GCN-Sem at SemEval-2019 Task 1: Semantic Parsing using Graph Convolutional and Recurrent Neural Networks

Shiva Taslimipoor   **Omid Rohanian**   **Sara Može**
Research Group in Computational Linguistics
University of Wolverhampton, UK
{shiva.taslimi, omid.rohanian, s.moze}@wlv.ac.uk

Abstract

This paper describes the system submitted to the SemEval 2019 shared task 1 'Cross-lingual Semantic Parsing with UCCA'. We rely on the semantic dependency parse trees provided in the shared task which are converted from the original UCCA files and model the task as tagging. The aim is to predict the graph structure of the output along with the types of relations among the nodes. Our proposed neural architecture is composed of Graph Convolution and BiLSTM components. The layers of the system share their weights while predicting dependency links and semantic labels. The system is applied to the CONLLU format of the input data and is best suited for semantic dependency parsing.

1 Introduction

Universal Conceptual Cognitive Annotation (UCCA) (Abend and Rappoport, 2013) is a semantically motivated approach to grammatical representation inspired by typological theories of grammar (Dixon, 2012) and Cognitive Linguistics literature (Croft and Cruse, 2004). In parsing, bi-lexical dependencies that are based on binary head-argument relations between lexical units are commonly employed in the representation of syntax (Nivre et al., 2007; Chen and Manning, 2014) and semantics (Hajič et al., 2012; Oepen et al., 2014; Dozat and Manning, 2018).

UCCA differs significantly from traditional dependency approaches in that it attempts to abstract away traditional syntactic structures and relations in favour of employing purely semantic distinctions to analyse sentence structure. The shared task, 'cross-lingual semantic parsing with UCCA' (Hershcovich et al., 2019) consists in parsing English, German, and French datasets using the UCCA semantic tagset. In order to enable multi-task learning, the UCCA-annotated data is automatically converted to other parsing formats, e.g. Abstract Meaning Representation (AMR) and Semantic Dependency Parsing (SDP), inter alia (Hershcovich et al., 2018).

Although the schemes are formally different, they have shared semantic content. In order to perform our experiments, we target the converted CONLLU format, which corresponds to traditional bi-lexical dependencies and rely on the conversion methodology which is provided in the shared task (Hershcovich et al., 2019) to attain UCCA graphs.

UCCA graphs contain both explicit and implicit units [1] However, in bi-lexical dependencies, nodes are text tokens and semantic relations are direct bi-lexical relations between the tokens. The conversion between the two format results in partial loss of information. Nonetheless, we believe that it is worth trying to model the task using one of the available formats (i.e. semantic dependency parsing) which is very popular among NLP researchers.

Typically, transition-based methods are used in syntactic (Chen and Manning, 2014) and semantic (Hershcovich et al., 2017) dependency parsing. By contrast, our proposed system shares several similarities with sequence-to-sequence neural architectures, as it does not specifically deal with parsing transitions. Our model uses word, POS and syntactic dependency tree representations as input and directly produces an edge-labeled graph representation for each sentence (i.e. edges and their labels as two separate outputs). This multi-label neural architecture, which consists of a BiLSTM and a Graph Convolutional Network (GCN), is described in Section 3.

---

[1] Explicit units (terminal nodes) correspond to tokens in the text, but implicit (semantic) units have no corresponding component in the text.

2 Related Work

A recent trend in parsing research is sequence-to-sequence learning (Vinyals et al., 2015b; Kitaev and Klein, 2018), which is inspired from Neural Machine Translation. These methods ignore explicit structural information in favour of relying on long-term memory, attention mechanism (content-based or position-based) (Kitaev and Klein, 2018) or pointer networks (Vinyals et al., 2015a). By doing so, high-order features are implicitly captured, which results in competitive parsing performance (Jia and Liang, 2016).

Sequence-to-sequence learning has been particularly effective in Semantic Role Labeling (SRL) (Zhou and Xu, 2015). By augmenting these models with syntactic information, researchers have been able to develop state-of-the-art systems for SRL (Marcheggiani and Titov, 2017; Strubell et al., 2018).

As information derived from dependency parse trees can significantly contribute towards understanding the semantics of a sentence, Graph Convolutional Network (GCN) (Kipf and Welling, 2017) is used to help our system perform semantic parsing while attending to structural syntactic information. The architecture is similar to the GCN component employed in Rohanian et al. (2019) for detecting gappy multiword expressions.

3 Methodology

For this task, we employ a neural architecture utilising structural features to predict semantic parsing tags for each sentence. The system maps a sentence from the source language to a probability distribution over the tags for all the words in the sentence. Our architecture consists of a GCN layer (Kipf and Welling, 2017), a bidirectional LSTM, and a final dense layer on top.

The inputs to our system are sequences of words, alongside their corresponding POS and named-entity tags.[2] Word tokens are represented by contextualised ELMo embeddings (Peters et al., 2018), and POS and named-entity tags are one-hot encoded. We also use sentence-level syntactic dependency parse information as input to the system. In the GCN layer, the convolution filters operate based on the structure of the dependency tree (rather than the sequential order of words).

Graph Convolution. Convolutional Neural Networks (CNNs), as originally conceived, are sequential in nature, acting as detectors of N-grams (Kim, 2014), and are often used as feature-generating front-ends in deep neural networks. Graph Convolutional Network (GCN) has been introduced as a way to integrate rich structural relations such as syntactic graphs into the convolution process.

In the context of a syntax tree, a GCN can be understood as a non-linear activation function f and a filter W with a bias term b:

$$c = f(\sum_{i \in r(v)} W x_i + b) \qquad (1)$$

where $r(v)$ denotes all the words in relation with a given word v in a sentence, and c represents the output of the convolution. Using adjacency matrices, we define graph relations as mask filters for the inputs (Kipf and Welling, 2017; Schlichtkrull et al., 2017).

In the present task, information from each graph corresponds to a sentence-level dependency parse tree. Given the filter W_s and bias b_s, we can therefore define the sentence-level GCN as follows:

$$C_s = f(W_s X^T A + b_s) \qquad (2)$$

where $X_{n \times v}$, $A_{n \times n}$, and $C_{o \times n}$ are tensor representation of words, the adjacency matrix, and the convolution output respectively.[3] In Kipf and Welling (2017), a separate adjacency matrix is constructed for each relation to avoid over-parametrising the model; by contrast, our model is limited to the following three types of relations: 1) the head to the dependents, 2) the dependents to the head, and 3) each word to itself (self-loops) similar to Marcheggiani and Titov (2017). The final output is the maximum of the weights from the three individual adjacency matrices.

The model architecture is depicted in Figure 1.

4 Experiments

Our system participated in the closed track for English and German and the open track for French. We exclusively used the data provided in the shared task. The system is trained on the training data only, and the parameters are optimised using the development set. The results are reported

---

[2]spaCy (Honnibal and Johnson, 2015) is used to generate POS, named-entity and syntactic dependency tags.

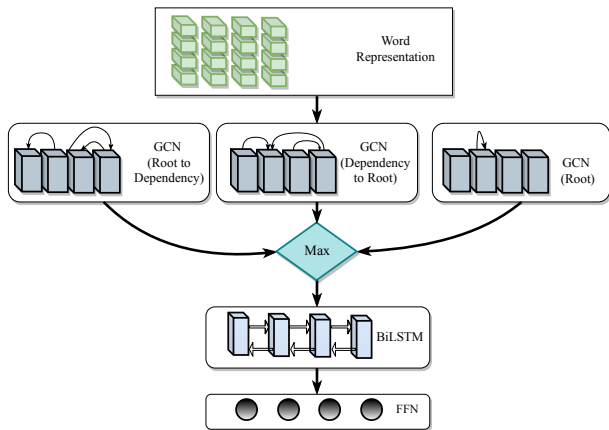[3]o: output dimension; v: word vectors dimension; n: sentence length

Figure 1: A GCN-based recurrent architecture.

on blind-test data in both in-domain and out-of-domain settings. We focus on predicting the primary edges of UCCA semantic relations and their labels.

4.1 Data

The datasets of the shared task are devised for four settings: 1) English in-domain, using the Wiki corpus; 2) English out-of-domain, using the Wiki corpus as training and development data, and 20K Leagues as test data; 3) German in-domain, using the 20K Leagues corpus; 4) French setting with no training data (except trial data), using the 20K Leagues corpus as development and test data.

Whilst the annotated files used by the shared task organisers are in the XML format, several other formats are also available. We decided to use CONLLU, as it is more interpretable. However, according to the shared task description,[4] the conversion between XML and CONLLU, which is a necessary step before evaluation, is lossy. Hershcovich et al. (2017) used the same procedure of performing dependency parsing methods on CONLLU files and converting the predictions back to UCCA.

4.2 Settings

We trained ELMo on each of the shared task datasets using the system implemented by Che et al. (2018). The embedding dimension is set to 1024. The number of nodes is 256 for GCN and 300 for BiLSTM, and we applied a dropout of 0.5 after each layer. We used the Adam optimiser for compiling the model.

We tested our model in four different settings,

as explained in Section 4.1. The parameters are optimised on the English Wiki development data (batch-size = 16 and number of epochs = 100) and used for all four settings. As no training data was available for French, the trained system on English Wiki was used to parse French sentences of 20K Leagues. For this reason the French model is evaluated within the open track.

4.3 Official Evaluation

Our model predicts two outputs for each dataset: primary edges and their labels (UCCA semantic categories). [5]

Table 1 shows the performance (in terms of precision, recall, and F1-score) for predicting primary edges in both labeled (i.e. with semantic tags) and unlabeled settings (i.e. ignoring semantic tags). Table 2 shows F1-scores for each semantic category separately. Although the overall performance of the system, as shown in the official evaluation in Table 1, is not particularly impressive, there are a few results worth reporting. These are listed in Table 2.

Our system is ranked second in predicting four relations, i.e. L (linker), N (Connector), R (Relator), and G (Ground), in all settings displayed in bold. A plausible explanation would be that these relations are somewhat less affected by the loss of information incurred as a result of the conversions between formats.

5 Discussion

Our neural model is applied to UCCA corpora, which are converted to bi-lexical semantic dependency graphs and represented in the CONLLU format. The conversion from UCCA annotations to CONLLU tags appears to have a distinctly negative impact on the system's overall performance. As reported in the shared task description, converting the English Wiki corpus to the CONLLU format and back to the standard format results in an F1-score of only 89.7 for primary labeled edges. This means that our system cannot go beyond this upper limit.

Since our system is trained on CONLLU files and the evaluation involves converting the CONLLU format back to the standard UCCA format,

---

dataset	track	labeled				unlabeled			
		Avg. F1	P	R	F1	Avg. F1	P	R	F1
UCCA_English-Wiki	closed	0.657	0.673	0.655	0.664	0.809	0.829	0.807	0.818
UCCA_English-20K	closed	0.626	0.632	0.642	0.637	0.8	0.808	0.821	0.814
UCCA_German-20K	closed	0.71	0.72	0.72	0.72	0.851	0.863	0.862	0.862
UCCA_French-20K*	open	0.438	0.443	0.447	0.445	0.690	0.698	0.705	0.702

Table 1: Official results of the shared task evaluation for predicting different semantic category labels. (* The results for French are for Post-Evaluation.)

dataset	(D)	(C)	(N)	(E)	(F)	(G)	(L)	(H)	(A)	(P)	(U)	(R)	(S)	(Terminal)
English-Wiki	0.7	0.708	**0.866**	0.738	0.801	**0.286**	**0.836**	0.289	0.582	0.451	0.948	**0.914**	0	0.997
English-20K	0.521	0.733	**0.776**	0.743	0.647	0.04	**0.719**	0.248	0.538	0.527	0.978	**0.844**	0	0.997
German-20K	0.691	0.813	**0.796**	0.82	0.845	**0.778**	0.834	0.375	0.697	0.561	0.997	**0.916**	0	0.998
French-20K*	0.223	0.569	0.579	0.551	0.378	0.000	0.536	0.118	0.314	0.358	0.987	0.711	0	0.993

Table 2: Official results of the shared task evaluation for predicting Primary edges and their labels. (* The results for French are for Post-Evaluation.)

the reported results for our system can be misleading. In order to further investigate this issue, we performed an evaluation using the English Wiki development data, comparing the predicted labels with the gold standard in development set in the CONLLU format. The average F1-score for labelled edges was 0.71 compared to the 0.685 score our system achieved on the development set using the official evaluation script.

This clearly demonstrates that our system fares significantly better if it receives its input in the form of bi-lexical dependency graphs. Therefore, the system is best suited for semantic dependency parsing, although we believe that promising results could also be achieved in UCCA annotation if the conversion between the CONLLU and UCCA formats is improved to map and preserve information more accurately.

6 Conclusion and Future Work

In this paper, we described the system we submitted to the SemEval-2019 Task 1: 'Semantic Parsing using Graph Convolutional and Recurrent Neural Networks'. The model performs semantic parsing using information derived from syntactic dependencies between words in each sentence. We developed the model using a combination of GCN and BiLSTM components. Due to the penalisation resulting from the use of lossy CONLLU files, we argue that the results cannot be directly compared with those of the other task participants. [6]

In the future, we would like to build on the work

presented in this paper by applying the architecture to the standard UCCA dataset, or possibly training the system to perform bi-lexical semantic dependency annotation.

References

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *Proc. of ACL*, pages 228–238.

Wanxiang Che, Yijia Liu, Yuxuan Wang, Bo Zheng, and Ting Liu. 2018. Towards better UD parsing: Deep contextualized word embeddings, ensemble, and treebank concatenation. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 55–64, Brussels, Belgium. Association for Computational Linguistics.

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750. Association for Computational Linguistics.

William Croft and D. A Cruse. 2004. *Cognitive linguistics*. Cambridge University Press.

Robert M. W Dixon. 2012. *Basic linguistic theory*. Oxford University Press.

Timothy Dozat and Christopher D. Manning. 2018. Simpler but more accurate semantic dependency parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 484–490. Association for Computational Linguistics.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková,

[6] The code is available at https://github.com/shivaat/GCN-Sem.

Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing prague czech-english dependency treebank 2.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3153–3160. ELRA, European Language Resources Association.

Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2017. A transition-based directed acyclic graph parser for ucca. In *Proc. of ACL*, pages 1127–1138.

Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2018. Multitask parsing across semantic representations. In *Proc. of ACL*, pages 373–385.

Daniel Hershcovich, Zohar Aizenbud, Leshem Choshen, Elior Sulem, Ari Rappoport, and Omri Abend. 2019. Semeval 2019 task 1: Cross-lingual semantic parsing with ucca.

Matthew Honnibal and Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.

Robin Jia and Percy Liang. 2016. Data recombination for neural semantic parsing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12–22. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*.

Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia. Association for Computational Linguistics.

Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515. Association for Computational Linguistics.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Glsen Eryigit, Sandra Kbler, Svetoslav Marinov, and Erwin Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(02).

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Yi Zhang. 2014. Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 63–72.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of NAACL*.

Omid Rohanian, Shiva Taslimipoor, Samaneh Kouchaki, Le An Ha, and Ruslan Mitkov. 2019. Bridging the gap: Attending to discontinuity in identification of multiword expressions.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling relational data with graph convolutional networks. *arXiv preprint arXiv:1703.06103*.

Emma Strubell, Patrick Verga, Daniel Andor, David Weiss, and Andrew McCallum. 2018. Linguistically-informed self-attention for semantic role labeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5027–5038. Association for Computational Linguistics.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015a. Pointer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2692–2700. Curran Associates, Inc.

Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. 2015b. Grammar as a foreign language. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2773–2781. Curran Associates, Inc.

Jie Zhou and Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1127–1137. Association for Computational Linguistics.