Using the Web to Validate Lexico-Semantic Relations

Hernani Costa[1], Hugo Gonçalo Oliveira[2] and Paulo Gomes

{hpcosta,hroliv,pgomes}@dei.uc.pt

Cognitive & Media Systems Group
CISUC, University of Coimbra

Lisbon, October, 2011

[1]supported by FCT sholarship BII/FCTUC/C2008/CISUC/2ndPhase.
[2]supported by FCT scholarship SFRH/BD/44955/2008.

1. Introduction
2. Web-based Similarity Measures
3. Experimentation
   - Datasets
   - Preliminary Analysis
   - Correlation Analysis
   - Select the correct instances
4. Final Remarks

Information extraction (IE) from text

1. *Hernani is a researcher at the University of Coimbra.*

Information extraction (IE) from text

1. *Hernani is a researcher at the University of Coimbra.*

- Entities
  - *Hernani*
  - *University of Coimbra*

Information extraction (IE) from text

1. *Hernani is a researcher at the University of Coimbra.*

- Entities
  - *Hernani*
  - *University of Coimbra*

- Binary relations
  - $t_1 = (Hernani, has_affiliation, University\ of\ Coimbra)$

Information extraction (IE) from text

1. *Hernani is a researcher at the University of Coimbra.*
2. *Animals, such as dogs.*

- Entities
  - *Hernani*
  - *University of Coimbra*
  - *animal*
  - *dog*
- Binary relations
  - $t_1 = (Hernani, has_affiliation, University\ of\ Coimbra)$

Information extraction (IE) from text

1. *Hernani is a researcher at the University of Coimbra.*
2. *Animals, such as dogs.*

- Entities
  - *Hernani*
  - *University of Coimbra*
  - *animal*
  - *dog*
- Binary relations
  - $t_1 = (Hernani, has_affiliation, University\ of\ Coimbra)$
  - $t_2 = (animal, hypernym_of, dog)$

Evaluation of semantic relations

- Generally ends up being done by humans...

Evaluation of semantic relations

- Generally ends up being done by humans...
  - Manual evaluation
    - Less prone to errors

Evaluation of semantic relations

- Generally ends up being done by humans...
  - Manual evaluation
    - Less prone to errors
  - But...
    - Hard to repeat
    - Time-consuming
    - (More) subjective

Evaluation of semantic relations

- Generally ends up being done by humans...
  - Manual evaluation
    - Less prone to errors
  - But...
    - Hard to repeat
    - Time-consuming
    - (More) subjective
- Approaches for automatic evaluation of domain ontologies
  - All have limitations
  - Not well-suited for broad-coverage open-domain knowledge!

Relations' confidence

- IE system with a pattern learning component
  - Higher recall
  - Lower precision

Relations' confidence

- IE system with a pattern learning component
  - Higher recall
  - Lower precision
- Words that occur in the **same contexts**, tend to have **similar meanings** [Harris, 1970]

Relations' confidence

- IE system with a pattern learning component
  - Higher recall
  - Lower precision
- Words that occur in the **same contexts**, tend to have **similar meanings** [Harris, 1970]
- Rank instances (and patterns)
  - Take advantage of **redundancy**
  - **Similarity measures**
  - Assign a **confidence** value

In this work...

- Relations are denoted in free text by **disciminating patterns**

In this work...

- Relations are denoted in free text by **disciminating patterns**
- Relations occurring **more times**, tend to be more relevant/correct

In this work...

- Relations are denoted in free text by **disciminating patterns**
- Relations occurring **more times**, tend to be more relevant/correct
- Compute the **confidence** on a lexico-semantic relation
  - **Frequent** discriminating patterns for a relation (eg. "*is a*" or "*and other*" for hyponymy)

In this work...

- Relations are denoted in free text by **disciminating patterns**
- Relations occurring **more times**, tend to be more relevant/correct
- Compute the **confidence** on a lexico-semantic relation
  - **Frequent** discriminating patterns for a relation (eg. "*is a*" or "*and other*" for hyponymy)
  - Distribution of related words connected by these patterns on the **Web**
  - **Similarity measures**

In this work...

- Relations are denoted in free text by **disciminating patterns**
- Relations occurring **more times**, tend to be more relevant/correct
- Compute the **confidence** on a lexico-semantic relation
  - **Frequent** discriminating patterns for a relation (eg. "*is a*" or "*and other*" for hyponymy)
  - Distribution of related words connected by these patterns on the **Web**
  - **Similarity measures**
- How suitable are these measures to validate lexico-semantic relations?
- Select the best measure(s) for this task

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$
- Confidence for relation $t = (e_1, r, e_2)$
  - π_{ri} is a discriminating pattern for relation r

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$
- Confidence for relation $t = (e_1, r, e_2)$
  - π_{ri} is a discriminating pattern for relation r
  - $P(e_1) \Rightarrow P(e_1\ \pi_{ri})$
  - $P(e_2) \Rightarrow P(\pi_{ri}\ e_2)$

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$
- Confidence for relation $t = (e_1, r, e_2)$
  - π_{ri} is a discriminating pattern for relation r
  - $P(e_1) \Rightarrow P(e_1\ \pi_{ri})$
  - $P(e_2) \Rightarrow P(\pi_{ri}\ e_2)$
  - $P(e_1 \cap e_2) \Rightarrow P(e_1\ \pi_{ri}\ e_2)$

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$
- Confidence for relation $t = (e_1, r, e_2)$
  - π_{ri} is a discriminating pattern for relation r
  - $P(e_1) \Rightarrow P(e_1\ \pi_{ri})$
  - $P(e_2) \Rightarrow P(\pi_{ri}\ e_2)$
  - $P(e_1 \cap e_2) \Rightarrow P(e_1\ \pi_{ri}\ e_2)$
- If e_1=$\{planet\}$, e_2=$\{Mars\}$ and π_{ri}=$\{such\ as\}$

Measures Application

- Web-based similarity measures
  - $P(q) \Rightarrow$ page count in a search engine for q
- Similarity of words e_1 and e_2
  - $P(e_1 \cap e_2) \Rightarrow P(\text{“}e_1\ AND\ e_2'')$
  - $P(e_1 \cup e_2) \Rightarrow P(e_1) + P(e_2)$
- Confidence for relation $t = (e_1, r, e_2)$
  - π_{ri} is a discriminating pattern for relation r
  - $P(e_1) \Rightarrow P(e_1\ \pi_{ri})$
  - $P(e_2) \Rightarrow P(\pi_{ri}\ e_2)$
  - $P(e_1 \cap e_2) \Rightarrow P(e_1\ \pi_{ri}\ e_2)$
- If e_1=$\{planet\}$, e_2=$\{Mars\}$ and π_{ri}=$\{such\ as\}$
  - $P(e_1)$=$\{planet\ such\ as\}$
  - $P(e_2)$=$\{such\ as\ Mars\}$
  - $P(e_1 \cap e_2)$=$\{planet\ such\ as\ Mars\}$

Measures used

- Used by [Bollegala et al., 2007] and [Cimiano and Wenderoth, 2007]
  - WebJaccard

Measures used

- Used by [Bollegala et al., 2007] and [Cimiano and Wenderoth, 2007]
  - WebJaccard
  - WebOverlap

Measures used

- Used by [Bollegala et al., 2007] and [Cimiano and Wenderoth, 2007]
  - WebJaccard
  - WebOverlap
  - WebDice

Measures used

- Used by [Bollegala et al., 2007] and [Cimiano and Wenderoth, 2007]
  - WebJaccard
  - WebOverlap
  - WebDice
  - WebPMI

Measures used

- Used by [Bollegala et al., 2007] and [Cimiano and Wenderoth, 2007]
  - WebJaccard
  - WebOverlap
  - WebDice
  - WebPMI
- NWS [Gracia and Mena, 2008]

Set-up

- Semantic relations can be expressed by several different textual patterns

Set-up

- Semantic relations can be expressed by several different textual patterns
- Select two sets:
  - $\Pi_h \Rightarrow$ frequent hyponymy patterns
  - $\Pi_p \Rightarrow$ frequent part-of patterns

Set-up

- Semantic relations can be expressed by several different textual patterns
- Select two sets:
  - $\Pi_h \Rightarrow$ frequent hyponymy patterns
  - $\Pi_p \Rightarrow$ frequent part-of patterns
- Computed the final scores:
  - NP $\Rightarrow$ no patterns, simple co-occurrence (baseline)

Set-up

- Semantic relations can be expressed by several different textual patterns
- Select two sets:
  - $\Pi_h \Rightarrow$ frequent hyponymy patterns
  - $\Pi_p \Rightarrow$ frequent part-of patterns
- Computed the final scores:
  - NP $\Rightarrow$ no patterns, simple co-occurrence (baseline)
  - B $\Rightarrow$ score of the best pattern

Set-up

- Semantic relations can be expressed by several different textual patterns
- Select two sets:
  - $\Pi_h \Rightarrow$ frequent hyponymy patterns
  - $\Pi_p \Rightarrow$ frequent part-of patterns
- Computed the final scores:
  - NP $\Rightarrow$ no patterns, simple co-occurrence (baseline)
  - B $\Rightarrow$ score of the best pattern
  - 2B $\Rightarrow$ average of the scores of the two best patterns
  - Av $\Rightarrow$ average of the scores of all patterns

Datasets

- Relations from WordNet 2.0[3]

[3] http://wordnet.princeton.edu

[4] eg. {*corporation.1, corp.1*} hyponym-of {*firm.1, house.2, business firm.1*} → {*corporation*, hyponym-of, *firm*}.

[5] $t = (e_1, r, e_2)$, $score(t) = log(P(e_1) + P(e_2))$

Datasets

- Relations from WordNet 2.0[3]
  1. Select all the relation instances between synsets which are the first sense of their most frequent word.

[3] `http://wordnet.princeton.edu`

[4] eg. {*corporation.1, corp.1*} hyponym-of {*firm.1, house.2, business firm.1*} → {*corporation*, hyponym-of, *firm*}.

[5] $t = (e_1, r, e_2)$, $score(t) = log(P(e_1) + P(e_2))$

Datasets

- Relations from WordNet 2.0[3]
  1. Select all the relation instances between synsets which are the first sense of their most frequent word.
  2. For each of the latter, define instances held by the first word in the connected synsets [4]

[3] `http://wordnet.princeton.edu`

[4] eg. {*corporation.1, corp.1*} hyponym-of {*firm.1, house.2, business firm.1*} → {*corporation*, hyponym-of, *firm*}.

[5] $t = (e_1, r, e_2)$, $score(t) = log(P(e_1) + P(e_2))$

Datasets

- Relations from WordNet 2.0[3]
  1. Select all the relation instances between synsets which are the first sense of their most frequent word.
  2. For each of the latter, define instances held by the first word in the connected synsets [4]
  3. Rank instances according to the frequency of their arguments in Google[5]

[3] `http://wordnet.princeton.edu`

[4] eg. {*corporation.1, corp.1*} hyponym-of {*firm.1, house.2, business firm.1*} → {*corporation*, hyponym-of, *firm*}.

[5] $t = (e_1, r, e_2)$, $score(t) = log(P(e_1) + P(e_2))$

Datasets

- Relations from WordNet 2.0[3]
  1. Select all the relation instances between synsets which are the first sense of their most frequent word.
  2. For each of the latter, define instances held by the first word in the connected synsets [4]
  3. Rank instances according to the frequency of their arguments in Google[5]
  4. Select the first 1,100 hyponymy instances (H) and 1,100 part-of instances (P)

[3] `http://wordnet.princeton.edu`

[4] eg. {*corporation.1, corp.1*} hyponym-of {*firm.1, house.2, business firm.1*} → {*corporation*, hyponym-of, *firm*}.

[5] $t = (e_1, r, e_2)$, $score(t) = log(P(e_1) + P(e_2))$

Datasets

- H and P contain only correct instances

Datasets

- H and P contain only correct instances
- A new set, I, was created with 1,010 random pairs of words, never related by hyponymy nor part-of

Datasets

- *H* and *P* contain only correct instances
- A new set, *I*, was created with 1,010 random pairs of words, never related by hyponymy nor part-of
- *WR* contains correct instances where the relation type was changed

Datasets

- *H* and *P* contain only correct instances
- A new set, *I*, was created with 1,010 random pairs of words, never related by hyponymy nor part-of
- *WR* contains correct instances where the relation type was changed

Classification	Examples	
Correct (C)	*fight* hyponym-of *conflict*	*hour* part-of *day*
Incorrect (I)	*towel* hyponym-of *engineer*	*ibuprofen* part-of *light*
Wrong Relation (WR)[6]	*eye* hyponym-of *face*	*hometown* part-of *town*

[6] Introducing noise: changing the type of relation in *H* and *P*

Page counts for hyponymy patterns

R	Textual Pattern (π_h)	Correct		WrongRel		Incorrect	
		Av	SD	Av	SD	Av	SD
Y hypo X	is a|an|one|the kind of	0.46	73.38	0.01	5.46	$9.90E^{-4}$	0.99
	is a|an|one|the	274.7	44560.8	8.09	1007.66	0.53	175.74
	is a|an|one|the variety of	0.01	5.28	0.0	0.0	0.0	0.0
	is a|an|one|the type of	0.77	191.2	0.07	23.87	0.0	0.0
	is a|an|one|the form of	0.99	510.96	$4.5E^{-3}$	3.31	0.0	0.0
	and|or other	66.9	15512.9	15.15	2748.7	0.36	23.34
X hyper Y	such as	27.48	6832.4	18.15	2620.2	0.16	6.3
	like	42.60	6486.2	14.29	3264.8	0.02	11.41
	including	26.47	7307.9	81.63	10414.3	$9.90E^{-4}$	8.74
	especially	2.79	570.8	21.1	4147.7	0.03	11.41

Page counts for part-of patterns

R	Textual Pattern (π_p)	Correct		WrongRel		Incorrect	
		Av	SD	Av	SD	Av	SD
Y part X	of	208.28	49053.4	313.58	65352.7	26.13	15971.7
	of a|an|one|the	546.28	119150.7	319.91	73777.7	3.99	1390.4
	from a|an|one|the	43.07	18873.5	71.42	43370.6	0.21	100.90
	in	646.38	43269.8	152.86	45098.7	9.12	6785.4
	is part of	2.08	418.33	0.11	47.64	$2.97E^{-3}$	2.23
	is member of	$2.72E^{-3}$	1.73	$2.73E^{-3}$	2.99	0.0	0.0
	part of a|an|one|the	1.45	251.7	1.72	161.95	0.13	120.98
	member of a|an|one|the	0.27	122.73	0.89	17.12	$4.95E^{-3}$	2.64
	is a|one|the part of	0.99	290.33	0.06	19.94	$1.98E^{-3}$	1.99
	is a|an|one|the member of	1.33	1439.3	0.01	7.19	0.0	0.0
	is a|an|one|the part of a|one|the	0.91	218.01	0.08	13.92	$1.98E^{-3}$	1.99
	is a|an|one|the member of a|one|the	0.42	301.84	0.12	64.59	0.0	0.0
X has Y	's	550.9	188279.1	243.92	159845.1	5.23	3233.0
	has a|an|one|the	9.17	1809.0	9.48	2577.7	0.25	177.89
	contains a|an|one|the	1.12	309.75	2.67	871.14	$9.90E^{-4}$	4.12
	consists of	0.61	111.26	2.36	1228.2	$6.93E^{-3}$	0.99
	is made of	0.02	8.74	0.11	68.13	0.0	0.0

Correlation Analysis

- Parallelism between correctness and the similarity scores

Relation		nHits	Jaccard	Overlap	Dice	PMI	NWS	hasHits
Hyponymy ($C + I$)	NP	0.11	0.14	0.15	0.14	0.13	0.63	**0.77**
	B	0.16	0.17	0.34	0.18	**0.93**	0.87	0.92
	2B	0.18	0.19	0.36	0.20	**0.92**	0.86	0.86
	Av	0.18	0.20	0.35	0.22	**0.78**	0.72	-
Hyponymy ($C + I + WR$)	NP	$-5.3E^{-3}$	-0.11	-0.11	-0.11	-0.17	-0.14	**0.39**
	B	0.04	0.16	0.29	0.17	**0.76**	0.74	**0.76**
	2B	0.04	0.17	0.32	0.19	0.73	**0.74**	0.69
	Av	0.07	0.20	0.34	0.21	**0.69**	0.67	-
Part-of ($C + I$)	NP	0.19	0.22	0.35	0.23	0.29	0.71	**0.76**
	B	0.16	0.18	0.21	0.23	**0.89**	0.85	0.85
	2B	0.17	0.19	0.34	0.23	**0.90**	0.86	0.88
	Av	0.18	0.21	0.26	0.25	**0.78**	0.72	-
Part-of ($C + I + WR$)	NP	0.13	0.24	0.33	0.25	0.33	**0.65**	0.39
	B	0.16	0.17	0.21	0.20	**0.82**	0.69	0.72
	2B	0.17	0.17	0.24	0.20	**0.82**	0.68	0.72
	Av	0.18	0.15	0.25	0.16	**0.57**	0.42	-

- nHits: just the number of page counts
- hasHits: correct if $P(e_1 \cap e_2) > 1$

Task: select the correct instances automatically

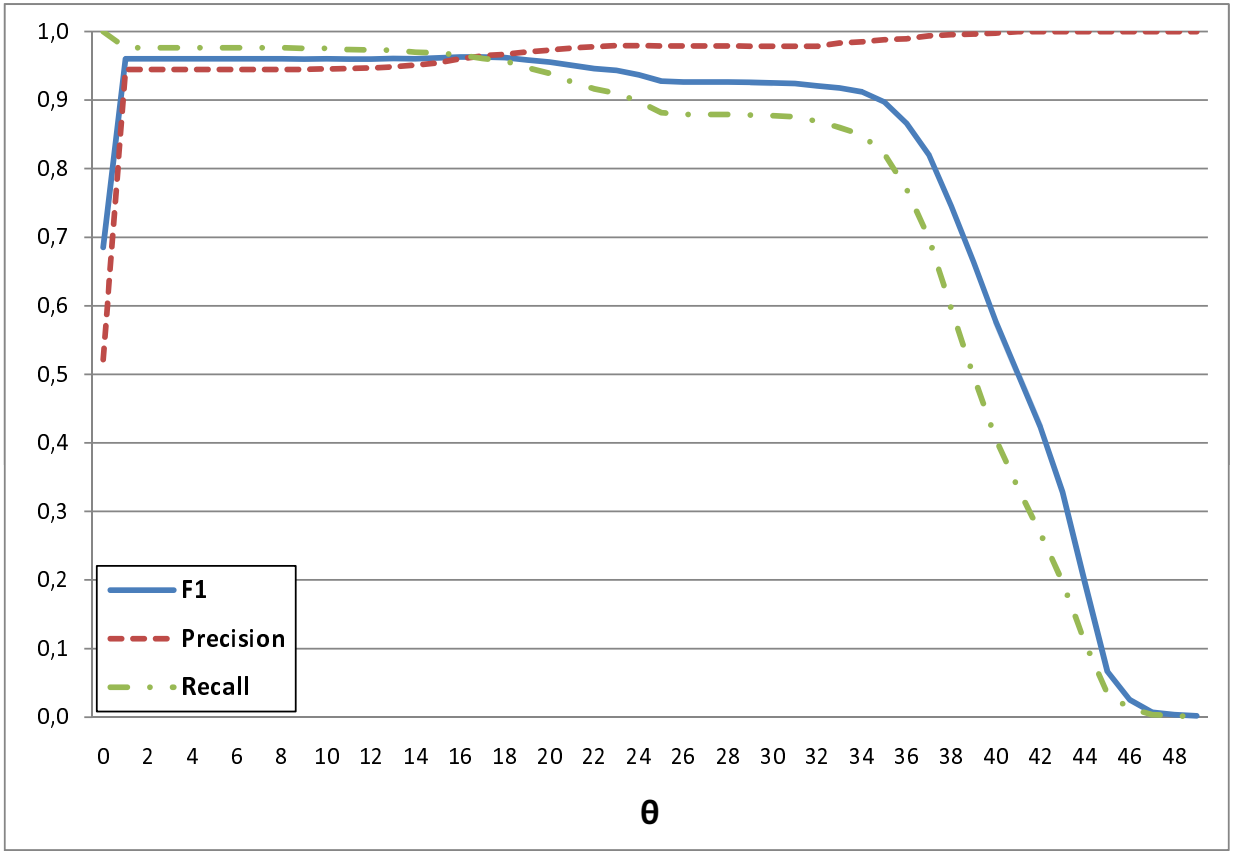
- Variable cut points (θ) in the scores
- Compute Precision, Recall and F_1

Task: select the correct instances automatically

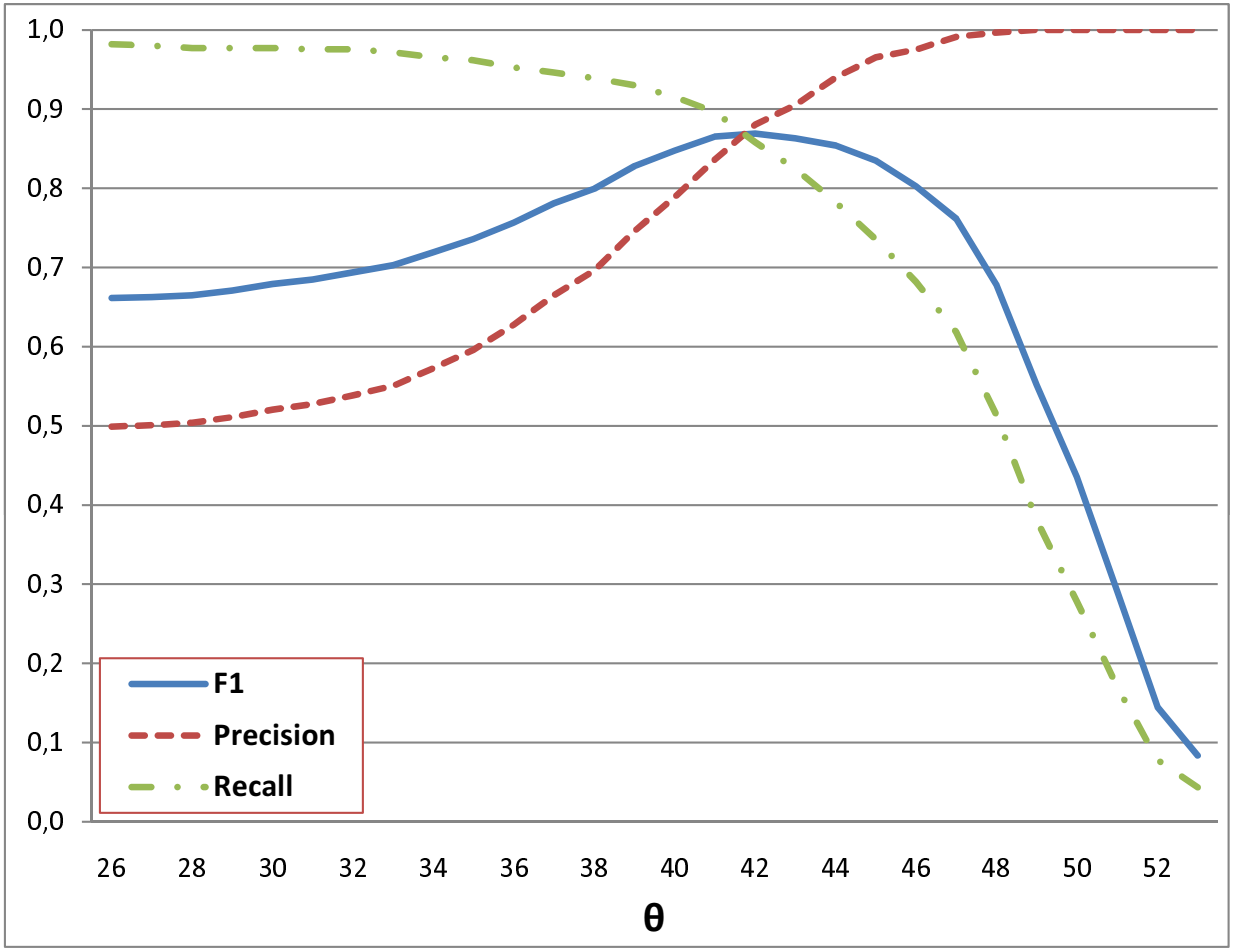
- Variable cut points (θ) in the scores
- Compute Precision, Recall and F_1
- Best F_1 measures and respective θ:

R		Jaccard		Overlap		Dice		PMI		NWS		hasHits
		F_1	θ	F_1	θ	F_1	θ	F_1	θ	F_1	θ	F_1
Hypo	NP	70.5	$1E^{-4}$	68.5	$1E^{-4}$	77.5	$2E^{-4}$	68.5	0	68.5	0	**89.0**
	B	94.9	$2E^{-4}$	68.5	0	95.4	$2E^{-4}$	**96.2**	26	96.1	0.15	96.0
	2B	94.1	$2E^{-4}$	68.5	0	95.1	$2E^{-4}$	**96.3**	16	96.0	0.05	92.6
	Av	86.3	$2E^{-4}$	68.5	0	90.9	$2E^{-4}$	**96.2**	3	91.5	0.05	-
Part	NP	91.5	$2E^{-4}$	68.5	$1E^{-4}$	**93.7**	$2E^{-4}$	68.5	0	91.6	0.05	88.9
	B	93.9	$2E^{-4}$	80.7	$2E^{-4}$	94.0	$2E^{-4}$	94.3	32	**94.7**	0.25	92.8
	2B	93.8	$2E^{-4}$	75.6	0.05	93.8	$2E^{-4}$	94.7	33	**94.9**	0.2	94.1
	Av	86.7	$2E^{-4}$	68.5	0	90.3	$2E^{-4}$	**94.5**	4	87.1	0.05	-
Including instances with wrong relation (*WR*)												
Hypo	NP	51.0	$1E^{-4}$	51.0	$1E^{-4}$	51.0	$1E^{-4}$	51.0	0	51.0	0	**61.8**
	B	**75.3**	$2E^{-4}$	54.7	$2E^{-4}$	**75.3**	$3E^{-4}$	69.9	28	75.2	0.25	69.5
	2B	74.8	$2E^{-4}$	51.1	0	**74.9**	$5E^{-4}$	69.9	16	74.1	0.25	68.2
	Av	72.7	$2E^{-4}$	51.1	0	**75.1**	$2E^{-4}$	71.9	4	74.8	0.05	-
Part	NP	70.6	$2E^{-4}$	51.0	$1E^{-4}$	70.7	$4E^{-4}$	62.1	1	**72.5**	0.05	61.5
	B	65.5	$2E^{-4}$	62.1	$2E^{-4}$	65.2	$2E^{-4}$	**86.9**	42	62.3	0.2	65.7
	2B	65.4	$2E^{-4}$	59.4	0.05	65.6	$2E^{-4}$	**85.5**	41	67.5	0.2	68.8
	Av	59.7	$2E^{-4}$	51.0	0	62.3	$2E^{-4}$	**68.6**	4	60.7	0.05	-

Identification of correct hyponymy instances with WebPMI (2B)



Identification of correct part-of instances with WebPMI (B)



Conclusion

1. The scores given by some measures are highly correlated to the correctness of instances

Conclusion

1. The scores given by some measures are highly correlated to the correctness of instances
2. High F_1 scores in the selection of correct instances
   - $>$ 96% for hyponymy
   - $>$ 94% for part-of

Conclusion

1. The scores given by some measures are highly correlated to the correctness of instances
2. High F_1 scores in the selection of correct instances
   - $>$ 96% for hyponymy
   - $>$ 94% for part-of
3. The best performing measures can be used as an alternative to manual evaluation of semantic relations

Conclusion

1. The scores given by some measures are highly correlated to the correctness of instances
2. High F_1 scores in the selection of correct instances
   - $>$ 96% for hyponymy
   - $>$ 94% for part-of
3. The best performing measures can be used as an alternative to manual evaluation of semantic relations

- Future: evaluate other kinds of relations, for other languages

Thank you!

References I

[Bollegala et al., 2007] Bollegala, D., Matsuo, Y., and Ishizuka, M. (2007).
Measuring semantic similarity between words using web search engines.
In *Proc. 16th International Conf. on the World Wide Web*, pages 757–766, New York, NY, USA. ACM.

[Cimiano and Wenderoth, 2007] Cimiano, P. and Wenderoth, J. (2007).
Automatic Acquisition of Ranked Qualia Structures from the Web.
In *Proc. 45th Annual Meeting of the Association of Computational Linguistics*, pages 888–895, Prague, Czech Republic. ACL.

[Gracia and Mena, 2008] Gracia, J. and Mena, E. (2008).
Web-based measure of semantic relatedness.
In *Proc. 9th International Conf. on Web Information Systems Engineering*, pages 136–150. Springer.

[Harris, 1970] Harris, Z. (1970).
Distributional structure.
In *Papers in Structural and Transformational Linguistics*, pages 775–794. D. Reidel Publishing Comp., Dordrecht, Holland.