Deep Bayesian Semi-Supervised Active Learning for Sequence Labelling.

Tomáš Šabata, Juraj Eduard Páll and Martin Holeňa

Department of Applied Mathematics, Faculty of Information technology, Czech Technical University in Prague

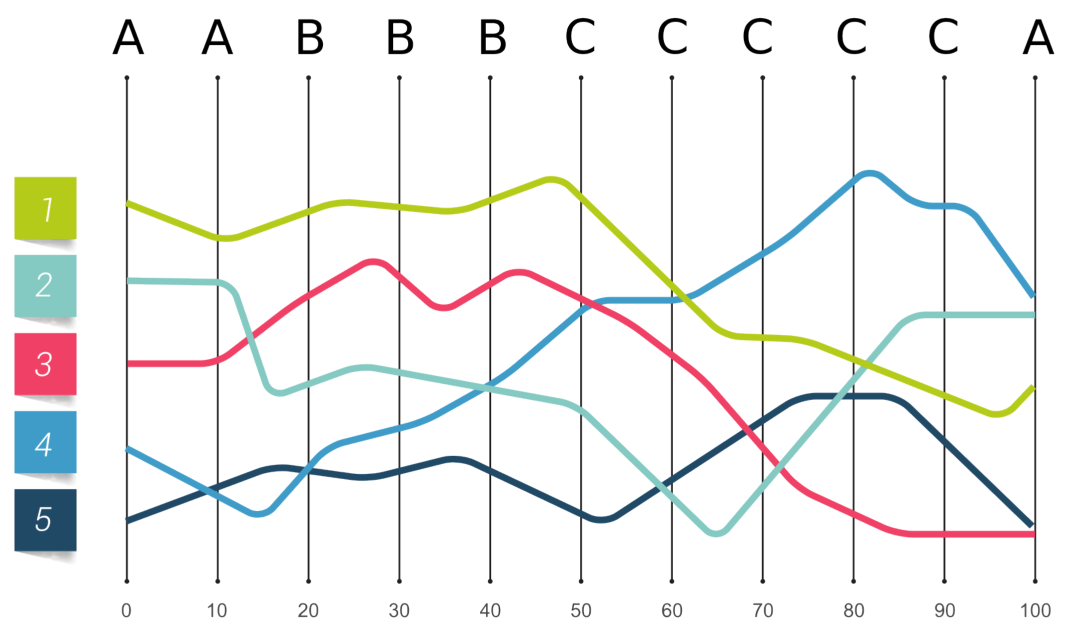
tomas.sabata@fit.cvut.cz

1. Introduction

In recent years, deep learning has shown supreme results in many sequence labelling tasks, especially in natural language processing. However, it typically requires a large training data set compared with statistical approaches. In areas where collecting of unlabelled data is cheap but labelling expensive, active learning can bring considerable improvement. Sequence learning algorithms require a series of token-level labels for a whole sequence to be available during the training process. Annotators of sequences typically label easily predictable parts of the sequence although such parts could be labelled automatically instead. In this paper, we introduce a combination of active and semi-supervised learning for sequence labelling.

2. Sequence labelling

A sequence labelling model assigns categorical labels to all elements of a sequence of observed values. In general, it considers the optimal label for a given element to be dependent on the choices of nearby elements.



3. Underlying models

The approach is designed for two neural network architectures. These architectures use LSTM RNN and differ in the last layer.

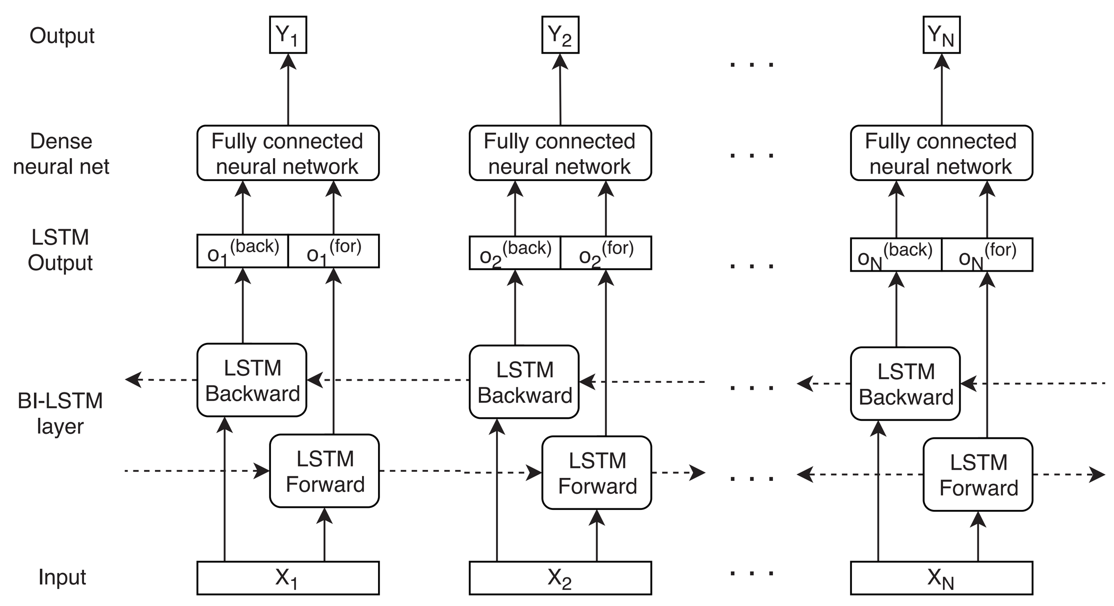


Figure 1: BI-LSTM-FCN

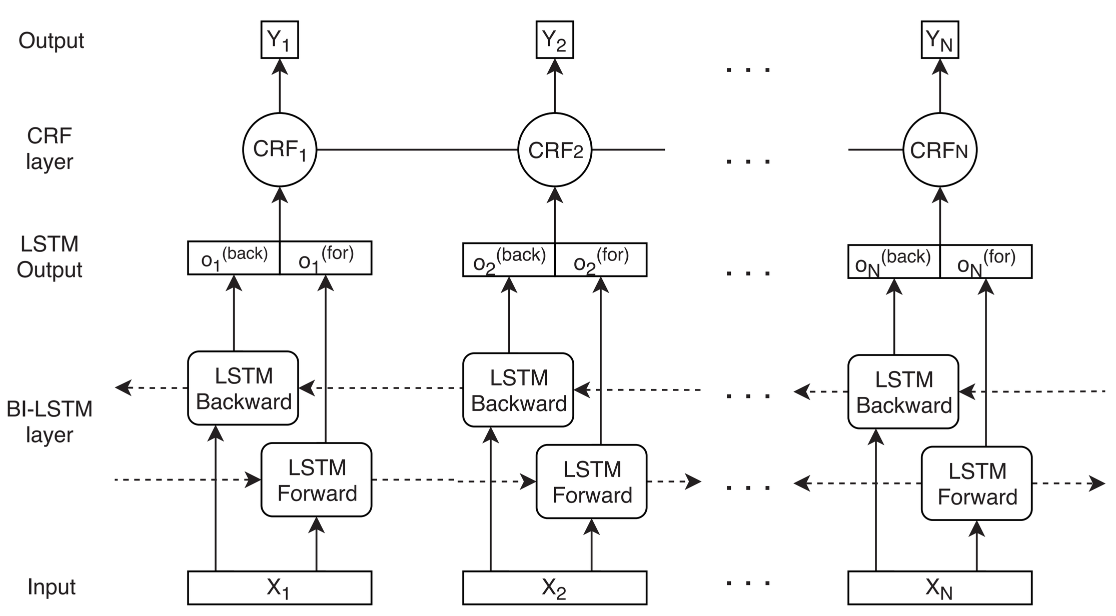
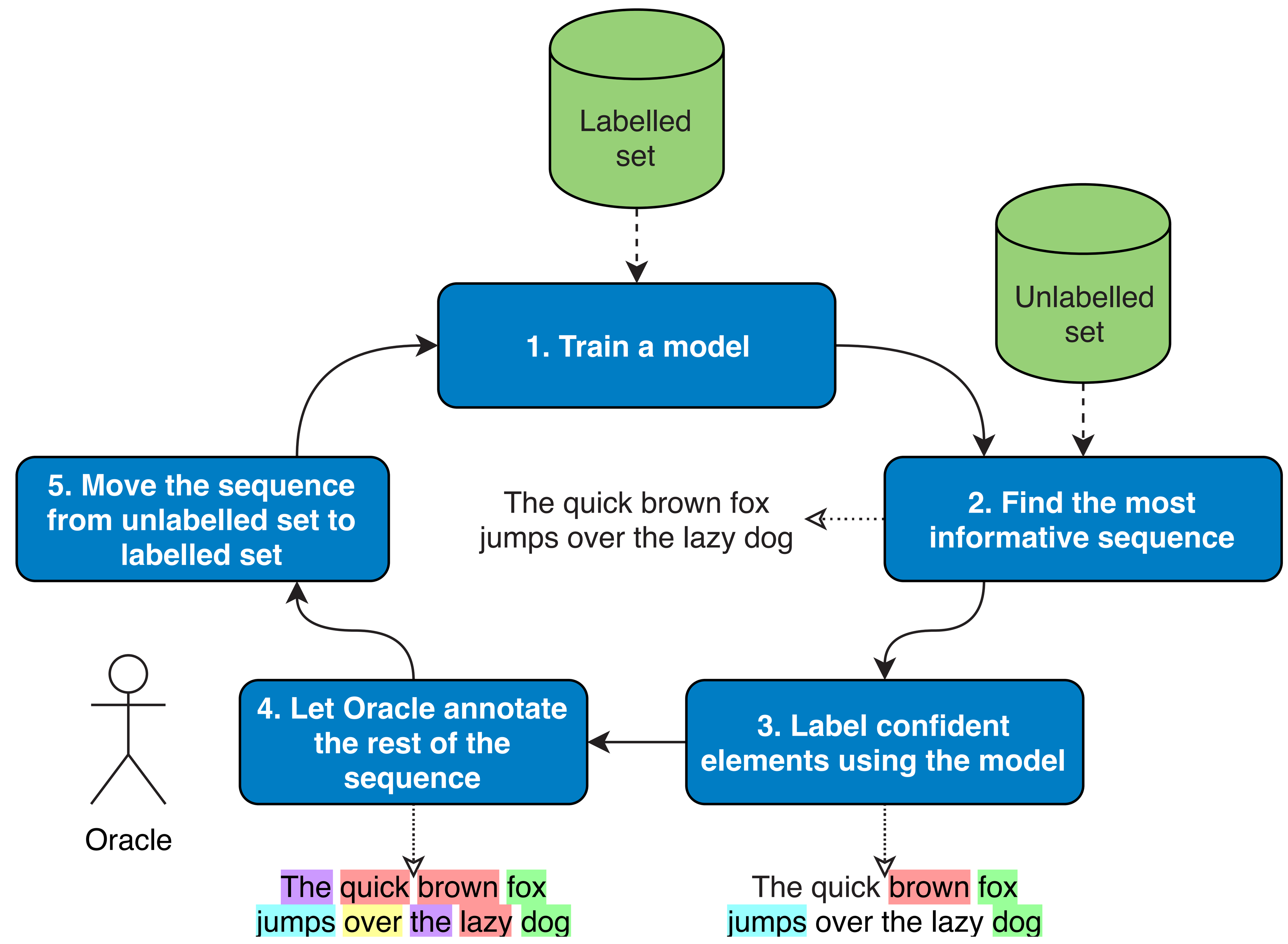


Figure 2: BI-LSTM-CRF

4. Deep Bayesian Semi-Supervised Active Learning

Our approach utilizes an approximation of Bayesian inference for neural nets using Monte Carlo dropout[1]. The approximation yields a measure of uncertainty that is needed in many active learning query strategies. We propose Monte Carlo token entropy and Monte Carlo N-best sequence entropy strategies. Furthermore, we use semi-supervised pseudo-labelling to reduce labelling effort.



5. Experimental results

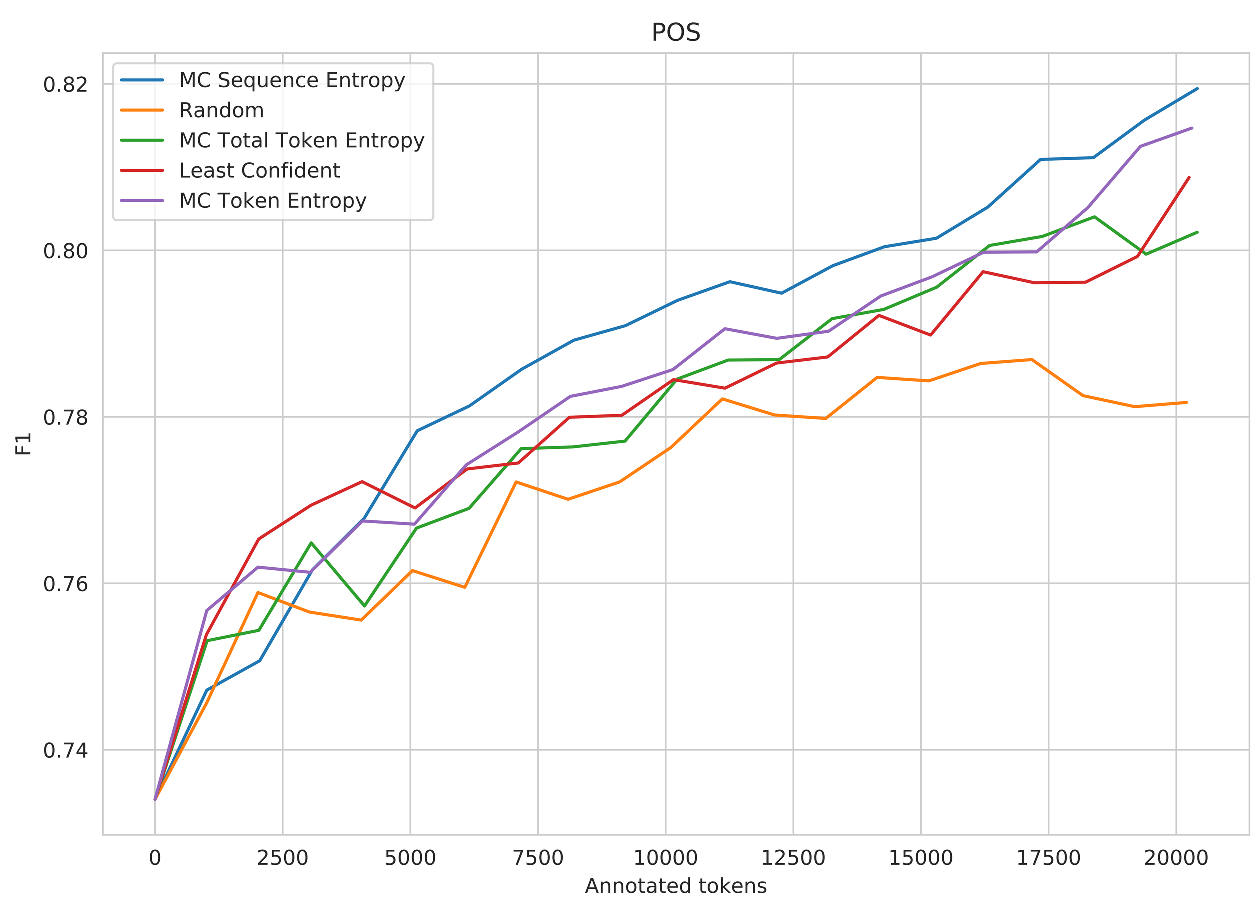
The proposed approach was evaluated on natural language processing tasks: named entity recognition (NER) and part of speech tagging (POS).

Table 1: Relative amount of pseudo-labelled tokens

Task type	NER		POS	
Allowed error	**0%**	**1%**	**0%**	**1%**
Least confident	77.3	84.3	72.4	82.5
Token entropy	79.5	83.7	72.2	77.9
Total token entropy	75.5	83.5	71.2	81.6

Table 2: Query strategies - F1 score

Task type	**NER**	**POS**
Random	76.5	78.1
Least Confident	76.9	80.8
MC Total Token Entropy	76.9	80.2
MC Sequence Entropy	**77.5**	**82.0**



6. Conclusions

1. The proposed query strategies have shown a substantial improvement over the until now used strategy in sequence labelling with deep neural networks, least confident.
2. Semi-supervised learning can rapidly reduce labelling effort for BI-LSTM-CRF.

7. Acknowledgements

The reported research was supported by the grant 18-18080S of the Czech Science Foundation (GACR).

8. References

[1] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059, 2016.