

ABSTRAK

Indonesia memiliki bahasa yang beragam di tiap daerahnya, salah satunya adalah bahasa Madura. Mayoritas masyarakat Madura menggunakan bahasa Madura sebagai bahasa utamanya. Sementara untuk kemampuan menggunakan bahasa nasionalnya sendiri, pada umumnya hanya didominasi oleh beberapa kalangan saja, hal ini tentunya akan membuat mereka mengalami kesulitan melakukan komunikasi dengan masyarakat daerah lainnya. Untuk mengatasi permasalahan tersebut, maka dibutuhkan suatu sarana yang dapat membantu masyarakat Madura dalam belajar bahasa Indonesia, salah satunya adalah mesin penerjemah. Dalam ruang lingkup mesin penerjemah, terdapat satu pendekatan yaitu *Neural Machine Translation* (NMT) dengan arsitektur *encoder-decoder*, pendekatan ini digunakan guna mengatasi permasalahan berubahnya susunan frase dari hasil terjemahan pada penggunaan mesin dengan penerjemahan perkata (leksikal). Dalam implementasinya, model yang cukup populer digunakan adalah *Long Short Term Memory* (LSTM). Arsitektur *encoder-decoder* memiliki kelemahan yaitu hilangnya informasi yang diterima *decoder* karena hanya menerima inputan dari *output* terakhir *encoder*. Oleh karena itu maka dibuat suatu metode yang disebut *Attention* guna menyelaraskan seluruh *output encoder* terhadap *decoder*. Berdasarkan penelitian ini, arsitektur *encoder-decoder* dengan *Attention* efektif digunakan saat menggunakan korpus yang cukup panjang (3-37 kata), dengan perolehan *score BLEU 1-4 gram* tertinggi rata-rata 0.0767 (merupakan *score BLEU* tertinggi pada penelitian ini), meskipun dengan jumlah data yang relatif sedikit (3.734 pasangan korpus). Namun pada saat korpus yang digunakan hanya memiliki panjang kalimat maksimal 2 kata, arsitektur *encoder-decoder* dengan *Attention* justru menjadi tidak efektif untuk digunakan, hal ini karena *score BLEU 1-4 gram* tertinggi rata-rata yang diperoleh turun menjadi hanya 0.0344 meskipun dengan jumlah data yang relatif jauh lebih banyak (28.247 pasangan korpus).

kata kunci: Bahasa Madura, *Attention*, LSTM, NMT, Mesin Penerjemah

ABSTRACT

Indonesia has a variety of languages in each region, one of which is the Madura language. The majority of Madurese people use Madurese as their main language. Meanwhile, the ability to use their own national language, in general, is only dominated by a few people, this will certainly make it difficult for them to communicate with other regional communities. To overcome this problem, a means is needed that can help the people of Madura in learning Indonesian, one of which is a machine translator. In the scope of machine translation, there is one approach, namely Neural Machine Translation (NMT) with an encoder-decoder architecture, this approach is used to overcome the problem of changing the arrangement of phrases from the translation results in the use of machine translation of words (lexical). In its implementation, a model that is quite popularly used is Long Short Term Memory (LSTM). The encoder-decoder architecture has the disadvantage of losing the information that the decoder receives because it only accepts input from the encoder's final output. Therefore, a method called Attention was created to align all encoder outputs with the decoder. Based on this study, the encoder-decoder architecture with Attention was effectively used when using a fairly long corpus (3-37 words), with the highest BLEU score of 1-4 grams on average 0.0767 (the highest BLEU score in this study), although with a relatively small amount of data (3,734 corpus pairs). However, when the corpus used only had a maximum sentence length of 2 words, the encoder-decoder architecture with Attention became ineffective to use, this was because the highest average BLEU score of 1-4 grams obtained dropped to only 0.0344 even though the amount of data was relatively much larger (28,247 corpus pairs).

keywords: *Madurese language, Attention, LSTM, NMT, Machine Translation*