Using Robustness to Learn to Order Semantic Properties in Referring Expression Generation

Pablo Duboue, Martín Domínguez

Facultad de Matemática, Astronomía y Física
Universidad Nacional de Córdoba
Córdoba, Argentina

Our Work at a Glance (AI)

- Our objective is to learn an **ordering** of properties.

- This ordering is used within an algorithm, we are using as a metric the number of errors this algorithm makes when using each potential ordering.

- We haven't succeeded (yet) but have encouraging signals.

Our Work at a Glance (NLP)

- Natural Language Generation (NLG)
  - Multiparagraph text construction from non-linguistic data.
- Referring Expression Generation (REG)
  - Distinguishing a given entity from a set of confusors.
- REG algorithm: Incremental Algorithm
  - Needs a Property Ordering (PO).
- This work: learning PO using on errors
  - On old data, produce a ref. expression, check whether holds.
- Intuitions
  - A good referring expression should refer to stable properties
- Results
  - Robustness helps to learn ordernings
  - But popularity on DBpedia is a stronger signal

Referring Expression Generation (REG)

- Classic NLG problem
  - **Input:** set of entities (with a distinguished element), set of triples pertaining to the entities.
  - **Output:** a Definite Description, i.e., a set of *positive triples* and *negative triples*.
  - Focus on running time **efficiency** and generating **succint** and **easily understandable** expressions.

- Example output
  - Task: {‘Eben_Moglen’(EB), ‘Lawrence_Lessig’(LL), ‘Linus_Torvalds’(LT)}

Referent	Incremental Algorithm	Gardent
EB	{ (EB *occupation* Software_Freedom_Law_Center) }	{ (EB *occupation* Software_Freedom_Law_Center) }
LL	{ (LL *birthPlace* United_States), (LL, *occupation* Harvard_Law_School) }	{ (LL *birthPlace* Rapid_City,_South_Dakota) }
LT	{ (LT *occupation* Software_engineer) }	{ (LT *nationality* Finnish_American) }

Incremental Algorithm (IA) – an established REG algo

- Introduced in [Dale and Reiter, 1995]
  - Greedy approach, use a **default ordering**: Preference Order (PO)
  - Iterates over PO and selects a type
  - Adds a triple of the given type one at a time
  - Removes from the confusor set C all entities ruled out by the new triple
  - Triples that do not eliminate any new entity from C are ignored
  - The algorithm terminates when C is empty.
- Many other algorithms
  - Graph
  - Full Brevity
  - Gardent's

Experiments With Wikinews-derived REG Tasks

- Wikinews, a news service operated as a wiki
  - Entities disambiguated by *interwiki* links.

```
Former [[New Mexico]] {{w|Governor of New
Mexico|governor}} {{w|Gary Johnson}} ended his
campaign for the {{w|Republican Party (United
States)|Republican Party}}
```

- Human-written Property Ordering [Pacheco et al., 2012]:

TYPE ORDERINOFFICE NATIONALITY COUNTRY PROFESSION BIRTHPLACE LEADERNAME^{-1} KEYPERSON^{-1} AUTHOR^{-1} COMMANDER^{-1} OCCUPATION KNOWNFOR INSTRUMENT SUCCESSOR MONARCH SUCCESSOR^{-1} PRIMEMINISTER^{-1} ACTIVEYEARSENDDATE PARTY DEATHDATE DEATHPLACE CHILD ALMAMATER ACTIVEYEARSSTARTDATE RELIGION SPOUSE PRESIDENT^{-1} NOTABLECOMMANDER^{-1} VICEPRESIDENT PRESIDENT PRIMEMINISTER AWARD MILITARYRANK CHILD^{-1} MILITARYCOMMAND SERVICESTARTYEAR OFFICE BATTLE SPOUSE^{-1} KNOWNFOR^{-1} PREDECESSOR FOUNDATIONPERSON^{-1} MONARCH^{-1} PREDECESSOR^{-1} ACTIVEYEARSSTARTYEAR ACTIVEYEARSENDYEAR STARRING^{-1} LIEUTENANT PARENT GOVERNOR^{-1} HOMEPAGE RESIDENCE APPOINTER^{-1} ...

Experiments With Wikinews-derived REG Tasks

- Wikinews, news articles with *interwiki* links.

```
Former [[New Mexico]] {{w|Governor of New Mexico|governor}}
{{w|Gary Johnson}} ended his campaign for the {{w|Republican
Party (United States)|Republican Party}}
```

- Focus on people and organizations

Category	Definition	Size
People	Entity has “birth date”?	3,051 tasks
Organizations	Entity has “creation date”?	2,370 tasks

Using Change to Simulate Errors

Duboue, Dominguez, Estrella. *On the Robustness of Standalone Referring Expression Generation Algorithms Using RDF Data.* WebNLG 2016.

- Three algorithms of REG on anachronistic input.
  - On old data, produce a referring expression, check whether holds on new data.
- We found poor results with marginal differences among the algorithms.
- Metrics
  - Dice: $\frac{2|T \cap T'|}{|T|+|T'|}$, where T is the target set (always $|T| = 1$) and T' is the obtained set.
  - Inclusion errors: the ref. exp. on old data added spurious referents
  - Exclusion errors: the ref. exp. on old data excluded the referent

Data: DBpedia

- DBpedia [Bizer et al., 2009] is an ontology curated from Wikipedia infoboxes
  - Infoboxes are the small tables containing structured information at the top of most Wikipedia pages.
  - Not to be confused with a new project targeting to provide structured information to Wikipedia, wikidata.
- Two Versions: Compared

Property	3.6	2014
Unique subjects (entities)	1,668,503	4,218,628
Unique objects (types)	250	547
Max objects per subject	6	16
Number of verbs	1,100	1,370
Number of triples	13,795,664	33,449,633

GA Details

- We define a GA to traverse the search space of all possible permutations of Preference Orders (POs), $92!$.
- (1) *individuals* are arrays of 92 elements.
- (2) *fitness function*: we tested robustness and popularity.
- (3) A strategy for evolution:
  - *mutation*: swap between two elements randomly selected.
  - *crossover*: Let $[g_1, \ldots, g_i, \ldots, g_n]$ and $[h_1, \ldots, h_i, \ldots, h_n]$be two individuals. Let $i, 0 \leq i \leq n$ then, the resulting individual is: $sub([g_1, \ldots, g_n], [h_{i+1}, \ldots, h_n]) \cdot [h_{i+1}, \ldots, h_n]$ **where** $(\cdot)$ appends arrays; $sub(x, y)$ deletes elements in x that are in y.
  - *crossover* 0.8 prob. of being applied, *mutation* 0.08.
  - selection strategy the *tournament selection*, 7 as parameter.
  - population of 200 individuals; it evolves for 50 generations.
- We used the *ECJ* java library to implement the GA.

Iberamia Metrics

- Measure learned POs against the hand-written PO
  - Kendall's τ [Lebanon and Lafferty, 2002]:

$$\tau = 1 - \frac{2(\text{number of inversion})}{N(N-1)/2}$$

    * too strict, moved to a metric that considers the REs being generated rather than the exact ordering
  - Dice over selected properties.
    * Seemingly very different POs produce comparable results
    * Metric of choice (we refer to it as “target”)
- Do observable variables change similarly to target?
  - Spearman's rho

First Experiments

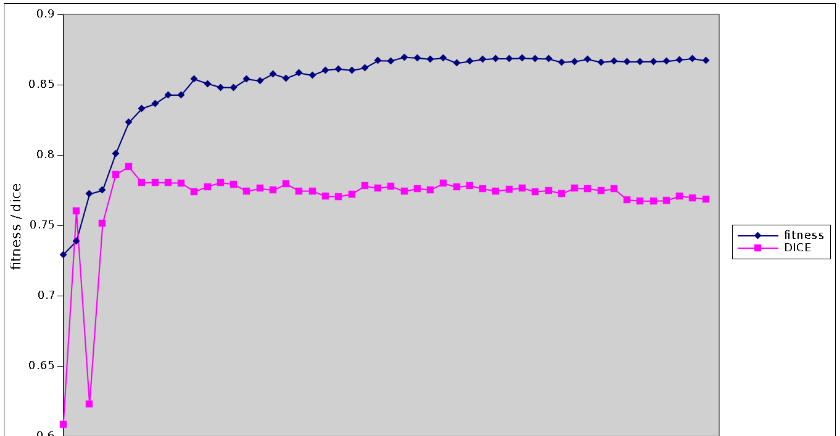
- Experiment: Correlations over People

Exp/metric	length	Dice	exclusion errors	inclusion errors
Hand-written	-0.018	-0.215	0.185	0.397
Popularity	-0.226	0.232	-0.258	0.394

- Experiment: overfit fitness function
  - A function that approximates the Dice for the hand-picked PO using the observable variables. Linear regression:

    $$target = -1.608 * length + 15.5279 * inclusion + 0.8787 * exclusion + 1.9403$$

  - Pearson's correlation coefficient of 0.872
- Experiment:
  Can the GA learn?



Main Experiments

- Experiment: Genetic Algorithm over organizations using fitness trained on peopled
  - Disappointment: after 50 generations, we get a target of only 0.435
  - When popularity PO achieves 0.93.
- This is our main negative result

Closing Experiments

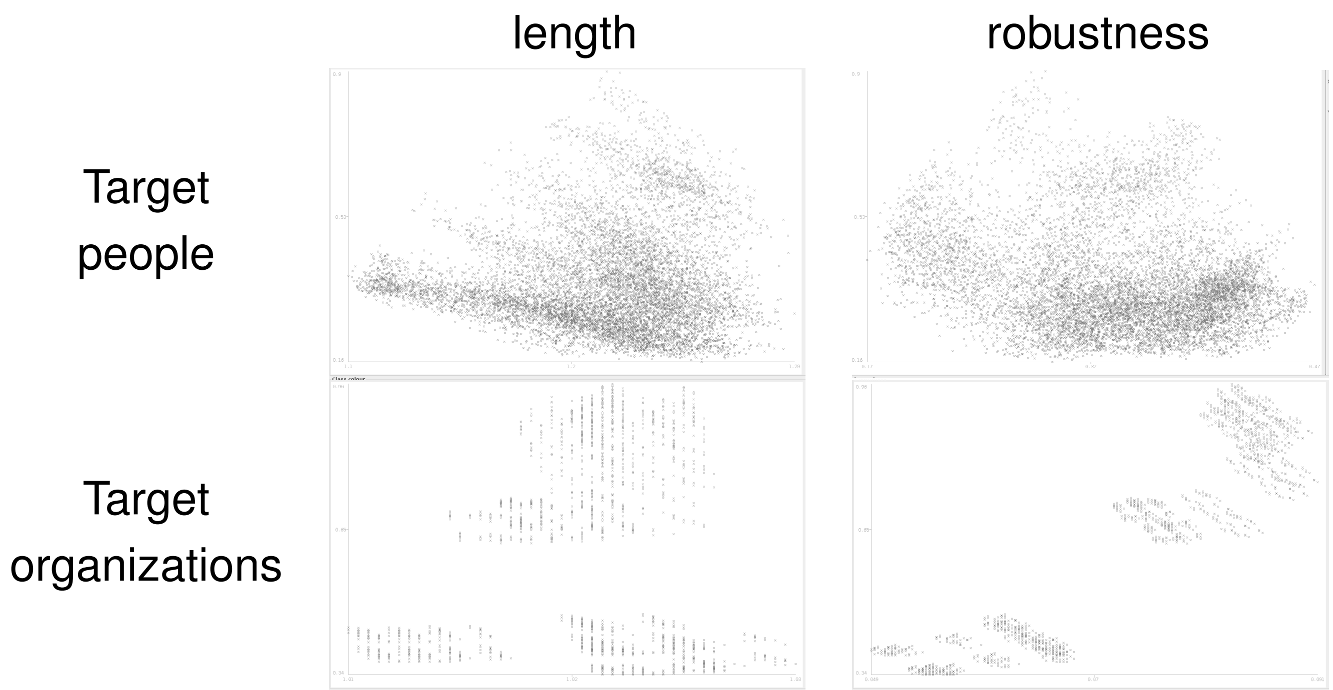
- Experiment: Correlations over Organizations
  - Strong Spearman's rho, but very different from people's numbers.

Exp/metric	length	Dice	exclusion errors	inclusion errors
Hand-written	0.059	0.832	-0.834	0.840
Popularity	-0.064	0.864	-0.866	0.841

- Experiment: Only length & inclusion errors
  - Preliminary result, insight obtained from looking at both tables (test set)
  - Trained on people, target on organizations of 0.906
  - Trained on organizations, target on people of 0.608
  - Below the popularity PO but more generalization strength
- Experiment: GA using only inclusion errors
  - Dice of 0.272 (people) and 0.361 (organizations)
  - Robustness alone is not enough, combining it with length is key.

Discussion

- Main result: correlation between hand-written PO and robustness
- Lack of generalization: organizations change differently from people (hypothesis)



Other Material

- 2014 course Aprendizaje Automático sobre Grandes Volúmenes de Datos
  - Slides in Spanish, under Creative Commons licence
  - Hadoop / Mahout sample code / YouTube videos
  - Three end-to-end case studies: recommendation, clustering and classification
  - `http://aprendizajengrande.net`
- Information Extraction for Open Data
  - Two IE pipelines, one in Apache UIMA using rules and CRFs
  - Thousands of documents in French and Spanish available for download
  - Instructional material (in English) from my ECI course
  - `http://ie4opendata.org`

Conclusions

- DBpedia/Wikinews is a suitable source for doing research on robust REG algorithms.
- Robustness has a correlation with a hand-written PO.
- People and organizations are too different to generalize from one another
- Where to go from here:
  - Focus on sub-types of entities (sub-types of organizations or people).
  - Focus on popularity ordering.

Backup Slides

Implementation Details

- Alusivo: an Open Source implementation of REG algorithms
  - `https://github.com/DrDub/Alusivo`
  - Java, Maven, RDF-based
- Interface
  - `public ReferringExpression resolve(URI referent, List<URI> confusors, RepositoryConnection repo)`
- Libraries
  - Sesame (RDF)
  - ChocoSolver (CSP)
  - jgrapht (Graph algorithms)

References

[Bizer et al., 2009] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009).

[Dale and Reiter, 1995] Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.

[Lebanon and Lafferty, 2002] Lebanon, G. and Lafferty, J. (2002). Combining rankings using conditional probability models on permutations. In Sammut, C. and A. Hoffmann, e., editors, *Proceedings of the 19th International Conference on Machine Learning*, San Francisco, CA. Morgan Kaufmann Publishers.

[Pacheco et al., 2012] Pacheco, F., Duboue, P. A., and Domínguez, M. A. (2012). On the feasibility of open domain referring expression generation using large scale folksonomies. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, pages 641–645, Stroudsburg, PA, USA. Association for Computational Linguistics.