Symmetry-Based Semantic Parsing

Chloé Kiddon and **Pedro Domingos**
Department of Computer Science & Engineering
University of Washington
Seattle, WA 98105
{chloe,pedrod}@cs.washington.edu

Abstract

Semantic parsing maps sentences to formal meaning representations, enabling question answering, natural language interfaces, and many other applications. However, there is no agreement on what the meaning representation should be, and constructing a sufficiently large corpus of sentence-meaning pairs for learning is extremely challenging. In this paper, we argue that both of these problems can be avoided if we adopt a new notion of semantics. For this, we take advantage of symmetry group theory, a highly developed area of mathematics concerned with transformations of a structure that preserve its key properties. We define a symmetry of a sentence as a syntactic transformation that preserves its meaning. Semantically parsing a sentence then consists of inferring its most probable orbit under the language's symmetry group, i.e., the set of sentences that it can be transformed into by symmetries in the group. The orbit is an implicit representation of a sentence's meaning that suffices for most applications. Learning a semantic parser consists of discovering likely symmetries of the language (e.g., paraphrases) from a corpus of sentence pairs with the same meaning. Once discovered, symmetries can be composed in a wide variety of ways, potentially resulting in an unprecedented degree of immunity to syntactic variation.

1 Introduction

The goals of natural language semantics are to represent the meanings of sentences formally and to relate those meanings. Semantic parsers map sentences into their formal meaning representations. Traditionally, semantic meanings have been characterized in terms of formal logical languages and benefitted from logical entailment (Montague, 1970). One challenge for semantic parsing is that there is little consensus on which meaning representations to choose. A second challenge is creating a sufficiently large training corpus of sentences labeled with their meaning representation for supervised learning methods; such data sets are developed at high human cost (Miller et al. 1994, Zettlemoyer and Collins 2005). We propose a new notion of semantics using insights from symmetry group theory that avoids these challenges.

Symmetry group theory studies the formal properties of *symmetry groups*, which are groups of transformations under which key properties of a structure are preserved (Miller Jr., 1972). We introduce the concept of a semantic symmetry group, which contains syntactic operations which when applied to a sentence preserve its meaning. Since symmetry groups are closed under composition, we can use semantic symmetries that were not seen at training time. A semantic symmetry group partitions the set of all sentences into *orbits*, sets of sentences with the same meaning. The semantic parse of a sentence is the orbit of which it is a member. Since natural language frequently contains ambiguities, we utilize a probabilistic approach to semantic symmetry and orbit membership. Properties of symmetry group theory allow the design of compact probabilistic models of meaning over which inference is efficient.

While symmetry group theory does not involve an explicit system of entailment, we hypothesize that symmetry-based semantic parsing is extensible to entailment, since entailment rules act like symmetries on a knowledge base.

Symmetry-based semantic parsers can learn from paraphrase corpora, which are easy and cheap to produce. Recent work has focused on utilizing data sets without explicitly paired meanings, such as data sets of question-answer pairs (Clarke et al., 2010) or conversation logs (Artzi and Zettlemoyer, 2011). Paraphrase corpora are just as easy to generate while allowing for a more direct form of supervision. Another related work is unsupervised semantic parsing (USP), which learns clusters of meaning-equivalent lambda forms (Poon and Domingos, 2009). Unlike USP, symmetry-based semantic parsers learn from supervised data and are easier to use.

Symmetry is foundational in modern physics (Wigner, 1967), and has been applied in many areas of computer science such as search (Crawford et al., 1996), model checking (Ip and Dill, 1996), and vision (Liu et al., 2010). The trend towards paraphrase-based semantic parsing shows the strength of using symmetry-like paraphrase transformations for semantics. However, to date, previous approaches have still targeted pre-defined meaning representations (Jurčíček et al. 2009, Fader et al. 2013) or have had limited forms of composition of transformations (MacCartney 2009, Stern et al. 2011). Most importantly, while transformation-based semantics has become increasingly popular, there is still limited understanding of how expressive it can be or what its formal properties are. We seek to develop a general framework of semantic symmetry that connects previous work and formalizes the use of syntactic transformations for modeling semantics.

2 Symmetry group theory

A *symmetry* of a structure x is a function that when applied leaves key properties unchanged. For example, a 120° clockwise rotation of an equilateral triangle is a symmetry for that triangle since its shape and orientation are unaffected. A *group* is an ordered pair $(G, \circ)$ where G is a set and $\circ$ is an operation which together satisfy the set of *group axioms*: (1) **Closure**: $\forall\ t_i, t_j \in G$, $t_i \circ t_j \in G$; (2) **Associativity**: $\forall\ t_i, t_j, t_k \in G$, $t_i \circ (t_j \circ t_k) = (t_i \circ t_j) \circ t_k$; (3) **Identity**: $\exists\ e \in G$ such that $\forall\ t_i \in G$, $e \circ t_i = t_i \circ e = t_i$; (4) **Inverses**: $\forall\, t_i \in G$, $\exists\ t_i^{-1} \in G$, such that $t_i \circ t_i^{-1} = t_i^{-1} \circ t_i = e$. A *symmetry group* of a structure is a group where the set contains symmetries and the group operation $\circ$ is function composition.

For example, a symmetry group for an equilateral triangle consists of clockwise rotations of 120° and 240°, reflections across the three medians, and doing nothing (i.e., the identify transformation). Every composition of symmetries is equivalent to a symmetry in the group; for example, a rotation and then a reflection of an equilateral triangle is equivalent to a different reflection.

The *orbit* of a structure x under a symmetry group G is the set of all structures that x can be transformed to by application of elements in G: $O(x) = \{g.x | g \in G\}$, where the period represents the application of the symmetry g on x.

We can compactly represent a symmetry group G in terms of a *generating set* T. G is generated by T if every symmetry in G can be expressed as a composition of a finite number of elements of T.

3 A symmetry group theory of semantics

A language L is a set of strings; for natural language, strings are sentences. Let C be the set of all constituents of sentences in L. Constituents may have subconstituents; for a constituent c, let $Ch(c)$ be the ordered list of c's subconstituents.

Definition 3.1 *A syntactic transformation on a language L is a function f: $C \to C$ such that, for every sentence in L with a constituent c, the sentence with c replaced by $t(c)$ is also in L.*

For example, t_1: *sunshine* $\to$ *pizza* is a valid syntactic transformation even through it will change the meaning of sentences. *sunshine* $\to$ *enjoys* is not a syntactic transformation as it would create invalid sentences. A syntactic transformation applies to all constituents in its specified language, but it only maps a certain set of constituents to different constituents. For all other constituents, it acts like the identity transformation. For example, t_1 leaves the sentence *William wore sunglasses* unchanged since the word *sunshine* is not present.

3.1 Semantic symmetries

Let M be the set of all constituent meanings and the *interpretation* i: $C \to M$ be a function from each constituent to its meaning. The meaning of a constituent c is $i(c)$.

Definition 3.2 *A symmetry t: $C \to C$ is a bijective syntactic transformation on a language L such that for every constituent c in L, $i(c) = i(t(c))$.*

Since *happy*, *glad*, and *jolly* are synonyms, there exists a symmetry t_1: *happy* $\to$ *glad* $\to$ *jolly* $\to$ *happy* that permutes these words. Applying t_1 to *William is happy* creates the sentence *William is glad*, without altering the meaning. Another symmetry may reposition a clause, changing *William is glad when it is sunny* to *When it* *is sunny, William is glad*. Every symmetry is reversible; t_1^{-1} is *happy* $\rightarrow$ *jolly* $\rightarrow$ *glad* $\rightarrow$ *happy*.

The symmetries of a language can be given as axioms or learned from data (e.g., a corpus of pairs of sentences with the same meaning). A symmetry group G of a language L is the group of all the symmetries of L under function composition.

Proposition 3.1 *If G is the symmetry group of a language L, there is a one-to-one correspondence between the orbits of G and the meanings of L.*

Since the semantic symmetries that generate an orbit for a constituent preserve its meaning, it follows that there is this one-to-one correspondence between meanings and orbits. The set of unique sentence orbits forms a partition of L. The meaning of a sentence is represented implicitly by membership in one of those orbits.

3.2 A generative model of sentences

Given the symmetry group G, the probability of a sentence s is given by $P(s) = \sum_{m \in M} P(m)P(s|m)$, where M is the set of meanings for the set of all sentences S. As a running example, let's assume a simple language L_{simple} where all sentences are subject-verb-object triples. The probability of a meaning is the product of the probabilities of an agent, an event, and a patient. Since there is a one-to-one correspondence between meanings and orbits, the generative model can be written as $P(s) = \sum_{o \in O} P(o)P(s|o)$, where O is the set of all orbits of a language L given its symmetry group G.

3.3 Orbit distributions

An orbit distribution assigns a probability to each element of an orbit o, such that $\sum_{c \in o} P(c \mid o) = 1$. We focus on orbit distributions that can be written compactly using a recursive definition.

The orbit of a word consists of itself and its synonyms. Assume the following orbits of words with their probabilities: $\{$(*William*, 0.2), (*Bill*, 0.8)$\}$, $\{$(*wore*, 0.6), (*donned*, 0.4)$\}$, and $\{$(*sunglasses*, 0.7), (*shades*, 0.3)$\}$. Since an orbit corresponds to a meaning, which is an agent-event-patient triple, we can construct all the sentences in the sentence orbit. For example, it is clear that there are eight sentences in the orbit of sentences about William wearing his sunglasses: to generate a sentence we select a word from the William orbit, the wearing orbit, and the sunglasses orbit independently and combine those selections (i.e., with a Cartesian product). The probability of a sentence in L_{simple} given an orbit decomposes into the product of the probabilities of the word choices given the agent, event, and patient orbits. (E.g., given the orbit under discussion, the probability of the sentence *Bill wore shades* is $0.8 * 0.6 * 0.3 = 0.144$.)

Proposition 3.2 *The orbit of constituent c under symmetry group G, $O_G(c)$, is the union of the Cartesian product of the orbits of $Ch(c')$ over all $c' \in O_G(c)$.*

In general, the orbit of a sentence will not decompose perfectly into a product. For example, if we add the notion of passive voice to L_{simple}, we add eight new sentences to the language (e.g., *Sunglasses were worn by Bill.*). However, in the passive sentences, the order of the agent and the patient orbits are swapped, so they will have a different decomposition than the active voice sentences. (The Cartesian product is not commutative.) The union of the Cartesian products of the child orbits from the active sentences and from the passive sentences does generate the whole sentence orbit.

The symmetry that swaps between active and passive voice applies to all sentences; if a sentence is in active voice, it becomes passive, and vice versa. We noted above that given an orbit of sentences we can split them between the sentences with active voice and those with passive voice. Once we select a voice, the subset of the orbit of sentences with that property can be decomposed into a product of the child orbits.

Given a symmetry $t \in G$ there is a set of constituents C_t that t acts on in a non-identity way. We postulate that for the orbit of $c \in C_t$ under G, $O_G(c)$, there exists a partitioning of $O_G(c)$ into subsets called *sub-orbits* such that for each sub-orbit $O_G(c)_i$, the Cartesian product of the orbits of $Ch(c')$ for all $c' \in O_G(c)_i$ are equivalent and equal to $O_G(c)_i$. For symmetries of active/passive voice, the sentences are partitioned by which voice they demonstrate. Another example involves actions where the set of child orbits is different among the subsets (e.g., here the orbits for *sold* and *bought* are children of only some of the sentences: **x** *sold* **y** *to* **z** $\rightarrow$ **z** *bought* **y** from **x**). The probability of choosing a constituent from a sub-orbit o_i given an orbit o can be calculated by summing the probabilities of the sub-orbit's constituents: $P(o_i|o) = \sum_{c \in o_i} P(c|o)$. For a partition of o, the probability of the sub-orbits sums to 1.

$P(c|o)$ is defined in two levels. The probability of a constituent given an orbit is the weighted sum of the probabilities of the constituent given an orbit's sub-orbits:

$$P(c|o) = \sum_{o_i} P(o_i|o)P(c|o_i).$$

The probability of a constituent given a sub-orbit is the product of the probabilities of the constituent's subconstituents given their orbits:

$$P(c|o_i) = \prod_{c' \in Ch(c)} P(c'|O_G(c'))$$

Note that the orbit distributions can share child orbits distributions. For example, the probability distribution over the orbit of $\{$*William*, *Bill*$\}$ can be shared among all sentences that involve William. Because different sentence orbits share in this way, the full probabilistic model over all possible orbits can be written compactly.

4 Symmetry-based semantic parsing

The goal of symmetry-based semantic parsing is to identify the most probable orbit, and thereby the implicit meaning representation, of a sentence.

Theorem 4.1 *The most probable orbit of a sentence s can be found in polynomial time.*

Proof sketch: We follow the proof of Theorem 1 in Gens and Domingos 2013 that states the partition function, and therefore also the MAP state, of a sum-product network (SPN) can be computed in time linear to the number of edges in the SPN. We can apply that proof here, because like an SPN, orbits are structured as levels of products and weighted sums.

This formulation requires a closed-world assumption on orbits; given a new sentence s, we find the best orbit for s from all previously seen orbits. Clearly, we need to be able to model never before seen orbits as well. We hypothesize we can extend the probabilistic model by adding a new unseen orbit o^* to the set of possible orbits. The prior $P(o^*)$ can be set to some constant depending on how likely it is to see a new orbit. The conditional probability $P(s|o^*)$ can be approximated by examining the symmetries that affect s. If o^* is the MAP semantic parse, the model will add o^* to the set of possible sentence orbits.

5 Learning symmetries

If the symmetry group G of a language L is provided, then the learning problem is simply to estimate the parameters of the orbit distributions. All of the orbit structure is defined solely by G acting on L. The main goal of learning, however, will be to discover the likely semantic symmetries of L.

Assume a training data set D of pairs of sentences with the same meaning. The sentences that are paraphrases of each other form a set of approximate orbits O. We seek a minimal generating set of symmetries T for G that maximizes the (penalized) likelihood of sentences given their approximate orbits:

$$\begin{aligned} L(S,O,\theta) &= P(S|O,\theta) \propto P(S,O,\theta) \\ &= \prod_{s \in S} \sum_{o \in O} P(o,\theta)P(s|o,\theta), \end{aligned}$$

where θ is the set of parameters for the orbit distributions. Given a generating set T for G, the probability distributions over the constituent orbits can be estimated empirically in closed form.

The sets of sentences with the same meaning give us a set of candidate symmetries based on the constituents that appear in meaning-equivalent sentences. For example, if we know that *Bill wore shades* has the same meaning as *Bill donned shades*, there is a high probability that *wore* $\rightarrow$ *donned* is a symmetry (and its inverse). To learn T, we score possible transformations on how likely they are to be symmetries and greedily add the highest scoring transformations to T. Because T is a minimal generating set, we may need to reduce symmetries in T as new symmetries are found. For example, if we first learn a symmetry *wore shades* $\rightarrow$ *donned sunglasses* and then learn *wore* $\rightarrow$ *donned*, we can infer that the first symmetry is a composition of the second symmetry and a new symmetry *shades* $\rightarrow$ *sunglasses*. T can then be altered to only include symmetries that are not compositions of other symmetries in T.

6 Conclusions and Future Work

We proposed symmetry-based semantic parsing as a new approach to semantic parsing that does not require a formal meaning representation or high cost labeled training corpora. Using symmetry group theory for semantics is a promising area of research still in its first stages. We plan to extend our theories to account for richer, more complex forms of sentence compositions, as well as symmetries at multiple levels of abstraction (e.g., under ontologies). Another important step will be examining how to incorporate entailment by viewing logical rules as symmetries of a knowledge base.

Acknowledgements

This research was partly funded by ARO grant W911NF-08-1-0242, ONR grants N00014-13-1-0720 and N00014-12-1-0312, and AFRL contract FA8750-13-2-0019. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ARO, ONR, AFRL, or the United States Government.

References

Y. Artzi and L. Zettlemoyer. 2011. Bootstrapping semantic parsers from conversations. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 421–432.

J. Clarke, D. Goldwasser, M. Chang, and D. Roth. 2010. Driving semantic parsing from world's response. In *2010 Conference on Natural Language Learning*.

J. Crawford, M. L. Ginsberg, E. Luks, and A. Roy. 1996. Symmetry-breaking predicates for search problems. In L. C. Aiello, J. Doyle, and S. Shapiro, editors, *KR96: Principles of Knowledge Representation and Reasoning*, pages 148–159.

A. Fader, L. Zettlemoyer, and O. Etzioni. 2013. Paraphrase-Driven Learning for Open Question Answering. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.

C. Ip and D. Dill. 1996. Better verification through symmetry. *Formal Methods in System Design*, 9(1/2):41–75.

J. Jurčíček, M. Gasic, S. Keizer, F. Mairesse, B. Thomson, and S. Young. 2009. Transformation-based learning for semantic parsing. In *Interspeech*.

Y. Liu, H. Hel-Or, C. S. Kaplan, and L. V. Gool. 2010. Computational symmetry in computer vision and computer graphics. *Foundations and Trends in Computer Graphics and Vision*, 5(1/2):1–195.

B. MacCartney. 2009. *Natural language inference*. Ph.D. thesis, Stanford University.

S. Miller, R. Bobrow, R. Ingria, and R Schwartz. 1994. Hidden understanding models of natural language. In *Proceedings of the Thirty-Second Annual Meeting of the Association for Computational Linguistics*, pages 25–32.

W. Miller Jr. 1972. *Symmetry Groups and Their Applications*. Academic Press, New York.

R. Montague. 1970. English as a formal language. In B. Visentini et al., editor, *Linguaggi nella Societa et nella Technica*, pages 188–221. Milan: Edizioni di Communita. Reprinted in Thomason (ed.) 1974, pages 188–221.

H. Poon and P. Domingos. 2009. Unsupervised semantic parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1–10.

H. Poon and P. Domingos. 2011. Sum-product networks: A new deep architecture. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*.

A. Stern, A. Lotan, S. Mirkin, E. Shnarch, L. Kotlerman, J. Berant, and I. Dagan. 2011. Knowledge and tree-edits in learnable entailment proofs. In *Proceedings of TAC*.

E. Wigner. 1967. *Symmetries and Reflections*. Indiana Univ. Press.

L. S. Zettlemoyer and M. Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*.