

Q1 Decision Tree Learning

1. Code files submitted
 - a. To run the program
 - i. Place the data files 'horseTrain.txt' and 'horseTest.txt' in the same directory as the code
 - ii. Execute 'decision_tree.py'
2. Picture of the decision – generated by code and saved in the file 'output.png'
3. How many of the training instances does the tree classify correctly?
 - a. Training accuracy is 100 %
 - b. The predictions on training data can be seen in the file 'predictions_training_data.txt'
4. How many of the test instances does the tree classify correctly?
 - a. Test accuracy is 100 %
 - b. The predictions on test data can be seen in the file 'predictions_test_data.txt'
5. Description of how you used the information metric
For choosing the best attribute, I chose the attribute with the highest value of information gain metric.

As per the equation of Information gain,

$$IG(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A)$$

IG is highest when remainder is lowest. So to maximize IG(A), I minimize remainder(A)

The remainder(A) is defined as:

$$remainder(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

Since for this assignment we only use binary tests for each node, I have split the examples based on a threshold for this attribute. Depending on the value of the threshold chosen, the remainder(A) value will differ. I consider every unique value from min to max (of the possible values taken by this attribute) as a candidate for threshold. Then I find the threshold that gives the least value for remainder and use that as the value of remainder(A)

Q2 Paper report

Selected Paper: Percy Liang, Learning executable semantic parsers for natural language understanding, Communications of the ACM, 59 (9), pp. 68–76, 2016

a) What are the motivations of this work?

In order to build a system in AI that can understand Natural Languages, Semantic Parsing has been observed as an important paradigm. With the rise of NLP, there were solutions for document classification, POS (part of speech tagging), but the limitation with such models was that they lacked end-to-end understanding. The Semantic Parsers map natural language into logical form and the Statistical semantic parsers try to overcome this problem of deep understanding to an extent. Although they face a couple of issues. Representation of semantics of natural language and its construction compositionally can be viewed as a linguistic challenge. Making semantic parsers learn from weak supervision and generalizing properly to examples can be categorized as a statistical challenge. Efficient search over combinatorially large space of possible logical forms is a computational issue. Overall, the authors wanted to address the deep understanding of Natural Languages, and in specific they've covered this problem in the article: Given an utterance x in a context c , output the desired action y .

b) What is the proposed solution?

To understand natural language properly, the article has proposed an approach called statistical semantic parsing, where the main idea is to have an intermediate logical form z that connects x (utterance) and y (desired action), in a context c . The proposed semantic framework has 5 components: Executor, Grammar, Model, Parser, and Learner. For any semantic parser, representation is in language of mathematics. So, **Executor** is the standard interpretation or denotation(actions), $y = [z]_c$. **Grammar** is a set of rules that define a connection between an utterance to possible candidate derivations of logical forms. In the proposed framework, they start with an input utterance and repeatedly apply rule in G , until the designated ROOT category is produced over the entire utterance. A **Model** specifies scoring of candidate derivations produced by the Grammar. The article uses log-linear model which defines a feature vector of x (input utterance), c (context), and d (derivation from Grammar). It specifies distribution over derivations, telling how good the derivation is. **Parser** calculates the derivation(s) with highest probability for an utterance x under trained model. The article uses chart parser for the current framework, that recursively builds derivations for each span of the utterance. This would result in exponentially huge number of derivations. Hence the article suggests to couple it with beam search as a heuristic to compute the derivations with high probability under their model. **Learner** solves the inverse problem, as opposed to Parser. Learner tries to maximize the likelihood of the training data using stochastic gradient descent, clubbing it with beam search results in heuristic approximation of the gradient. The components are loosely coupled, which allows scope to improve on each component in isolation.

c) What is the evaluation of the proposed solution?

The article has not really evaluated their framework but has suggested that datasets are the main driver for statistical approaches. Hence, they've surveyed existing datasets. *Geo880* dataset has 880 questions and database queries about US Geography. Utterances - compositional but language is clean. Learning from logical forms and answers both achieve 90% accuracy. The *ATIS-3* dataset contains more disfluencies in utterances compared with *Geo880*, but they are logically simpler. The best reported result is based on semantic parsing and obtains 84.6% accuracy. The *Regexp824* has many logically equivalent regular expressions. The semantic unification for logical forms equivalence test obtains 65.6% accuracy. *Free917* and *WebQuestions* are both on Freebase. The difference is the former being compositional and latter being questions independent of Freebase, asked by people on the web. State-of-the-art accuracy for the former being 68% and for the latter being 52.5%. *WikiTableQuestions* extends questions beyond Freebase to HTML tables on Wikipedia, semi-structured. Since the tables are new at test-time, methods must learn how to generalize to new predicates. Result is 37.1%. The dataset using crowdsourcing to generate semantic pairing datasets achieved 58.8% accuracy.

d) What are the contributions?

The article is a survey-based paper that has suggested that statistical semantic parsing is one of the best possible solutions to solve the problem of deep-understanding of natural languages. They have proposed a framework of semantic parsing, but also contributed by suggesting some improvements on the individual components used in the framework namely, Executor, Grammar, Model, Parser and Learner. **Executor** is nothing but description of language of the logical form. Typically, the executor in a semantic parsing describes the language as first-order logic, which will need all the variables involved in a sentence mapping to their truth values. The logical language suggested in the article which would be more powerful than first-order logic is lambda calculus based language, called lambda dependency-based semantics(DCS). Lambda DCS combines functions from objects to truth values(sets) rather than truth values, also partially eliminating the need for variables. Combinatory Category Grammar(CCG) instead of the conventional **Grammar** suggested in the framework. It is coupled with logical forms in lambda calculus. It describes Nouns(N) and Noun Phrases(NP) as objects in a sentence, and directionality of slashes ($(N \backslash N) / NP$) indicate the location of the arguments. Another approach suggested is cruder grammar rules employed in lambda DCS, as compared to CCG, to avoid complications. And a third Grammar suggested is floating grammar, the article suggests that one can mix-and-match them to obtain best results depending upon the domain. **The statistical model** calculates a function score (x, c, d) that judges any derivation d , in the current framework it uses log-linear model. For improvisation of model, the article suggests more expressive features (for example defining a few features from large corpora like 'birthplace' and 'born in'). Another way to produce features can be done by including type constraints on ill-formed logical forms can be pruned. The article also suggests that adopting a non-linear scoring function, not requiring any domain-knowledge about semantic parsing. An example would be a simple one-layer neural network, which takes weighted combination of m non-linear basis functions. For **Parsers**, instead of chart parsing, the solution suggested was shift-reduce parsing. Parses from left-to-right, and new sub-derivations can depend arbitrarily on the sub-derivations constructed thus far, to support incremental contextual interpretations. Agenda-based parsing to save wastage of resources on

non-promising spans. **Learner** in the current framework does not do everything from scratch, and requires hard-coded grammar rules as training examples. A more generic induction algorithm based on higher-order unification doesn't require any initial Grammar. Regularization term can generalize better by promoting feature weights to become zero.

e) What are the future direction for this research?

The article suggests that in future, there is still a lot of scope for research to address the problems of "Alternative Semantic Representation" and "Alternative Supervision".

The problem of divergence between structure of natural language and logical form can be addressed by using general para-phrasing models to map input utterances to canonical utterances of logical forms. Also use of domain-general logical forms such as Abstract meaning representation (AMR) parsers which offers broad coverage, or use of *executable* semantic parsing that operates on the full pipeline from utterance to task output, requires additional work for solving downstream understanding tasks and retaining domain-generalizability respectively. Designing the best general representation that supports many downstream tasks remains an open challenge.

Providing alternative sources of supervision is also an improvement that can be undertaken by future research. For instance, employing a large corpus of text to exploit even weaker supervision or more generally, extending this to reinforcement learning setting where an agent who presented with an utterance in some context performs some action, and receives a corresponding reward signal. Also, more work can be done on recurrent neural networks(RNN) and their extensions for semantic parsing. These methods require more data and further investigation needs to be done to evaluate whether they can learn to perform complex logical reasoning in a generalizable way.

Also, the authors argue that due to the recent surge in interest in natural language interfaces (e.g. Siri, Google Home etc.) there is a growing demand for deeper understanding of semantic parsing and natural language understanding. Further research can be done for generating new insights into the nature of language and learning.