بسم الله الرحمن الرحيم

Discovery of Inference Rules, Their Selectional Preferences and Directionality

Ehsan Khoddammohammadi

Computational Semantics, 2012

Saarland University

Chapters

1. Discovery of Inference Rules from Text
2. Inferential Selectional Preferences
3. Learning Directionality of Inference Rules

Chapter 1:

Discovery of Inference Rules from Text
(DIRT)

Outline

- What is Relation?
- What is Inference Rule and why is it important?
- What is Unsupervised discovery of Inference Rules?
- How is DIRT working?
- How is performance of DIRT?

Inference Rules

- Relations
  - X is parent of Y
  - Y is child of X
  - X wrote Y
  - X is the author of Y

- Inference Rules
  - X is parent of Y $\Leftrightarrow$ Y is child of X
  - S eats T $\Rightarrow$ S likes T
  - X wrote Y $\Leftrightarrow$ X is the author of Y

What kind of Inference Rules could DIRT find?

- Mostly Paraphrases
  - X is author of $Y \Leftrightarrow X$ wrote Y

- And also other type of Rules
  - X manufactures $Y \Leftrightarrow X'$s Y factory

What is the Application of Finding Inference Rules?

- In Information Retrieval

- In Question/Answering

- In Summarization

Outline Of The Algorithm

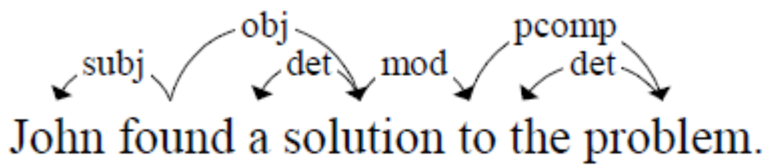
Objective: given a relation, find its paraphrases

Solution:

1. Extract dependency paths that share at least one feature with the relation
2. Prune them
3. For each candidate count common features and calculate the similarity
4. Report most K-similar paths as paraphrases

An Extension to Harris' Distributional Hypothesis

- Internal structure of sentences could be shown by dependency relations.



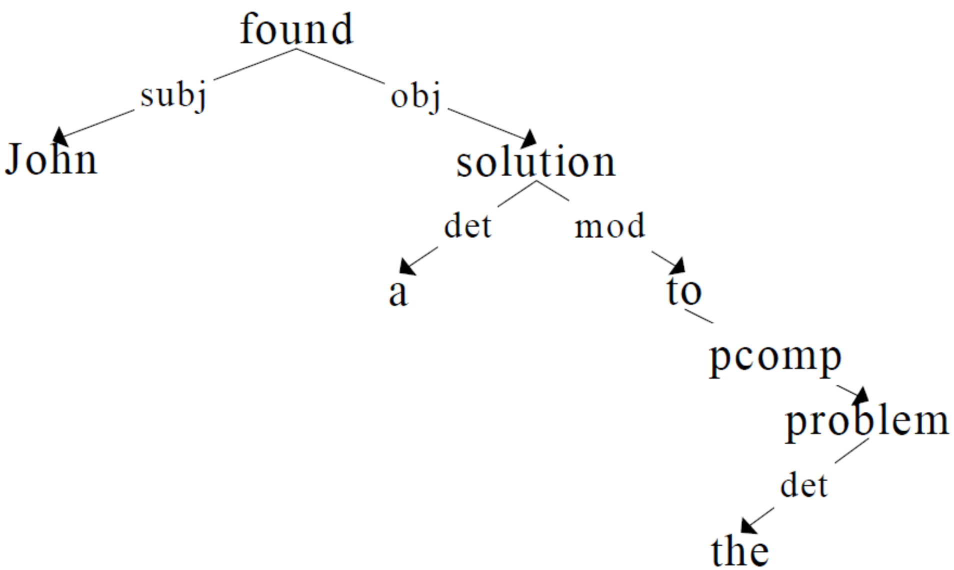
- we can assign meaning to paths.
- We can formulate our task to "finding paths with similar meaning".

Extended Distributional Hypothesis:

If two paths tend to occur in similar contexts, the meanings of the path tend to be similar.

Definition of path

John found a solution to the problem.



$X\ finds\ solution\ to\ Y$

$N\!:subj\!:V \leftarrow find \rightarrow V\!:obj\!:N \rightarrow solution \rightarrow N\!:to\!:N$

Pruning Dependency Trees

- Slot fillers must be nouns
- Dependency relation should connect two content words
- Frequency count of an internal relation must be greater than a certain threshold

An Example

"X finds a solution to Y"		*"X solves Y"*	
SlotX	*SlotY*	*SlotX*	*SlotY*
commission	strike	committee	problem
committee	civil war	clout	crisis
committee	crisis	government	problem
government	crisis	he	mystery
government	problem	she	problem
he	problem	petition	woe
legislator	budget deficit	researcher	mystery
sheriff	dispute	sheriff	murder

A Short Reminder about Pointwise Mutual Information

- What is Pointwise Mutual Information?

$$pmi(x,y) = log\frac{P(x,y)}{P(x)P(y)}$$

- Here:

$$mi(p,slot,w) = log\frac{P(p,Slot,w)}{P(Slot)P(p|Slot)P(w|Slot)}$$

How to measure similarity of paths?

- Similarity function for slots:

$$sim(slot_1, slot_2) = \frac{\sum_{w \in T(p_1,s) \cap T(p_2,s)} mi(p_1, s, w) + mi\ (p_2, s, w)}{\sum_{w \in T(p_1,s)} mi(p_1, s, w) + \sum_{w \in T(p_2,s)} mi(p_2, s, w)}$$

- Similarity function for relations:

$$S(p_1, p_2) = \sqrt{sim(SlotX_1, SlotX_2) \times sim(SlotY_1, SlotY_2)}$$

How good is it working?

- Compared to human-generated paraphrases of the first six questions in the TREC-8 QA:

Q#	PATHS	MAN.	DIRT	INT.	ACC.
Q_1	X is author of Y	7	21	2	52.5%
Q_2	X is monetary value of Y	6	0	0	N/A
Q_3	X manufactures Y	13	37	4	92.5%
Q_4	X spend Y	7	16	2	40.0%
	spend X on Y	8	15	3	37.5%
Q_5	X is managing director of Y	5	14	1	35.0%
Q_6	X asks Y	2	23	0	57.5%
	asks X for Y	2	14	0	35.0%
	X asks for Y	3	21	3	52.5%

Chapter 2:

Learning Inferential Selectional Preferences (ISP)

Learning Inferential Selectional Preferences (ISP)

- Main motivation:
  - Improve automatic discovery of inference rules (DIRT)

- Approach:
  - Filtering out incorrect IR that their arguments are not in the same semantic classes that a relation imposes on.

- Resources:
  - Relies on having a bank of semantic classes:
    - Manual collections: WordNet, FrameNet,…
    - Automatic induction: CBC,…

Task definition

- For each binary relation determine which semantic classes are valid to be its arguments
- <X,p,Y>, p=relation, X,Y=arguments
- c(X) and c(Y) are valid semantic classes for relation p

$$Given\ an\ inference\ rule\ p_i \Rightarrow p_j\ and\ the\ instance$$
$$< x, p_i, y >, our\ task\ is\ to\ determine$$
$$if < x, p_j, y > is\ valid.$$

Outline of method

- Make a repository of semantic classes
- Extract relational SP for each relation
- Decide whether two IR shares same SP

How to create semantic classes repository?

Either

- Run CBC clustering algorithm on a corpus and used generated noun concepts

Or

- Extract semantic classes from the WordNet by truncating the noun synset hierarchy.

How to extract Relational Selectional Preferences for a Rule?

- For each instance of p increase the count of correspondent semantic class of its arguments:
  - <x,p,y> → ++ <c(x),p,c(y)> (JRM)
  - <x,p,y> → ++ <c(x),p,*> and ++ <*,p,c(y)> (IRM)
- Rank based on the strength of association between C(x) and C(y):
  - pmi of c(x) and c(y) given p (JRM)
  - Rank based on conditional probability of c(x) or c(y) given p (IRM)

How to find Inferential Selectional Preferences?

- Given two relations p_i and p_j we want to find common SPs
- For each inference rule like $p_i \Rightarrow p_j$:
  - find the intersection of relational SP of p_i and p_j.
    use min, max or average of same classes
- Filter out inferences by top τ percent :
  - **ISP.JIM**
  - **ISP.IIM.**$\wedge$
  - **ISP.IIM.**$\vee$

Joint Inferential Model (JIM)

$$p_i = \text{X is charged by Y}$$

$< Person, p_i, Law\ Enforcement\ Agent >$ = 1.45
$< Person, p_i, Law\ Enforcement\ Agency >$ = 1.21
$< Bank\ Account, p_i, Organization > = 0.97$

$$p_j = \text{Y announced the arrest of X}$$

$< Law\ Enforcement\ Agent, p_j,\ Person >$ = 2.01
$< Law\ Enforcement\ Agency, p_j, Person >$ = 1.61
$< Reporter, p_j, Person > = 0.97$

Minimum	Maximum	Average
1.45	2.01	1.73
1.21	1.61	1.41

Independent Inferential Model (IIM)

$$p_i = \text{X is charged by Y}$$

$< Law\ Enforcement\ Agent, p_i, * > = 3.43$
$< * \ p_i, Person > = 2.17$
$< *, p_i, Organization > = 1.24$

$$p_j = \text{Y announced the arrest of X}$$

$< *, p_j,\ Person > = 2.87$
$< Law\ Enforcement\ Agent, p_j, * > = 1.61$
$< Reporter, p_j, * > = 0.89$

Minimum	Maximum	Average
1.61	3.43	2.52
2.17	2.87	2.52

Evaluation criteria

		GOLD STANDARD	
		True	**False**
SYSTEM	**True**	A	B
	False	C	D

1. **Sensitivity**: probability of accepting correct inferences $\frac{A}{A+C}$

2. **Specificity:** probability of rejecting incorrect inferences $\frac{D}{B+D}$

3. **Accuracy:** probability of a filter being correct $\frac{A+D}{A+B+C+D}$

Evaluation

System		Parameters Selected from DEV Set		Sensitivity (95% conf)	Specificity (95% conf)	Accuracy (95% conf)
		Ranking Strategy	τ (%)			
B0		-	-	0.00±0.00	1.00±0.00	0.50±0.04
B1		-	-	1.00±0.00	0.00±0.00	0.49±0.04
Random		-	-	0.50±0.06	0.47±0.07	0.50±0.04
CBC	**ISP.JIM**	**maximum**	**100**	**0.17±0.04**	**0.88±0.04**	**0.53±0.04**
	ISP.IIM.∧	maximum	100	0.24±0.05	0.84±0.04	0.54±0.04
	ISP.IIM.∨	**maximum**	**90**	**0.73±0.05**	**0.45±0.06**	**0.59±0.04**†
WordNet	ISP.JIM	minimum	40	0.20±0.06	0.75±0.06	0.47±0.04
	ISP.IIM.∧	minimum	10	0.33±0.07	0.77±0.06	0.55±0.04
	ISP.IIM.∨	minimum	20	0.87±0.04	0.17±0.05	0.51±0.05

Confusion Matrix for Best Methods

ISP.IIM.V		GOLD STANDARD	
		True	False
SYSTEM	True	184	139
	False	63	114

Best Accuracy

ISP.JIM		GOLD STANDARD	
		True	False
SYSTEM	True	42	28
	False	205	225

Best Specificity

Chapter 3:

Learning Directionality of Inference Rules (LEDIR)

Learning Directionality of Inference Rules (LEDIR)

- Rules inference by DIRT are all bidirectional ($\Leftrightarrow$, symmetric)
- DIRT is not finding strict logical entailments
- DIRT is based on finding plausible Inference Rules relied on mutual co-occurrence and context similarity
- One should decide about the directionality of rules.

LEDIR is using …

- Distributional hypothesis (DIRT)
- Selectional preferences (ISP)
- Directional hypothesis:
  - **”specific relation implies general relation”**

Directionality Hypothesis

•

$$\mathrm{p_i} \Rightarrow p_j \stackrel{\text{def}}{=} x \in V_m(p_i)\ then\ x \in V_m\left(p_j\right)$$
$$V_m(p_i) \subset V_m(p_j)$$
$$|V_m(p_i)| < |V_m(p_j)|$$

How does LEDIR work?

- we have computed SP of relations . $< C_x, p_i, C_y >$
- We have Inferential SP.
- We can measure the similarity of the antecedent and the consequent of each IR by e.g. Overlap Coefficient.
- If similarity is more than a certain threshold we consider it as plausible IR.

Plausibility

- Overlap Coefficient:

$$sim(p_i, p_j) = \frac{|< C_x, p_i, C_y > \cap < C_x, p_j, C_y >|}{\min(| < C_x, p_i, C_y >, < C_x, p_j, C_y > |)}$$

- Rule:

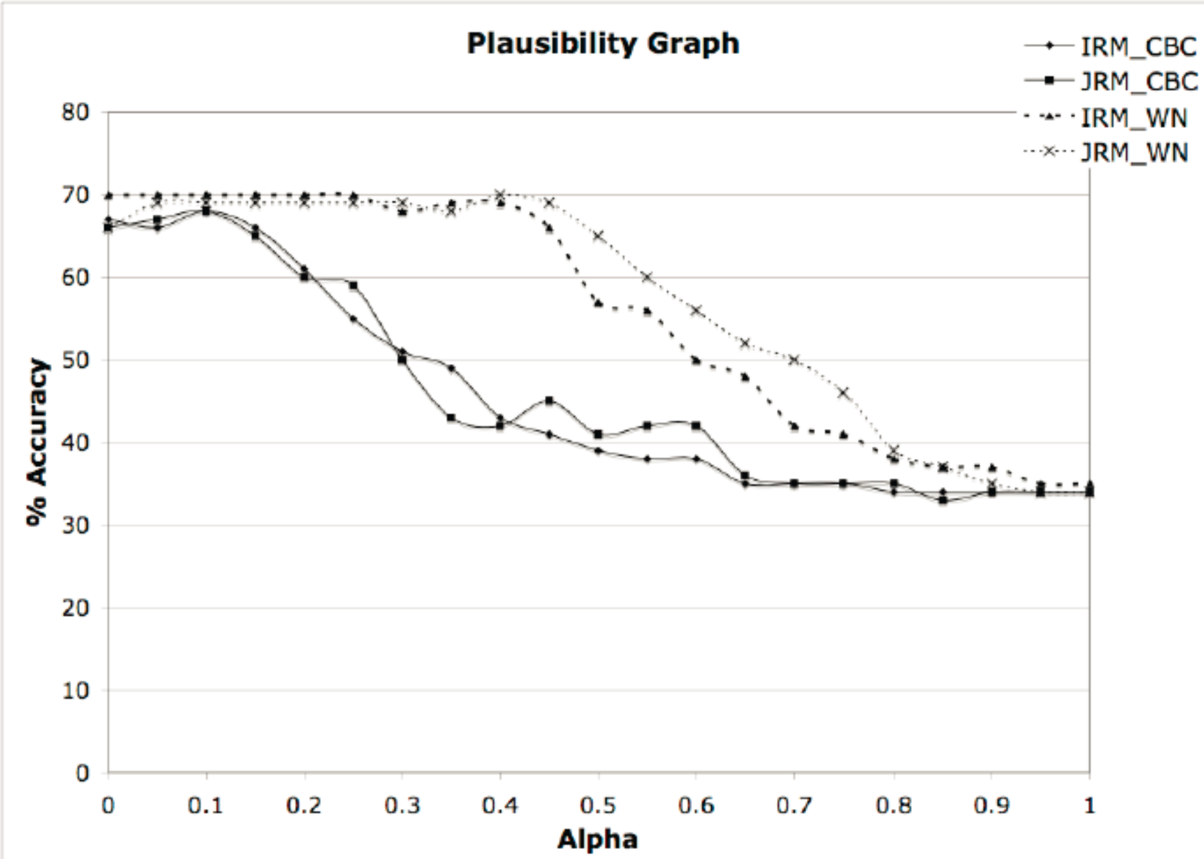
$$\begin{aligned} &if \;\; sim(p_i, p_j) \geq \alpha: \\ &\qquad inference\ is\ plausible \\ &else: \\ &\qquad inference\ is\ not\ plausible \end{aligned}$$

Directionality

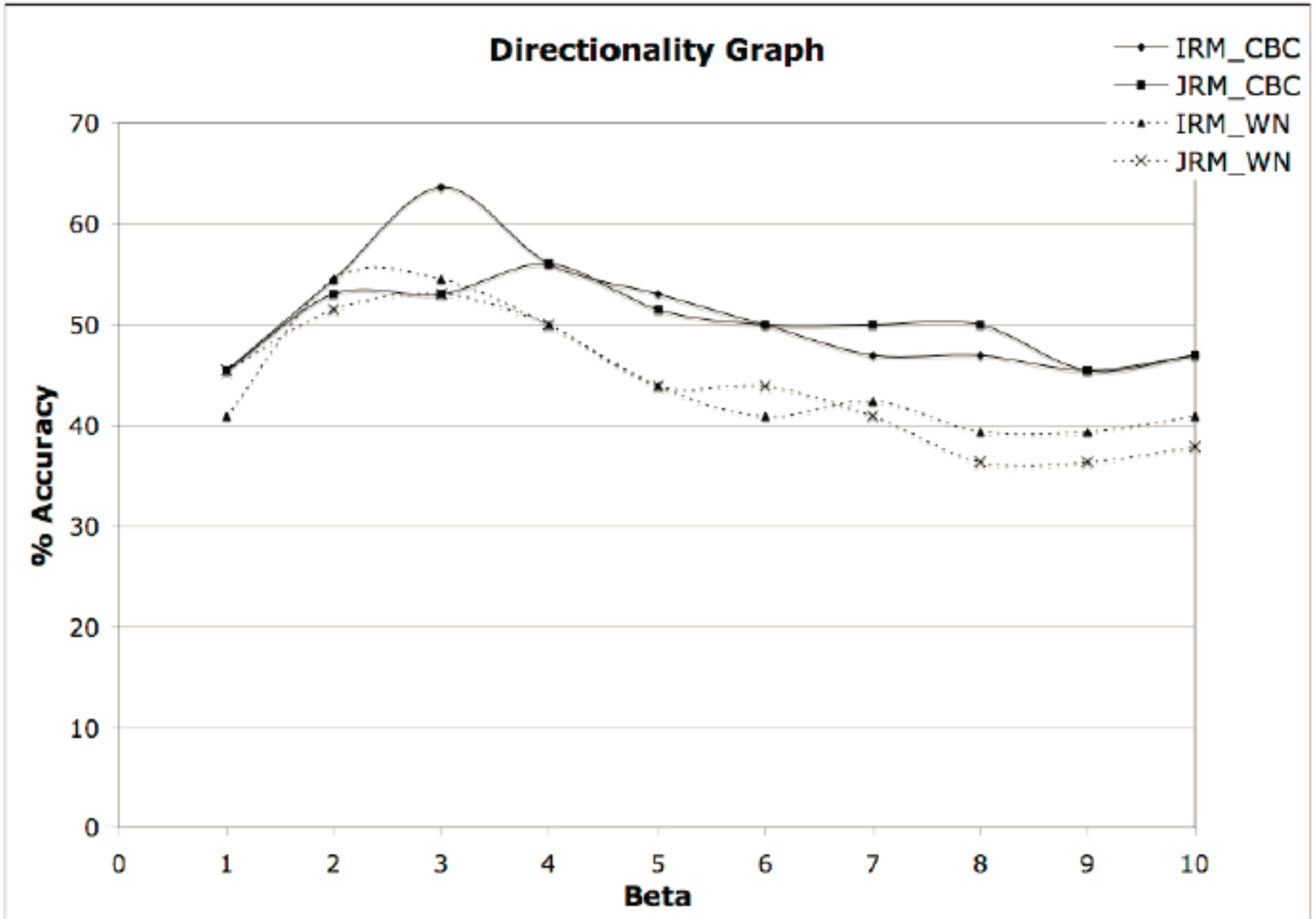
- As we said before the direction is from specific relation to general direction. A relation with broader semantic classes is considered to be more general.
- Use cardinality of RSP of each relation to determine the directionality
- $\frac{|C_x, p_j, C_y|}{|C_x, p_i, C_y|} > \beta \; then \; p_i \Rightarrow p_j$

Evaluation: Plausibility

- Baseline : 66%
- Systems: 68% -70%



Evaluation: Directionality



Evaluation: all

Model		**α**	**β**	**Accuracy (%)**
B-random		-	-	25
B-frequent		-	-	34
B-DIRT		-	-	25
JRM	CBC	0.15	2	38
	WN	0.55	2	38
IRM	**CBC**	**0.15**	**3**	**48**
	WN	0.45	2	43

Review on DIRT

- Finding Inference Rules from corpus and their equivalents (paraphrases).
- Actually it finds equivalent binary semantic relations.
- Relation is a verb and slot fillers are two nouns.
- Inference rule is equivalency relation.
- Extension on Harris' Distributional Hypothesis:
  - "If two paths tend to occur in similar contexts, the meanings of the paths tend to be similar."
- Here context is dependency path between words.

Problems of DIRT

- Low Precision:
  1. For some relations there is a constraint on the semantic class of their slot filler which is not captured by DIRT. Due to different senses of arguments and relations. (Solution: ISP)
  2. Inference rules are always bidirectional which is not true for many cases. (Solution: LEDIR)
  3. Antonym paths will be easily confused.

References

- Dekang Lin, Patrick Pantel: DIRT,discovery of inference rules from text. KDD 2001: 323-328
- Patrick Pantel et al: ISP: Learning Inferential Selectional Preferences. HLT-NAACL 2007: 564-571
- Rahul Bhagat, Patrick Pantel, Eduard H. Hovy: LEDIR, An Unsupervised Algorithm for Learning Directionality of Inference Rules. EMNLP-CoNLL 2007: 161-170
- Dekang Lin, Patrick Pantel: Discovery of inference rules for question-answering. Natural Language Engineering 7(4): 343-360 (2001)