Cognitive Computational Models of Pronoun Resolution

Olga Seminck

Université Paris Diderot (Paris 7)
Laboratoire Linguistique Formelle, CNRS

olga.seminck@cri-paris.org

Paris, 23 November 2018

Presentation Outline

Introduction

Experiment

Conclusion

Introduction

Cognitive Computational Models

Goal:
Simulate human behaviour

For example:
Predict how a human would rate a sentence on a scale of 0 to 10.
Predict how long it takes for a human to read a sentence.

Usefulness:
Development and evaluation of theories
Inspiration for artificial intelligence

Anaphora Resolution

Pronoun resolution is a form of anaphora resolution.

NP α_1 takes NP α_2 as its anaphoric antecedent if α_1 depends on α_2 for its interpretation.
(Van Deemter and Kibble 2000)

A **secret**'s worth depends on the people from whom **it** must be kept.

The Shadow of the Wind, Carlos Ruiz Zafón

Pronoun resolution is the process of finding the antecedent of an anaphoric pronoun.

Cognitive Computational Models that Simulate Human Pronoun Resolution

Specify theoretical claims in details

Corpus data $\Rightarrow$ Natural text

Multiple factors

Experiment:

1. A cost metric for pronoun resolution

&

2. Evidence from eye-tracking data

1. An Information Theoretical Cost Metric for Pronoun Resolution

Cost Metrics

Cost Metric: formula that predicts processing cost

- Translates hypothesis into prediction

Example: surprisal

- Hypothesis: unexpected events are harder to process
- Cost metric: $\text{Difficulty(event)} = -\log(P(event))$

Information Theory inspired cost metrics for linguistic processes

Cost metric for pronoun resolution

- Based on entropy

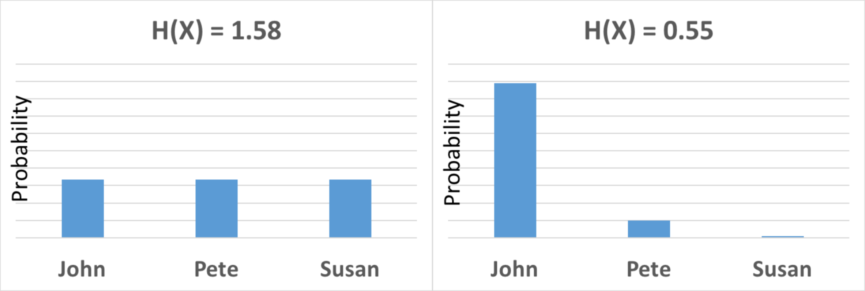
A cost metric to predict the difficulty of pronouns

Prediction for pronouns resolution:
More uncertainty about the antecedent → more processing cost

Entropy: measure of uncertainty
$H(X) = -\sum_{j \in X} p(X = j) \cdot \log_2(p(X = j))$

Entropy

Applies to a random variable: antecedent of a pronoun



Relative Entropy

Entropy increases with the number of antecedent candidates.

- Keep scores comparable through the text
- 'Normalise' the entropy

Normalisation: relative entropy
'Distance' between actual probability distribution & flat distribution

$$H_{relative}(P||Q) = \sum_{i \in P \wedge i \in Q} P(i) \log \frac{P(i)}{Q(i)} \quad (1)$$

Larger distance $\Rightarrow$ less uncertainty $\Rightarrow$ less processing cost

An NLP-system gives the probability distribution

The Red House tells the story of **a mysterious, tormented individual** who breaks into **toy shops and museums** to steal **dolls and puppets**. Once they are in his power...

1. Probability distribution from parameters of resolution system
2. Calculate relative entropy over this probability distribution

Antecedent of they	Probability	Relative entropy
The Red House	0.05	
a mysterious, tormented individual	0.04	
toy shops and museums	0.31	0.83
dolls and puppets	0.58	
$\emptyset$	0.02	

Implementation

Obtain probabilities from the best NLP-system (Lee et al. 2017)

Lee et al.'s system:

- End to end: without pre-processing
- Neural-network architecture
- Ranking system

2. The study of Pronoun Resolution on a Reading Corpus

The Dundee Eye-Tracking Corpus (Kennedy et al. 2003)

Eye-movements of 10 native English speakers
Reading 65 texts
From the Independent (newspaper)
Total: 50 000 tokens
Annotated with part of speech (Frank 2010) and dependency relations (Barrett et al. 2015)

Annotation of the antecedent of all 1 109 anaphorical pronouns.
A data-set to study pronoun resolution in natural data.

Reading Behaviour

Reading: a sequence of fixations on a text

Each fixation has a duration, expressed in milliseconds.

The eye jumps from fixation to fixation.

(Rayner 1998)

Reading Behaviour

216

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

156

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

227

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

187

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

182

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

96

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

232

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

335

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

168

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

Are tourists enticed by these attractions threatening their very existence ?

173

Reading Behaviour

188

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

Are tourists enticed by these attractions threatening their very existence ?

88

Reading Behaviour

174

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

168

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

170

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

271

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

88

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

232

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

202

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

222

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

157

Are tourists enticed by these attractions threatening their very existence ?

Reading Behaviour

157

Are tourists enticed by these attractions threatening their very existence ?

The pattern of fixations is used in various reading metrics.

Reading times: sum of fixation durations in one region

Are | tourists | enticed | by | these | attractions | threatening | their | very | existence?

Often, one word is one region.

Assumption: Longer reading time ⇒ more processing difficulty

(Rayner 1998)

Example: first pass reading time & total reading time

$$\underbrace{\text{Are}}_{1}\ \underbrace{\text{tourists}}_{2}\ \underbrace{\text{enticed}}_{3,13}\ \text{by}\ \underbrace{\text{these}}_{4,14,15}\ \underbrace{\text{attractions}}_{5,16,17}\ \underbrace{\text{threatening}}_{6,7,18}\ \underbrace{\text{their}}_{9}\ \underbrace{\text{very}}_{8,19}\ \underbrace{\text{existence?}}_{10,11,12,20,21}$$

First pass: $\sum$ durations of fixations 10, 11 and 12

Total: $\sum$ durations of fixations 10, 11, 12, 20 and 21

Measuring reading time for pronouns: a hard problem

Pronouns are fixated only 20 - 30% of the time.
(Ehrlich and Rayner 1983)

Pronouns are very short.

Spill-over effects

Previous experiment:
Take a window of words around the pronoun.

- ... at a time **[when they are at greatest risk]**, and then ...
- ... on it; **[but it would seriously degrade the]** quality ...

Problems:

- Need multiple models
- Few data-points per pronoun

Solution: binomial metric

A simpler reading metric was more suited:
Is the pronoun fixated or not?

Binomial outcome: yes/no answer.

Advantages:

- More data points
- There is only one point to measure

“a word is skipped because it has been identified on the previous fixation” (Brysbaert and Vitu 1998)

Hypothesis: a fixated pronoun indicates more processing difficulty.

Statistical Model

A statistical model predicted whether the pronoun is fixated or not.

Is the relative entropy of importance to this prediction?

Mixed effects model:
fixated $\sim$ length + frequency + comma + punctuation + rel_ent
+ (1 + rel_ent | participant) + (1 | dundee_tokens)

Result

The entropy cost metric predicts reading behaviour

The relative entropy was a predictor in reading behaviour.

A lower distance between the entropy and the maximal entropy
$\Rightarrow$ more participants fixating the pronoun

Estimate: -0.07 (95% Credible interval = [-0.01, -0.13])

Conclusion:
Information Theory is also relevant to pronoun resolution.

Conclusion

A lot of exploratory work

Positive and negative outcomes:

- Example positive outcome: entropy cost metric
- After the thesis, confirm the positive results on other data-sources.

Take home messages:

- Probabilistic NLP-systems can serve as models to simulate human behaviour.
  - Hypothesis: The reason is that humans are also sensitive to statistical phenomena in language; this affects language processing.
- Cognitive computational models can handle the multi-factor problem of pronoun resolution.
  - Possibility to test theories on natural data.
  - Help to develop theories.

Future Work

Compare different cognitive computational models to evaluate their plausibility

- Parallel function

Evaluate existing theories of pronoun resolution using corpus data

- Bayesian Theory of Pronoun Resolution (Kehler and Rohde 2013)

Develop cognitive computational models of processing difficulty that integrate multiple linguistic levels

- Syntax, semantics, discourse

Bibliography I

Barrett, Maria, eljko Agi, and Anders Søgaard (2015). "The Dundee Treebank". In: *The 14th International Workshop on Treebanks and Linguistic Theories (TLT 14)*.

Brysbaert, Marc and Françoise Vitu (1998). "Word skipping: Implications for theories of eye movement control in reading". In: *Eye guidance in reading and scene perception*. Elsevier, pp. 125–147.

Ehrlich, Kate and Keith Rayner (1983). "Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing". In: *Journal of Verbal Learning and Verbal Behavior* 22.1, pp. 75–87.

Frank, Stefan L (2010). "Uncertainty reduction as a measure of cognitive processing effort". In: *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics*. Association for Computational Linguistics, pp. 81–89.

Bibliography II

Kehler, Andrew and Hannah Rohde (2013). "A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation". In: *Theoretical Linguistics* 39.1-2, pp. 1–37.

Kennedy, Alan, Robin Hill, and Joël Pynte (2003). "The dundee corpus". In: *Proceedings of the 12th European conference on eye movement*.

Lee, Kenton, Luheng He, Mike Lewis, and Luke Zettlemoyer (2017). "End-to-end Neural Coreference Resolution". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 188–197.

Rayner, Keith (1998). "Eye movements in reading and information processing: 20 years of research." In: *Psychological bulletin* 124.3, p. 372.

Bibliography III

Van Deemter, Kees and Rodger Kibble (2000). "On coreferring: Coreference in MUC and related annotation schemes". In: *Computational linguistics* 26.4, pp. 629–637.