

Neural Class Expression Synthesis

N'Dah Jean Kouagou^{1,*}, Stefan Heindorf¹, Caglar Demir¹ and
Axel-Cyrille Ngonga Ngomo¹

¹Department of Computer Science, Paderborn University, Warburger Str. 100, Paderborn, 33098, Germany

Abstract

Many applications require explainable node classification in knowledge graphs. Towards this end, a popular “white-box” approach is class expression learning: Given sets of positive and negative nodes, class expressions in description logics are learned that separate positive from negative nodes. Most existing approaches are search-based approaches generating many candidate class expressions and selecting the best one. However, they often take a long time to find suitable class expressions. In this paper, we cast class expression learning as a translation problem and propose a new family of class expression learning approaches which we dub neural class expression synthesizers. Training examples are “translated” into class expressions in a fashion akin to machine translation. Consequently, our synthesizers are not subject to the runtime limitations of search-based approaches. We study three instances of this novel family of approaches based on LSTMs, GRUs, and set transformers (STs), respectively. An evaluation of our approach on four benchmark datasets suggests that it can effectively synthesize high-quality class expressions with respect to the input examples in approximately one second on average. Moreover, a comparison to state-of-the-art approaches suggests that we achieve better F-measures on large datasets. For reproducibility purposes, we provide our implementation as well as pretrained models in our public GitHub repository at <https://github.com/dice-group/NeuralClassExpressionSynthesis>

Keywords

Neural network, Concept learning, Class expression learning, Learning from examples, NCES

Class expression learning (CEL) approaches learn a class expression that describes individuals provided as positive examples. They are applied in a wide range of domains, including ontology engineering, bio-medicine, and Industry 4.0. Several methods have been proposed to address CEL. The state of the art consists of approaches based on refinement operators [1, 2], and evolutionary algorithms [3]. However, the majority of these approaches suffer from scalability issues because they explore an infinite conceptual space for each learning problem. We propose a new family of self-supervised neuro-symbolic approaches dubbed *neural class expression synthesis* (NCES) approaches for CEL. NCES [4] instances view CEL as a machine translation problem by translating from the language of example embeddings to that of description logics (or any other logic for that matter).

Overall, neural class expression synthesizers work as follows: First, a given knowledge base is converted into a set of triples (s, p, o) and then embedded into a continuous vector space such

NeSy 2023, 17th International Workshop on Neural-Symbolic Learning and Reasoning, Certosa di Pontignano, Siena, Italy

*Corresponding author.

✉ ndah.jean.kouagou@upb.de (N. J. Kouagou); heindorf@upb.de (S. Heindorf); caglar.demir@upb.de (C. Demir); axel.ngonga@upb.de (A. N. Ngomo)

🆔 0000-0002-4217-897X (N. J. Kouagou); 0000-0002-4525-6865 (S. Heindorf); 0000-0001-8970-3850 (C. Demir); 0000-0001-8970-3850 (A. N. Ngomo)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Table 1

Evaluation results per approach and dataset. The star (*) indicates the Wilcoxon statistical significance between NCES and the best search-based approach. NCES uses the ensemble of GRU, LSTM, and ST.

	F ₁ (%) ↑				Runtime (sec.) ↓			
	Carcinogen.	Mutagenesis	Sem. Bible	Vicodi	Carcinogen.	Mutagenesis	Sem. Bible	Vicodi
CELOE	37.92±44.25	82.95±33.48	93.18±17.52*	35.66±42.06	239.58±132.59	92.46±125.69	135.30±139.95	289.95±103.63
EvoLearner	91.48±14.30	93.27±12.95	91.88±10.14	92.74±10.28	54.73±25.86	48.00±31.38	17.16±9.20	213.78±81.03
NCES	97.06±13.06*	<u>91.39±22.91</u>	87.11±24.05	95.51±12.14*	0.27*±0.00	0.31*±0.00	0.15*±0.00	0.15*±0.00

as \mathbb{R}^d . Next, learning problems are generated automatically from the input knowledge base using a refinement operator and an instance checker. Finally, the synthesizers are trained to translate the embeddings of positive/negative examples to the corresponding class expressions in the training data. If necessary, the second and third step can be iterated until a stopping criterion (e.g., convergence) is fulfilled.

We compared NCES to state-of-the-art algorithms for CEL, including CELOE [1] and EvoLearner [3], on the popular benchmarks Carcinogenesis, Mutagenesis, Semantic Bible, and Vicodi. We measure the performance of each approach in terms of runtime and number of positive/negative examples covered/ruled out by the computed solution. Table 1 gives the results of our experiments on a total of 380 learning problems. These results suggest that post training, NCES instances are over 300 times faster on average than search-based approaches. Moreover, they perform particularly well on the largest datasets Carcinogenesis and Vicodi with up to 5.5% absolute improvement in F-measure.

Acknowledgments This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No 860801 and the European Union’s Horizon Europe research and innovation programme under the grant No 101070305. This work has also been supported by the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW) within the project SAIL under the grant No NW21-059D and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 318/1 2021 – 438445824.

References

- [1] J. Lehmann, S. Auer, L. Bühmann, S. Tramp, Class expression learning for ontology engineering, *J. Web Semant.* 9 (2011) 71–81.
- [2] N. J. Kouagou, S. Heindorf, C. Demir, A. N. Ngomo, Learning concept lengths accelerates concept learning in ALC, in: *ESWC*, volume 13261 of *LNCS*, Springer, 2022, pp. 236–252.
- [3] S. Heindorf, L. Blübaum, N. Düsterhus, T. Werner, V. N. Golani, C. Demir, A. Ngonga Ngomo, Evolearner: Learning description logics with evolutionary algorithms, in: *WWW*, ACM, 2022, pp. 818–828.
- [4] N. J. Kouagou, S. Heindorf, C. Demir, A.-C. Ngonga Ngomo, Neural class expression synthesis, in: *ESWC*, volume 13870 of *LNCS*, Springer, 2023, pp. 209–226.