

A Transformer-Based Neural Machine Translation System for Multilingual Text Translation to Tamil

S. Hamsanandhini

Assistant Professor, Department of Artificial Intelligence, Kongu Engineering College (Autonomous), Perundurai, Erode, India

Saliniyan P

Students, Department of Artificial Intelligence and Machine Learning, Kongu Engineering College (Autonomous), Perundurai, Erode, India

Sanjay A

Students, Department of Artificial Intelligence and Machine Learning, Kongu Engineering College (Autonomous), Perundurai, Erode, India

Yeshwanth A

Students, Department of Artificial Intelligence and Machine Learning, Kongu Engineering College (Autonomous), Perundurai, Erode, India

Abstract:

The explosive growth in the field of digital information sources has greatly intensified the need for effective translation mechanisms. Tamil, which is one of the most ancient languages with more than 80 million speakers around the world, continues to struggle in accessing digital information across the globe because of constant language barriers. In this study, we design a NMT system utilizing Transformer that translates texts from Hindi, Telugu, and Spanish into Tamil via pivot-based translation. Since parallel corpora between Tamil and various other languages are rare, English is used as a pivot language intermediary. This model leverages a transformer-based encoder-decoder framework that uses multi-headed self-attention systems and translates in two phases in sequence, as follows:

Source Language → English, and (ii) English → Tamil. A large number of multilingual parallel datasets, along with the Samanantar dataset is utilized for training the models. The results of the experiments conducted on chrF++, SBERT Cosine Similarity, Translation Edit Rate (TER) and Adequacy scores show competitive results; for example, the direct English–Tamil translation pipeline yields 30.88, 0.7626, 84.24, and 0.655 for chrF++, SBERT Cosine, TER, and Adequacy, respectively.

Keywords:

Neural Machine Translation, Transformer Architecture, Pivot Translation, Multilingual Translation, Tamil Language, Samanantar, SBERT, TER.