy, 2001]
- dependency parse trees [Moschitti, 2006]
- feature-enriched tree kernel [Sun & Han, 2014]
- directed acyclic graphs [Suzuki & al., 2003]

Algorithms for Relation Learning (4)

Tree kernels

- Similarity between two trees is the (normalized) sum of similarities between their subtrees
- Similarity between subtrees based on similarities between roots and children (leaf nodes or subtrees)
- Similarity between leaf (word) nodes can be 0/1 or based on semantic similarity using e.g., clusters or word embeddings [Plank and Moschitti, 2013; Nguyen et al., 2015]

Algorithms for Relation Learning (5)

Neural networks

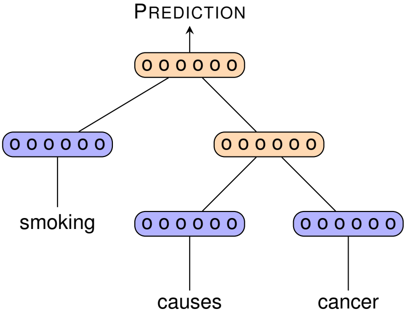
Recursive networks create a bottom-up representation for a tree context by recursively combining representations of siblings [Socher et al., 2012]

Convolutional networks create a representation by sliding a window over the context and pooling the representations at each step [Zeng et al., 2014]

Recurrent networks create a representation for a sequence context by processing each item in the sequence and updating the representation at each step [Not yet investigated?]

Algorithms for Relation Learning (6)

Recursive neural networks [Socher et al., 2012]



Word vectors (can be pretrained)

Compositional vectors (RNN):
$v_{parent} = f(W_l v_l + W_r v_r + b)$

Compositional vectors and matrices (MV-RNN):
$v_{parent} = f(W_{Vl} M_r v_l + W_{Vr} M_l v_r + b)$
$M_{parent} = W_{Ml} M_l + W_{Mr} M_r$

Algorithms for Relation Learning (7)

Convolutional neural networks [Zeng et al., 2014, Liu et al., 2015, dos Santos et al., 2015]

Algorithms for Relation Learning (8)

Sequential labelling methods

- HMMs / MEMMs / CRFs
  [Bikel & al., 1999; Lafferty & al., 2001; McCallum & Li, 2003]
- useful for
  - argument identification
    - *e.g., born-in* holds between *Person* and *Location*
  - relation extraction
    - argument order matters for some relations

Algorithms for Relation Learning (9)

Sequential labelling: argument identification

- *words*: individual words, previous/following two words, word substrings (prefixes, suffixes of various lengths), capitalization, digit patterns, manual lexicons (*e.g.,* of days, months, honorifics, stopwords, lists of known countries, cities, companies, and so on)
- *labels*: individual labels, previous/following two labels
- combinations of *words* and *labels*

George	*W.*	*Bush*	,	*son*	*of*	*the*	*Republican*	*president*	*George*	*H.*	*W.*
B-PER	I-PER	I-PER	O	O	O	O	O	O	B-PER	I-PER	I-PER

Bush	,	*was*	*born*	*in*	*New*	*Haven*	,	*Connecticut*	.
I-PER	O	O	O	O	B-LOC	I-LOC	I-LOC	I-LOC	O

Algorithms for Relation Learning (10)

Sequential labelling: relation extraction

- when one argument is known: the task becomes argument identification
  - *e.g.*, this GeneRIF is about *COX-2*
    - **COX-2** expression is significantly more common in **endometrial adenocarcinoma** and **ovarian serous cystadenocarcinoma**, but not in cervical squamous carcinoma, compared with normal tissue.
- some relations come in order
  - *e.g.*, *Party*, *Job* and *Father* below

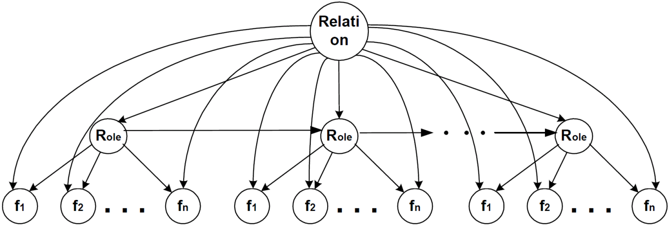
George	*W.*	*Bush*	*,*	*son*	*of*	*the*	*Republican*	*president*	*George*	*H.*	*W.*
B-Target	I-Target	I-Target	O	O	O	O	B-Party	B-Job	B-Father	I-Father	I-Father

Bush	*,*	*was*	*born*	*in*	*New*	*Haven*	*,*	*Connecticut*	*.*
I-Father	O	O	O	O	B-BirthPlace	I-BirthPlace	I-BirthPlace	I-BirthPlace	O

Algorithms for Relation Learning (11)

Sequential labelling: relation extraction

- HMMs, CRFs [Culotta & al., 2006; Bundschus & al., 2008]
- Dynamic graphical model [Rosario & Hearst, 2004]



Beyond Binary Relations (1)

Non-binary relations

- Some relations are not binary
  - **Purchase** (*Purchaser, Purchased_Entity, Price, Seller*)
- Previous methods generally apply
- **but** there are some issues
  - Features: not easy to use the words between entity mentions, or the dependency path between mentions, or the least common subtree
  - Partial mentions
    - *Sparks Ltd. bought 500 tons of steel from Steel Ltd.*
    - *Steel Ltd. bought 200 tons of coal.*

Beyond Binary Relations (2)

Non-binary relations

- Coping with partial mentions
  - treat partial mentions as negatives
  - ignore partial mentions
  - train a separate model for each combination of arguments
  - McDonald & al. (2005)
    1. predict whether two entities are related to each other
    2. use strong argument typing and graph-based global optimization to compose *n*-ary predictions
  - many solutions for *Semantic Role Labeling* [Palmer et al., 2010]

Supervised Methods: Practical Considerations (1)

Some very general advice

- Favour high-performing algorithms such as SVM, logistic regression or CRF
  (CRF only if it makes sense as a sequence-labelling problem)
- entity and relational features are almost always useful
- the value of background features varies across tasks
  - *e.g.*, for noun compounds, background knowledge is key, while context is not very useful

Supervised Methods: Practical Considerations (2)

Performance depends on a number of factors

- the number and nature of the relations used
- the distribution of those relations in data
- the source of data for training and testing
- the annotation procedure for data
- the amount of training data available
- …

Conservative conclusion: state-of-the-art systems perform well above random or majority-class baseline.

Supervised Methods: Practical Considerations (3)

Performance at SemEval

- SemEval-2007 Task 4
  - winning system: F=72.4%, Acc=76.3%, using resources such as WordNet
    [Beamer & al., 2007]
  - later: similar performance, using corpus data only
    [Davidov & Rappoport, 2008; Ó Séaghdha & Copestake, 2008; Nakov & Kozareva, 2011]
- SemEval-2010 Task 8
  - winning system: F=82.2%, Acc=77.9%, using many manual resources
    [Rink & Harabagiu, 2010]
  - later: improvement (F=84.1%), neural network with corpus data only
    [dos Santos et al., 2015]

Supervised Methods: Practical Considerations (4)

Performance at ACE

- Different task
  - full documents rather than single sentences
  - relations between specific classes of named entities
- F-score
  - low-to-mid 70s [Jiang & Zhai, 2007; Zhou & al., 2007, 2009]
- Granularity matters
  - moving from <10 ACE relation types to >20 relation subtypes (on the same data!) decreases F1 by about 20%

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. **Unsupervised Methods**
6. Embeddings
7. Wrap-up

Mining Very Large Corpora (1)

Very large corpora

- examples
  - GigaWord (news texts)
  - PubMed (scientific articles)
  - World-Wide Web
- contain massive amounts of data
  - cannot all be encoded to train a supervised model

Mining Very Large Corpora (2)

Very large corpora

- suitable for unsupervised relation mining
- useful in extracting relational knowledge
  - Taxonomic
    - *e.g., What kinds of animals exist?*
  - Ontological
    - *e.g., Which cities are located in the United Kingdom?*
  - Event
    - *e.g., Which companies have bought which other companies?*
- needed because manual knowledge bases are inherently incomplete, *e.g.,* Cyc and Freebase

Mining Very Large Corpora (3)

Example

- Swanson (1987) discovered a connection between migraines and magnesium
- *Swanson linking*
  - publication 1: illness *A* is caused by chemical *B*
  - publication 2: drug *C* reduces chemical *B* in the body
  - linking: connection between illness *A* and drug *C*

Mining Very Large Corpora (4)

Challenges

- a lot of irrelevant information
- high precision is key
- a supervised model might not be feasible
  - new relations, not seen in training
  - deep features too expensive

Mining Very Large Corpora (5)

Historically important: Crafted patterns

- very high precision
- low recall
  - not a problem because of the scale of corpora
- low coverage
  - cover only a small number of relations

Mining Very Large Corpora (6)

Brief history

- pioneered by Hearst (1992)
- initially, taxonomic relations – the backbone of any taxonomy or ontology
  - is-a: hyponymy/hypernymy
  - part-of: meronymy/holonymy
- gradually expanded
  - more relations
  - larger scale of corpora – Web-scale now within reach
    - the *Never-Ending Language Learner* project
    - the *Machine Reading* project

Early Work: Mining Dictionaries (1)

Extracting taxonomic relations from dictionaries

- popular in 1980s
  - [Ahlswede & Evens, 1988; Alshawi, 1987; Amsler, 1981; Chodorow & al., 1985; Ide & al., 1992; Klavans & al., 1992]
- focus on is-a
  - hypenymy/hyponymy
  - subclass/superclass
- used dictionaries such as Merriam-Webster
- pattern-based

Early Work: Mining Dictionaries (2)

Merriam-Webster: GROUP and related concepts

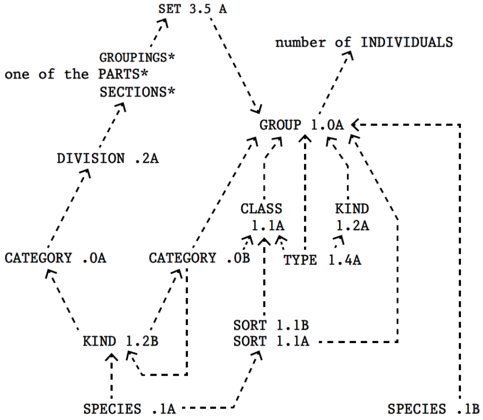
[Amsler, 1981]

- GROUP 1.0A – a number of individuals related by a common factor (as physical association, community of interests, or blood)
- CLASS 1.1A – a group of the same general status or nature
- TYPE 1.4A – a class, kind, or group set apart by common characteristics
- KIND 1.2A – a group united by common traits or interests
- KIND 1.2B – CATEGORY
- CATEGORY .0A – a division used in classification
- CATEGORY .0B – CLASS, GROUP, KIND
- DIVISION .2A – one of the parts, sections, or groupings into which a whole is divided
- *GROUPING <== W7 – a set of objects combined in a group
- SET 3.5A – a group of persons or things of the same kind or having a common characteristic usu. classed together
- SORT 1.1A – a group of persons or things that have similar characteristics
- SORT 1.1B - CLASS
- SPECIES .IA – SORT, KIND
- SPECIES .IB – a taxonomic group comprising closely related organisms potentially able to breed with one another

Early Work: Mining Dictionaries (3)

Merriam-Webster: GROUP and related concepts

[Amsler, 1981]



Early Work: Mining Dictionaries (4)

Mining dictionaries: summary

- PROs
  - short, focused definitions
  - standard language
  - limited vocabulary
- CONs
  - circularity
  - hard to identify the key terms
    - group of persons
    - number of individuals
  - limited coverage

Mining Relations with Patterns (1)

Relation mining patterns

- when matched against a text fragment, identify relation instances
- can involve
  - lexical items
  - wildcards
  - parts of speech
  - syntactic relations
  - flexible rules, *e.g.,* as in regular expressions
  - ...

Mining Relations with Patterns (2)

Hearst's (1992) lexico-syntactic patterns

- NP such as {NP,}* {(or|and)} NP
  "... bow lute, such as Bambara ndang ..."
  → (bow lute, Bambara ndang)
- such NP as {NP,}* {(or|and)} NP
  "... works by such authors as Herrick, Goldsmith, and Shakespeare"
  → (authors, Herrick); (authors, Goldsmith); (authors, Shakespeare)
- NP {, NP}* {,} (or|and) other NP
  "... temples, treasuries, and other important civic buildings ..."
  → (important civic buildings, temples); (important civic buildings, treasuries)
- NP{,} (including|especially) {NP,}* (or|and) NP
  "... most European countries, especially France, England and Spain ..."
  → (European countries, France); (European countries, England); (European countries, Spain)

Mining Relations with Patterns (3)

Hearst's (1992) lexico-syntactic patterns

- designed for very high precision, but low recall
- only cover *is-a*
- later, extended to other relations, *e.g.,*
  - part-of [Berland & Charniak, 1999]
  - protein-protein interactions
    [Blaschke & al., 1999; Pustejovsky & al., 2002]
    - *N1 inhibits N2*
    - *N2 is inhibited by N1*
    - *inhibition of N2 by N1*
- unclear if such patterns can be designed for *all* relations

Mining Relations with Patterns (4)

Hearst's (1992) lexico-syntactic patterns

- ran on Grolier's American Academic Encyclopedia
  - small by today's standards
  - still, large enough: 8.6 million tokens
- very low recall
  - extracted just 152 examples (but with very high precision)
- increase recall
  - bootstrapping

Bootstrapping (1)

Semantic Relation	*Surface Patterns*	*$Hits_{est}$ for patterns*
narrow similarity (2)*	X ie Y Xed ie Yed	219,480
broad similarity (2)*	Xed and Yed to X and Y	154,518,326
strength (8)	X even Y Xed even Yed X and even Y Xed and even Yed Y or at least X Yed or at least Xed not only Xed but Yed not just Xed but Yed	1,016,905
enablement (4)	Xed * by Ying the Xed * by Ying or to X * by Ying the to X * by Ying or	2,348,392
antonymy (7)	either X or Y either Xs or Ys either Xed or Yed either Xing or Ying whether to X or Y Xed * but Yed to X * but Y	18,040,916
happens-before (12)	to X and then Y to X * and then Y Xed and then Yed Xed * and then Yed to X and later Y Xed and later Yed to X and subsequently Y Xed and subsequently Yed to X and eventually Y Xed and eventually Yed	8,288,871

Bootstrapping (2)

Require:
- $\mathcal{P}$—a set of seed patterns
- $\mathcal{R}$—a set of seed relation instances
- $\mathcal{C}$—a corpus
- N—maximum number of iterations

Ensure: $\mathcal{R}$—a set of relation instances

for $i = 1..N$ **do**
    $\mathcal{P}' = \{\}$
    $\mathcal{R}' = \{\}$
    for $p \in \mathcal{P}$ **do**
        match p in $\mathcal{C}$
        add matched pairs: $\mathcal{R}' = \{(np_i, np_j)\} \cup \mathcal{R}'$
    end for
    $\mathcal{R} = \mathcal{R} \cup Top_k(rankInstances(\mathcal{R}'))$
    for $(np_i, np_j) \in \mathcal{R}$ **do**
        match $np_i(.*)np_j$ in $\mathcal{C}$
        add matched pattern $(.*)$ to $\mathcal{P}'$
    end for
    $\mathcal{P} = \mathcal{P} \cup Top_k(rankPatterns(\mathcal{P}'))$
end for
return $\mathcal{R}$

Bootstrapping (3)

Bootstrapping

- Initialization
  - few seed examples
  - *e.g.*, for *is-a*
    - *cat-animal*
    - *car-vehicle*
    - *banana-fruit*
- Expansion
  - new patterns
  - new instances
- Several iterations
- Main difficulty
  - semantic drift

Bootstrapping (4)

Bootstrapping

- Context-dependency
  - not good for context-dependent relations
    - in one newspaper: “Manchester United defeated Chelsea”
    - six months later: “Chelsea defeated Manchester United”
- Specificity
  - good for specific relations such as *birthdate*
  - cannot distinguish between fine-grained relations
  - *e.g.*, different kinds of *Part-Whole* – maybe *Component-Integral_Object, Member-Collection, Portion-Mass, Stuff-Object, Feature-Activity and Place-Area* – would share the same patterns

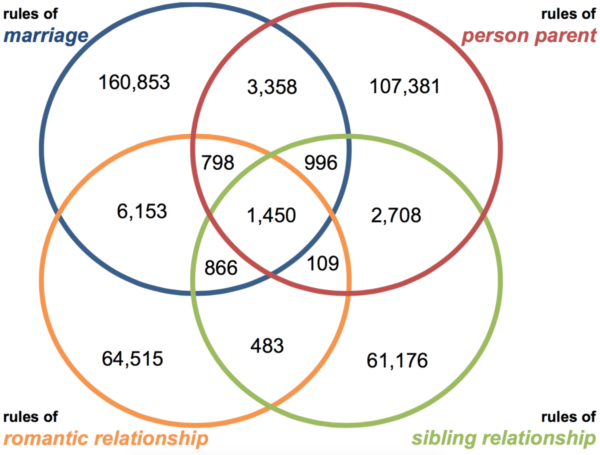
Tackling Semantic Drift (1)

Example of semantic drift

Seeds		Patterns		Added examples
London	→	`mayor of X`	→	California
Paris		`lives in X`		Europe
New York		...		...

Tackling Semantic Drift (2)

Example: Euler diagram for four people-relations [Krause&al.,2012]



Tackling Semantic Drift (3)

Some strategies

- Limit the number of iterations
- Select a small number of patterns/examples per iteration
- Use semantic types, *e.g.*, the SNOWBALL system

$\langle$*Organization*$\rangle$*'s headquarters in* $\langle$*Location*$\rangle$
$\langle$*Location*$\rangle$*-based* $\langle$*Organization*$\rangle$
$\langle$*Organization*$\rangle$*,* $\langle$*Location*$\rangle$

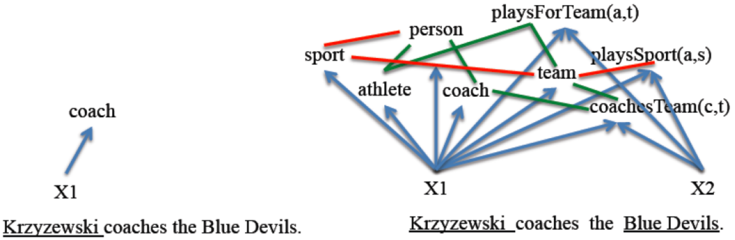
Tackling Semantic Drift (4)

More strategies

- scoring patterns/instances
  - specificity: prefer patterns that match less contexts
  - confidence: prefer patterns with higher precision
  - reliability: based on PMI
- argument type checking
- coupled training

Tackling Semantic Drift (5)

Coupled training [Carlson & al., 2010]



(A) A difficult semi-supervised learning problem

(B) An easier semi-supervised learning problem

Used in the *Never-Ending Language Learner*

Distant Supervision (1)

Distant supervision

- Issue with bootstrapping: starts with a *small number* of seeds
- Distant supervision uses a huge number
  [Craven & Kumlien, 1999]
  1. Get huge seed sets, *e.g.,* from WordNet, Cyc, Wikipedia infoboxes, Freebase
  2. Find contexts where they occur
  3. Use these contexts to train a classifier

Distant Supervision (2)

Example: experiments of Mintz & al. [2009]

- 102 relations from Freebase, 17,000 seed instances
- mapped them to Wikipedia article texts
- extracted
  - 1.8 million instances
  - connecting 940,000 entities
- Assumption: all co-occurrences of a pair of entities express the same relation
  - Riedel & al. [2010] assume that *at least one* context expresses the target relation (rather than *all*)
  - Ling & al. [2013] assume that *a certain percentage* (which can vary by relation) of the contexts are true positive

Distant Supervision (3)

Dr. Henry Walton "Indiana" Jones, Jr., Ph.D.[12] is a fictional character and the protagonist of the *Indiana Jones* franchise. George Lucas and Steven Spielberg created the character in homage to the action heroes of 1930s film serials. The character first appeared in the 1981 film *Raiders of the Lost Ark*, to be followed by *Indiana Jones and the Temple of Doom* in 1984, *Indiana Jones and the Last Crusade* in 1989, *The Young Indiana Jones Chronicles* from 1992 to 1996, and *Indiana Jones and the Kingdom of the Crystal Skull* in 2008. Alongside the more widely known films and television programs, the character is also featured in novels, comics, video games, and other media. Jones is also featured in the theme park attraction *Indiana Jones Adventure*, which exists in similar forms at Disneyland and Tokyo DisneySea.

Jones is most famously played by Harrison Ford and has also been portrayed by River Phoenix (as the young Jones in *The Last Crusade*), and in the television series *The Young Indiana Jones Chronicles* by Corey Carrier, Sean Patrick Flanery, and George Hall. Doug Lee has supplied Jones's voice to two LucasArts video games, *Indiana Jones and the Fate of Atlantis* and *Indiana Jones and the Infernal Machine*, while David Esch supplied his voice to *Indiana Jones and the Emperor's Tomb*.

Particularly notable facets of the character include his iconic look (bullwhip, fedora, and leather jacket), sense of humor, deep knowledge of many ancient civilizations and languages, and fear of snakes.

Indiana Jones remains one of cinema's most revered movie characters. In 2003, he was ranked as the second greatest movie hero of all time by the American Film Institute.[13] He was also named the sixth greatest movie character by *Empire* magazine.[14] *Entertainment Weekly* ranked Indy 2nd on their list of *The All-Time Coolest Heroes in Pop Culture*.[15] *Premiere* magazine also placed Indy at number 7 on their list of *The 100 Greatest Movie Characters of All Time*.[16] Since his first appearance in *Raiders of the Lost Ark*, he has become a worldwide star. On their list of the *100 Greatest Fictional Characters*, Fandomania.com ranked Indy at number 10.[17] In 2010, he ranked #2 on *Time* Magazine's list of the greatest fictional characters of all time, surpassed only by Sherlock Holmes.[citation needed]

Contents [hide]

Henry Jones, Jr.

Indiana Jones character

Harrison Ford as Indiana Jones in *Raiders of the Lost Ark*

First appearance	*Raiders of the Lost Ark*
Created by	George Lucas Steven Spielberg
Portrayed by	**Films:** Harrison Ford (ages 36–58) River Phoenix (age 13) **TV series:**

- training sentences
  1. positive: with the relation
  2. negative: without the relation
- train a two-stage classifier:
  1. identify the sentences with a relation instance
  2. extract relations from these sentences

Distant Supervision (4)

False negatives

- Knowledge bases used to provide distant supervision are incomplete
  1. avoid false negatives [Min et al. 2013]
  2. fill in gaps [Xu et al. 2013]

Distant Supervision (5)

Distant and partial supervision

- Choose representative and useful training examples to maximize performance
  1. active learning [Angeli et al. 2014]
  2. infusion of labeled data [Pershina et al. 2014]
  3. semantic consistency [Han and Sun, 2014]

Unsupervised Relation Extraction

- Other issues with bootstrapping
  - uses multiple passes over a corpus
    - often undesirable/unfeasible, *e.g.,* on the Web
  - if we want to extract *all* relations
    - no seeds for all of them
- Possible solution
  - unsupervised relation extraction
  - no pre-specified list of relations, seeds or patterns

Extracting *is-a* Relations (1)

Pantel & Ravichandran [2004]

- cluster nouns using cooccurrence as in [Pantel & Lin, 2002]
  - *Apple, Google, IBM, Oracle, Sun Microsystems, ...*
- extract hypernyms using patterns
  - Apposition (N:appo:N), *e.g.,* . . . **Oracle**, a **company** known for its progressive employment policies . . .
  - Nominal subject (-N:subj:N), *e.g.,* . . . **Apple** was a hot young **company**, with Steve Jobs in charge . . .
  - Such as (-N:such as:N), *e.g.,* . . . **companies** such as **IBM** must be weary . . .
  - Like (-N:like:N), *e.g.,* . . . **companies** like **Sun Microsystems** do not shy away from such challenges . . .
- *is-a* between the hypernym and each noun in the cluster

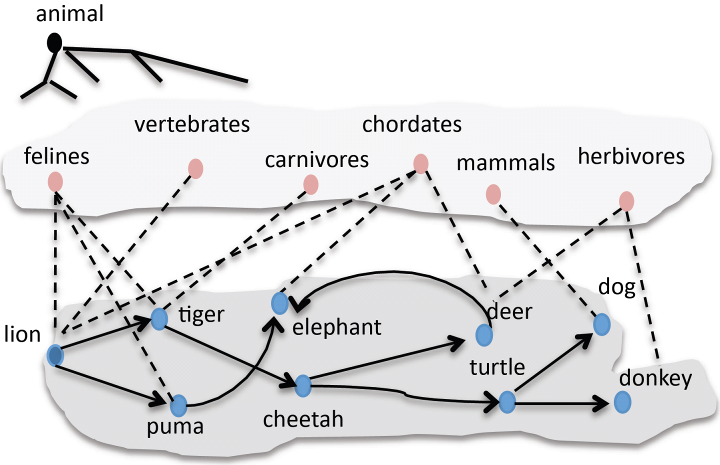
Extracting *is-a* Relations (2)

[Kozareva & al., 2008]

- uses a doubly-anchored pattern (DAP)
  - "*sem-class* such as $term_1$ and *"
- similar to the Hearst pattern
  - NP_0 *such as* $\{NP_1, NP_2, \ldots, (and \mid or)\}$ NP_n
- but different
  - exactly two arguments after *such as*
  - *and* is obligatory
- prevents sense mixing
  - *cats–jaguar–puma*
  - *predators–jaguar–leopard*
  - *cars–jaguar–ferrari*

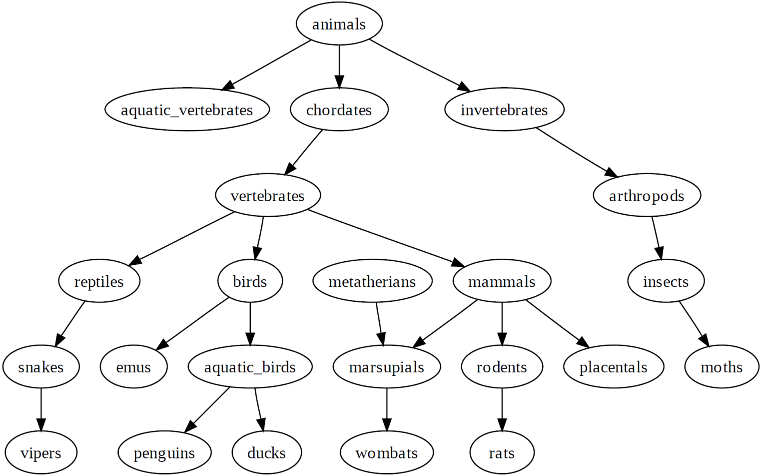
Extracting *is-a* Relations (3)

[Kozareva & Hovy, 2010]: DAPs can yield a taxonomy



Extracting *is-a* Relations (4)

[Kozareva & Hovy, 2010]: DAPs can yield a taxonomy



Emergent Relations (1)

Emergent relations in open relation extraction

- no fixed set of relations
- need to identify novel relations
  - use verbs, prepositions
    - different verbs, same relation: *shot against the flu*, *shot to prevent the flu*
    - verb, but no relation: “It rains.” or “I do.”
    - no verb, but relation: *flu shot*
  - use clustering
    - string similarity
    - distributional similarity

Emergent Relations (2)

Clustering with distributional similarity

- using paraphrases from dependency parses
  [Lin & Pantel, 2001; Pasca, 2007]
  - *e.g.*, DIRT for *X solves Y*
    - *Y is solved by X, X resolves Y, X finds a solution to Y, X tries to solve Y, X deals with Y, Y is resolved by X, X addresses Y, X seeks a solution to Y, X does something about Y, X solution to Y, Y is resolved in X, Y is solved through X, X rectifies Y, X copes with Y, X overcomes Y, X eases Y, X tackles Y, X alleviates Y, X corrects Y, X is a solution to Y, X makes worse Y, X irons out Y*
- extracted shared property model
  [Yates & Etzioni, 2007]
  - *e.g.*, if *(lacks, Mars, ozone layer)* and *(lacks, Red Planet, ozone layer)*, then *Mars* and *Red Planet* share the property *(lacks, *, ozone layer)*

Emergent Relations (3)

[Davidov & Rappoport, 2008]

Prefix CW_1 Infix CW_2 Postfix

label	patterns
(pets, dogs)	{such X as Y, X such as Y, Y and other X}
(phone, charger)	{buy Y accessory for X!, shipping Y for X, Y is available for X, Y are available for X, Y are available for X systems, Y for X}

These (CW_1, CW_2) clusters are efficient as background features for supervised models.

Self-Supervised Relation Extraction (1)

Self-supervision

- algorithm
  1. parse a small corpus
  2. extract and annotate relation instances: *e.g.,* based on heuristics and the connecting path between entity mentions
  3. train relation extractors on these instances
     - not guided by or assigned to any particular relation type
     - features: shallow lexical and POS, dependency path
- applicable on the Web
- used in the Machine Reading project at U Washington

Self-Supervised Relation Extraction (2)

Self-supervision

- Issues with the extracted relations
  - not coherent
    - *e.g., The Mark 14 was central to the torpedo scandal of the fleet.* $\rightarrow$ ***was central torpedo***
  - uninformative
    - *e.g., ... is the author of ...* $\rightarrow$ ***is***
  - too specific
    - *e.g.,* ***is offering only modest greenhouse gas reductions targets at***

Self-Supervised Relation Extraction (3)

Self-supervision

- Improving the relation quality
  - constraints: syntactic, positional and frequency [Fader & al., 2011]
  - focus on functional relations, *e.g., birthplace* [Lin & al., 2010]
  - use redundancy: the "KnowItAll hypothesis" [Downey & al., 2005, 2010] – extractions from more distinct sentences in a corpus are more likely to be correct
    - high frequency is not enough though:
    - `"Elvis killed JFK"` yields 19,300 hits (in October 2012)
    - still, `"Oswald killed JFK"` had 39,200 hits

Web-Scale Relation Extraction (1)

Two large-scale knowledge acquisition projects that harvest the Web continuously

- Never-Ending Language Learner (NELL)
  - at Carnegie-Mellon University
  - `http://rtw.ml.cmu.edu/rtw/`
- Machine Reading
  - at the University of Washington
  - `http://ai.cs.washington.edu/projects/open-information-extraction`

Web-Scale Relation Extraction (2)

Never-Ending Language Learner [Mohamed & al., 2011]

- starting with a seed ontology
  - 600 categories and relations
  - each with 20 seed examples
- learns
  - new concepts
  - new concept instances
  - new instances of the existing relations
  - novel relations
- approach: bootstrapping, coupled learning, manual intervention, clustering
- learned (as of September 2012)
  - 15 million confidence-scored relations (beliefs)
  - 1.4 million with high confidence scores, 85% precision

Web-Scale Relation Extraction (3)

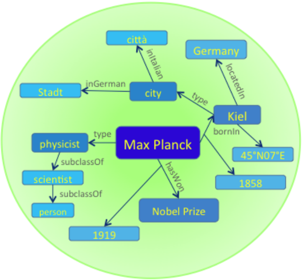
Machine Reading at U Washington

- KnowItAll [Etzioni & al., 2005] – bootstrapping using Hearst patterns
- TextRunner [Banko & al., 2007] – self-supervised, specific relation models from a small corpus, applied to a large corpus
- Kylin [Wu & Weld, 2007] and WPE [Hoffmann & al., 2010] bootstrapping starting with Wikipedia infoboxes and associated articles
- WOE [Wu & Weld, 2010] extends Kylin to open information extraction, using part-of-speech or dependency patterns
- ReVerb [Fader & al., 2011] – lexical and syntactic constraints on potential relation expressions
- OLLIE [Mausam & al., 2012] – extends WOE with better patterns and dependencies (*e.g.,* some relations are true for some period of time, or are contingent upon external conditions)

Other Large-Scale Knowledge Acquisition Projects (1)

YAGO-NAGA [Hoffart&al., 2015]

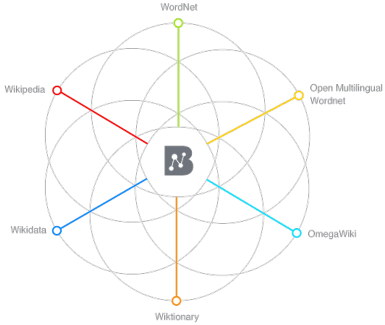
- harvest, search, and rank knowledge from the Web
- large-scale, highly accurate, machine-processible
- integration with Wikipedia and WordNet
- started in 2016, several subprojects



Other Large-Scale Knowledge Acquisition Projects (2)

BabelNet [Navigli&Ponzetto, 2012]

- multilingual semantic network
- integrates several knowledge sources
- no additional Web mining (just integration)



Unsupervised Methods: Summary

Unsupervised relation extraction

- good for
  - large text collections or the Web
  - context-independent relations
- methods
  - bootstrapping (but semantic drift)
  - distant supervision
  - semi-supervision
  - self-supervision
- applications
  - continuous open information extraction
    - NELL
    - Machine Reading

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. **Embeddings**
7. Wrap-up

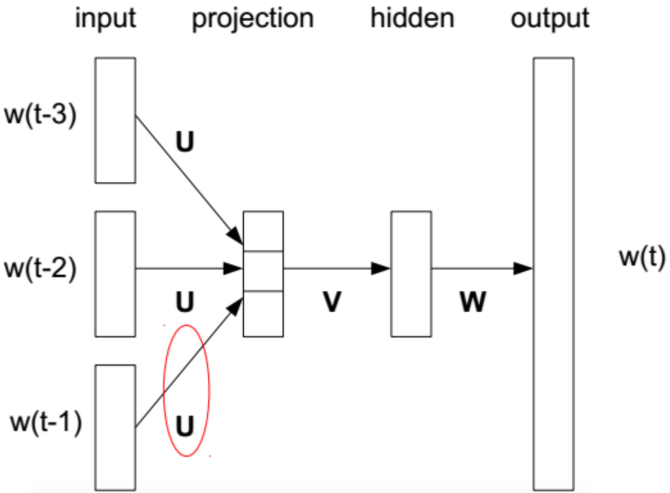
Word Embeddings (1)

Word Embedding

- What is it?
  - mapping words to vectors of real numbers in a low dimensional space
- How is it done?
  - neural networks (e.g., CBOW, skip-gram) [Mikolov&al.2013a]
  - dimensionality reduction (e.g., LSA, LDA, PCA)
  - explicit representation (words in the context)
- Why should we care?
  - useful for a number of NLP tasks
  - . . . including semantic relations

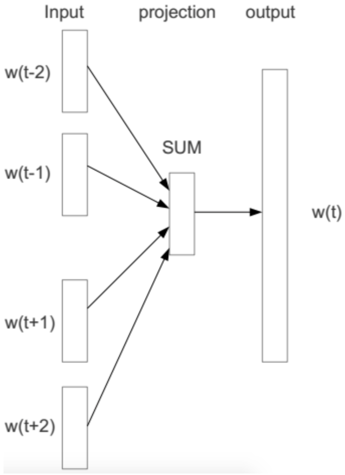
Word Embeddings (2)

Word Embeddings from a Neural LM [Bengio &al.2003]



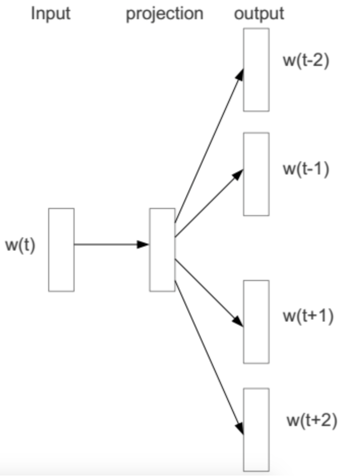
Word Embeddings (3)

Continuous Bag of Words ("predict word") [Mikolov &al.2013a]



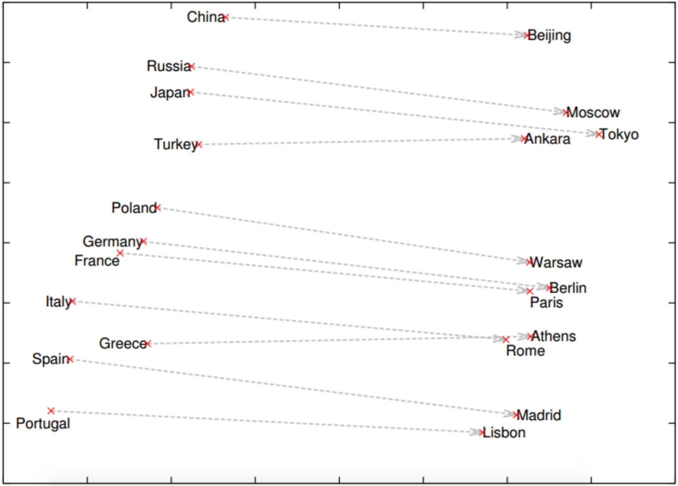
Word Embeddings (4)

Skip-gram (“predict context”) [Mikolov &al.2013a]



Word Embeddings (5)

Skip-gram: projection with PCA



Word Embeddings (6)

Skip-gram: properties [Mikolov&al.2013a]

- Word embeddings have linear structure that enables analogies with vector arithmetics
- Due to training objective: input and output (before softmax) are in a linear relationship

Word Embeddings (7)

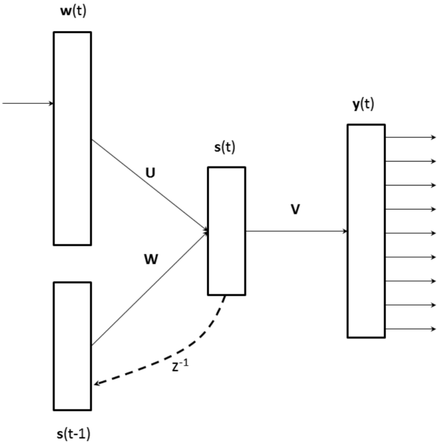
Skip-gram: vector arithmetics

Expression	*Nearest token*
Paris - France + Italy	Rome
bigger - big + cold	colder
sushi - Japan + Germany	bratwurst
Cu - copper + gold	Au
Windows - Microsoft + Google	Android
Montreal Canadiens - Montreal + Toronto	Toronto Maple Leafs

Word Embeddings (8)

Recurrent Neural Network Language Model (RNNLM)

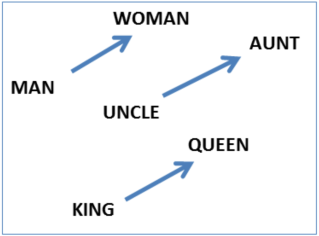
[Mikolov&al.2013b]

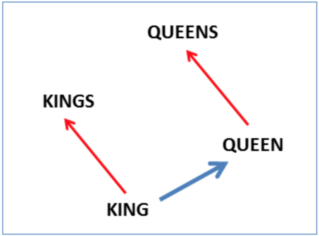


Word Embeddings (9)

RNNLM: beyond semantic relations [Mikolov&al.2013b]

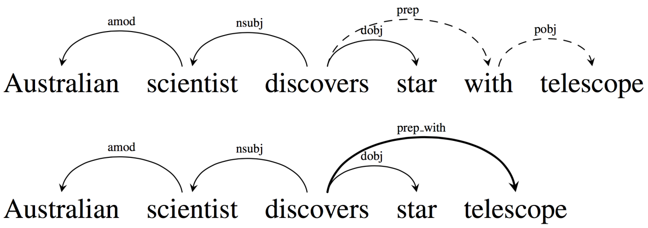
- gender, number, etc.





Syntactic Word Embeddings (1)

Dependency-based embeddings [Levy&Goldberg,2014a]



WORD	CONTEXTS
australian	scientist/amod^{-1}
scientist	australian/amod, discovers/nsubj^{-1}
discovers	scientist/nsubj, star/dobj, telescope/prep_with
star	discovers/dobj^{-1}
telescope	discovers/prep_with^{-1}

Syntactic Word Embeddings (2)

Dependency- vs. word-based embeddings [Levy&Goldberg,2014a]

- **Words: topical**
- **Dependencies: functional**
  - also true for explicit representations [Lin,1998; Padó&Lapata,2007]

- Example: *Turing*
  - **Words:** *nondeterministic, non-deterministic, computability, deterministic, finite-state*
  - **Dependencies:** *Pauling, Hotelling, Heting, Lessing, Hamming*

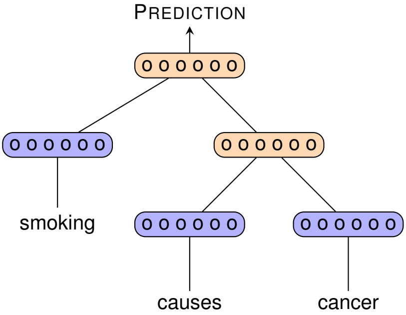
Word Embeddings: Should We Care?

Embeddings vs. Explicit Representations

- embeddings are better across many tasks [Baroni&al., 2014]
  - semantic relatedness
  - synonym detection
  - concept categorization
  - selectional preferences
  - analogy
- BUT explicit representation can be as good on analogies, with a better objective [Levy&Goldberg,2014b]

Embeddings for Relation Extraction (1)

Recursive Neural Networks (RNN) [Socher&al., 2012]



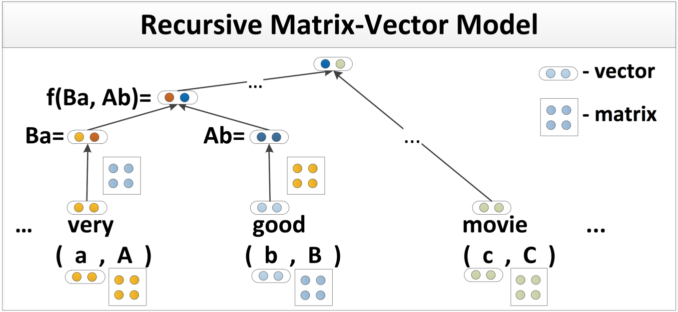
Word vectors (can be pretrained)

Compositional vectors (RNN):
$v_{parent} = f(W_l v_l + W_r v_r + b)$

Compositional vectors and matrices (MV-RNN):
$v_{parent} = f(W_{Vl} M_r v_l + W_{Vr} M_l v_r + b)$
$M_{parent} = W_{Ml} M_l + W_{Mr} M_r$

Embeddings for Relation Extraction (2)

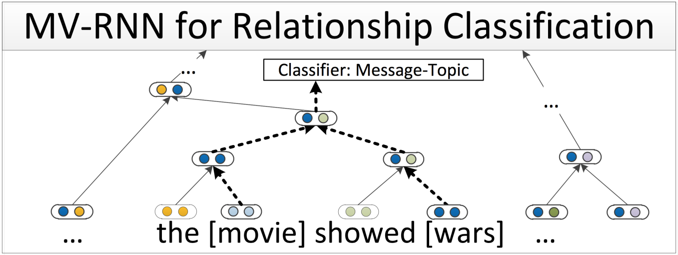
Matrix-Vector Recursive Neural Networks (MV-RNN) [Socher&al., 2012]



- vectors: for compositionality
- matrices: for operator semantics

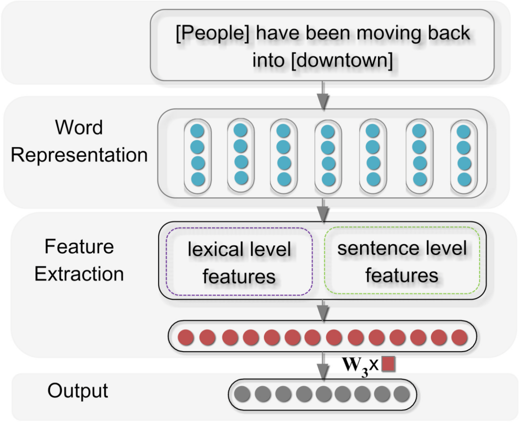
Embeddings for Relation Extraction (3)

MV-RNN for Relation Classification [Socher&al., 2012]



Embeddings for Relation Extraction (4)

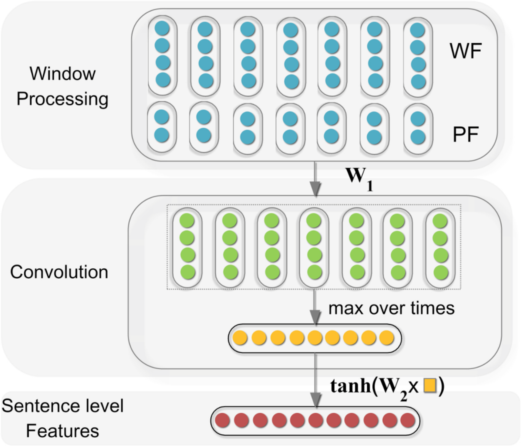
CNN: Convolutional Deep Neural Network [Zeng&al., 2014]



Embeddings for Relation Extraction (5)

CNN (sentence level features) [Zeng&al., 2014]

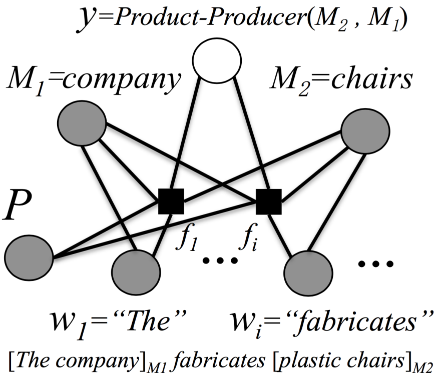
- WF: word vectors; PF: position vectors (distance to e_1, e_2)



Embeddings for Relation Extraction (6)

FCM: Factor-based Compositional Embed. Model [Yu&al., 2014]

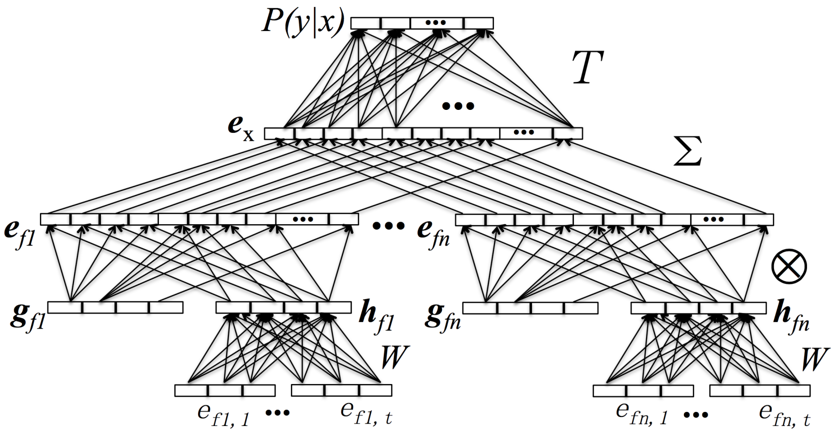
- Extension of the model will be presented at EMNLP'2015 [Gormley&al., 2015]



Embeddings for Relation Extraction (7)

FCM (continued) [Yu&al., 2014]

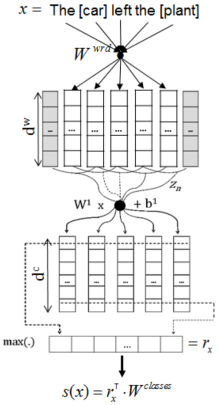
- extension of the model at EMNLP'2015! [Gormley&al., 2015]



Embeddings for Relation Extraction (8)

CR-CNN: Classification by Ranking CNN [dos Santos&al., 2015]

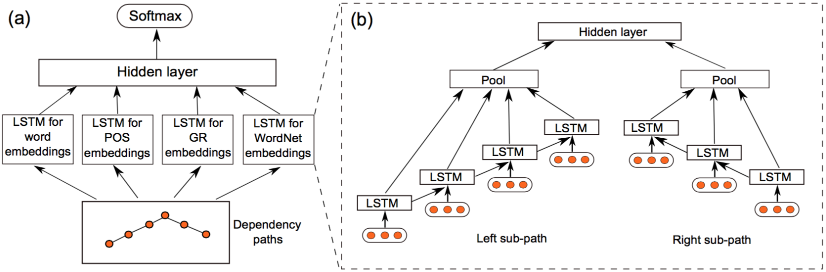
- pairwise ranking loss
- word, class, position, sentence embeddings



Embeddings for Relation Extraction (9)

SDP-LSTM: Shortest dependency path LSTM [Yan Xu&al., 2015]

- to be presented at EMNLP'2015!



Embeddings for Relation Extraction (10)

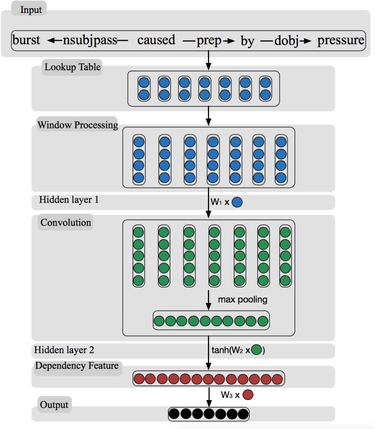
Comparison, testing on SemEval-2010 Task 8 [Yan Xu&al., 2015]

Classifier	Feature set	F_1
SVM	POS, WordNet, prefixes and other morphological features, depdency parse, Levin classes, PropBank, FanmeNet, NomLex-Plus, Google n-gram, paraphrases, TextRunner	82.2
RNN	Word embeddings	74.8
	Word embeddings, POS, NER, WordNet	77.6
MVRNN	Word embeddings	79.1
	Word embeddings, POS, NER, WordNet	82.4
CNN	Word embeddings	69.7
	Word embeddings, word position embeddings, WordNet	82.7
Chain CNN	Word embeddings, POS, NER, WordNet	82.7
FCM	Word embeddings	80.6
	Word embeddings, depedency parsing, NER	83.0
CR-CNN	Word embeddings	82.8†
	Word embeddings, position embeddings	82.7
	Word embeddings, position embeddings	**84.1**†
SDP-LSTM	Word embeddings	82.4
	Word embeddings, POS embeddings, WordNet embeddings, grammar relation embeddings	**83.7**

Embeddings for Relation Extraction (11)

depLCNN: Dependency CNN (w/ neg. sampling) [Kun Xu&al., 2015]

- to be presented at EMNLP'2015!



Embeddings for Relation Extraction (12)

Another comparison on SemEval-2010 Task 8 [Kun Xu&al., 2015]

Method	Feature Sets	F1
SVM	16 types of features	82.2
RNN	-	74.8
	+POS, NER, WordNet	77.6
MVRNN	-	79.1
	+POS, NER, WordNet	82.4
CNN	-	78.9
(Zeng et al., 2014)	+WordNet,words around nominals	82.7
DepNN	+NER	83.6
depCNN	-	81.3
depLCNN	-	81.9
depLCNN	+WordNet,words around nominals	83.7
depLCNN+NS	-	**84.0**
	+WordNet,words around nominals	**85.6**

Outline

1. Introduction
2. Semantic Relations
3. Features
4. Supervised Methods
5. Unsupervised Methods
6. Embeddings
7. Wrap-up

Lessons Learned

Semantic relations

- are an open class
- just like concepts, they can be organized hierarchically
- some are ontological, some idiosyncratic
- the way we work with them depends on
  - the application
  - the method

Lessons Learned

Learning to identify or discover relations

- investigate many detailed features in a (small) fully-supervised setting, and try to port them into an open relation extraction setting
- set an inventory of targeted relations, or allow them to emerge from the analyzed data
- use (more or less) annotated data to bootstrap the learning process
- exploit resources created for different purposes for our own ends (Wikipedia!)

Extracting Relational Knowledge from Text

The bigger picture: NLP finds knowledge in a lot of text and then gets the deeper meaning of a little text

- Manual construction of knowledge bases
  - PROs: accurate (insofar as people who do it do not make mistakes)
  - CONs: costly, inherently limited in scope
- Automated knowledge acquisition
  - PROs: scalable, *e.g.*, to the Web
  - CONs: inaccurate, *e.g.*, due to semantic drift or inaccuracies in the analyzed text
- Learning relations
  - PROs: reasonably accurate
  - CONs: needs relation inventory and annotated training data, does not scale to large corpora

The Future

Hot research topics and future directions

- embeddings, deep learning
- Web-scale relation mining
- continuous, never-ending learning
- distant supervision
- use of large knowledge sources such as Wikipedia, DBpedia
- semi-supervised methods
- combining symbolic and statistical methods
    - *e.g.*, ontology acquisition using statistics

Relevant Literature is Huge! (1)

Relevant papers at EMNLP'2015

- [Li&al., 2015] compare *recursive* (based on syntactic trees) vs. *recurrent* (inspired by LMs) neural networks on four tasks, including semantic relation extraction
- [Kun Xu&al., 2015] learn robust relation representations from shortest dependency paths through a convolution neural network using simple negative sampling
- [Yan Xu&al., 2015] use long short term memory networks along shortest dependency paths for relation classification
- [Gormley&al., 2015] propose a feature-rich compositional embedding model for relation extraction, which combines (unlexicalized) hand-crafted features with learned word embeddings

Relevant Literature is Huge! (2)

Relevant papers at EMNLP'2015

- [Li&Jurafsky, 2015] propose a multi-sense embedding model based on Chinese Restaurant Processes, which they evaluate on a number of tasks, including semantic relation identification
- [D'Souza&Ng, 2015] use expanding parse trees with sieves for spatial relation extraction
- [Gupta&al., 2015] use distributional vectors for fine-grained semantic attribute extraction
- [Bovi&al., 2015] perform knowledge base relation unification via sense embeddings and disambiguation
- [Su&al., 2015] use bilingual correspondence recursive autoencoder to model bilingual phrases in translation

Relevant Literature is Huge! (3)

Relevant papers at EMNLP'2015

- [Garcia-Duran&al.,2015] perform link prediction in knowledge bases by composing relationships with translations in the embedding space
- [Zhong&al., 2015] perform link predictions in KBs and relational fact extraction by aligning knowledge and text embeddings by entity descriptions
- [Batista&al., 2015] use word embeddings and bootstrapping for relation extraction
- [Luo&al, 2015] infer binary relation schemas for open information extraction
- [Lin&al., 2015] embed entities and relations using a path-based representation for knowledge base completion and relation extraction

Relevant Literature is Huge! (4)

Relevant papers at EMNLP'2015

- [Augenstein&al., 2015] extract relations between non-standard entities using distant supervision and imitation learning
- [Grycner&al., 2015] mine relational phrases and their hypernyms
- [Tuan&al., 2015] use trustiness and collective synonym/contrastive evidence into taxonomy construction
- [Mitra&Baral, 2015] extract relations to automatically solve logic grid puzzles
- [Seo&al., 2015] extract relations from text and visual diagrams to solve geometry problems

Relevant Literature is Huge! (5)

Relevant papers at EMNLP'2015

- [Gardner&Mitchell, 2015] extract relations using subgraph feature selection for knowledge base completion
- [Toutanova&al., 2015] learn joint embeddings of text and knowledge bases for knowledge base completion
- [Kloetzer&al., 2015] acquire entailment pairs of binary relations on a large-scale
- [Luo&al., 2015] present context-dependent knowledge graph embedding for link prediction and triple classification

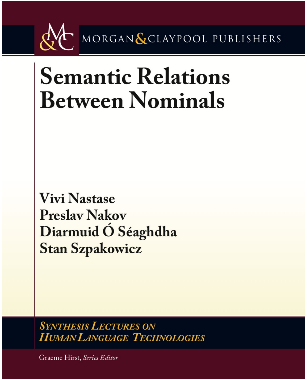
Relevant Literature is Huge! (6)

Relevant papers at EMNLP'2015

- and much more...
- ...and even more beyong EMNLP'2015...

Read the Book!

doi:10.2200/S00489ED1V01Y201303HLT019



Thank you!

Questions?

Bibliography I

Thomas Ahlswede and Martha Evens.
Parsing vs. text processing in the analysis of dictionary definitions.
In *Proc. 26th Annual Meeting of the Association for Computational Linguistics, Buffalo, NY, USA*, pages 217–224, 1988.

Hiyan Alshawi.
Processing dictionary definitions with phrasal pattern hierarchies.
Americal Journal of Computational Linguistics, 13(3):195–202, 1987.

Robert Amsler.
A taxonomy for English nouns and verbs.
In *Proc. 19th Annual Meeting of the Association for Computational Linguistics, Stanford University, Stanford, CA, USA*, pages 133–138, 1981.

Gabor Angeli, Julie Tibshirani, Jean Wu, and Christopher D. Manning.
Combining distant and partial supervision for relation extraction.
In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1556–1567, Doha, Qatar, October 2014. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/D14-1164`.

Isabelle Augenstein, Andreas Vlachos, and Diana Maynard.
Extracting relations between non-standard entities using distant supervision and imitation learning.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 747–757, 2015.

Bibliography II

Michele Banko, Michael Cafarella, Stephen Sonderland, Matt Broadhead, and Oren Etzioni.
Open information extraction from the Web.
In *Proc. 22nd Conference on the Advancement of Artificial Intelligence, Vancouver, BC, Canada*, pages 2670–2676, 2007.

Ken Barker and Stan Szpakowicz.
Semi-automatic recognition of noun modifier relationships.
In *Proc. 36th Annual Meeting of the Association for Computational Linguistics, Montréal, Canada*, pages 96–102, 1998.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski.
Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors.
In *Proc. of the Annual Meeting of the Association for Computational Linguistics*, pages 238–247, 2014.

David S. Batista, Bruno Martins, and Mário J. Silva.
Semi-supervised bootstrapping of relationship extractors with distributional semantics.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 499–504, 2015.

Brandon Beamer, Suma Bhat, Brant Chee, Andrew Fister, Alla Rozovskaya, and Roxana Girju.
UIUC: a knowledge-rich approach to identifying semantic relations between nominals.
In *Proc. 4th International Workshop on Semantic Evaluations (SemEval-1), Prague, Czech Republic*, pages 386–389, 2007.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin.
A neural probabilistic language model.
J. Mach. Learn. Res., 3:1137–1155, March 2003.

Bibliography III

Matthew Berland and Eugene Charniak.
Finding parts in very large corpora.
In *Proc. 37th Annual Meeting of the Association for Computational Linguistics, College Park, MD, USA*, pages 57–64, 1999.

Daniel M. Bikel, Richard Schwartz, and Ralph M. Weischedel.
An algorithm that learns what's in a name.
Machine Learning, 34(1-3):211–231, February 1999.
URL `http://dx.doi.org/10.1023/A:1007558221122`.

Christian Blaschke, Miguel A. Andrade, Christos Ouzounis, and Alfonso Valencia.
Automatic extraction of biological information from scientific text: protein-protein interactions.
In *Proc. 7th International Conference on Intelligent Systems for Molecular Biology (ISMB-99)*, Heidelberg, Germany, 1999.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan.
Latent Dirichlet allocation.
Journal of Machine Learning Research, 3:993–1022, 2003.

Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai.
Class-Based n-gram Models of Natural Language.
Computational Linguistics, 18:467–479, 1992.

Bibliography IV

Razvan Bunescu and Raymond J. Mooney.
A shortest path dependency kernel for relation extraction.
In *Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP-05)*, Vancouver, Canada, 2005.

Cristina Butnariu and Tony Veale.
A concept-centered approach to noun-compound interpretation.
In *Proc. 22nd International Conference on Computational Linguistics*, pages 81–88, Manchester, UK, 2008.

Michael Cafarella, Michele Banko, and Oren Etzioni.
Relational Web search.
Technical Report 2006-04-02, University of Washington, Department of Computer Science and Engineering, 2006.

Nicola Cancedda, Eric Gaussier, Cyril Goutte, and Jean-Michel Renders.
Word-sequence kernels.
Journal of Machine Learning Research, 3:1059–1082, 2003.
URL `http://jmlr.csail.mit.edu/papers/v3/cancedda03a.html`.

Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., and Tom M. Mitchell.
Coupled semi-supervised learning for information extraction.
In *Proc. Third ACM International Conference on Web Search and Data Mining (WSDM 2010)*, 2010.

Bibliography V

Joseph B. Casagrande and Kenneth Hale.
Semantic relationships in Papago folk-definition.
In Dell H. Hymes and William E. Bittleolo, editors, *Studies in southwestern ethnolinguistics*, pages 165–193. Mouton, The Hague and Paris, 1967.

Roger Chaffin and Douglas J. Herrmann.
The similarity and diversity of semantic relations.
Memory & Cognition, 12(2):134–141, 1984.

Eugene Charniak.
Toward a model of children's story comprehension.
Technical Report AITR-266 (hdl.handle.net/1721.1/6892), Massachusetts Institute of Technology, 1972.

Martin S. Chodorow, Roy Byrd, and George Heidorn.
Extracting semantic hierarchies from a large on-line dictionary.
In *Proc. 23th Annual Meeting of the Association for Computational Linguistics, Chicago, IL, USA*, pages 299–304, 1985.

Massimiliano Ciaramita, Aldo Gangemi, Esther Ratsch, Jasmin Šarić, and Isabel Rojas.
Unsupervised learning of semantic relations between concepts of a molecular biology ontology.
In *Proc. 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland*, pages 659–664, 2005.

Bibliography VI

Michael Collins and Nigel Duffy.
Convolution kernels for natural language.
In *Proc. 15th Conference on Neural Information Processing Systems (NIPS-01)*, Vancouver, Canada, 2001.
URL `http://books.nips.cc/papers/files/nips14/AA58.pdf`.

M. Craven and J. Kumlien.
Constructing biological knowledge bases by extracting information from text sources.
In *Proc. Seventh International Conference on Intelligent Systems for Molecular Biology*, pages 77–86, 1999.

Dmitry Davidov and Ari Rappoport.
Classification of semantic relationships between nominals using pattern clusters.
In *Proc. 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Columbus, OH, USA*, pages 227–235, 2008.

Ferdinand de Saussure.
Course in General Linguistics.
Philosophical Library, New York, 1959.
Edited by Charles Bally and Albert Sechehaye. Translated from the French by Wade Baskin.

Claudio Delli Bovi, Luis Espinosa Anke, and Roberto Navigli.
Knowledge base unification via sense embeddings and disambiguation.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 726–736, 2015.

Cícero Nogueira dos Santos, Bing Xiang, and Bowen Zhou.
Classifying relations by ranking with convolutional neural networks.
In *Proceedings of ACL-15*, Beijing, China, 2015.

Bibliography VII

Doug Downey, Oren Etzioni, and Stephen Soderland.
A probabilistic model of redundancy in information extraction.
In *Proc. 9th International Joint Conference on Artificial Intelligence, Edinburgh, UK*, pages 1034–1041, 2005.

Doug Downey, Oren Etzioni, and Stephen Soderland.
Analysis of a probabilistic model of redundancy in unsupervised information extraction.
Artificial Intelligence, 174(11):726–748, 2010.

Pamela Downing.
On the creation and use of English noun compounds.
Language, 53(4):810–842, 1977.

Jennifer D'Souza and Vincent Ng.
Sieve-based spatial relation extraction with expanding parse trees.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 758–768, 2015.

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates.
Unsupervised named-entity extraction from the web: an experimental study.
Artificial Intelligence, 165(1):91–134, June 2005.
ISSN 0004-3702.

Anthony Fader, Stephen Soderland, and Oren Etzioni.
Identifying relations for open information extraction.
In *Proc. Conference of Empirical Methods in Natural Language Processing (EMNLP '11)*, Edinburgh, Scotland, UK, July 27-31 2011.

Bibliography VIII

Christiane Fellbaum, editor.
WordNet – An Electronic Lexical Database.
MIT Press, 1998.

Timothy Finin.
The semantic interpretation of nominal compounds.
In *Proc. 1st National Conference on Artificial Intelligence, Stanford, CA, USA*, 1980.

Gottlob Frege.
Begriffschrift.
Louis Nebert, Halle, 1879.

Alberto Garcia-Duran, Antoine Bordes, and Nicolas Usunier.
Composing relationships with translations.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 286–290, 2015.

Jean Claude Gardin.
SYNTOL.
Graduate School of Library Service, Rutgers, the State University (Rutgers Series on Systems for the Intellectual Organization of Information, Susan Artandi, ed.), New Brunswick, New Jersey, 1965.

Matt Gardner and Tom Mitchell.
Efficient and expressive knowledge base completion using subgraph feature extraction.
In *Proc. Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2015.

Bibliography IX

Roxana Girju.
Improving the Interpretation of Noun Phrases with Cross-linguistic Information.
In *Proc. 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic*, pages 568–575, 2007.

Roxana Girju, Adriana Badulescu, and Dan Moldovan.
Learning semantic constraints for the automatic discovery of part-whole relations.
In *Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Edmonton, Alberta, Canada*, 2003.

Roxana Girju, Dan Moldovan, Marta Tatu, and Daniel Antohe.
On the semantics of noun compounds.
Computer Speech and Language, 19:479–496, 2005.

Matthew R. Gormley, Mo Yu, and Mark Dredze.
Improved relation extraction with feature-rich compositional embedding models.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1774–1784, 2015.

Adam Grycner, Gerhard Weikum, Jay Pujara, James Foulds, and Lise Getoor.
RELLY: Inferring hypernym relationships between relational phrases.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 971–981, 2015.

Abhijeet Gupta, Gemma Boleda, Marco Baroni, and Sebastian Padó.
Distributional vectors encode referential attributes.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 12–21, 2015.

Bibliography X

Xianpei Han and Le Sun.
Semantic consistency: A local subspace based method for distant supervised relation extraction.
In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 718–724, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/P14-2117`.

Roy Harris.
Reading Saussure: A Critical Commentary on the Cours le Linquistique Generale.
Open Court, La Salle, Ill., 1987.

Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka.
Task-oriented learning of word embeddings for semantic relation classification.
arXiv preprint arXiv:1503.00095, 2015.

Marti Hearst.
Automatic acquisition of hyponyms from large text corpora.
In *Proc. 15th International Conference on Computational Linguistics, Nantes, France*, pages 539–545, 1992.

Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum.
YAGO2: A spatially and temporally enhanced knowledge base from wikipedia.
Artif. Intell., 194:28–61, January 2013.

Raphael Hoffmann, Congle Zhang, and Daniel Weld.
Learning 5000 relational extractors.
In *Proc. 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden*, pages 286–295, 2010.

Bibliography XI

Nancy Ide, Jean Veronis, Susan Warwick-Armstrong, and Nicoletta Calzolari.
Principles for encoding machine-readable dictionaries.
In *Fifth Euralex International Congress*, pages 239–246, University of Tampere, Finland, 1992.

Jing Jiang and ChengXiang Zhai.
Instance Weighting for Domain Adaptation in NLP.
In *Proc. 45th Annual Meeting of the Association for Computational Linguistics*, ACL '07, pages 264–271, Prague, Czech Republic, 2007.
URL `http://www.aclweb.org/anthology/P07-1034`.

Karen Spärck Jones.
Synonymy and Semantic Classification.
PhD thesis, University of Cambridge, 1964.

Su Nam Kim and Timothy Baldwin.
Automatic Interpretation of noun compounds using WordNet::Similarity.
In *Proc. 2nd International Joint Conference on Natural Language Processing, Jeju Island, South Korea*, pages 945–956, 2005.

Su Nam Kim and Timothy Baldwin.
Interpreting semantic relations in noun compounds via verb semantics.
In *Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, Sydney, Australia*, pages 491–498, 2006.

Bibliography XII

Judith L. Klavans, Martin S. Chodorow, and Nina Wacholder.
Building a knowledge base from parsed definitions.
In George Heidorn, Karen Jensen, and Steve Richardson, editors, *Natural Language Processing: The PLNLP Approach*. Kluwer, New York, NY, USA, 1992.

Julien Kloetzer, Kentaro Torisawa, Chikara Hashimoto, and Jong-Hoon Oh.
Large-scale acquisition of entailment pattern pairs by exploiting transitivity.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1649–1655, 2015.

Zornitsa Kozareva and Eduard Hovy.
A Semi-Supervised Method to Learn and Construct Taxonomies using the Web.
In *Proc. 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, USA*, pages 1110–1118, 2010.

Zornitsa Kozareva, Ellen Riloff, and Eduard Hovy.
Semantic class learning from the Web with hyponym pattern linkage graphs.
In *Proc. 46th Annual Meeting of the Association for Computational Linguistics ACL-08: HLT*, pages 1048–1056, 2008.

Sebastian Krause, Hong Li, Hans Uszkoreit, and Feiyu Xu.
Large-scale learning of relation-extraction rules with distant supervision from the web.
In *Proc. International Conference on The Semantic Web*, pages 263–278, 2012.

Bibliography XIII

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira.
Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
In *Proc. Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
ISBN 1-55860-778-1.
URL http://dl.acm.org/citation.cfm?id=645530.655813.

Maria Lapata.
The disambiguation of nominalizations.
Computational Linguistics, 28(3):357–388, 2002.

Mirella Lapata and Frank Keller.
The Web as a baseline: Evaluating the performance of unsupervised Web-based models for a range of NLP tasks.
In *Proc. Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 121–128, Boston, USA, 2004.

Mark Lauer.
Designing Statistical Language Learners: Experiments on Noun Compounds.
PhD thesis, Macquarie University, 1995.

Judith N. Levi.
The Syntax and Semantics of Complex Nominals.
Academic Press, New York, 1978.

Bibliography XIV

Omer Levy and Yoav Goldberg.
Dependency-based word embeddings.
In *Proc. 52nd Annual Meeting of the Association for Computational Linguistics*, pages 302–308, 2014a.

Omer Levy and Yoav Goldberg.
Linguistic regularities in sparse and explicit word representations.
In *Proc. Conference on Computational Natural Language Learning*, pages 171–180, 2014b.

Jiwei Li and Dan Jurafsky.
Do multi-sense embeddings improve natural language understanding?
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1722–1732, 2015.

Jiwei Li, Thang Luong, Dan Jurafsky, and Eduard Hovy.
When are tree structures necessary for deep learning of representations?
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 2304–2314, 2015.

Dekang Lin.
An information-theoretic definition of similarity.
In *Proc. International Conference on Machine Learning*, pages 296–304, 1998.

Dekang Lin and Patrick Pantel.
Discovery of inference rules for question-answering.
Natural Language Engineering, 7(4):343–360, 2001.
ISSN 1351-3249.

Bibliography XV

Thomas Lin, Mausam, and Oren Etzioni.
Identifying functional relations in web text.
In *Proc. 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1266–1276, Cambridge, MA, October 2010.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu.
Modeling relation paths for representation learning of knowledge bases.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 705–714, 2015.

Xiao Ling, Peter Clark, and Daniel S. Weld.
Extracting meronyms for a biology knowledge base using distant supervision.
In *Proceedings of Automated Knowledge Base Construction (AKBC) 2013: The 3rd Workshop on Knowledge Extraction at CIKM 2013*, San Francisco, CA, October 27-28 2013.

Kangqi Luo, Xusheng Luo, and Kenny Zhu.
Inferring binary relation schemas for open information extraction.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 555–560, 2015a.

Yuanfei Luo, Quan Wang, Bin Wang, and Li Guo.
Context-dependent knowledge graph embedding.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1656–1661, 2015b.

Tuan Luu Anh, Jung-jae Kim, and See Kiong Ng.
Incorporating trustiness and collective synonym/contrastive evidence into taxonomy construction.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1013–1022, 2015.

Bibliography XVI

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni.
Open language learning for information extraction.
In *Proc. 2012 Conference on Empirical Methods in Natural Language Processing, Jeju Island, Korea*, pages 523–534, 2012.

Andrew McCallum and Wei Li.
Early results for Named Entity Recognition with Conditional Random Fields, feature induction and Web-enhanced lexicons.
In *Proc. 7th Conference on Natural Language Learning at HLT-NAACL 2003 – Volume 4*, CONLL '03, pages 188–191, 2003.
doi: 10.3115/1119176.1119206.
URL http://dx.doi.org/10.3115/1119176.1119206.

John McCarthy.
Programs with common sense.
In *Proc. Teddington Conference on the Mechanization of Thought Processes*, 1958.

Ryan McDonald, Fernando Pereira, Seth Kulik, Scott Winters, Yang Jin, and Pete White.
Simple Algorithms for Complex Relation Extraction with Applications to Biomedical IE.
In *Proc. 43rd Annual Meeting of the Association for Computational Linguistics (ACL-05)*, Ann Arbor, MI, 2005.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean.
Distributed representations of words and phrases and their compositionality.
In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. 2013a.

Bibliography XVII

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig.
Linguistic regularities in continuous space word representations.
In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751, Atlanta, Georgia, 2013b.

Bonan Min, Ralph Grishman, Li Wan, Chang Wang, and David Gondek.
Distant supervision for relation extraction with an incomplete knowledge base.
In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 777–782, Atlanta, Georgia, June 2013. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/N13-1095`.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky.
Distant supervision for relation extraction without labeled data.
In *Proc. Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2*, ACL '09, pages 1003–1011, 2009.

Arindam Mitra and Chitta Baral.
Learning to automatically solve logic grid puzzles.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1023–1033, 2015.

Thahir Mohamed, Estevam Hruschka Jr., and Tom Mitchell.
Discovering relations between noun categories.
In *Proc. 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK*, pages 1447–1455, 2011.

Bibliography XVIII

Dan Moldovan, Adriana Badulescu, Marta Tatu, Daniel Antohe, and Roxana Girju.
Models for the semantic classification of noun phrases.
In *Proc. HLT-NAACL Workshop on Computational Lexical Semantics*, pages 60–67. Association for Computational Linguistic, 2004.

Alessandro Moschitti.
Efficient convolution kernels for dependency and constituent syntactic trees.
Proc. 17th European Conference on Machine Learning (ECML-06), 2006.
URL `http://dit.unitn.it/~moschitt/articles/ECML2006.pdf`.

Preslav Nakov.
Improved Statistical Machine Translation using monolingual paraphrases.
In *Proc. 18th European Conference on Artificial Intelligence, Patras, Greece*, pages 338–342, 2008.

Preslav Nakov and Marti Hearst.
UCB: System description for SemEval Task #4.
In *Proc. 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pages 366–369, Prague, Czech Republic, 2007.

Preslav Nakov and Marti Hearst.
Solving relational similarity problems using the Web as a corpus.
In *Proc. 6th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Columbus, OH, USA*, pages 452–460, 2008.

Bibliography XIX

Preslav Nakov and Zornitsa Kozareva.
Combining relational and attributional similarity for semantic relation classification.
In *Proc. International Conference on Recent Advances in Natural Language Processing, Hissar, Bulgaria*, pages 323–330, 2011.

Vivi Nastase and Stan Szpakowicz.
Exploring noun-modifier semantic relations.
In *Proc. 6th International Workshop on Computational Semantics, Tilburg, The Netherlands*, pages 285–301, 2003.

Roberto Navigli and Simone Paolo Ponzetto.
BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network.
Artificial Intelligence, 193:217–250, 2012.

Thien Huu Nguyen and Ralph Grishman.
Employing word representations and regularization for domain adaptation of relation extraction.
In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 68–74, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/P14-2012`.

Diarmuid Ó Séaghdha and Ann Copestake.
Semantic classification with distributional kernels.
In *Proc. 22nd International Conference on Computational Linguistics*, pages 649–656, Manchester, UK, 2008.

Bibliography XX

Marius Paşca.
Organizing and searching the World-Wide Web of facts – step two: harnessing the wisdom of the crowds.
In *16th International World Wide Web Conference, Banff, Canada*, pages 101–110, 2007.

Sebastian Padó and Mirella Lapata.
Dependency-based construction of semantic space models.
Computational Linguistics, 33(2):161–199, 2007.

Martha Palmer, Daniel Gildea, and Nianwen Xue.
Semantic Role Labeling.
Synthesis Lectures on Human Language Technologies. Morgan & Claypool, 2010.

Patrick Pantel and Dekang Lin.
Discovering word senses from text.
In *Proc. 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada*, pages 613–619, 2002.

Patrick Pantel and Deepak Ravichandran.
Automatically labeling semantic classes.
In *Proc. Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Boston, MA, USA*, pages 321–328, 2004.

Bibliography XXI

Siddharth Patwardhan and Ellen Riloff.
Effective information extraction with semantic affinity patterns and relevant regions.
In *Proc. 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Language Learning, Prague, Czech Republic*, pages 717–727, 2007.

Charles Sanders Peirce.
Existential graphs (unpublished 1909 manuscript).
In Justus Buchler, editor, *The philosophy of Peirce: selected writings*. Harcourt, Brace & Co., 1940.

Jeffrey Pennington, Richard Socher, and Christopher Manning.
Glove: Global vectors for word representation.
In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, 2014.

Maria Pershina, Bonan Min, Wei Xu, and Ralph Grishman.
Infusion of labeled data into distant supervision for relation extraction.
In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 732–738, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/P14-2119`.

James Pustejovsky, José M. Castaño, Jason Zhang, M. Kotecki, and B. Cochran.
Robust relational parsing over biomedical literature: Extracting *inhibit* relations.
In *Proc. 7th Pacific Symposium on Biocomputing (PSB-02)*, Lihue, HI, USA, 2002.

Bibliography XXII

M. Ross Quillian.
A revised design for an understanding machine.
Mechanical Translation, 7:17–29, 1962.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik.
A comprehensive grammar of the English language.
Longman, 1985.

Deepak Ravichandran and Eduard Hovy.
Learning surface text patterns for a Question Answering system.
In *Proc. 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA< USA*, pages 41–47, 2002.

Sebastian Riedel, Limin Yao, and Andrew McCallum.
Modeling relations and their mentions without labeled text.
In *Proc. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD '10)*, volume 6232 of *Lecture Notes in Computer Science*, pages 148–163. Springer, 2010.

Bryan Rink and Sanda Harabagiu.
UTD: Classifying Semantic Relations by Combining Lexical and Semantic Resources.
In *Proc. 5th International Workshop on Semantic Evaluation*, pages 256–259, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
URL http://www.aclweb.org/anthology/S10-1057.

Bibliography XXIII

Barbara Rosario and Marti Hearst.
Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy.
In *Proc. 2001 Conference on Empirical Methods in Natural Language Processing, Pittsburgh, PA< USA*, pages 82–90, 2001.

Barbara Rosario and Marti Hearst.
The descent of hierarchy, and selection in relational semantics.
In *Proc. 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA*, pages 247–254, 2002.

Minjoon Seo, Hannaneh Hajishirzi, Ali Farhadi, Oren Etzioni, and Clint Malcolm.
Solving geometry problems: Combining text and diagram interpretation.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1466–1476, 2015.

Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng.
Semantic compositionality through recursive matrix-vector spaces.
In *Proc. 2012 Conference on Empirical Methods in Natural Language Processing*, Jeju, Korea, 2012.

Vivek Srikumar and Dan Roth.
Modeling semantic relations expressed by prepositions.
2013.

Jinsong Su, Deyi Xiong, Biao Zhang, Yang Liu, Junfeng Yao, and Min Zhang.
Bilingual correspondence recursive autoencoder for statistical machine translation.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1248–1258, 2015.

Bibliography XXIV

Le Sun and Xianpei Han.
A feature-enriched tree kernel for relation extraction.
In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 61–67, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
URL `http://www.aclweb.org/anthology/P14-2011`.

Jun Suzuki, Tsutomu Hirao, Yutaka Sasaki, and Eisaku Maeda.
Hierarchical directed acyclic graph kernel: Methods for structured natural language data.
In *Proce. 41st Annual Meeting of the Association for Computational Linguistics (ACL-03)*, Sapporo, Japan, 2003.

Don R. Swanson.
Two medical literatures that are logically but not bibliographically connected.
JASIS, 38(4):228–233, 1987.

Diarmuid Ó Séaghdha.
Designing and Evaluating a Semantic Annotation Scheme for Compound Nouns.
In *Proc. 4th Corpus Linguistics Conference (CL-07)*, Birmingham, UK, 2007.
URL `www.cl.cam.ac.uk/~do242/Papers/dos_cl2007.pdf`.

Diarmuid Ó Séaghdha and Ann Copestake.
Co-occurrence contexts for noun compound interpretation.
In *Proc. ACL Workshop on A Broader Perspective on Multiword Expressions*, pages 57–64. Association for Computational Linguistics, 2007.

Bibliography XXV

Lucien Tesnière.
Éléments de syntaxe structurale.
C. Klincksieck, Paris, 1959.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon.
Representing text for joint embedding of text and knowledge bases.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1499–1509, 2015.

Stephen Tratz and Eduard Hovy.
A taxonomy, dataset, and classifier for automatic noun compound interpretation.
In *Proc. 48th Annual Meeting of the Association for Computational Linguistics*, pages 678–687, Uppsala, Sweden, 2010.

Peter Turney.
Similarity of semantic relations.
Computational Linguistics, 32(3):379–416, 2006.

Peter Turney and Michael Littman.
Corpus-based learning of analogies and semantic relations.
Machine Learning, 60(1-3):251–278, 2005.

Lucy Vanderwende.
Algorithm for the automatic interpretation of noun sequences.
In *Proc. 15th International Conference on Computational Linguistics, Kyoto, Japan*, pages 782–788, 1994.

Bibliography XXVI

Beatrice Warren.
Semantic Patterns of Noun-Noun Compounds.
In *Gothenburg Studies in English 41, Goteburg, Acta Universtatis Gothoburgensis*, 1978.

Joseph Weizenbaum.
ELIZA – a computer program for the study of natural language communication between man and machine.
Communications of the ACM, 9(1):36–45, 1966.

Terry Winograd.
Understanding natural language.
Cognitive Psychology, 3(1):1–191, 1972.

Fei Wu and Daniel S. Weld.
Autonomously Semantifying Wikipedia.
In *Proc. ACM 17th Conference on Information and Knowledge Management (CIKM 2008), Napa Valley, CA, USA*, pages 41–50, 2007.

Fei Wu and Daniel S. Weld.
Open information extraction using Wikipedia.
In *Proc. 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden*, pages 118–127, 2010.

Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao.
Semantic relation classification via convolutional neural networks with simple negative sampling.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 536–540, 2015a.

Bibliography XXVII

Wei Xu, Raphael Hoffmann, Le Zhao, and Ralph Grishman.
Filling knowledge base gaps for distant supervision of relation extraction.
In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 665–670, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
URL http://www.aclweb.org/anthology/P13-2117.

Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin.
Classifying relations via long short term memory networks along shortest dependency paths.
In *Proc. Conference on Empirical Methods in Natural Language Processing*, pages 1785–1794, 2015b.

Alexander Yates and Oren Etzioni.
Unsupervised resolution of objects and relations on the Web.
In *Proc. Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, USA*, pages 121–130, 2007.

Mo Yu, Matthew R. Gormley, and Mark Dredze.
Factor-based compositional embedding models.
In *The NIPS 2014 Learning Semantics Workshop*, 2014.

Mo Yu, Matthew R. Gormley, and Mark Dredze.
Combining word embeddings and feature embeddings for fine-grained relation extraction.
In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1374–1379, Denver, Colorado, May–June 2015. Association for Computational Linguistics.
URL http://www.aclweb.org/anthology/N15-1155.

Bibliography XXVIII

Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella.
Kernel methods for relation extraction.
Journal of Machine Learning Research, 3:1083–1106, 2003.

Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao.
Relation classificaiton via convolutional deep neural network.
In *Proceedings of COLING-14*, Dublin, Ireland, 2014.

Guo Dong Zhou, Min Zhang, Dong Hong Ji, and Qiao Ming Zhu.
Tree kernel-based relation extraction with context-sensitive structured parse tree information.
In *Proc. 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-07)*, pages 728–736, Prague, Czech Republic, 2007.

Guo Dong Zhou, Long Hua Qian, and Qiao Ming Zhu.
Label propagation via bootstrapped support vectors for semantic relation extraction between named entities.
Computer Speech and Language, 23(4):464–478, 2009.

Karl E. Zimmer.
Some General Observations about Nominal Compounds.
Working Papers on Language Universals, Stanford University, 5, 1971.