Automatic Acquisition of Qualia Structure from Corpus Data

by Ichino Yamada, Timothy Baldwin, *Nonmembers*, Hideki Sumiyoshi, Masahiro Shibata, and Nobuyuki Yagi, *Members*

Presenter: Yu-Yun Chang

Graduate Institute of Linguistics,
National Taiwan University

April 30, 2014

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim
- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - qua formal and constitutive roles have enough attention due to previous studies
  - still not available in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim
- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - cuz formal and constitutive roles have enough attention due to previous studies
  - still not availabe in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim

- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - cuz formal and constitutive roles have enough attention due to previous studies
  - still not availabe in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim

- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - cuz formal and constitutive roles have enough attention due to previous studies
  - still not availabe in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim

- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - cuz formal and constitutive roles have enough attention due to previous studies
  - still not availabe in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Introduction

motiv It's a hard work in manually maintaining and updating lexical relations

aim

- automatically discover lexical knowledge from corpus, e.g. Maximum Entropy Learning
- focus on the telic and agentive roles of nouns
  - cuz formal and constitutive roles have enough attention due to previous studies
  - still not availabe in automatic acquisition and large-scale lexical resources

goal To generate a ranked list of verbs for a given noun for each of the telic and agentive roles

Qualia Structure

- from Generative Lexicon by Pustejovsky (1995)
- captures our understanding of an object or a relation in the world
- the qualia structure of a lexical item includes four roles:
  - formal role: the basic category of which distinguishes the meaning of a word within a larger domain, e.g. shape, color, magnitude, etc.
  - constitutive role: the internal constitution of the entity, e.g. component elements
  - telic role: the typical function of the entity
  - agentive role: the origin of the entity or its coming into being, e.g. creator, artifact, natural kind, etc.

Qualia Structure

An Example ...

ps. take a note on telic and agentive roles

the qualia structure of *book*

- formal role: publication
- constitutive role: text
- telic role: read, *study*, *publish* (predicates)
- agentive role: write, *study*, *publish* (predicates)

application of telic and agentive roles

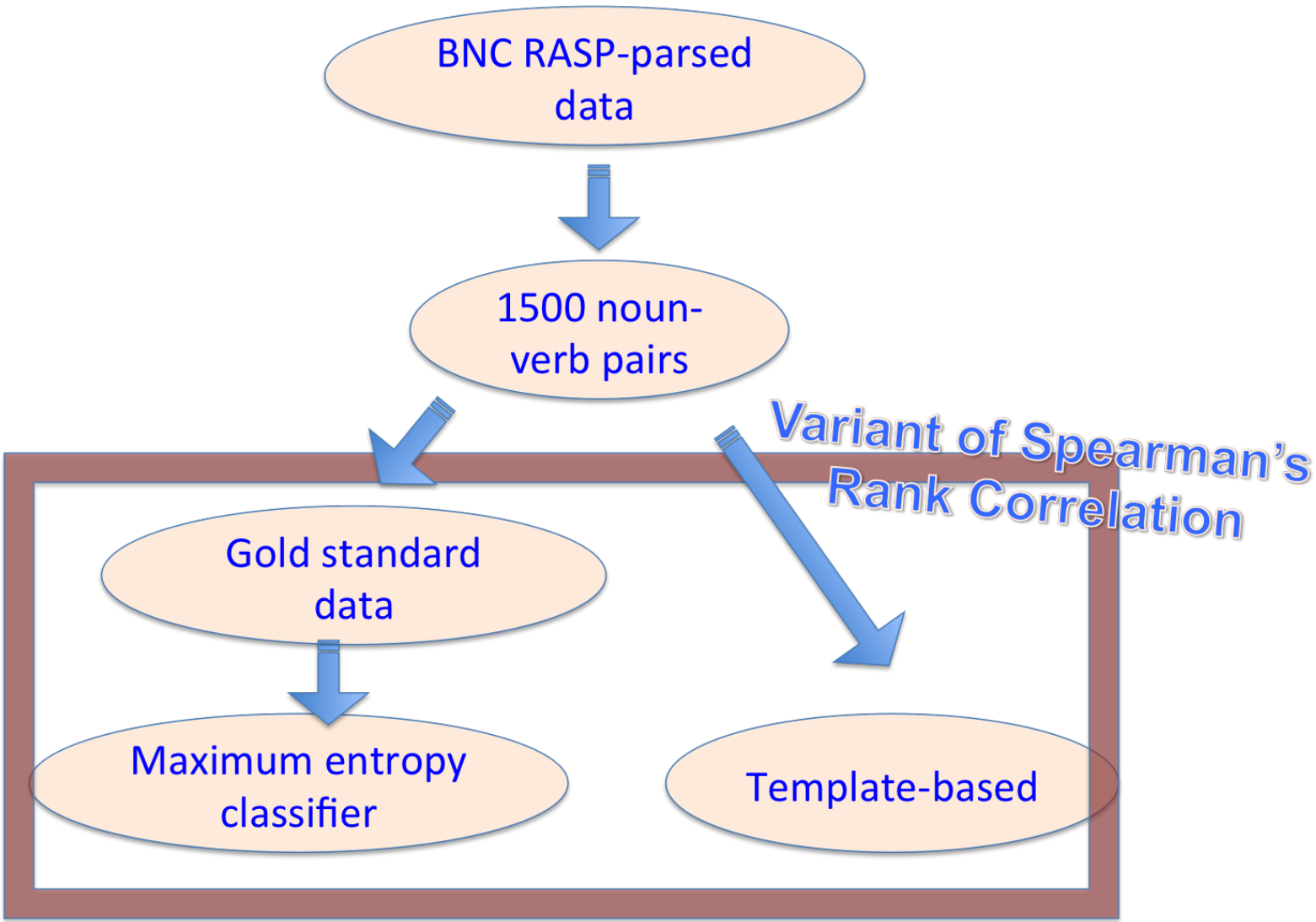
- Mary finished her beer.
- Mary finished drinking her beer.

Qualia Pairs: noun-verb pairs

The noun-verb pairs on telic and agentive roles ...

- Bouillon et al. use inductive logic programming to identify noun-verb pairs from corpus data
- Cimiano and Wenderoth use POS-tagged web data and Google counts for an quantitative analysis
- In this study,
  - identify the qualia roles of an arbitrary noun
  - prefer a qualitative analysis
  - for each telic and agentive role ...
    generate a ranked list of verbs given a noun

Procedures


BNC RASP-parsed data
1500 noun-verb pairs
Variant of Spearman's Rank Correlation
Gold standard data
Maximum entropy classifier
Template-based

Resources I

corpus BNC tokenized sentences (dependency-parsed)

- use CLAW-2 tagset to tag raw sentences
- use tag-sequence grammar over word-level tag
- use RASP to tag relations (23 relations)
- output: dependency tuples - a head + dependents + a relation

ncmod(_, ticket_NN1, airplane_NN2) *airplane tickets*
dobj(read_VV0, book_NN2, _) *read books*

Resources II

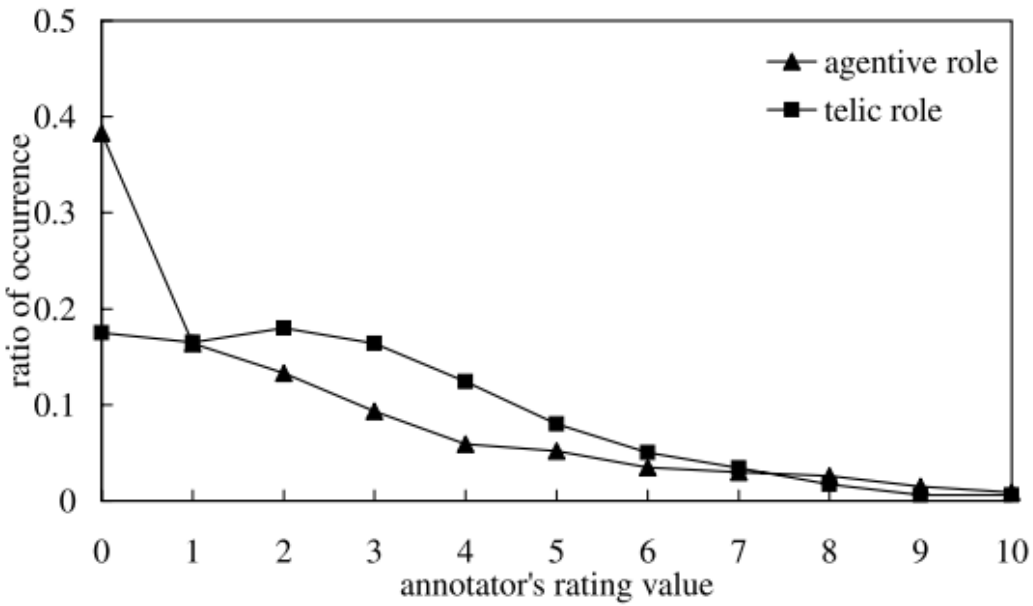
word list 30 nouns * 50 verbs = 1500 noun-verb pairs

- 30 nouns: 10 from literature + 20 randomly selected
- use 30 nouns to find all the dependency tuples with the nouns
- 50 verbs: hand-chosen and randomly selected from the found dependency tuples, and manually tag with telic or agentive roles (gold standard)

evaluate gold standard data

- two native English speakers
- evaluate the noun-verb pairs for the telic and agentive roles
- a scale from 0-10
- take the mean of evaluation between two annotators

Resources III



- most of the verbs are regarded as being unrelated to the noun under the given qualia relation
- based on the variance, the annotators judgments are more clear-cut for the agentive role than the telic role

Machine Learning and Benchmark

- Machine learning: Maximum Entropy Classifier
- Benchmark: Hand-generated templates

Machine Learning I

Maximum Entropy Learning ...

- categorize the data from gold standard data
  - positive: 7-10 (the noun-verb pair is adequate)
  - negative: 0 (the noun-verb pair is impossible)
- 29 nouns for training and 1 noun for testing
- based on the above selected noun-verb pair, extract all the relevant parsed-BNC sentences
- based on the RASP relations, extract the information all the given noun-verb pairs, noun, verb, etc.
- organize the above step to generate featuresets

Machine Learning II

- algorithms:
  - Probability: given the noun-verb pair occurrences, the probability of assigning telic or agentive role
  - Mutual Information: to reduce the effects caused by counting the noun-verb pair occurrences
  - Maximum Entropy: the larger the value, the better the adequacy

Machine Learning

Role	Positive instances	Negative instances
Agentive	print, publish, write, make, compile, design, start	abandon, appear, destroy, dispose, follow, hand, hold, keep, lay, pack, remove, return, snatch, suit, throw, thrust, ithdraw
Telic	read, browse	call, end, appear, suit

(|ncsubj| |have:3_VH0| |I:2_PPIS1| _)
(|dobj| |have:3_VH0| |**book**:5_NN1| _)
(|ncsubj| |**read**:7_VV0| |I:2_PPIS1| _)
(|xcomp| |to:6_TO| |**book**:5_NN1| |**read**:7_VV0|)
(|detmod| _ |**book**:5_NN1| |a:4_AT1|)
(|aux| _ |have:3_VH0| |Can:1_VM|)

- Can I have a book to read?
- Extract the featuresets
- Probabilities → MI → Maximum Entropy

Hand-Generated Templates I

- used as a benchmark for Maximum Entropy Classifier
- all templates assume that the noun will occur as the deep object of a transitive verb
  - telic role: 8 constructional templates
  - agentive role: 1 constructional template

Hand-Generated Templates II

Table 2 Templates for the telic role.

Template	Example
N (be |ϕ) (worth |deserving |meriting)(V[+ing] |V[+nom])	*(a) book worth reading*
N BE worthy of V[+nom]	*(the) book is worthy of reading*
N (deserves |merits) V[+nom]	*(the) book merits reading*
Adverb-V[+en] N	*(a) well-read book*
Adverb V[+en] N	*(a) well read book*
N BE Adverb-V[ed]	*(the) book is well-read*
V[+ing] Noun	*(I enjoy) reading books*
N to V	*(a) book to read*

Table 3 Template for the agentive role.

Template	Example
N BE V[+en]	*(the) book was written (by Kim)*

- count the relative frequency and rank the verbs based on scores

Evaluation on Ranked Lists

Table 4 Top-8 verbs for the agentive role of *book*.

Rank	ME-based method (Value of score_ME)	Template-based method (Value of score_template)	Gold-standard data (Mean value of rating)
1	dedicate (1.084)	publish (0.157)	write (10.0)
2	publish (0.898)	write (0.102)	publish (8.0)
3	compile (0.651)	read (0.019)	compile (8.0)
4	dispose (0.605)	call (0.015)	print (7.5)
5	write (0.438)	dedicate (0.011)	make (7.5)
6	browse (0.408)	print (0.008)	start (7.0)
7	borrow (0.399)	keep (0.007)	design (7.0)
8	print (0.386)	compile (0.006)	translate (6.0)

Table 5 Top-8 verbs for the telic role of *book*.

Rank	ME-based method (Value of score_ME)	Template-based method (Value of score_template)	Gold-standard data (Mean value of rating)
1	read (2.814)	read (0.316)	read (10.0)
2	write (2.221)	write (0.112)	browse (9.0)
3	compile (2.115)	publish (0.079)	think (6.5)
4	dedicate (1.982)	buy (0.036)	buy (6.0)
5	buy (1.775)	keep (0.016)	provide (6.0)
6	borrow (1.695)	appear (0.015)	borrow (5.5)
7	throw (1.682)	make (0.014)	return (5.5)
8	publish (1.656)	provide (0.014)	start (5.5)

Evaluation on Ranked Lists

- Variant of Spearman's Rank Correlation
  - cuz most verbs could not be construed as fulfilling the telic or agentive roles of a given noun
  - value 0-1 (1 represents highly correlated)

Results and Conclusion

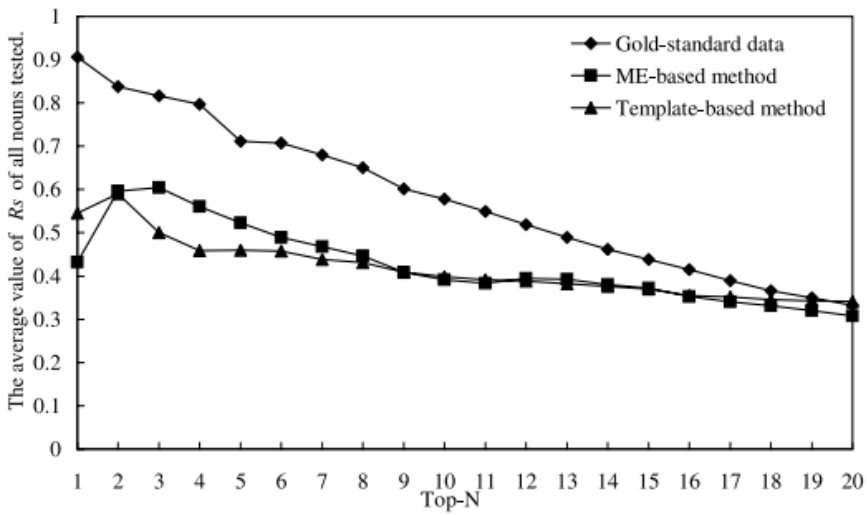


Fig. 2 Rank correlation for the agentive role.

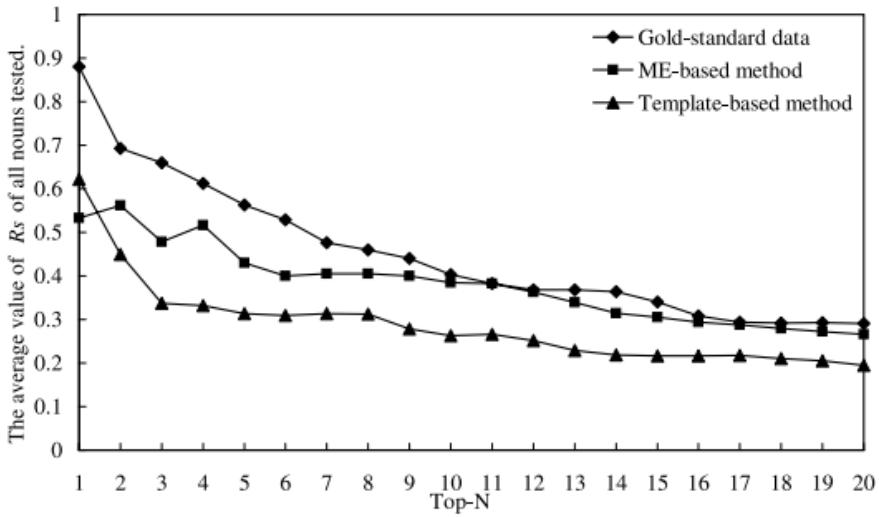


Fig. 3 Rank correlation for the telic role.

- the gold standard data has high agreement at higher ranks
- the gold standard has greater variation in interpreting the telic role.
- in Top-1, templates $>$ maximum entropy
- generally speaking, maximum entropy $>$ templates

Results and Conclusion

	ME	Templates	Gold
agentive	0.605	0.5	0.816
telic	0.479	0.337	0.659

- Maximum Entropy classifier is relatively successful at identifying qualia structure than the traditional template-based approach.

– End of Presentation –
Thank you