From Narrative Text to VerbNet-Based DRSes

Introduction

Computational linguists have long studied various logic forms for capturing essential semantic information carried by narratives. Among these, discourse representation structure (DRS) [Kamp and Reyle, 1993] form is designed to acquire the entities, entities' property, events, event types, the occurring time of events, and event arguments. In this paper, we describe a system called Text2DRS that takes English narrative as an input and outputs a DRS in Neo-Davidsonian style. In this regard, it is similar to Boxer [Bos, 2008] which is an open-domain NLP tool for semantic analysis of a text and produces a DRS for a given narrative. Unlike Boxer, Text2DRS can capture and provide the missing information. Furthermore, Text2DRS relies on lexical resource VerbNet [Kipper-Schuler, 2005, Palmer, 2006] for annotating the specific relations between relevant entities and events mentioned in the narrative using the verb classes and thematic roles of VerbNet.

Text2DRS Details

Text2DRS is implemented on the top of the LTH system [Johansson and Nugues, 2007] and the Standford CoreNLP system [Manning et al., 2014]. The LTH is a semantic role labeler for unrestricted text in English that uses predicates and semantic roles from PropBank [Palmer et al., 2005]. The Standford CoreNLP system provides a set of NLP tools including the coreference resolution system. Text2DRS utilizes semantic role labeler function from LTH system and coreference resolution function from CoreNLP system to process the narrative. Additionally, Text2DRS also includes SemLink [Bonial et al., 2013] that maps predicates and semantic roles of PropBank to verb classes and thematic roles of VerbNet.

For example, given a narrative: "*John travelled to the hallway. Sandra journeyed to the hallway.*" Text2DRS generates output:

r1 r2 r3 e1 e2
entity(r1) entity(r2) entity(r3)
property(r1, John) property(r2, hallway) property(r3, Sandra)
event(e1) event(e2)
eventType(e1, run-51.3.2-1) eventTime(e1, 0)
eventArgument(e1, theme, r1) eventArgument(e1, location, r2)
eventType(e2, run-51.3.2-1) eventTime(e2, 1)
eventArgument(e2, theme, r3) eventArgument(e2, location, r2)

Table 1: DRS for the given narrative

The first block displays the entities and events introduced by the narrative.The entities are represented as *r1, r2, r3* ("r" stands for a referent), and the events are represented as *e1, e2*. The second block shows descriptive details about the entities and events of the narrative. The property is a mapping of an entity referent and its original text in the narrative. The eventType is a relation between an event referent and its corresponding VerbNet class. In this example, both verb "travelled" and "journeyed" are mapped to the verb class "run-51.3.2-1." An eventArgument relation presents information about events. For instance, eventArgument(e2,location,r2) says that entity *r2* (that has a property of being a "hallway") plays a thematic role "location" of event *e2* (that belongs to VerbNet class "run-51.3.2-1").

Figure 1 presents the Text2DRS's system architecture. The entity reference generator creates an entity coreference map from the CoreNLP output and uses this map to assign a reference ID to each distinct entity. From the LTH output, Text2DRS looks up the related VerbNet class information in the SemLink and returns the verb class number along with the respective thematic roles. However, sometime SemLink maps one PropBank predicate into multiple VerbNet classes. In this case, We pick the first verb class from the data list and use it in the final output. After entity reference generator and event reference generator complete process the data, the DRS generator merges the data and outputs the DRS for the given narrative.

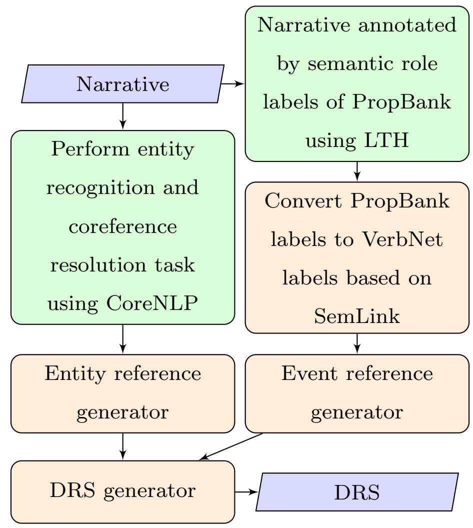

Figure 1: System architecture

Future Work

Furthermore, SemLink doesn't always contain the mapping of semantic role from the PropBank to the thematic role in the VerbNet. In some cases, we extended SemLink with missing mappings, and it is a future work effort to enhance SemLink further.

References

Claire Bonial, Kevin Stowe, and Martha Palmer. *SemLink*, 2013. `https://verbs.colorado.edu/semlink/` [Accessed: 2017].

Johan Bos. Wide-coverage semantic analysis with boxer. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, STEP '08, pages 277–286, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. URL `http://dl.acm.org/citation.cfm?id=1626481.1626503`.

Richard Johansson and Pierre Nugues. *Language Technology at LTH*, 2007. `http://nlp.cs.lth.se/` [Accessed: 2017].

Hans Kamp and Uwe Reyle. *From discourse to logic*, volume 1,2. Kluwer, 1993.

Karin Kipper-Schuler. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania, 2005.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014. URL `http://www.aclweb.org/anthology/P/P14/P14-5010`.

Martha Palmer. *VerbNet*, 2006. `https://verbs.colorado.edu/ mpalmer/projects/verbnet.html` [Accessed: 2017].

Martha Palmer, Daniel Gildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March 2005. ISSN 0891-2017. doi: 10.1162/0891201053630264. URL `http://dx.doi.org/10.1162/0891201053630264`.