autoBOT: Evolving Neuro-symbolic Representations for Explainable Low Resource Text Classification

Blaž Škrlj [1,2] and Matej Martinc [1,2] and
Nada Lavrač [1] and Senja Pollak [1]

[1]Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

[2]Jožef Stefan Institute, Ljubljana, Slovenia

University/organization {blaz.skrlj, jan.kralj, matej.martinc,nada.lavrac,senja.pollak}@ijs.si

Abstract

Learning from texts has been widely adopted throughout industry and science. While state-of-the-art neural language models have shown very promising results for text classification, they are expensive to (pre-)train, require large amounts of data and tuning of hundreds of millions or more parameters. This paper explores how automatically evolved text representations can serve as a basis for explainable, low-resource branch of models with competitive performance that are subject to automated hyperparameter tuning. We present autoBOT (automatic Bags-Of-Tokens), an autoML approach suitable for low resource learning scenarios, where both the hardware and the amount of data required for training are limited. The proposed approach consists of an evolutionary algorithm that jointly optimizes various sparse representations of a given text (including word, subword, POS tag, keyword-based, knowledge graph-based and relational features) and two types of document embeddings (non-sparse representations). The key idea of autoBOT is that, instead of evolving at the learner level, evolution is conducted at the representation level. The proposed method offers competitive classification performance on fourteen real-world classification tasks when compared against a competitive autoML approach that evolves ensemble models, as well as state-of-the-art neural language models such as BERT and RoBERTa. Moreover, the approach is explainable, as the importance of the parts of the input space is part of the final solution yielded by the proposed optimization procedure, offering potential for meta-transfer learning.

1 Main Objectives and Problem Formulation

Rationale for this study

Our goal was to develop an autoML engine capable of performing automated neuro-symbolic text classification. We emphasized the neuro-symbolic aspect as it is strongly related to the overall method's interpretability. By exploiting the best of both worlds, we hypothesized one can obtain better, more interpretable models, albeit not necessarily at the performance level of contemporary neural language models. These models, however, contain orders of magnitude more parameters than the autoBOT's final outputs.

The created autoBOT is a simple-to-use library, accessible in a few lines of code[1]. It also supports user-defined text transformations. Ultimately, we aimed for the user to specify only one parameter: evolution time.

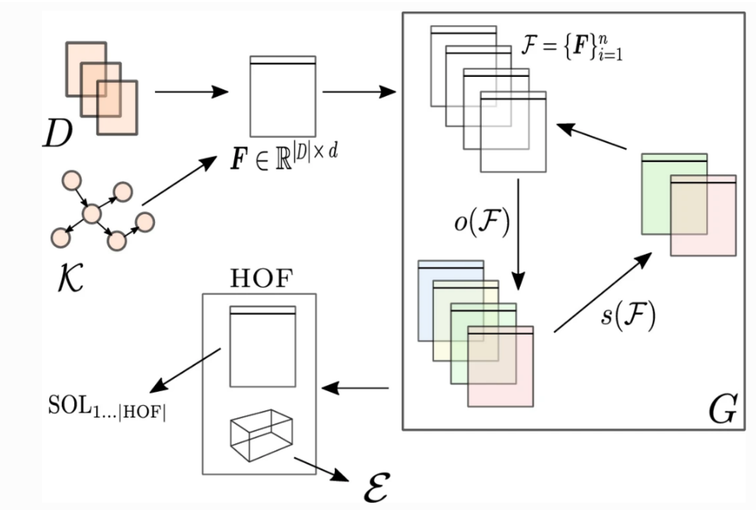
```
import autoBOTLib
import pandas as pd

dataframe = pd.read_csv("train.tsv", sep="\t")
train_sequences = dataframe['texts']
train_targets = dataframe['labels']
autoBOTLibObj = autoBOTLib.GAlearner(train_sequences, train_targets).evolve()
dataframe2 = pd.read_csv("test.tsv", sep="\t")
test_sequences = dataframe2['texts']
predictions = autoBOTLibObj.predict(test_sequences)
```

2 Methodology

Key ideas

The developed autoBOT explores the following key ideas. Sparse representations at the word, sub-word and character level; dense contextual and non-contextual sub-symbolic representations; Evolution based representation selection; Comparison of final weight spaces with implications to transfer learning Inclusion of triplet-based background knowledge as one of the feature spaces.



Key elements

Schematic overview of autoBOT. The input is a collection of documents D alongside a knowledge graph. The feature space is constructed based on the information from both sources. Next, G generations of representation evolution are conducted. Here, the o represents the application of different operators to solution vectors representing weights of feature subspaces (e.g., word, character etc.), followed by selection, where the next generation of solutions is chosen. Once the optimization finishes, the best solutions (HOF - Hall Of Fame) are used for the final set of predictions. The SOL space denotes the individual solutions, used for construction of final classifiers, and $\mathcal{E}$ represents the set of explanations – feature-value associations. As the solutions encode both the weights at the feature subspace level, as well as weights of individual features, autoBOT offers two distinct views of feature importances.

3 Experimental setting

The experimental setting includes a large collection of text classification benchmarks. The baselines included (strong) neural language models and a competitive autoML (tpot), but also weaker baselines (e.g., TF-IDF-only learners). An overview of the considered data sets is shown below. Data sets from the domains such as biomedical, news, spam/bot detection and others were considered.

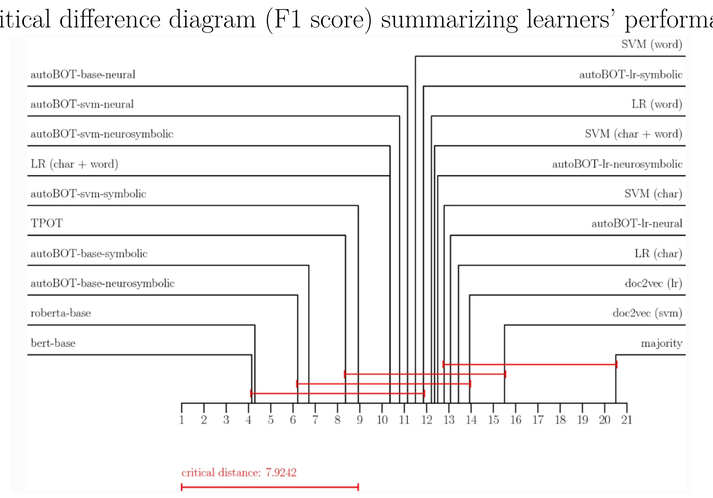
Data set	Documents	Unique tokens	Unique labels	Task	Source
kenyan	462	46189	2	News source prediction	Pollak et al. (2011)
semeval-2017-sentiment	1156	5144	4	Sentiment prediction	Nakov et al. (2013) [a]
bbc	2225	73491	4	News category prediction	Greene and Cunningham (2006)
subjects	1786	132996	4	Topic prediction	[b]
fox-news	2107	220063	7	News topic prediction	Qian and Zhai (2014)
insults	3946	36021	2	insult prediction	[c]
questions	5452	13279	6	Question types	Li and Roth (2002)
mbti	8675	572269	16	Personality type prediction	Myers (1962)[d]
yelp	10000	125446	5	Review prediction	[e]
hatespeech	10868	30555	4	Hate speech prediction	[f]
semeval2019	13240	53693	2	Offensive language prediction	Zampieri et al. (2019) [g]
sentimix	17000	89694	3	Sentiment prediction	[h]
articles	19990	285167	20	Objectivity prediction	Hajj et al. (2019)
sarcasm	28619	58779	2	Sarcasm prediction	Misra and Arora (2019)

4 Results

Performance

The results indicate the competitive performance of autoBOT across the wide spectrum of data sets. Even though it does not yet quite reach the performance of larger neural language models, it remains near the optimum when considering also the model size (parameter count), which is orders of magnitude smaller for the autoBOT's final models (linear, regularized learners).

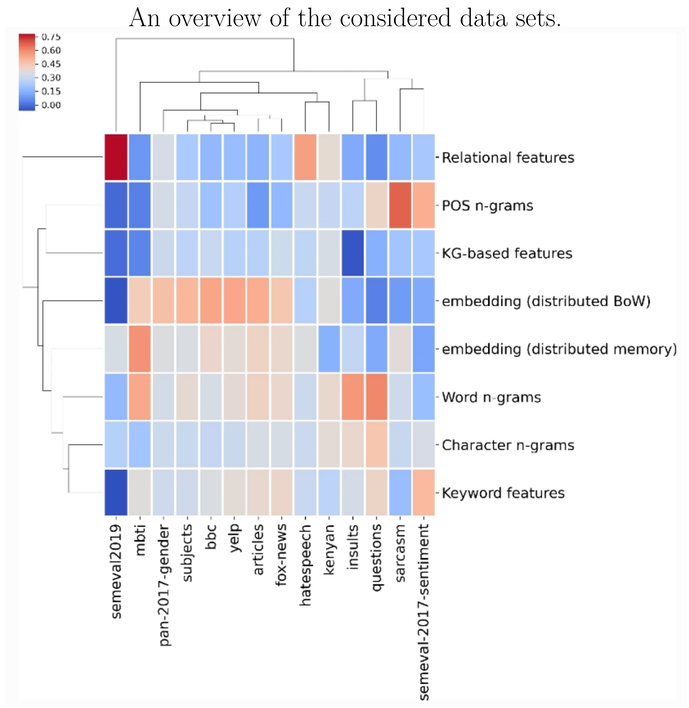
The critical difference diagram (F1 score) summarizing learners' performances.



Meta transfer

An interesting side-product of evolution are the feature type-specific weight vectors, which we hypothesized, could be compared to identify similar tasks. This indeed seems to be the case, which has implications for meta-transfer learning.

An overview of the considered data sets.



References

[1] Blaž Škrlj, Matej Martinc, Nada Lavrač, and Senja Pollak. autobot: evolving neuro-symbolic representations for explainable low resource text classification. Machine Learning, Apr 2021.

Acknowledgements

The work of the first author was funded by the Slovenian Research Agency through a young researcher grant. The work of other authors was supported by the Slovenian Research Agency (ARRS) core research programme Knowledge Technologies (P2-0103), an ARRS funded research project Semantic Data Mining for Linked Open Data (financed under the ERC Complementary Scheme, N2-0078) and European Union's Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). We also gratefully acknowledge the support of NVIDIA Corporation for the donation of Titan-XP GPU. This research was also partially supported by TAILOR (a project funded by the EU Horizon 2020 research and innovation programme under GA No 952215) and AI4EU (GA No 825619). We would also like to thank the reviewers for their valuable comments.

[1]https://github.com/SkBlaz/autobot