Abstract Rule Learning with Neural Networks

Radha Kopparti[1] and Tillman Weyde[1]

Research Centre for Machine Learning, Department of Computer Science, City, University of London, UK.
{radha.kopparti, t.e.weyde}@city.ac.uk

Abstract. Despite the successes achieved with deep neural networks over recent years, there has been an increasing awareness recently that there are tasks that still elude neural network learning, specifically the generalisation from patterns to rules. We tackle the problem of learning abstract rules by introducing Relation Based Patterns (RBP) which model equality relationships. RBP creates an inductive bias in the neural networks that leads to the learning of generalisable solutions. We observe that integration of RBP leads to almost perfect generalisation in abstract rule learning tasks with synthetic data and to improvements in neural language modelling on real-world data.

Keywords: Rule Learning · Neural Networks · Systematicity.

1 Introduction

Humans are very effective at extracting abstract relations from sensory input, often after very brief exposure. In a well-known study, [3] showed that 7-month old infants learn to recognize patterns defined by simple grammatical rules. These rules define abstract patterns based on equality relations, creating structures such as ABA or ABB. The infants learn the patterns from a small number of examples after just two minutes of familiarization, but when the same experiment was applied to recurrent neural networks, they failed to recognise the abstract patterns. More recently, a lack of linguistic systematicity has been in deep learning for machine translation [2] and shortcomings in visual abstraction have been identified [4].

We tackle the problem of learning abstract rules in neural networks by introducing Relation Based Patterns (RBP) to model equality relationships. The RBP model has been designed as a set of additional neurons with a rectified difference activation and fixed-weight connections that connect them to standard networks. The RBP structure creates an inductive bias in the neural networks that favours the learning of generalisable solutions. We observe that integration of RBP leads to almost perfect generalisation in abstract rule learning tasks with synthetic data and to improvements in prediction tasks on text and music [7, 6].

Type	Standard FFNN	Early Fusion	Mid Fusion
1. ABA vs other	50 (1.86)	65 (1.26)	100 (0.00)
2. ABB vs other	50 (1.83)	65 (1.29)	100 (0.00)
3. ABA-BAB vs other	50 (1.73)	75 (1.22)	100 (0.05)
4. ABA vs ABB	50 (1.81)	55 (1.18)	100 (0.00)
5. ABC vs other	50 (1.68)	65 (1.04)	100 (0.00)

Table 1: Accuracy (in %) and standard deviation over 10 simulations (in brackets) using different models for Abstract Pattern Learning (ABA vs other, ABB vs other, ABA-BAB vs other, ABA vs ABB, ABC vs other).

2 Relation Based Patterns (RBP)

The RBP model is based on an input that consists of multiple items, where each can be represented by a vector of input neurons. We use the DR units to compare corresponding neurons in different vectors. For the comparison we introduce differentiator-rectifier (DR) units, which calculate the absolute difference of two inputs: $f(x, y) = |x - y|$. We create one DR unit for every pair of corresponding input unit with the weights from the inputs to the DR units fixed at 1. We have multiple vector comparisons that correspond to the different equality relations in patters, e.g. equality of the pairs of vectors in positions (1,2), (1,3), (2,3), to recognise patterns of the forms ABA, ABB, AAB, ABC etc.

We integrate DR units in *Early Fusion* (added to the input layer) and *Mid Fusion* (added to the hidden layer).

3 Experiments and Results

3.1 Learning Abstract Relations

In this task, triples of the forms ABA, ABB, ABC, AAB and BAB are given to the network as a supervised formulation of [3] with some variants. We use a 75/25 train/test split with separate vocabulary between them. The results of the experiments are given in Table 1. We can observe that without RBP networks never improve above chance level (50%), and Mid Fusion leads to almost perfect results.

3.2 Character and Melody Prediction

For character prediction we use recurrent neural networks and their gated variants (LSTM and GRU) on a subset of the Gutenberg electronic book collection[1], consisting of 42252 words. We use 2 hidden layers with 50 neurons each, an initial learning rate of 0.01 and the network training converged after 30 epochs. The results using context size 5 are summarised in Table 2 show a a consistent improvement with RBP.

[1] https://www.gutenberg.org/

Type	RNN	GRU	LSTM
Simple Network	3.8281	3.8251	3.8211
Early Fusion	3.8254	3.8163	3.8162
Mid Fusion	**3.8112**	3.8134	3.8148

Table 2: Average Cross Entropy Loss per predicted character using context length 5.

Type	RNN	GRU	LSTM
Simple Network	2.6994	2.5702	2.5589
Early Fusion	2.6992	2.5714	2.5584
Mid Fusion	2.6837	2.5623	**2.5483**

Table 3: Average Cross Entropy Loss per note for Melody Prediction Task using context length 5.

In another experiment, we tested RBP as a pitch prediction in melodies as an extension of the work done by [1] with the dataset taken from the Essen Folk Song Collection [5] using recurrent neural networks and their gated variants (LSTM and GRU). We performed a grid search for each context length for hyper parameter tuning, with [10,30,50,100] as the size of the hidden layer and [30,50] epochs with learning rate set to 0.01, with one hidden layer and context length of size 5. The results in Table 3 summarize the results and show a consisten reduction in cross entropy with RBP.

4 Conclusions

We have shown that 'Relation Based Patterns' (RBP) as an inductive bias enable the learning of equality rules with neural networks. We observed that networks with suitable RBP structure learn abstract grammar patterns with almost 100% accuracy and also lead to improvements in character prediction and melody prediction tasks. In future, we will extend this work towards improving the performance of neural networks in providing better abstractions and generalizations.

References

1. Kopparti, R., Weyde, T.: Evaluating repetition based melody prediction over different context lengths. ICML Joint Workshop on Music and Machine Learning (2018)
2. Lake, B.M., Baroni, M.: Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks (2017), https://arxiv/abs/1711.00350
3. Marcus, G.F., Vijayan, S., Rao, S., Vishton, P.: Rule learning by seven-month-old infants. Science, 283 (5398), 77–80 (1999)
4. Santoro, A., Hill, F., Barrett, D., Morcos, A., Lillicrap, T.: Measuring abstract reasoning in neural networks. In: International Conference on Machine Learning. pp. 4477–4486 (2018)
5. Schaffrath, H.: The essen folksong collection in the humdrum kern format. Database edited by David Huron, CCAHR, Menlo Park, CA. (1995), http://kern.ccarh.org/cgi-bin/ksbrow
6. Weyde, T., Kopparti, R.: Feed-forward neural networks need inductive bias to learn equality relations (2018), https://arxiv.org/abs/1812.01662
7. Weyde, T., Kopparti, R.: Modelling identity rules with neural networks. Journal of Applied Logics (2019), https://arxiv.org/abs/1812.02616