Multitask Learning for Twitter Sentiment Analysis

Université Grenoble-Alpes, CNRS, Grenoble INP*, LIG
Georgios.Balikas@imag.fr, Simon.Moura@imag.fr, Massih-Reza.Amini@imag.fr

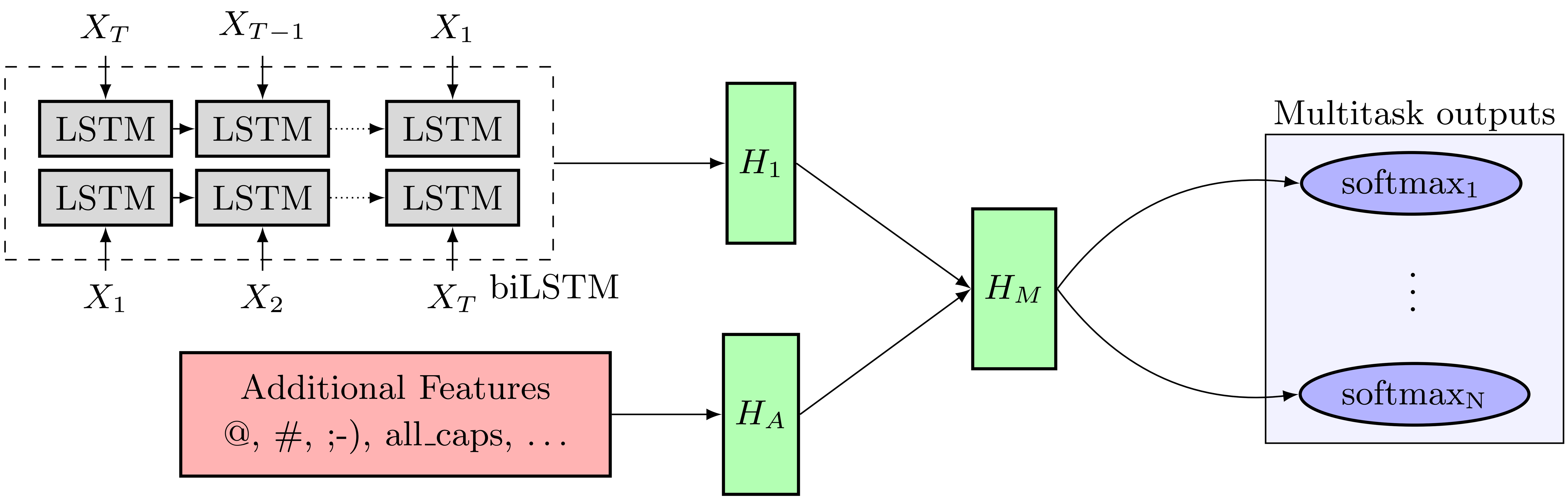
1. Motivation

- In machine learning, usually we focus on a single task. We optimize our model over a loss or a metric for the task at hand, either using a single model or an ensemble, whose hyperparameters we tune.
- As a result, we ignore information that may be useful, for instance training signals from related tasks.
- **Hypothesis:** We can enable our model to generalize better on the original task, by sharing representations between related tasks.

From wax on, wax off…

… to karate!

2. Multi-task Learning with BiLSTM



The neural network architecture for multitask learning using hard weight sharing between the N tasks. A bidirectional LSTM is used for learning a tweet representation that can be combined with "Additional Features". These can be domain-specific features or prior knowledge, are extracted in the preprocessing step and enable the network to integrate additional knowledge.

3. The Experimental Framework: Twitter Sentiment Classification

We experiment with two sentiment classification tasks:

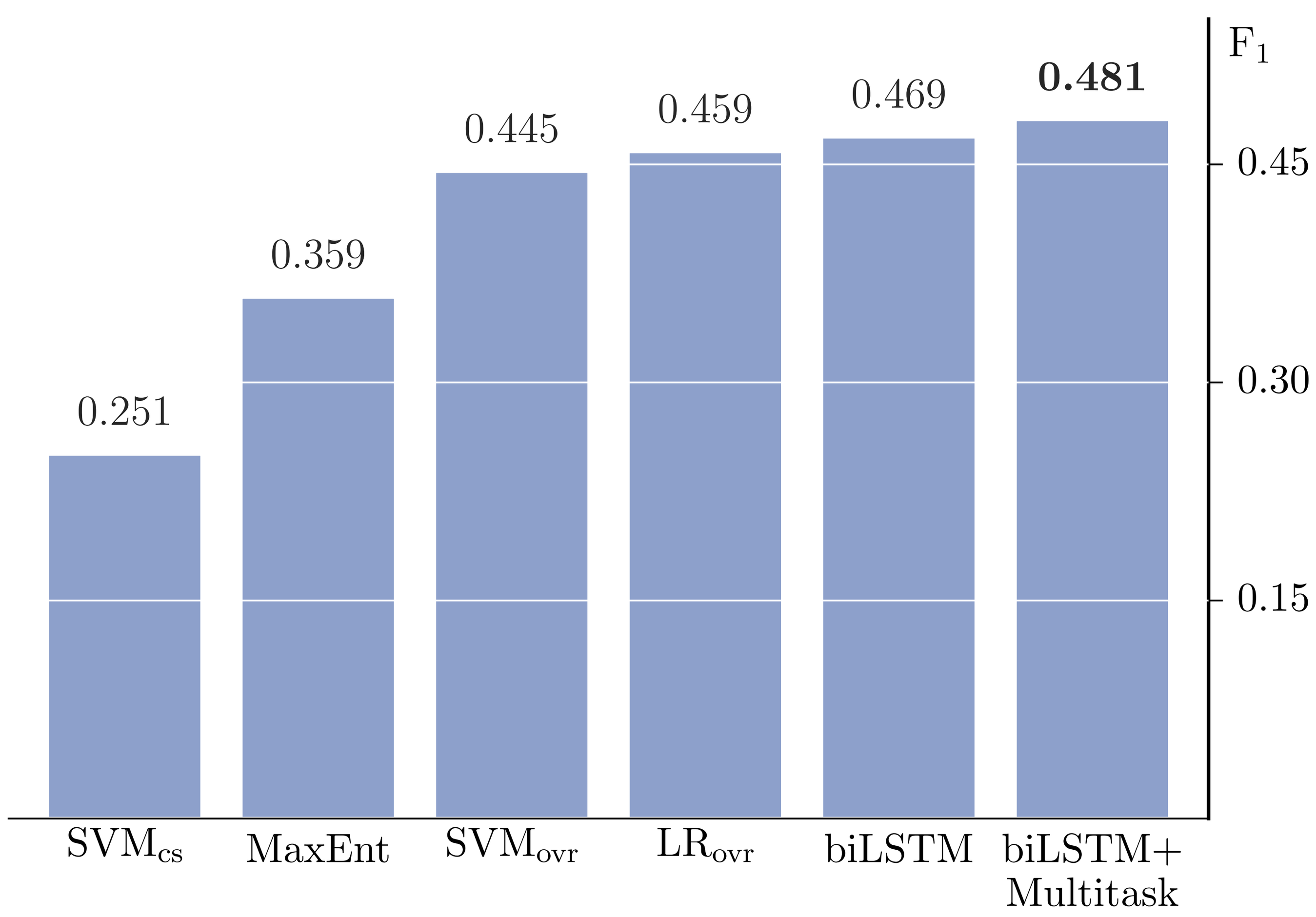
- Fine-grained: VeryPositive, …, VeryNegative. The primary task for this work.
- Ternary: Positive, Neutral, Negative

	nbow	nbow+
SVM_{ovr}	0.840	0.714
SVM_{cs}	0.946	0.723
LR_{ovr}	0.836	0.712
MaxEnt	0.842	0.715
Balikas et al. [1] (SOTA)	-	0.719
biLSTM (single task)	0.827±0.017	0.694±0.04
biLSTM+Multitask	0.786±0.025	**0.685**±0.024

The scores on macro-averaged mean absolute error. The best (lowest) score is shown in bold and is achieved in the multitask setting with the biLSTM architecture. `nbow` uses only text (GloVe embeddings), while `nbow+` incorporates the "Additional Features".

The performance improvements of may be due to:

1. Data Augmentation: Expands the set of training examples
2. Inductive bias: Prefer a hypothesis that performs well across tasks
3. Feature attention focusing: Decrease the weights on noisy features



F_1 scores with `nbow+`: Multitask learning achieves the best score.

References

[1] G. Balikas and MR Amini: "TwiSE at SemEval-2016 Task 4: Twitter Sentiment Classification", SemEval, (2016)
[2] R. Caruana: "Multitask Learning", Machine Learning, (1997)
[3] I. Goodfellow, Y. Bengio and A. Courville: "Deep Learning", MIT Press, (2016)

Open Source Implementation: `https://github.com/balikasg/sigir2017`

Acknowledgements

This work is partially supported by the CIFRE N 28/2015. Part of the SIGIR 2017 conference registration is covered by a SIGIR student grant. The models are implemented with Scikit-Learn and Keras.