Machine Learning for Semantic Parsing in Review

Ahmad Aghaebrahimian

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
ebrahimian@ufal.mff.cuni.cz

Filip Jurčíček

Charles University in Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
jurcicek@ufal.mff.cuni.cz

Abstract

Spoken Language Understanding (SLU) and more specifically, semantic parsing is an indispensable task in each speech-enabled application. In this survey, we review the current research on SLU and semantic parsing with emphasis on machine learning techniques used for these tasks. Observing the current trends in semantic parsing, we conclude our discussion by suggesting some of the most promising future research trends.

Keywords: Spoken Language Understanding, Semantic Parsing, Machine Learning

1. Introduction

Around half a century ago, Alan Turing proposed the first model for machine understanding. From that time on, there has been a growing interest in developing technologies for machine understanding in different modalities. Speech, specifically, as one of the most elementary and natural communication medium, has attracted much attention. It brings together seemingly unrelated areas of science, from speech processing in electrical engineering and natural language processing in computer science to machine learning in artificial intelligence all together to challenge the task of Natural Language Understanding (NLU). In this survey, we focus on speech-enabled NLU or simply Spoken Language Understanding (SLU) as well as semantic parsing.

NLU and SLU are both designed to find conceptual representation of natural language sentences and utterances. What distinguishes SLU and NLU is the fact that SLU's input is an utterance, which in addition to prosody and speaker identity contains situational and contextual information. In plain words, the function of SLU component is to convert an utterance to a representation of the user's intention which is referred to as semantic representation. Semantic representations are usually in logical forms. Logical forms are detailed, context-independent, fully specified and unambiguous expressions which cover the utterance's arguments and predicates using an artificial language. This artificial language or technically formalism can be of different sorts, such as λ-Calculus (Church, 1936, 1941; Carpenter, 1997), first order fragments (Bird et al., 2009), robot controller language (Matuszek et al., 2012), etc.

SLU components are mostly used in Spoken Dialogue Systems (SDS) (Jurčíček et al., 2014; Young et al., 2013; Lee and Eskenazi, 2013), spoken information retrieval systems (speech mining) (De Jong et al., 2000) and automated speech translation (Xiaodong et al., 2013). The current expansion in the market of smart phones, smart watches (e.g. Google's and Apple's watches) and myriad of other speech-enabled gadgets such as digital personal assistants, home entertainment, and security systems opens up a great opportunity for more challenging efforts in SLU research and development. Apple's Siri, Google Now, Amazon's Echo, Microsoft's Cortana, Clarity Lab's Sirius, Nuance's Dragon Go and many other successful speech understanding applications are already true samples of such efforts. SLU has been the overall or partial focus of numerous research projects including ATIS (1989-1995), Verbmobile (1996-2000), How may I help you? (2001), AMITIES (2001-2004), TALK (2009-2011), Classic (2008-2011), Parlance (2011-2014), Carnegie Mellon Communicator (1999-2002), Companions (2006-2010) and Alex (2012-current), to name a few.

We organized the remaining of this survey into the following sections. After having an overview on the subject matter in Section 2, we review SLU sub-tasks in Section 3 and machine learning techniques for them in Section 4. Before we conclude in Section 6, we introduce two widely used parsing techniques and formalisms in Section 5.

2. SLU Overview

SLU components can typically be categorized into three systems as hand-crafted, data-driven or a combination thereof in hybrid systems. Hand-crafted systems are based on symbolic analysis of language while data-driven and more specifically, statistical systems are mostly based on the notions of information theory.

The hand-crafted paradigm is mostly advocated in artificial intelligence. In this paradigm, a common approach is to build intelligent agents which mimic the human brain for language understanding. Hand-crafted systems still have some strong proponents and basically many commercial applications today make use of them extensively. However, the advocates of the data-driven paradigm argue that mimicking biological organisms in mechanical machines is a fruitless effort. Fred Jelinek, the pioneer of statistical methods, once made an analogy between birds and air planes by arguing that airplanes do not flap their wings like birds and they still can fly.

Data-driven approaches benefit from a learning model which makes them language independent. Language independency is a significant advantage for data-driven systems over the hand-crafted ones. Efficient language-independent models in one language will work in another language provided that the features representing the new

language are fed into the model through a set of training data.

Although current data-driven approaches are proved more efficient in terms of development efforts and accuracy, similar to the hand-crafted ones, they are mostly domain-dependent, i.e. designed to recognize and work with specific notions and functions relating to a single topic. For instance, the air travel domain is a limited domain in which only a few notions such as departure_time or flight_number and a few functions such as flight_booking are recognized. However, open-domain systems capable of performing conversations in wider range of topics are more desirable. In open-domain applications, the number of the notions and the functions is practically unbounded. This is the reason why the design of Meaning Representation Language (MRL) for such systems poses a big challenge

In essence, data-driven and fully statistical SLU components provide us with an approach for easier adaptation to new domains and more robust performance as well as less deployment cost. In this survey, we concentrate on statistical and data-driven methods.

3. SLU Sub-Tasks

SLU implementations generally include three main components including domain detection, intent determination and slot filling components (Li et al., 2012).

3.1. Domain Detection

The Meaning Representation Languages (MRL) in most semantic parsers is designed to represent notions limited to a specific domain. Therefore, given an utterance, the first task of an SLU component is to recognize the relevant domain. This process is applied for multi-domain SLUs, while for single-domain SLUs, it suffices to detect out-of-domain utterances and handle them appropriately.

3.2. Intent Determination

In each domain, there is a set of pre-defined intentions which are specific to that domain. For instance, in the air traffic domain, 'ticket reservation' is a common intention and the intent determination component is expected to recognize this intention in user's utterance.

Until recently, all MRLs used for semantic parsing in different SLU components were limited to domain-dependent representations (Zelle and Mooney, 1996). It was because of their inherent limitation on the number of the lexicon they recognized and the number of the grammatical rules they realized to compose their lexicons into valid sentences. However, currently, there is huge interest in academics and in industry for open-domain MRLs and systems with unbound lexicons and grammar.

3.3. Slot Filling

SLUs may be used as a component in SDS (Black and Eskenazi, 2009), speech information retrieval (Hafen and Henry, 2012), spoken language translation (García et al., 2012) or spoken question answering system (Zettlemoyer and Collins, 2005), each of which interprets the task of slot filling in slightly different way. In the case of question answering for instance, slot filling is the task of mapping natural language utterances into their appropriate logical slots to form its logical representation (Zelle and Mooney, 1996)

Due to limited expressive power of current MRLs, these logical forms are limited to specific domains and expanding logical forms for larger domains or, ideally, for open-domain applications have been attracting special interest recently. It requires a learning capacity which makes systems able to recognize unbounded number of lexicons and grammar. Such broad coverage and learning capacity paves the way for semantic parsing in open-domain dialogue systems, open-domain question answering and open-domain information extraction.

4. Machine Learning Techniques

In the history of semantic parsing, different models, grammars, and techniques are examined for different SLU sub-tasks including generative models (e.g., Hidden Markov Models (HMM) (Schwartz et al. 1996)), discriminative techniques (Wang and Acero, 2006), or probabilistic context free grammars (Ward and Issar, 1994). All of these models and techniques have their own pros and cons; while generative models, for instance, are flexible in mapping individual words to their corresponding semantic tree nodes, they are limited in modeling local dependencies and co-related features. In contrast, latent-variable discriminative models are able to learn representations with long dependencies. To have more elaboration on such techniques, we briefly review some of the most practical machine learning methods for semantic parsing in the following sections.

4.1. CRF

State-of-the-art semantic parsing approaches use statistical machine learning techniques such as Conditional Random Fields (CRF) (Lafferty et al., 2001). CRFs are a class of discriminative sequence labeling models. They are essentially undirected graphical models. They are flexible in dealing with overlapping features and for this reason, CRFs usually outperform generative models such as HMM (Yao et al., 2013).

In contrast to Maximum Entropy Markov Model (MEMM) (Ratnaparkhi, 1998) classifiers which are left-to-right probabilistic models for sequence labeling, CRFs-based models provide global conditional distribution. Using this feature in CRFs, Xu and Sarikaya (2013 a, b) proposed a joint model for domain detection and intent determination which enhanced the SLU performance by decreasing the effect of error propagation in the SLU pipeline.

4.2. Neural Networks

Artificial Neural Networks (NN) have been recently successfully applied in SLU tasks. Compared to previous discriminative state-of-the-art models such as CRFs and SVMs, simple NNs such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) have significantly higher performance in some tasks, such as sequence labeling. (Collobert et al., 2011; Yao et al., 2014)

An RNN models a short-term memory and saves the current state of the network using recurrent connections. Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) is an advanced RNN model which has a forgetting gate with linear activation function and a

memory cell in addition to input, output and hidden layers in regular RNNs. Thanks to the memory cells, LSTMs can find and exploit long-range dependencies in the data better than simple RNNs.

The performance of NNs in many Natural Language Processing (NLP) applications is very promising and nowadays different architectures of NNs are being used for different NLP tasks. These tasks include but not limited to POS tagging, semantic parsing, dependency parsing, machine translation, etc.

NNs have shown a promising capacity in automatic feature learning, too (Mesnil et al., 2013); non-NN SLU approaches typically use handcrafted features (Liu and Sarikaya, 2014), such as N-gram patterns or bag-of-words. These features need to be designed by hand before the model can be trained. The problem of using such features is that for a typical utterance containing 10 words for instance, there would be tens of thousands features. Processing such an immense amount of features is very computationally expensive.

Instead, NNs can automatically learn efficient and dense features in the form of word embeddings. A word embedding projects a high-dimensional word representation vector into a dense low-dimensional one. This capacity of NNs in feature generation can be integrated into classical models, and some studies show that it enhances their performances significantly. For an instance, Xu and Sarikaya (2013a) exploited feature learning mechanism in neural networks and applied it to their joint model of intent detection and slot filling in SLU.

NNs have also gained vast popularity in classification, sequence labeling tasks and other tasks in SLU components. Deoras and Sarikaya (2013) and Sarikaya et al. (2014b) suggested the use of Deep Belief Networks (DBN) as a variant of NNs in semantic parsing. DBNs are stacks of Restricted Boltzmann Machines (RBM) which are basically probabilistic models. RBMs are trained in a process called contrastive divergence, in which each RBM layer is trained by using previous layers' hidden unit as its input unit. This process provides RBM layers with their initial weights. DBN is then tuned using the back-propagation algorithm. In contrast to CRFs, which estimate the global probability of a sequence of words, in DBN-based approaches, the conditional probability of a slot sequence is decomposed into a product of local probability functions. Each of these probabilities models the distribution of a particular slot tag given the context at that time stance. Deoras et al., (2013) expanded the DBN model in (Deoras and Sarikaya, 2013) for joint intent detection and slot filling.

4.3. Deep Learning (DL)

The DL approach encompasses many layers of non-linear information processing in Deep Neural Networks (DNN) with many hidden layers. A DNN is basically a stack of large number of simple NNs stacked on top of each other.

Deep Learning techniques is shown to improve the state-of-the-art in some SLU tasks such as slot filling (Sarikaya et al., 2014a,b; Mesnil et al., 2013). Recently Mesnil et al., (2013) investigated the task of slot filling in RNN by implementing different architectures (Elman-type, Jordan type and their variants) and they showed that their RNN architectures outperforms state-of-the-art CRF models for slot filling task.

4.4. Unsupervised / Distributional Methods

Supervised approaches in SLU have their own drawbacks. First, most supervised SLU systems today limit their application to narrowly defined domains. Second, high accuracy in supervised SLU methods is dependent on annotated samples which in many cases are costly to produce and subsequently, very sparse. Therefore, the trend of research on statistical SLU in recent years has shifted toward semi-supervised (Tür et al., 2005, Celikyilmaz et al., 2013) or unsupervised learning (Lorenzo et al., 2013) learning.

To continue our discussion about unsupervised learning in semantic parsing, we concentrate on distributional semantics theory. In Section 4.2, we talked about making use of embeddings as a model of feature representation in vector space. Here we use the notion of distributional semantics to define word meaning representation in vector space. Vector space representation in DNN is optimized to maximize the performance given the objective function. Vector Space Representation has recently received a growing body of research efforts and has been proved successful in many tasks. Notably in NLP tasks, Collobert et al. (2011) showed that low-dimensional word vectors learned by Neural Network Language Models (NNLM) are beneficial for many NLP tasks. However, low-dimensional distributional representation of meaning suffers from scalability issues due to data sparseness. To address such scalability issues in vector space, Mikolov et al. (2013) proposed highly scalable high-dimensional word vectors: They introduced Word2Vec for continuous word representation, which takes text as input and learns features of the words in the text –the word embeddings- as latent variables.

Word embeddings contain rich information about words linguistically and conceptually. They contain syntactic and semantic relationships among constituents of the text. However, they lack domain-dependent information which is of much importance in domain-constrained natural language queries. To address this, Celikyilmaz et al. (2015) extended the Mikolov et al. (2013)'s model by supplementing it with information about the relationship between entities extracted from a knowledge graph. They showed that their model improved performance of semantic tagging. They also extended the model by adding prior information to its objective function using synonymy relations between words from WordNet.

5. Parsing Approaches

Semantic representation for natural languages like Knowledge Representation (KR) for the phenomena in natural world is a hard task. There are a number of semantic representation formalisms, however, for the sake of brevity, we concentrate on λ-calculus as a widely used formalism for semantic representation in SLU components. λ-calculus represents the meaning of real world objects using a set of ontologies, and it uses syntactic rules for composition (Church, 1941). It has been widely used for semantic parsing whether directly as λ-expressions in (Zettlemoyer and Collins, 2005) or some other variants such as λ-Dependency Compositional

Semantics (λ-DCS) (Liang, 2013). We briefly introduce CCG and DCS which both use a variant of λ-calculus in their meaning representation formalism.

5.1. CCG

Compositional Categorial Grammar or CCG (Steedman, 1996, 2000) is a parsing scheme that combines lexicons and constitutionality in an elegant way. CCG uses λ-calculus as MRL and is considered a mildly context-sensitive and lexicalized grammar. It consists of a collection of lexicons and combinators which combine these lexicons to build up the meaning of a sentence.

CCG parsers are used in variety of applications from question answering (Kushman, 2014) to robot control (Krishnamurthy and Kollar, 2013) and semantic parsing (Kwiatkowski et al., 2011).

5.2. DCS and λ-DCS

DCS represents the formal semantics of an utterance using a tree structure. In DCS tree structures, lexicons are in the tree nodes and the dependencies among lexicons are captured through edges between the nodes.

By integrating λ-calculus into DCS, Liang (2013) introduced λ-DCS. Like CCG, DCS can be trained using annotated logical forms and map new utterances to their equivalent logical forms. But mapping utterances to their logical forms is not the only way of language understanding. In addition, this method suffers from two major limitations. First, it needs large amounts of annotated logical forms which are costly to produce. Second, it is limited to the concepts in the domain as well as the number of logical predicates (Wong and Moony 2007). A more desirable approach which is mostly referred to as 'Grounded Language Learning' exploits a more natural way of mapping utterances directly to real world responses like mapping questions to their correct answers. Liang and Potts (2015) investigated semantic parsing by directly mapping utterances to their denotations. These denotations can be of various kinds, such as the entries in a database, the responses from a real environment or even the responses generated by a cognitive model.

6. Conclusion

We provided a review of semantic representation formalisms and machine learning techniques used in current SLU research. In current SLU problems, moving from domain-dependent to open-domain systems is a great research and development leap. Building open-domain SLU modules through deeper linguistic understanding will widen the functionality of systems to large number of applications. Use of Big Data technologies via knowledge extraction techniques from large knowledge graphs is one active direction for this purpose. Another active direction is deep learning and the related technologies. Still, scaling-up techniques such as bootstrapping or search query logs or web search (for either structured or unstructured documents) would be among interesting tracks in the future. In addition to SDSs, focus of SLU has changed from human-machine interaction and database query to general information retrieval tasks such as voice search. At the same time SLUs are improving to cover more challenging tasks including multi-human and human-human spontaneously generated speech, interaction with live ASR, use of multimodal features in Dialogue Act (DA) detection, handling different modalities including gesture and geo-location and handling multi language for cross-lingual functionality.

Acknowledgments

This research was partly funded by the Ministry of Education, Youth and Sports of the Czech Republic under the grant agreement LK11221, core research funding of Charles University in Prague. This work has been using language resources distributed by the LINDAT/CLARIN project of the Ministry of Education, and Sports of the Czech Republic (project LM2010013).

Reference

Bird, S., Klein, E. and Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media

Black, A., and Eskenazi, M. (2009). *The Spoken Dialogue Challenge*. SIGDIAL. Association for Computational Linguistics (ACL).

Carpenter B. (1997). *Type-Logical Semantics*. The MIT Press.

Celikyilmaz, A., Hakkani-Tür, D., Tür, G., and Sarikaya, R. (2013). *Semi-Supervised Semantic Tagging for Conversational Understanding Using Markov Topic Regression.* Association for Computational Linguistics (ACL).

Celikyilmaz, A., Hakkani-Tür, D., Pasupat, P. and Sarikaya, R. (2015). *Enriching Word Embeddings Using Knowledge Graph for Semantic Tagging in Conversational Dialog Systems*. AAAI

Church, A. (1936). *An unsolvable problem of elementary number theory*, American Journal of Mathematics, 58

Church, A. (1941). *The Calculi of Lambda Conversion*. Princeton University Press.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa. P. (2011). *Natural Language Processing (Almost) from Scratch*. Machine Learning Research.

Deoras, A., Tür, G., Sarikaya, R., and Hakkani-Tür, D. (2013). *Joint Discriminative Decoding of Word and Semantic Tags for Spoken Language Understanding.* IEEE Transactions on Audio, Speech, and Language Processing.

Deoras, A. and Sarikaya, R. (2013). *Deep Belief Network Based Semantic Taggers for Spoken Language Understanding*. INTERSPEECH.

De Jong, F., Gauvain, J. L., Hiemstra, D., Netter, K. (2000). Language-based multimedia information retrieval. 6th RIAO Conference.

García, F., Hurtado, L-F., Segarra, E., Sanchis, E. and Riccardi, G. (2012). *Combining Multiple Translation Systems for Spoken Language Understanding Portability*. IEEE-SLT

Hafen, R. and Henry, M. (2012). *Speech information retrieval*: a review. Springer-Verlag

Hochreiter, S., and Schmidhuber, J. (1997). *Long Short-Term Memory*. Neural Computation.

Jurčíček, F., Dušek, O., Plátek, O., Žilka, L. (2014). *Alex: A Statistical Dialogue Systems Framework.* Text, Speech and Dialogue.

Krishnamurthy, J. and Kollar, T. (2013). *Jointly Learning to Parse and Perceive: Connecting Natural Language to the Physical World.* Transactions of the Association for Computational Linguistics

Kushman, N., Artzi, Y., Zettlemoyer, L. and Barzilay, R. (2014). *Learning to Automatically Solve Algebra Word Problems*. Association for Computational Linguistics (ACL).

Kwiatkowski, T., Zettlemoyer, L., Goldwater, S., and Steedman, M. (2011). *Lexical Generalization in CCG Grammar Induction for Semantic Parsing*. EMNLP.

Lafferty, J., McCallum, A., and Pereira, F. (2001). *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*., the 18th International Conference on Machine Learning (ICML).

Lee, S., and Eskenazi, M. (2013). Recipe For Building Robust Spoken Dialog State Trackers: Dialog State Tracking Challenge System Description, SIGDIAL.

Liang. P. (2013). *Lambda Dependency-based Compositional Semantics*. Technical report, ArXiv

Liang, P. and Potts, C. (2015). *Bringing Machine Learning and Compositional Semantics Together*, Annual Reviews of Linguistics.

Liu, X., and Sarikaya, R. (2014). A Discriminative Model Based Entity Dictionary Weighting Approach for Spoken Language Understanding. IEEE.

Li, Deng., Tur, G., Xiaodong, He., Hakkani-Tur, D. (2012). Use of kernel deep convex networks and end-to-end learning for spoken language understanding. Spoken Language Technology Workshop (SLT)

Lorenzo, A., Rojas-Barahona, L., and Cerisara, C. (2013). Unsupervised Structured Semantic Inference for Spo-ken Dialog Reservation Tasks. SIGDIAL.

Matuszek, C., Herbst, E., Zettlemoyer, L., Fox, D. (2012). *Learning to Parse Natural Language Commands to a Robot Control System.* International Symposium on Experimental Robotics

Mesnil, G., He, X., Deng, L. and Bengio, Y. (2013). *Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Language Understanding*. INTERSPEECH.

Mikolov, T., Chen, K., Corrado, G., and Dean. J. (2013). *Efficient Estimation of Word Representations in Vector Space*. ICLR.

Ratnaparkhi, A. (1998). *Maximum Entropy Models for Natural Language Ambiguity Resolution*, Ph.D. dissertation, University of Pennsylvania.

Sarikaya, R., Celikyilmaz, A., Deoras, A., and Jeong, M. (2014a). *Shrinkage Based Features for Slot Tagging with Conditional Random Fields*. ISCA.

Sarikaya, R., Hinton, G., and Deoras, A. (2014b). *Application of Deep Belief Networks for Natural Language Understanding*. IEEE.

Schwartz, R., Miller, S., Stallard, d., and Makhoul, J. (1996). Language understanding using hidden understanding Models. ICSLP'96.

Steedman, M. (1996). *Surface Structure and Interpretation*. MIT Press.

Steedman, M. (2000). *The Syntactic Process*, MIT Press.

Tür, G., Hakkani-Tür, D. and Chotimongkol, A. (2005). *Semi-supervised Learning for Spoken Language Understanding Using Semantic Role Labeling.* IEEE. ASRU.

Wang, L., Heck, L., and Hakkani-Tür, D. (2014). *Leveraging Semantic Web Search and Browse Sessions for Multi-Turn Spoken Dialog Systems*. IEEE. Acoustics, Speech, and Signal Processing.

Wang, Y. and Acero, A. (2006). *Discriminative Models for Spoken Language Understanding*. INTERSPEECH

Ward, W., and Issar, S. (1994). *Recent Improvements in the CMU Spoken Language Understanding System*. ARPA. HLT.

Wong, Y. and Mooney. R. (2007). *Generation by Inverting a Semantic Parser That Uses Statistical Machine Translation*. NAACL/HLT.

Xiaodong, He., Li, Deng., Hakkani-Tur, D., Tur, G. (2013). Multi-style adaptive training for robust cross-lingual spoken language understanding. Acoustics, Speech and Signal Processing (ICASSP)

Xu, P., and Sarikaya, R. (2013 a). *Joint Intent Detection and Slot Filling with Convolutional Neural Networks*., IEEE ASRU.

Xu, P., and Sarikaya, R. (2013 b). *Convolutional Neural Network Based Triangular CRF for Joint Intent Detection and Slot Filling*. IEEE. ASRU.

Yao, K., Peng, B., Zhang, Y., Yu, D., Zweig G., and Shi, Y. (2014). *Spoken Language Understanding Using Long Short-Term Memory Neural Networks*, IEEE.

Yao, K., Zweig, J., Hwang, M., Shi, Y., and Yu, D. (2013). *Recurrent Neural Networks for Language Understanding.* INTERSPEECH.

Young, S., Gasic, M., Thomson B., and Williams, J. (2013). *POMDP-based Statistical Spoken Dialogue Systems: a Review*. IEEE

Zelle, J. and Mooney, R. (1996). *Learning to Parse Database Queries Using Inductive Logic Programming*. Department of Mathematics and Computer Science Drake University.

Zettlemoyer, L. and Collins M. (2005). *Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorial Grammars*. Uncertainty in Artificial Intelligence.