Superiority of shallow neural networks over deep neural networks for morphological modelling- a case study for low resource languages

Anonymous ACL submission

Abstract

Modelling inflectional morphological system for a given word paradigm using neural networks has been an active area of research since the last decade or so. In this work, we study the inflectional morphology for the negatively polarised verb paradigm of an Indo-Aryan language, Assamese. The motivation for this study lies in the fact that the inflectional morphology for Assamese is relatively underexplored owing to the lack of data for this language. In this work, we precisely focus on the tense and person paradigm for the negative-polarity verbs in the said language and test the ability of the neural models on performing this task. Our experimental results on modeling the aforesaid morphology reveal that the Recurrent Neural Network (RNN) models slightly outperform the transformer models albeit with a small margin.

1 Introduction

Inflectional morphology is the process by which a word (such as noun or verb) is modified by applying certain affixation to express different grammatical aspects such as gender, tense and person.[1] Given a word paradigm, this inflectional phenomena can be either be **regular** or **irregular**. This sort of duality makes it extremely challenging to generalize inflectional morphology (for new words) in alignment with human cognition. Training a neural model to learn inflectional morphology dates back to 1987 when McClelland and Rumelhart (1987) came up with a neural network model capable of mapping English present tense verbs to their past tense in the case of both regular and irregular verbs. However, the study of such architectures in modeling the morphology of Indic languages such as *Assamese* remains severely underexplored.

Assamese is an Indo-Aryan language spoken in the North-Eastern part of India. Even though the language is spoken by nearly 15 million native speakers, the language is considered to be a low-resource language due to the lack of computable data (Singh et al., 2024).[3] From the linguistic point of view, Assamese is categorised as an agglutinative language owing to its relatively simple morphological inflectional forms compared to fusional languages (Socolof et al., 2022). However, in the case of verbs with negative polarity, the inflection significantly depends on the tense and the person. Concretely, two distinct inflectional phenomena occur for the paradigm above. In the case of present or future tense, the inflectional morpheme for a verb with the negative polarity is determined by a prefix to the inflected verb (e.g., নু) for first and third person while for the past tense, the inflectional morpheme for the same verb is realized as a suffix to the inflected verb (e.g., নাছিলে). For the Assamese word, শুন (*transl.* 'listen'), we show the inflectional forms in Table 1 with the abbreviations having the following meanings: `V`:verb; `IND`: indictive; `HAB`: habitual; `PRS`: present; `FUT`: future; `PST`: past; `1,3`: person.

meaning	lemma	form
V;IND;1;PRS	শুন	নুশুনোঁ
V;IND;1;FUT	শুন	নুশুনোঁ
V;IND;3;PRS	শুন	নুশুনে
V;IND;3;FUT	শুন	নুশুনে
V;IND;1;PST	শুন	নুশুনিলোঁ
V;HAB;IND;1;PST	শুন	শুনা নাছিলোঁ
V;HAB;IND;3;PST	শুন	শুনা নাছিলে

Table 1: Morphological inflection for the negatively polarised verb শুন based on tense and person. Underlined are inflections as prefixes or suffixes .[2]

Mechanistic interpretability explores the computational ability of a neural network aiming for a

[1] https://nagelhout.faculty.unlv.edu/AGiC/s4d.html
[2] https://unimorph.github.io/
[3] https://en.wikipedia.org/wiki/Assamese_language

thorough and precise understanding of model behavior (Bereska and Gavves, 2024). One may explore the model intepretability by various approaches, for instance probing (Belinkov, 2022), sparse autoencoders (Olshausen and Field, 1997), activation patching (Nainani, 2024). Probing tries to establish whether a representation captures specific information or not by embedding examples that exhibit certain information by training a neural model (Chowdhury and Allan, 2024) on the frozen features. As a part of our future work, we will build upon the present study and plan to apply this method to obtain a formal explanation for our findings through the lens of mechanistic interpretability.

1.1 Implementation Details

We implemented all the models using the OpenNMT library (Klein et al., 2020) in line with previous approaches for modelling language-based inflections (McCurdy et al., 2020; Dankers et al., 2021).[4] Following the setup of McCurdy et al. (2020), we implement the RNN architectures using 2 layers of LSTM as encoders and 2 layers of LSTM as decoders. We use 300-dimensional character embeddings and 100-dimensional hidden layers in both the encoder and decoder layers. We use the Adadelta optimizer with a batch size 20 and a dropout rate of 0.3. In the case of transformer networks, we follow the setup in Ma and Gao (2022). We use 2 layers of encoder and decoder layers, 4 attention heads, 128 dimensions of input features, and 512 as the dimension of the feed-forward network while using *self-attention*. While decoding the outputs, we set the beam size to 12. We additionally employ early stopping and obtain the best performance on the validation set with 5 epochs while iterating over a grid of learning rates from $1-10$ epochs. We report all the averaged results over random initializations of 3 seeds. All the models are trained on a single NVIDIA RTX 1080 GPU.

2 Results and Discussion

From Table 2 it is evident that The RNN model outperforms the transformer model on the test set. However, the performance of the transformer model is superior on the training and validation sets of our task. This reflects the ability of the transformer models to memorize the training data, as shown in Fig 1 in terms of the perplexity scores.

[4] https://opennmt.net/

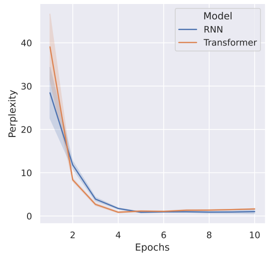

Figure 1: Perplexity plots for RNN and Transformer models.

Model	Train	Val	Test
RNN	98.1	92.2	**81.2**
Transformer	99.1	92.4	80.8

Table 2: Accuracy scores of different models

To analyze the gap in performance on the test set for both models, we perform an error analysis on the test set predictions. Both the models struggle to model the inflections within the <3><p> paradigm, which also has the lowest frequency in our dataset This paradigm also has the widest variety of inflections, accounting for 7 different types of patterns. The inflection frequently occurs as prefixes: নে, নু, না, নো, নি and sometimes as suffixes such as: নক ৰে, নাপাতে. The RNN model outperforms the transformer models by a small margin in identifying these suffixes (precision: 0.45 vs 0.38).

The performance breakdown of the RNN model is shown in Table 3. Evidently, the RNN models frequently predict the inflection with a prefix, নে, which has a higher occurrence in our corpus

3 Future Work

As a part of future work, we propose to explore the superiority of RNN over transformer model (as discussed above) employing the method of probing.

Paradigm	P	R	F1
`<1><t> lemma`	0.92	0.87	0.89
`<1><f> lemma`	0.93	0.80	0.86
`<1><p> lemma`	0.91	0.88	0.90
`<3><t> lemma`	0.92	0.89	**0.91**
`<3><p> lemma`	0.45	0.59	0.51
`<3><f> lemma`	0.89	0.87	0.88

Table 3: Performance breakup for our best-performing model (RNN) based on Precision (P), recall (R), and F1 scores.

References

Yonatan Belinkov. 2022. Probing classifiers: Promises, shortcomings, and advances. *Computational Linguistics*, 48(1):207–219.

Leonard Bereska and Efstratios Gavves. 2024. Mechanistic interpretability for ai safety–a review. *arXiv preprint arXiv:2404.14082*.

Tanya Chowdhury and James Allan. 2024. Probing ranking llms: Mechanistic interpretability in information retrieval. *arXiv preprint arXiv:2410.18527*.

Verna Dankers, Anna Langedijk, Kate McCurdy, Adina Williams, and Dieuwke Hupkes. 2021. Generalising to german plural noun classes, from the perspective of a recurrent neural network. In *Proceedings of the 25th conference on computational natural language learning*, pages 94–108.

Guillaume Klein, François Hernandez, Vincent Nguyen, and Jean Senellart. 2020. The opennmt neural machine translation toolkit: 2020 edition. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 102–109.

Xiaomeng Ma and Lingyu Gao. 2022. How do we get there? evaluating transformer neural networks as cognitive models for english past tense inflection. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1101–1114.

James L. McClelland and David E. Rumelhart. 1987. *On Learning the Past Tenses of English Verbs*.

Kate McCurdy, Sharon Goldwater, and Adam Lopez. 2020. Inflecting when there's no majority: Limitations of encoder-decoder neural networks as cognitive models for german plurals. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1745–1756.

Jatin Nainani. 2024. Evaluating brain-inspired modular training in automated circuit discovery for mechanistic interpretability. *arXiv preprint arXiv:2401.03646*.

Bruno A Olshausen and David J Field. 1997. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325.

Anushka Singh, Ananya Sai, Raj Dabre, Ratish Puduppully, Anoop Kunchukuttan, and Mitesh Khapra. 2024. How good is zero-shot MT evaluation for low resource Indian languages? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 640–649, Bangkok, Thailand. Association for Computational Linguistics.

Michaela Socolof, Jacob Louis Hoover, Richard Futrell, Alessandro Sordoni, and Timothy J. O'Donnell. 2022. Measuring morphological fusion using partial information decomposition. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 44–54, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.