Transitivity in Semantic Relation Learning

Francesca Fallucchi
Università Telematica
“G. Marconi”

Fabio Massimo Zanzotto
Department of Computer Science
University of Rome “Tor Vergata”

Motivation

- **Semantic networks** are applied to understand texts in a variety of NLP and KE applications
- *Manually built* **semantic networks** often lack coverage when used in generic and in specific knowledge domains

Common Goal in NLP and KE

Learning semantic networks from text collections

Observations

- Semantic resources are ultimately exploited in natural language understanding systems as ***networks of words***
- Semantic networks *have structural properties* of the target relations such as **transitivity**, simmetry, and so on.

Background

Learning semantic networks from text collections

Two basic classes of models:

- Distributional Hypothesis (DH) (Harris, 1964)

  "Words that tend to occur in the same contexts tend to have similar meanings."

  $\text{sim}(w_1,w_2) \approx \text{sim}(C(w_1), C(w_2))$

- Lexico-Syntactic Patterns (LSP) (originally used in (Robison, 1970))

  "It is possible to extract relevant semantic relations with some pattern."

  w_1 is in a relation **r** with w_2 if the context ***pattern***(w_1, w_2) is significantly frequent in the corpus

Background

Extensions on Distributional Models

Distributional Inclusion Hypothesis (DIH) (Geffet&Dagan, 2005) and **Formal Concept Analysis** (FCA) (as used in (Cimiano, 2005))

> "the word *a* is the generalization of the word *b* if the properties representing the contexts of *a* are included in those representing the contexts of *b*"

Some particular Lexico-Syntactic Patterns

- Heart's ISA patterns, e.g., "X as well as Y" →X ISA Y
- Player wins → win entails play (Zanzotto et al., 2006)

Background

Summing up

- Distributional Models (DH)

Target Relations	Hyperonymy (IS_A) Cotopy (Similarity)	✗
Use of structural properties	***Transitivity*** is implicitly exploited	✓

- Lexico-Syntactic Pattern Models (LSP)

Target Relations	All possible semantic relations	✓
Use of structural properties	***Transitivity*** is ***NOT*** exploited	✗

Our research Goal

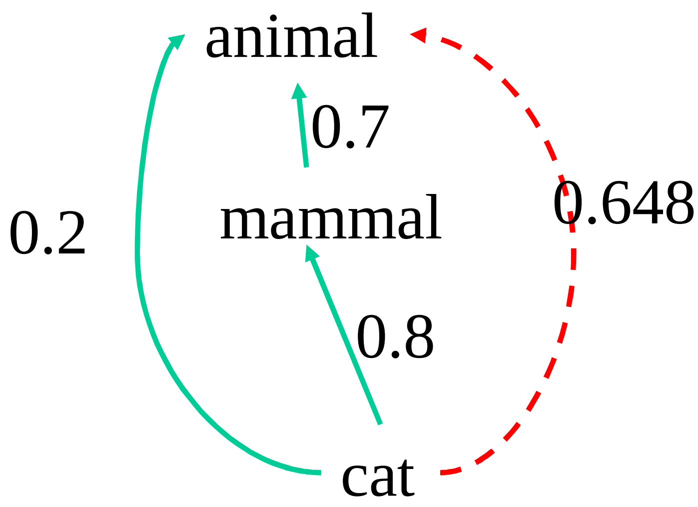
Target Relations	All possible semantic relations
Use of structural properties	***Transitivity*** is effectively exploited

Exploiting **Transitivity** within **Lexico-Syntactic Pattern Models**

- we exploit structural properties of target relations to determine the probability
- we focus on the transitivity to reinforce or lower the probability

Our research Goal

isa relation



$$P(R_{animal,cat} \mid E)$$

Direct Probabilities for Corpus Observation (E) with Lexico-Syntactic Patterns

$$P(\hat{R}_{animal,cat} \mid E)$$

Induced Probabilities

Outline of the rest of the talk

- Probability Propagation in a Transitive Semantic Network
  - Direct Probabilistic Model
  - Induced Probabilistic Model
    - Induced Intensional Model
    - Induced Extensional Model
- Experimental Evaluation
  - Experimental Set-Up
  - Result
- Conclusions and Future Works

Direct Probabilistic Model

Given:

- A set of training examples **R** representing a relation **T**
- A corpus **C** and a set of evidences **E** over the examples **R** extracted from the corpus **C**

We want to extract the probability distribution

$$P(R_{i,j} \in T|E)$$

where $R_{i,j} \in T$ stands for ***i is in relation T with j.***

e.g., $R_{animal,cat}$ where R is "generalizes" stands for:

animal generalizes a cat or a *cat is an animal*

Direct Probabilistic Model

… **dog**, as other **animals**,…

(dog,animal)

Corpus

, as other	1
, as	1
as other	1
,	1
as	1
other	1

Direct Probabilistic Model

$\ldots\ \mathbf{X_1}\ w_1\ w_2\ \ldots\ w_k\ \mathbf{Y_1}, \ldots$

Corpus

$$
\begin{array}{c}
 \\
f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ \\ f_n
\end{array}
\begin{array}{c}
\begin{array}{cccc}
(\mathbf{X_1},\mathbf{Y_1}) & (\mathbf{X_2},\mathbf{Y_2}) & \quad\quad & (\mathbf{X_m},\mathbf{Y_m})
\end{array} \\
\left[\begin{array}{cccc}
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
 & & & \\
\bullet & \bullet & & \bullet
\end{array}\right]
\end{array}
= E
$$

Evidence Matri

Direct Probabilistic Model

We estimate

$$P(R_{i,j} \in T|E)$$

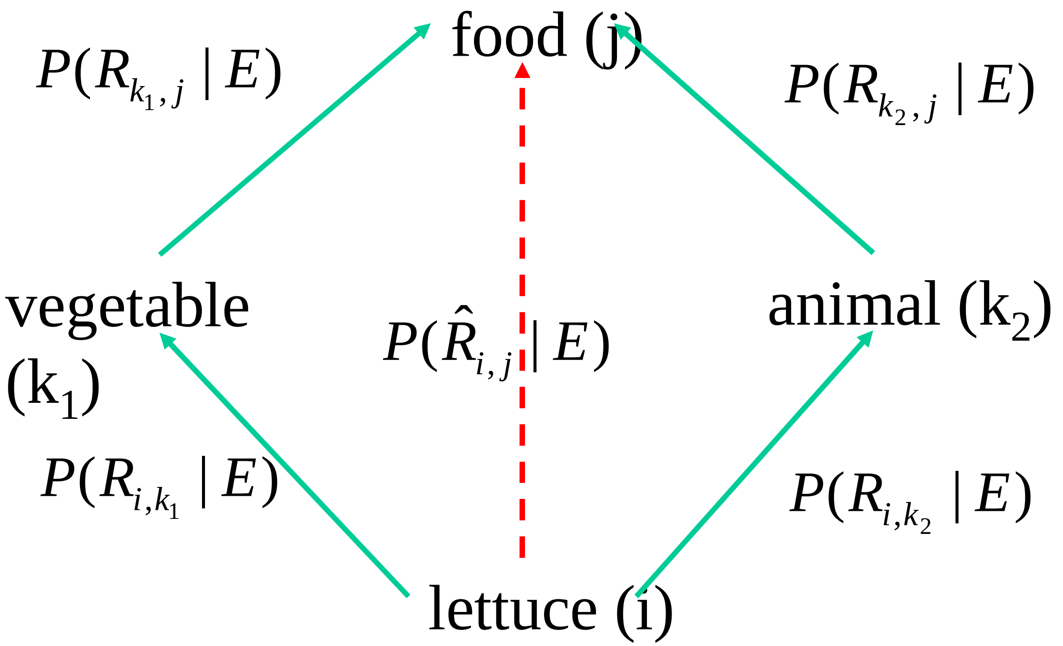
using logistic regression as:

- $R_{i,j} \in T$ *is a variable with two values (yes,no)*
- $P(R_{i,j} \in T|E) = P(R_{i,j} \in T|e_{i,j})$ where $e_{i,j}$ are the evidences E related to the pair (i,j)

Induced Probabilistic Model

- Goal: Exploiting direct probabilities and transitivity over the building semantic network to extract induced probabilies
- We prospose three models:
  - Induced Intensional Model
  - Induced Extensional Model
  - Mixed Induced Model

Induced Intensional Probabilistic Model



$$P(\hat{R}_{i,j} \mid E) = P(R_{i,j} \cup (R_{i,k_1} \cap R_{k_1,j}) \cup (R_{i,k_2} \cap R_{k_2,j}) \mid E)$$

Induced Intensional Probabilistic Model

$$P(\hat{R}_{i,j} \mid E) = P(R_{i,j} \cup (R_{i,k_1} \cap R_{k_1,j}) \cup (R_{i,k_2} \cap R_{k_2,j}) \mid E) =$$

(inclusion-exclusion principle)

$$= P(R_{i,j} \mid E) + P(R_{i,k_1} \cap R_{k_1,j} \mid E) + P(R_{i,k_2} \cap R_{k_2,j} \mid E) +$$
$$- P(R_{i,j} \cap R_{i,k_1} \cap R_{k_1,j} \mid E) - P(R_{i,j} \cap R_{i,k_2} \cap R_{k_2,j} \mid E) +$$
$$+ P(R_{i,j} \cap R_{i,k_1} \cap R_{k_1,j} \cap R_{i,k_2} \cap R_{k_2,j} \mid E)$$

($R_{i,j}$ indepence assumption)

$$= P(R_{i,j} \mid E) + P(R_{i,k_1} \mid E) P(R_{k_1,j} \mid E) + P(R_{i,k_2} \mid E) P(R_{k_2,j} \mid E) +$$
$$- P(R_{i,j} \mid E) P(R_{i,k_1} \mid E) P(R_{k_1,j} \mid E) - P(R_{i,j} \mid E) P(R_{i,k_2} \mid E) P(R_{k_2,j} \mid E) +$$
$$+ P(R_{i,j} \mid E) P(R_{i,k_1} \mid E) P(R_{k_1,j} \mid E) P(R_{i,k_2} \mid E) P(R_{k_2,j} \mid E)$$

Induced Intensional Probabilistic Model

General Equation

$$P(\widehat{R}_{i,j}|E)=P(R_{i,j}\cup\bigcup_{k\in K}(R_{i,k}\cap R_{k,j})|E)$$

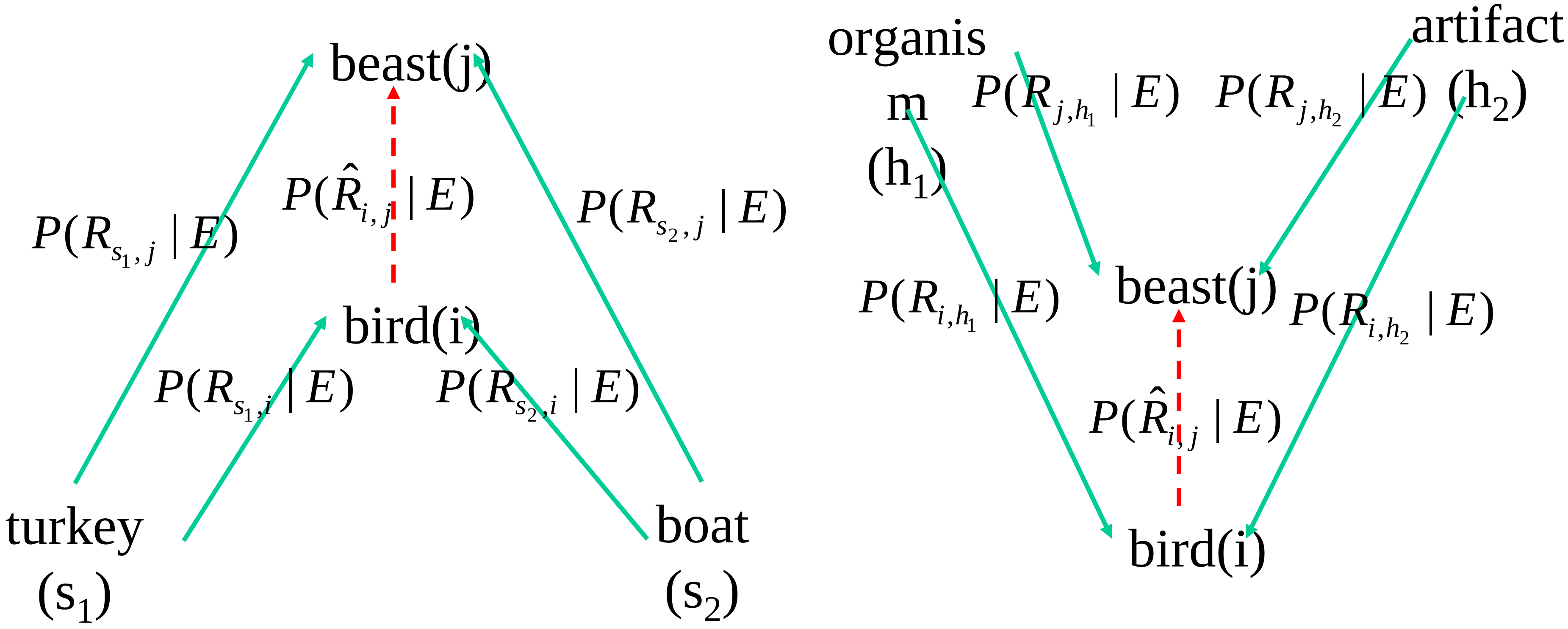
(inclusion-exclusion principle)

$$P(\widehat{R}_{i,j}|E)=\sum_{\emptyset=J\subseteq\{\epsilon,k_1,\ldots,k_n\}}(-1)^{|J|-1}P(R_J|E)$$

($R_{i,j}$ indepence assumption)

$$P(R_J|E) = \prod_{k\in J} P(R_k|E)$$

Induced Extensional Probabilistic Model



$$P(\hat{R}_{i,j} | E) = P(R_{i,j} \cup (R_{s_1,i} \cap R_{s_1,j}) \cup (R_{s_2,i} \cap R_{s_2,j}) \cup (R_{i,h_1} \cap R_{j,h_1}) \cup (R_{i,h_2} \cap R_{j,h_2}) | E)$$

Experimental Investigation

Question:

- Can our induced models effectively exploit transitivity when expanding or building semantic networks of words?

- Experimental set-up
- Results

Experimental Set-up

- Two target relations: generalization and meronymy (part-of)
- Corpus: English Web as Corpus, **ukWaC** (Ferraresi et al.,2008)
- Source of training and testing examples: WordNet

Test		*Set Description*	*Initial Size*	*Retreived Pairs*
isa	TR_p	$\mathcal{H}/\mathcal{H}_{ts}$	1983197	212076
	TR_n	$\overline{\mathcal{H}}/\overline{\mathcal{H}}_{ts}$	5594387	315428
	TS_p	$\mathcal{H}_{ts}$	506	150
	TS_n	$\overline{\mathcal{H}}_{ts}$	80436	258
partof	TR_p	$\mathcal{M}/\mathcal{M}_{ts}$	14333	8077
	TR_n	$\overline{\mathcal{H}}/\mathcal{M}_{ts}$	623616	318679
	TS_p	$\mathcal{M}_{ts}$	408	101
	TS_n	$\overline{\mathcal{M}}_{ts}$	34214	1713

Experimental Set-up

Two different regressors:

- Support Vector Machines

- Moore-Penrose pseudoinverse Q+ (Penrose, 1955)

$$Q^{+} = U\Sigma^{+}V^{T}$$

This second model gives the possibility of naturally using SVD as feature selector (Fallucchi&Zanzotto, 2009)

Results

Relative recall of the three methods

	direct		*intensional*		*extensional*	
	PI	SVM	PI	SVM	PI	SVM
100	30.67	30.00	4.00	1.33	37.33	35.33
200	56.67	49.33	27.33	26.00	60.67	61.33
300	74.67	74.67	64.00	65.33	81.33	78.67

	direct		*intensional*		*extensional*	
	PI	SVM	PI	SVM	PI	SVM
500	28.71	28.71	32.67	32.67	33.66	33.66
1000	44.55	70.30	54.46	70.3	49.5	72.28

Conclusions and Future Work

- We presented a model to exploit structural properties of the relation during semantic network learing
- The obtained resource contains semantic relations between words along with their probability obtained on a corpus
- Open issues:
  - Evaluation of the obtained resource with respect to a task: word sense disambiguation, textual entailment recognition, etg.