The Challenge of Composition in Distributional and Formal Semantics

Ran Tian
Tohoku University, Japan
tianran@ecei.tohoku.ac.jp

Koji Mineshima
Ochanomizu University, Japan
mineshima.koji@ocha.ac.jp

Pascual Martínez-Gómez
AIRC, AIST, Japan
pascual.mg@aist.go.jp

1 Tutorial Description

The principle of compositionality states that the meaning of a complete sentence must be explained in terms of the meanings of its subsentential parts; in other words, each syntactic operation should have a corresponding semantic operation. In recent years, it has been increasingly evident that distributional and formal semantics are complementary in addressing composition; while the distributional/vector-based approach can naturally measure semantic similarity (Mitchell and Lapata, 2010), the formal/symbolic approach has a long tradition within logic-based semantic frameworks (Montague, 1974) and can readily be connected to theorem provers or databases to perform complicated tasks. In this tutorial, we will cover recent efforts in extending word vectors to account for composition and reasoning, the various challenging phenomena observed in composition and addressed by formal semantics, and a hybrid approach that combines merits of the two.

For introduction, we briefly review the syntax-semantics interface and word vectors, the two starting points of this tutorial.

Then, we discuss vector-based models for composition, in which word vectors are combined into phrase/sentence vectors according to some syntactic structure (Hashimoto et al., 2014; Pham et al., 2015; Tian et al., 2016). The word vectors and composition operations are jointly learned in an unsupervised manner in these models. We mention but do not focus on approaches disregarding syntax, such as recurrent neural networks.

Next, we move to recent advances in machine learning theory of the most fundamental composition operation, the additive composition (Tian et al., 2017). We explain why additive composition works, how to train additive compositional vectors and how to use them in semantic composition.

As a final part of the vector-based approach, we discuss applications of vector-based composition related to database and reasoning, such as Guu et al. (2015) and Rocktäschel et al. (2015).

We introduce the symbolic approach to composition, also under the principles behind the syntax-semantics interface; its symbolic nature allows the use of theorem provers to perform natural language inferences. As an example, we show how this semantic composition takes place over syntactic derivations in Combinatory Categorial Grammar (CCG) (Steedman, 2000). We also demonstrate how different semantic theories can be implemented over a variety of syntactic grammars (not only CCG) using ccg2lambda (Martínez-Gómez et al., 2016), an open-sourced general framework for compositional semantics.

We then introduce the problem of Recognizing Textual Entailment (RTE) (Dagan et al., 2009), where we test whether a text T entails a hypothesis H. We compare different logic frameworks, including first-order logic, higher-order logic (Mineshima et al., 2015, 2016), and natural logic (Abzianidze, 2015), and discuss semantically challenging constructions such as generalized quantifiers, adjectival modification and intensional operators, drawing on the English dataset FraCaS and the Japanese dataset JSeM for RTE.

The solution to many RTE problems requires the use of external linguistic knowledge such as synonyms, antonyms, and paraphrases. Since vector representations naturally encode semantic similarities between words and phrases, here we expect a great synergy between the formal and distributional approaches. In the closing section of this tutorial, we introduce a widely adopted hybrid approach toward RTE, in which semantic similarities between words and phrases are explicitly converted to logic rules as linguistic knowledge used in inference (Tian et al., 2014; Beltagy et al., 2016;

Proceedings of the The 8th International Joint Conference on Natural Language Processing, pages 16–17,
Taipei, Taiwan, November 27 – December 1, 2017.

Martínez-Gómez et al., 2017). This approach has the merit that all knowledge is explicit, and it can easily integrate existing linguistic ontologies such as WordNet. We demonstrate how the distributional approach can overcome the low coverage of linguistic resources by composing phrase vectors faithful to meaning and compatible with logic, and how the formal approach can reduce computational complexity in logical inference by identifying the need of linguistic knowledge between specific concepts and constructing axioms on-demand.

2 Tutorial Outline

- Introduction
  - The syntax-semantics interface
  - Word vectors
- Vector-based approach
  - Vector-based composition models
  - Theory of additive composition
  - Vector-based reasoning
- Symbolic approach
  - ccg2lambda: compositionality for your favorite semantic theory
  - Logic systems for RTE
  - RTE datasets for formal semantics: FraCaS and JSeM
- A hybrid approach toward RTE

Ran Tian is a Project Assistant Professor at Tohoku University. He works on theory of vector-based composition, and integration of distributional semantics into logical inference.

Koji Mineshima is a Project Associate Professor at Ochanomizu University. He is working on Formal Semantics, Logic and Computational Linguistics. His recent work has focused on CCG semantic parsing, anaphora resolution and recognizing textual entailment.

Pascual Martínez-Gómez is a Research Scientist at the Artificial Intelligence Research Center, AIRC-AIST. He investigates extensions of tree transducers and their applications to recognizing textual entailment, fact validation and question-answering over large knowledge bases.

References

L. Abzianidze. 2015. A tableau prover for natural logic and language. In *EMNLP*.

I. Beltagy, S. Roller, P. Cheng, K. Erk, and R. J. Mooney. 2016. Representing meaning with a combination of logical and distributional models. *Computational Linguistics*, 42(4).

I. Dagan, B. Dolan, B. Magnini, and D. Roth. 2009. Recognizing textual entailment: Rational, evaluation and approaches. *Natural Language Engineering*, 15.

K. Guu, J. Miller, and P. Liang. 2015. Traversing knowledge graphs in vector space. In *EMNLP*.

K. Hashimoto, P. Stenetorp, M. Miwa, and Y. Tsuruoka. 2014. Jointly learning word representations and composition functions using predicate-argument structures. In *EMNLP*.

P. Martínez-Gómez, K. Mineshima, Y. Miyao, and D. Bekki. 2016. ccg2lambda: A compositional semantics system. In *ACL*.

P. Martínez-Gómez, K. Mineshima, Y. Miyao, and D. Bekki. 2017. On-demand injection of lexical knowledge for recognising textual entailment. In *EACL*.

K. Mineshima, P. Martínez-Gómez, Y. Miyao, and D. Bekki. 2015. Higher-order logical inference with compositional semantics. In *EMNLP*.

K. Mineshima, R. Tanaka, P. Martínez-Gómez, Y. Miyao, and D. Bekki. 2016. Building compositional semantics and higher-order inference system for a wide-coverage Japanese CCG parser. In *EMNLP*.

J. Mitchell and M. Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8).

R. Montague. 1974. English as a formal language. In R. Montague, editor, *Formal philosophy*. Yale University Press.

N. T. Pham, G. Kruszewski, A. Lazaridou, and M. Baroni. 2015. Jointly optimizing word representations for lexical and sentential tasks with the c-phrase model. In *ACL*.

T. Rocktäschel, S. Singh, and S. Riedel. 2015. Injecting logical background knowledge into embeddings for relation extraction. In *NAACL*.

M. Steedman. 2000. *The Syntactic Process*. MIT Press.

R. Tian, Y. Miyao, and T. Matsuzaki. 2014. Logical inference on dependency-based compositional semantics. In *ACL*.

R. Tian, N. Okazaki, and K. Inui. 2016. Learning semantically and additively compositional distributional representations. In *ACL*.

R. Tian, N. Okazaki, and K. Inui. 2017. The mechanism of additive composition. *Machine Learning*.