Exploring Neural Architectures And Techniques For TypologicallyDiverse Morphological Inflection

Pratik Jayarao, Siddhanth Pillay, Pranav Thombre, Aditi Chaudhary
CARNEGIE MELLON UNIVERSITY - CMU TARTANS

Abstract

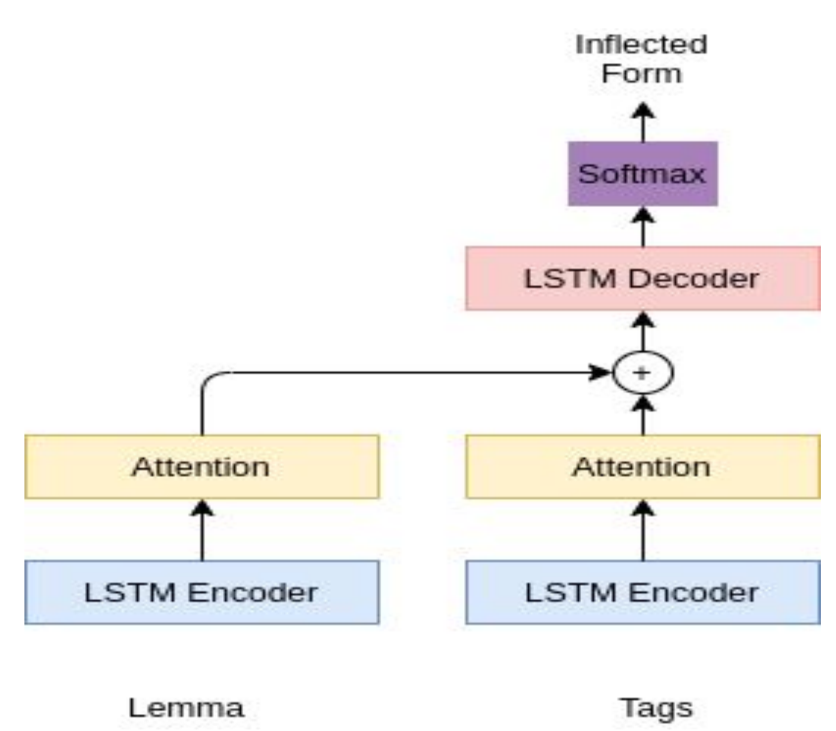
We describe our attention-based encoder-decoder approach that we implement using LSTMs and Transformers as the base units. We also describe the ancillary techniques that we experimented with, such as hallucination, language vector injection, sparsemax loss and adversarial language network alongside our approach to select the related language(s) for training.We present the results we generated on the constrained as well as unconstrained SIGMORPHON 2020 dataset.

Methodology

We implemented four variants of Sequence to Sequence architectures to tackle the problem of morphological injection. We primarily utilize LSTM and Transformers to construct our models.

LSTM Encoder Decoder (LSTM)

We prototyped an elementary LSTM sequence tosequence model with two encoders each encoder taking the input as the Lemma and Tags respectively.

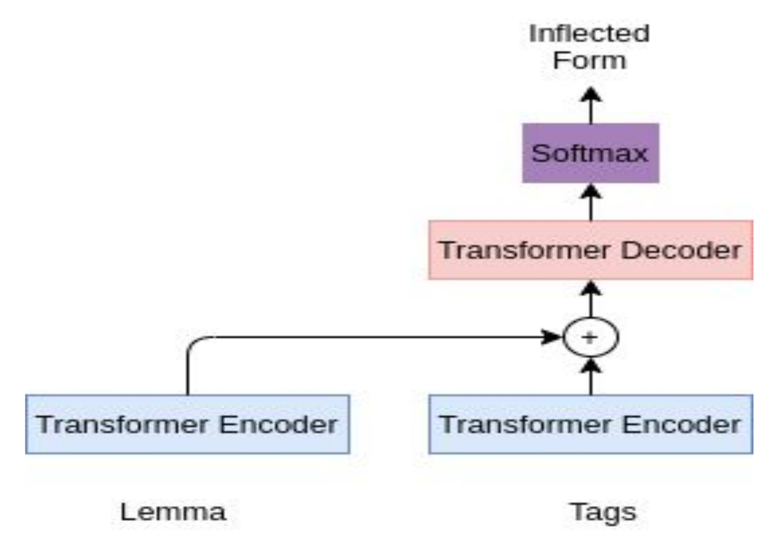


Transformer Encoders LSTM Decoder (TELD)

Recently, the use of attention has shown improvement in the performance of such models. Thus we replace the LSTM encoders in the previous modules with Transformer encoder.

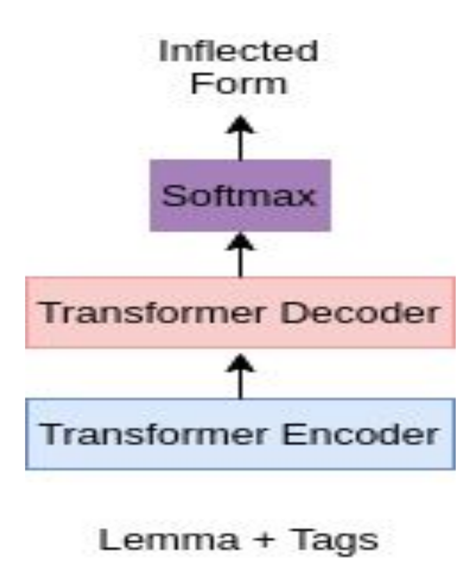
Transformer Encoders Transformer Decoder (TETD)

We further replace the LSTM Decoder with a Transformer Decoder



Joint Transformers (TJ)

The final architecture we implement is an end-to-end Transformer model. We concatenate the Lemma and the Tag and feed it to the Transformer.

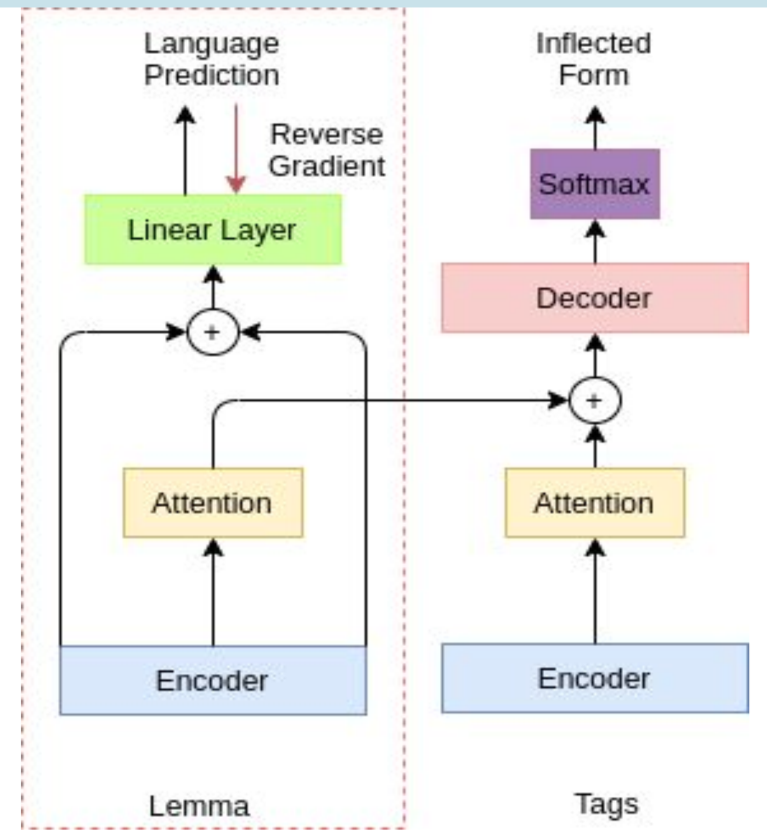


Hallucination (HALL)

We incorporated Hallucination techniques and observed a performance boost in their system.

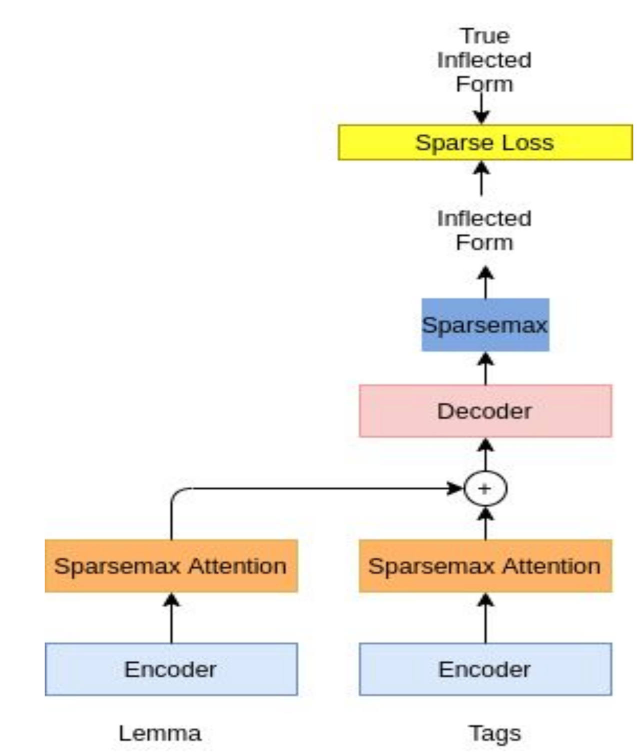
Adversarial Language Network (ADV-LANG)

Trying to transfer knowledge between related language(s) and a target language it is sometimes useful to learn language agnostic representations. Thus we implement a Language Adversarial Network.



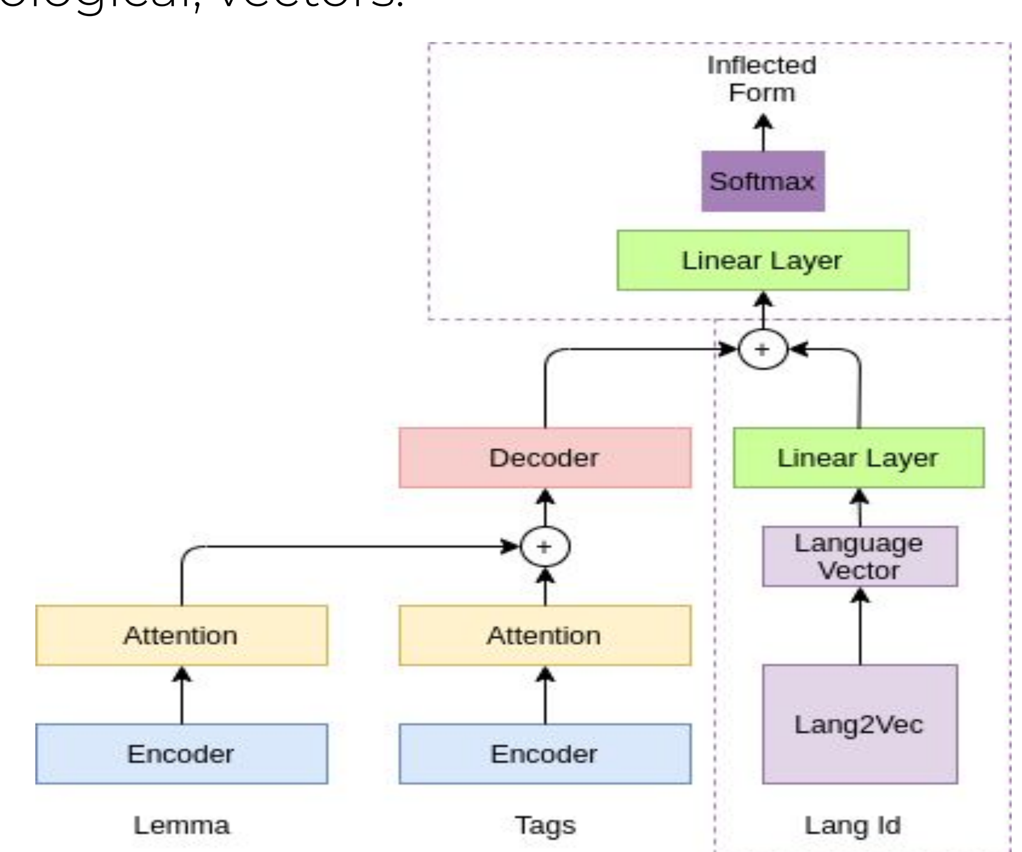
Sparse-Max and Sparse Loss (SPARSE)

Sparsemax assigns exactly zero attention weight to irrelevant source token sand implausible hypotheses.



Language Vector Injection (LVI)

The lang2vec represent languages using rich typological, vectors.



Results

We performed a comparative analysis of all the techniques implemented in our paper.

Target Language	Related Language(s)	Model	L1+L2	ADV-LANG	SPARSE	LVI
Zulu	gaa,lug,aka	LSTM	0.81	0.83	0.81	0.83
		TELD	0.83	0.84	0.83	**0.86**
		TETD	0.81	0.84	0.83	0.83
Chichica pan Zapotec	azg,cly	LSTM	0.84	0.83	0.87	0.84
		TELD	0.87	**0.88**	**0.88**	**0.88**
		TETD	0.85	0.86	0.85	**0.88**
Yoloxóchitl Mixtec	gmh,ang	LSTM	0.86	**0.89**	0.87	0.88
		TELD	0.84	0.84	0.84	0.83
		TETD	0.81	0.79	0.81	0.79
Sotho	nya,dan	LSTM	**1.0**	**1.0**	**1.0**	**1.0**
		TELD	0.98	**1.0**	0.96	**1.0**
		TETD	0.94	0.96	0.90	0.96
Luganda	lin,zul,ceb	LSTM	0.90	0.90	0.90	0.89
		TELD	0.91	0.90	**0.91**	0.90
		TETD	0.82	0.83	0.82	0.82
Livonian	gmh,ang, kon,swa	LSTM	0.91	0.91	0.92	0.91
		TELD	0.92	0.92	**0.94**	0.92
		TETD	0.67	0.71	0.71	0.70
Classical Syriac	ang	LSTM	**0.94**	0.92	0.83	**0.94**
		TELD	0.92	0.91	0.93	0.94
		TETD	0.93	0.92	**0.94**	0.93
Kannada	nob	LSTM	0.79	0.78	**0.83**	0.80
		TELD	0.79	0.79	0.79	0.80
		TETD	0.77	0.75	0.79	0.57
Swiss German	mlg,dan	LSTM	0.87	0.86	0.87	0.87
		TELD	0.85	**0.88**	0.86	0.85
		TETD	0.78	0.77	0.76	0.78

Conclusion

This paper presents a detailed description of the models that we implemented to undertake the along with different supporting techniques that we implemented, such as hallucination, language vector injection, adversarial language traning and sparsemax