Towards Word Sense Disambiguation by Reasoning

Javier Álvez Itziar Gonzalez-Dios German Rigau

LoRea & IXA Groups
Computer Languages and Systems Department
University of the Basque Country (UPV/EHU)

The fifth Vampire Workshop — FLoC 2018, Oxford, UK

December 10, 2018

Word Sense Disambiguation (WSD) (Agirre and Edmonds, 2007)

- Natural Language Processing (NLP) research line
- Choosing the intended sense of a word in a context
- Example:
  - *turtle* is either a "*a sweater or jersey with a high close-fitting collar*" or "*any of various aquatic and land reptiles having a bony shell and flipper-like limbs for swimming*"
  - Contexts:
    - "*The black fleece is soft as cotton candy, and is second on my preferred turtle list*"
    - "*Some turtles lay eggs in the sand and leave them to hatch on their own. The young turtles make their way to the top of the sand and scramble to the water while trying to avoid predators*"

Knowledge Resources I

- BLESS data (Baroni-Lenci Evaluation of Semantic Similarity) (Baroni and Lenci, 2011)
  - Designed for the differentiation between semantic relations and compositionality
  - It includes 200 concrete nouns (*targets*) (100 animate and 100 inanimate nouns) from different classes (e.g., *tools*, *clothing*, *vehicles*, *animals*, etc.)
  - Each target is associated with a set of other words (nouns, verbs or adjectives) via:
    - *hyperonymy*
    - *cohyponymy*
    - *meronymy*
    - *attribute*
    - *event*
    - *random*
  - Example:
    - *turtles* are *amphibians* (hyperonymy)
    - *turtles* and *frogs* are coordinate (cohyponymy)
    - *turtles* have *legs* (meronymy)
    - *turtles* are *slow* (attribute)
    - *turtles walk* (event)

Knowledge Resources II

- WordNet (Fellbaum, 1998) is a large lexical database of English
- Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (*synsets*): lexical concepts that are synonyms
  - Example: $turtleneck^1_n$, $turtle^1_n$, $polo\text{-}neck^1_n$
- Synsets can be identified by its *offset*
  - Example: $turtle^1_n$ is identified by 04502197-n
- Synsets are related among them: hyponymy, meronymy, ...

Knowledge Resources III

- Adimen-SUMO (Álvez et al., 2012) is a FOL ontology about common-sense knowledge
- It has been obtained by applying a reengineering process to SUMO (Niles and Pease, 2001)
  - With the help of ATPs (*Automated Theorem Provers*)
  - Around an 88% of the *core* of SUMO (top and middle levels) is translated
  - Domain ontologies are not used (by now)
- The resulting ontology can be used in tasks that involve reasoning with commonsense knowledge
- Example:

```
( =>
   ($instance ?LIMB Limb)
   (exists (?VERTEBRATE)
      (and
         ($instance ?VERTEBRATE Vertebrate))
         (properPart ?LIMB ?VERTEBRATE))))
```

Knowledge Resources IV

- The mapping between WordNet and SUMO (Niles and Pease, 2003)
- It connects synsets of WordNet to terms of SUMO using 3 relations:
  - *equivalence* (=)
  - *subsumption* (+)
  - *instance* (@)
- Some examples:

$$\begin{array}{rcl} \langle turtle_n^1 \rangle & : & [Clothing_c+] \\ \langle turtle_n^2 \rangle & : & [Reptile_c+] \\ \langle clothing_n^1 \rangle & : & [Clothing_c=] \\ \langle Snake_n^4 \rangle & : & [AstronomicalBody_c@] \end{array}$$

$Snake_n^4$ is "*a long faint constellation in the southern hemisphere near the equator stretching between Virgo and Cancer*"

New WSD Proposal I

- Our objective is to demonstrate the practical capabilities of Adimen-SUMO for NLP tasks using Vampire (Kovács and Voronkov, 2013)
- On the basis of our black-box testing method (Álvez et al., 2015, 2017b)
  - Adaptation of the methodology for the design and evaluation of ontologies introduced in (Grüninger and Fox, 1995)
  - Based on the use of Competency Questions (CQs):
    - Problems that an ontology is expected to answer
  - CQs are automatically created on the basis of few Question Patterns (QPs) by exploiting WordNet and its mapping into SUMO
  - Classification of (dual) problems (truth- and falsity-tests) by means of ATPs:
    - *Passing*: the ATPs are able to demonstrate a truth-test
    - *Non-passing*: the ATPs are able to demonstrate a falsity-test
    - *Unknown*: the ATPs produce no answer within a time limit

New WSD Proposal II

- We consider all the synsets to which words belong
  - Targets and related words
- We apply several predefined QPs to the resulting synset pairs
- For each target, we choose the most likely synset:
  - According to the difference between passing and non-passing CQs
- Example: *turtles* (2 synsets) have (meronymy) *legs* (9 synsets) $\rightarrow$ in total 18 synset pairs for this word pair
  - $turtle_n^2$ "any of various aquatic and land reptiles having a bony shell and flipper-like limbs for swimming", $Reptile_c+$
  - leg_n^1 "a structure in animals that is similar to a human leg and used for locomotion", $Limb_c+$
  - Resulting CQ:

```
(exists (?X ?Y)
   (and
      ($instance ?X Reptile)
      ($instance ?Y Limb)
      (properPart ?Y ?X)))
```

New WSD Proposal III

- We consider all the synsets to which words belong
  - Targets and related words
- We apply several predefined QPs to the resulting synset pairs
- For each target, we choose the most likely synset:
  - According to the difference between passing and non-passing CQs
- Example: *turtles* (2 synsets) have (meronymy) *legs* (9 synsets) -> in total 18 synset pairs for this word pair
  - $turtle_n^1$ "a sweater or jersey with a high close-fitting collar", $Clothing_c+$
  - leg_n^3 "one of the supports for a piece of furniture", $Artifact_c+$
  - Resulting CQ:

```
(exists (?X ?Y)
   (and
      ($instance ?X Clothing)
      ($instance ?Y Artifact)
      (properPart ?Y ?X)))
```

New WSD Proposal IV

- There are 14,400 BLESS pairs (excluding *random*)
- We obtain 386,891 synset pairs, from which 127,918 synset pairs yield to incorrect formulas
- From the remaining 258,973 synset pairs, the resulting CQs are:

Pattern	**#1**	**#2**	**#3**	**#4**	Total
hyperonymy	2,559	458	–	–	**3,017**
cohyponymy	6,243	1,147	1,128	204	**8,722**
meronymy	6,669	1,852	1,556	393	**10,470**
attribute	3,092	600	685	109	**4,486**
event	12,575	3,220	3,357	721	**19,873**
Total	–	–	–	–	**46,568**

Summary

- Adimen-SUMO v2.6
  - Including the *Closed World Assumption* (CWA) (Reiter, 1978) applied to *$subclass* and *$disjoint*
- Vampire v4.2.2 in an Intel® Xeon® CPU E5-2640v3@2.60GHz
  - CASC portfolio
  - Resource limits: 300 seconds and 2GB of memory
- Raw results:

Pattern	**#1**		**#2**		**#3**		**#4**		Solved	
hyperonymy	+1,027	-944	+173	-157	–	–	–	–	**1,201**	**39.81%**
cohyponymy	+712	-364	+189	-23	+216	-14	+27	-4	**1,549**	**17.76%**
meronymy	+785	-6	+37	-0	+109	-0	+0	-0	**937**	**8.95%**
attribute	+90	-84	+8	-1	+4	-3	+0	-0	**190**	**4.24%**
event	+0	-0	+0	-0	+0	-0	+0	-0	**0**	**0.00%**
Total	–	–	–	–	–	–	–	–	**3,877**	**8.33%**

WSD Results and Comparison

- We have manually constructed a *gold standard* for the 200 targets

WSD Results and Comparison

- We have manually constructed a *gold standard* for the 200 targets
  - 20 targets have been disambiguated to 2 senses, and 3 targets to 3 senses
  - Even a difficult task for humans, or impossible taken into account in the context words BLESS

dress (clothing)

- GOLD $dress_n^1$ $frock_n^2$ "*a one-piece garment for a woman; has skirt and bodice*", $Clothing_c+$
- $attire_n^1$ $garb_n^1$ $dress_n^2$ "*clothing of a distinctive style or for a particular occasion*", $Clothing_c+$
- $apparel_n^1$ $wearing\_apparel_n^1$ $dress_n^3$ $clothes_n^1$ "*clothing in general*", $Clothing_c+$

villa (building)

- $Villa_n^1$ $Pancho\_Villa_n^1$ $Francisco\_Villa_n^1$ $Doroteo\_Arango_n^1$ "*Mexican revolutionary leader (1877-1923)*", $Man_c@$
- GOLD $villa_n^2$ "*detached or semidetached suburban house*", $Building_c+$
- GOLD $villa_n^3$ "*country house in ancient Rome consisting of residential quarters and farm buildings around a courtyard*", $Building_c+$
- GOLD $villa_n^4$ "*pretentious and luxurious country residence with extensive grounds*", $StationaryArtifact_c+$

WSD Results and Comparison

- We have manually constructed a *gold standard* for the 200 targets
- We have disambiguated the 200 targets using UKB (Agirre et al., 2018), which is state-of-the-art in WSD
- We take as baseline a random disambiguator
- Results:

WSD	#	%
Baseline	–	55.38%
UKB	180	90.00%
Adimen-SUMO	142	71.00%

- UKB disambiguates correctly 39 targets that Adimen-SUMO does not
- Adimen-SUMO disambiguates correctly 14 targets that UKB does not
- The upper bound is 97.00% when using both UKB and Adimen-SUMO

Some Examples

Why are we not able to disambiguate *donkey* (ground_mammal) and UKB is?

- $donkey^1_n$ "*the symbol of the Democratic Party; introduced in cartoons by Thomas Nast in 1874*", $Icon_c+$
- GOLD $domestic\_ass^1_n$ $donkey^2_n$ $Equus\_asinus^2_n$ "*domestic beast of burden descended from the African wild ass; patient but stubborn*", $HoofedMammal_c+$
- **Context** mero: neck, ear, leg... hyper: animal, mammal, vertebrate... event: die, eat, live... coord: fox, lion, pig... and attri: big, large, stubborn...
- Why???

Some Examples II

Why are we able to disambiguate *fighter* (vehicle) and UKB is not?

- $combatant^1_n$ $battler^1_n$ $belligerent^1_n$ $fighter^1_n$ $crapper^1_n$ "*someone who fights (or is fighting)*", $SocialRole_c+$
- GOLD $fighter^2_n$ $fighter\_aircraft^1_n$ $attack\_aircraft^1_n$ "*a high-speed military or naval airplane designed to destroy enemy aircraft in the air*", $AirCraft_c+$
- $champion^2_n$ $fighter^3_n$ $hero^3_n$ $paladin^1_n$ "*someone who fights for a cause*", $Human_c+$
- **Context** mero: missile, seat, metal... hyper: plane, vehicle, transport... event: leave, go, run... coord: car, bus, train... and attri: big, destructive, lethal...
- Why???

Some Examples III

Why no approach is able to disambiguate *herring* (water_animal)?

- $herring_n^1$ "*valuable flesh of fatty fish from shallow waters of northern Atlantic or Pacific; usually salted or pickled*", $Meat_c+$
- GOLD $herring_n^2$ $Clupea\_harangus_n^1$ "*commercially important food fish of northern waters of both Atlantic and Pacific*", $Fish_c+$
- **Context** mero: eye, skin, tail... hyper: **fish, food,** animal... event: **cook, live**, eat (eat or to be eaten?)... coord: salmon, tuna, cod... and attri: edible, fresh, small...
- Both meanings (animal and flesh) are mixed in BLESS dataset

Error Analysis Remarks

- No possible disambiguation if both synsets belong to the same SUMO concept
- Context in BLESS can also be ambiguous (or not so fine grained as in WordNet), e.g. the attri *edible* applied to animals does not help to decide weather it is the animal or its flesh or the trees and fruit or vegetables

Conclusions

- Adimen-SUMO is subject of improvement
  - Mapping and knowledge errors
  - Lack of information (e.g. about events)
- Vampire v4.2.2 is able to work with Adimen-SUMO
  - Specialized portfolios
  - Increase resource limits
- Anyway, Adimen-SUMO can be applied to NLP tasks in its current state

Ongoing Work

- We are correcting Adimen-SUMO
  - Using black-box and white-box techniques (Álvez et al., 2017a)
  - Cross-checking Adimen-SUMO with other knowledge resources such as WordNet (Álvez and Rigau, 2018; Álvez et al., 2018)
- We are improving the formalization of Adimen-SUMO
  - By providing an stronger axiomatization: restrictions at the level of classes and not objects
  - By applying CWA and *Unique Name Assumption* (UNA) (Russell and Norvig, 2010)
- We are trying other ATPs like E (Schulz, 2002), CVC4 (Deters et al., 2014) or iProver (Korovin, 2013)
- We are implementing some optimizations of the resulting FOL formula
  - Removing non-constant function symbols
  - Preprocessing

Bibliography I

E. Agirre and P. Edmonds. *Word sense disambiguation: Algorithms and applications*, volume 33. Springer Science & Business Media, 2007.

E. Agirre, O. L. de Lacalle, and A. Soroa. The risk of sub-optimal use of open source NLP software: UKB is inadvertently state-of-the-art in knowledge-based WSD. *CoRR*, abs/1805.04277, 2018. URL `http://arxiv.org/abs/1805.04277`.

J. Álvez and G. Rigau. Cross-checking WordNet and SUMO using meronymy. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis, and T. Tokunaga, editors, *Proc. of the 11th Int. Conf. on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), may 2018.

J. Álvez, P. Lucio, and G. Rigau. Adimen-SUMO: Reengineering an ontology for first-order reasoning. *Int. J. Semantic Web Inf. Syst.*, 8(4):80–116, 2012.

Bibliography II

J. Álvez, P. Lucio, and G. Rigau. Improving the competency of first-order ontologies. In K. Barker and J. M. Gómez-Pérez, editors, *Proc. of the 8th Int. Conf. on Knowledge Capture (K-CAP 2015)*, pages 15:1–15:8. ACM, 2015. ISBN 978-1-4503-3849-3. doi: 10.1145/2815833.2815841.

J. Álvez, M. Hermo, P. Lucio, and G. Rigau. Automatic white-box testing of first-order logic ontologies. *CoRR*, abs/1705.10219, 2017a. URL `http://arxiv.org/abs/1705.10219`.

J. Álvez, P. Lucio, and G. Rigau. Black-box testing of first-order logic ontologies using WordNet. *CoRR*, abs/1705.10217, 2017b. URL `http://arxiv.org/abs/1705.10217`.

J. Álvez, I. Gonzalez-Dios, and G. Rigau. Validating WordNet meronymy relations using Adimen-SUMO. *CoRR*, abs/1805.07824, may 2018.

M. Baroni and A. Lenci. How we blessed distributional semantic evaluation. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 1–10. Association for Computational Linguistics, 2011.

Bibliography III

M. Deters, A. Reynolds, T. King, C. W. Barrett, and C. Tinelli. A tour of CVC4: how it works, and how to use it. In *Proc. of the Formal Methods in Computer-Aided Design (FMCAD 2014)*, page 7, 2014. URL `https://doi.org/10.1109/FMCAD.2014.6987586`.

C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

M. Grüninger and M. S. Fox. Methodology for the design and evaluation of ontologies. In *Proc. of the Workshop on Basic Ontological Issues in Knowledge Sharing (IJCAI 1995)*, 1995.

K. Korovin. Inst-Gen - A modular approach to instantiation-based automated reasoning. In *Programming Logics - Essays in Memory of Harald Ganzinger*, pages 239–270, 2013. doi: 10.1007/978-3-642-37651-1_10.

L. Kovács and A. Voronkov. First-order theorem proving and Vampire. In N. Sharygina and H. Veith, editors, *Computer Aided Verification*, LNCS 8044, pages 1–35. Springer, 2013. ISBN 978-3-642-39798-1.

Bibliography IV

I. Niles and A. Pease. Towards a standard upper ontology. In Guarino N. et al., editor, *Proc. of the 2nd Int. Conf. on Formal Ontology in Information Systems (FOIS 2001)*, pages 2–9. ACM, 2001. doi: 10.1145/505168.505170.

I. Niles and A. Pease. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In H. R. Arabnia, editor, *Proc. of the IEEE Int. Conf. on Inf. and Knowledge Engin. (IKE 2003)*, volume 2, pages 412–416. CSREA Press, 2003. ISBN 1-932415-08-4.

R. Reiter. *On Closed World Data Bases*, pages 55–76. Springer, Boston, MA, 1978. ISBN 978-1-4684-3384-5. doi: 10.1007/978-1-4684-3384-5_3. URL https://doi.org/10.1007/978-1-4684-3384-5_3.

S. J. Russell and P. Norvig. *Artificial Intelligence - A Modern Approach (3. internat. ed.)*. Pearson Education, 2010. ISBN 978-0-13-207148-2. URL http://vig.pearsoned.com/store/product/1,1207,store-12521_isbn-0136042597,00.html.

S. Schulz. E - A brainiac theorem prover. *AI Communications*, 15(2-3):111–126, 2002. ISSN 0921-7126.