

## Introduction to the special issue on learning semantics

**Antoine Bordes · Léon Bottou · Ronan Collobert ·  
Dan Roth · Jason Weston · Luke Zettlemoyer**

Received: 2 April 2013 / Accepted: 3 May 2013 / Published online: 7 June 2013  
© The Author(s) 2013

A key ambition of AI is to render computers able to evolve and interact with the real world. This can be made possible only if the machine is able to produce an interpretation of its available modalities (image, audio, text, etc.) which can be used to support reasoning and taking appropriate actions. Computational linguists use the term *semantics* (Lewis 1970) to refer to the possible interpretations of natural language sentences. A growing number of efforts to develop machine learning approaches for semantic analysis now aim to find (in an automated way) these interpretations (Miller et al. 1996; Zelle and Mooney 1996; Zettlemoyer and Collins 2005; Mitchell and Lapata 2008; Bordes et al. 2010; Chen and Mooney 2011; Liang et al. 2011). However, the need for semantic analysis is not restricted to natural

---

A. Bordes  
CNRS—Heudiasyc UMR 7253, Université de Technologie de Compiègne, Compiègne, France  
e-mail: [antoine.bordes@utc.fr](mailto:antoine.bordes@utc.fr)

L. Bottou  
Microsoft Research, New York, NY, USA  
e-mail: [leon@bottou.org](mailto:leon@bottou.org)

R. Collobert  
IDIAP, Martigny, Switzerland  
e-mail: [ronan.collobert@idiap.ch](mailto:ronan.collobert@idiap.ch)

D. Roth  
University of Illinois at Urbana-Champaign, Urbana, IL, USA  
e-mail: [danr@illinois.edu](mailto:danr@illinois.edu)

J. Weston  
Google, New York, NY, USA  
e-mail: [jweston@google.com](mailto:jweston@google.com)

L. Zettlemoyer (✉)  
University of Washington, Seattle, WA, USA  
e-mail: [lsz@cs.washington.edu](mailto:lsz@cs.washington.edu)

language (and speech), but is also crucial for visual understanding (Farhadi et al. 2009; Felzenszwalb and McAllester 2010; Gupta et al. 2010). Automatically recovering visual concepts and relationships would provide a leap forward in scene analysis and image parsing. Our hypothesis here is that there are many common problems in semantic analysis for language, vision, and other modalities, and that organizing a special issue will enable a unified view of a wide range of approaches that will enhance all of the individual efforts.

Progress in learning semantics has been slow mainly because this involves sophisticated models which are hard to train, especially since they seem to require large quantities of training data annotated with detailed semantic representations (Miller et al. 1996; Zelle and Mooney 1996; Zettlemoyer and Collins 2005). However, recent advances in learning with weak, limited and indirect supervision led to the emergence of a new body of research in semantics based on multi-task/transfer learning (Lampert et al. 2009; Socher et al. 2011; Collobert et al. 2011), on semi-supervised learning or learning with ambiguous/indirect supervision (Kate and Mooney 2007; Cour et al. 2009; Clarke et al. 2010; Artzi and Zettlemoyer 2011; Bordes and Glorot 2012; Matuszek et al. 2012) or even with no supervision at all (Poon and Domingos 2010; Goldwasser et al. 2011). Hence, this topic is gaining importance in the machine learning community. The NIPS'11 workshop on Learning Semantics<sup>1</sup> attracted substantial numbers of attendees (around 80) coming from different backgrounds and with different viewpoints. This special issue was created in order to collect and expose some of this disparate work. Our goal is to provide a snapshot of the current state of this emerging field, and to serve as a springboard for future directions.

A total of 15 submissions was received, 7 of which were finally accepted for this special issue. Each accepted paper has gone through two to three rounds of reviewing, each round with three to four referees. Among the 15 submissions, 4 were co-authored by guest editors of the special issue. These 4 papers have been edited separately by Ronan Collobert and Luke Zettlemoyer, without any involvement of the other editors. The contents of this special issue cover different modalities such as images, natural and robot control languages, or graphs and present different machine learning approaches, from reinforcement learning or clustering to deep learning. Interestingly, all papers consider weak supervision settings such as ambiguous, indirect or unsupervised learning and attempt to tackle complicated problems with realistic amounts of labeled data.

The definition of semantics and hence of reasoning within a statistical machine learning context is not well defined and can be quite controversial. The first paper of the special issue, “From Machine Learning to Machine Reasoning” by Léon Bottou is an essay which attempts to bridge trainable systems, like neural networks, and sophisticated “all-purpose” inference mechanisms, such as logical or probabilistic inference. By defining reasoning as “algebraically manipulating previously acquired knowledge in order to answer a new question”, the paper shows that there exists a conceptual continuity between these inference systems and simple manipulations, such as the mere concatenation of trainable learning systems. It is then proposed to enrich the set of manipulations applicable to training systems in order to build reasoning capabilities from the ground up.

The concept of semantics in machine learning primarily concerns learning to map natural language utterances to logical forms expressing their meaning. The special issue contains four papers related to this topic, with learning frameworks utilizing differing levels of supervision.

---

<sup>1</sup><http://nips.cc/Conferences/2011/Program/event.php?ID=2520>.

The paper “Learning Perceptually Grounded Word Meanings from Unaligned Parallel Data” by Stefanie Tellex, Pratiksha Thaker, Joshua Joseph and Nicholas Roy describes an approach to map natural language commands to actions for a forklift control task. The goal of the system is to learn to interpret natural language commands to a forklift operator such as going to a particular location or picking up an object. While the previous algorithm required annotating the correspondence between linguistic constituents and their groundings, the newly presented method only requires a demonstration of the high-level actions corresponding to each instruction. By using a reward function based on whether a particular grounding results in the specified action, the algorithm uses a policy gradient method to learn a mapping between language and perceptual features of the environment.

The paper “Interactive Relational Reinforcement Learning of Concept Semantics” by Matthias Nickles and Achim Rettinger presents a Relational Reinforcement Learning (RRL) approach for learning denotational concept semantics using symbolic interaction of artificial agents with human users. Unlike standard approaches aiming at learning word senses and other language aspects using text corpora, their approach allows for the interactive learning of concepts using a dialog of human and agent as supervision. The novelty of this paper is the use of a human-agent interaction component as part of a RRL framework, termed Interactive RRL, and the use of an Answer Set Programming implementation of Event Calculus for efficient use of formal rules as background knowledge. This new model is studied in depth in a block-world domain augmented with dialog actions.

The paper “Towards Natural Instruction-based Learning” by Dan Goldwasser and Dan Roth also studies the problem of allowing a human user to interact with an artificial system using natural instructions. In their setting, a human teacher wants to communicate the relevant domain expertise to an artificial learner with no prior information about the internal representations used for the machine learning process. The paper presents a new learning algorithm for the instruction interpretation problem that relies on feedback from its performance on a final task only, and which can hence be trained with only human-level task expertise.

The paper “A Semantic Matching Energy Function for Learning with Multi-relational Data” by Antoine Bordes, Xavier Glorot, Jason Weston and Yoshua Bengio also considers the mapping of natural language to logical meaning representations but only via an application of word-sense disambiguation. Its main goal is to define new algorithms for embedding knowledge bases (represented as multi-relational graphs) into vector spaces. The goal is for those vector representations to encode some of the inherent semantics of the original data. Several variants of neural network able to conveniently embed large-scale knowledge bases are presented and evaluated for the tasks of link prediction and word-sense disambiguation.

In addition to the papers related to language and knowledge representations, the special issue also presents two works on learning semantics for computer vision.

The paper “Learning What is Where from Unlabeled Images: Joint Localization and Clustering of Foreground Objects” by Ashok Chandrashekar, Lorenzo Torresani and Richard Granger studies the problem of automatically discovering salient objects in the observed environment. They tackle this challenging problem by proposing a generative model of object formation. They describe an efficient algorithm to automatically learn the parameters of the model from a collection of unlabeled images. Their proposal ends up being capable of partitioning a given collection of images into disjoint clusters, depending on their main displayed object and of localizing this object in the image. Their method is shown to reach state-of-the-art results on the problem of unsupervised foreground localization and clustering.

The paper “Learning Semantic Representations of Objects and their Parts” by Grégoire Mesnil, Antoine Bordes, Jason Weston, Gal Chechik and Yoshua Bengio mixes natural language and vision tasks by trying to jointly learn representations for object labels and image representations. More specifically, they are interested in jointly learning representations both for the objects in an image, and the parts of those objects, with the objective that such deeper representations could be useful in image retrieval or browsing. They propose a method able to learn to jointly label objects and parts without requiring exhaustively labeled data by using a proxy supervision obtained by combining standard image annotation with semantic part-based within-label relations.

This special issue would not have been possible without the contributions of many people. We wish to sincerely thank all the authors for submitting their work to this special issue. We wish to express our gratitude to all the referees for their expertise and dedication in providing invaluable comments and suggestions. We are also grateful to MLJ Editor-in-Chief, Peter Flach, for his encouraging support, and the editorial office for their consistent help.

## References

- Artzi, Y., & Zettlemoyer, L. (2011). Bootstrapping semantic parsers from conversations. In *Proceedings of the conference on empirical methods in natural language processing*.
- Bordes, A., & Glorot, X. (2012). Joint learning of words and meaning representations for open-text semantic parsing. In *Proc. of the 17th intern. conf. on artif. intel. and stat.*
- Bordes, A., Usunier, N., Collobert, R., & Weston, J. (2010). Towards understanding situated natural language. In *Proc. of the 13th intern. conf. on artif. intel. and stat.*
- Chen, D. L., & Mooney, R. J. (2011). Panning for gold: finding relevant semantic content for grounded language learning. In *Proceedings of symposium on machine learning in speech and language processing (MLSPLP 2011)*.
- Clarke, J., Goldwasser, D., Chang, M., & Roth, D. (2010). Driving semantic parsing from the world’s response. In *Proceedings of the conference on computational natural language learning*.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, 2493–2537.
- Cour, T., Sapp, B., Jordan, C., & Taskar, B. (2009). Learning from ambiguously labeled images. In *IEEE conference on computer vision and pattern recognition (CVPR 2009)* (pp. 919–926). New York: IEEE.
- Farhadi, A., Endres, I., Hoiem, D., & Forsyth, D. (2009). Describing objects by their attributes. In *IEEE conference on computer vision and pattern recognition (CVPR 2009)* (pp. 1778–1785). New York: IEEE.
- Felzenszwalb, P., & McAllester, D. (2010). *Object detection grammars* (Computer Science TR-2010-02, tech. rep.). University of Chicago.
- Goldwasser, D., Reichart, R., Clarke, J., & Roth, D. (2011). Confidence driven unsupervised semantic parsing. In *Proceedings of the Association of Computational Linguistics*.
- Gupta, A., Efros, A., & Hebert, M. (2010). Blocks world revisited: image understanding using qualitative geometry and mechanics. In *Computer vision—ECCV 2010* (pp. 482–496).
- Kate, R. J., & Mooney, R. J. (2007). Learning language semantics from ambiguous supervision. In *Proceedings of AAAI* (Vol. 22, p. 895). Menlo Park: AAAI Press.
- Lampert, C. H., Nickisch, H., & Harmeling, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In *IEEE conference on computer vision and pattern recognition (CVPR 2009)* (pp. 951–958). New York: IEEE.
- Lewis, D. (1970). General semantics. *Synthese*, 22, 18–67. doi:10.1007/BF00413598.
- Liang, P., Jordan, M. I., & Klein, D. (2011). Learning dependency-based compositional semantics. In *Proceeding of the Association for Computational Linguistics (ACL)*.
- Matuszek, C., FitzGerald, N., Zettlemoyer, L., Bo, L., & Fox, D. (2012). A joint model of language and perception for grounded attribute learning. In *Proceedings of the international conference on machine learning*.
- Miller, S., Stallard, D., Bobrow, R., & Schwartz, R. (1996). A fully statistical approach to natural language interfaces. In *Proceedings of the Association for Computational Linguistics*.
- Mitchell, J., & Lapata, M. (2008). Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT* (pp. 236–244).

- Poon, H., & Domingos, P. (2010). Unsupervised ontology induction from text. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics* (pp. 296–305). Stroudsburg: Association for Computational Linguistics.
- Socher, R., Lin, C. C., Ng, A. Y., & Manning, C. D. (2011). Parsing natural scenes and natural language with recursive neural networks. In *Proceedings of the 26th international conference on machine learning (ICML)*.
- Zelle, J., & Mooney, R. (1996). Learning to parse database queries using inductive logic programming. In *Proceedings of the national conference on artificial intelligence*.
- Zettlemoyer, L., & Collins, M. (2005). Learning to map sentences to logical form: structured classification with probabilistic categorial grammars. In *Proceedings of the conference on uncertainty in artificial intelligence*.