

ABSTRAK

Mesin penerjemah merupakan cabang penting dari pemrosesan bahasa alami yang bertujuan untuk menerjemahkan bahasa alami menggunakan komputer. Penerapan mesin penerjemah (MT) di Indonesia sudah dan masih banyak dilakukan dengan berbasis statistik (SMT) khususnya dalam eksperimen penerjemahan bahasa daerah. Dalam beberapa tahun terakhir, mesin penerjemah jaringan saraf tiruan (NMT) telah mencapai kesuksesan yang luar biasa dan menjadi metode pilihan baru dalam praktik mesin penerjemah bahkan sampai sekarang sudah diterapkan di google. Pemanfaatan mesin penerjemah dapat digunakan sebagai upaya dalam melestarikan bahasa daerah untuk komputasi digital. Mesin penerjemah jaringan saraf tiruan pada penelitian ini menggunakan mekanisme *attention* Bahdanau dalam bahasa Indonesia ke bahasa Tiochiu Pontianak dengan data korpus paralel sebanyak 5000 baris kalimat. Mekanisme *attention* muncul sebagai peningkatan atas mesin penerjemah jaringan saraf tiruan dalam menentukan seberapa kuat sebuah kata yang terhubung dengan kata yang lain, yang kemudian kata tersebut membantu jaringan saraf tiruan menyimpulkan seberapa penting hubungan sebuah kata dalam konteks kalimat tersebut. Perlunya dilakukan penelitian terhadap mekanisme *attention* untuk nilai akurasi tertinggi berdasarkan nilai skor BLEU dan ahli bahasa. Proses pengujian dilakukan dengan menguji nilai skor BLEU dari mekanisme *attention* menggunakan metode penambahan secara konsisten dengan jumlah *epoch* 10 sebanyak tujuh kali pengujian dan metode pembagian data uji dengan *k-fold cross validation*. Setelah didapatkan mekanisme *attention* yang menghasilkan akurasi tertinggi berdasarkan dari metode pembagian data uji dengan *k-fold cross validation* maka data uji tersebut akan dilakukan pengujian lagi dengan *k-fold cross validation* untuk mengambil sampel dari data uji yang memiliki akurasi tertinggi kemudian sampel tersebut akan dilakukan pengujian manual oleh ahli bahasa. Hasil pengujian berdasarkan metode pembagian data uji dengan *k-fold cross validation* didapatkan hasil nilai rata-rata sebesar 18,41% dengan *attention* Bahdanau, sedangkan pengujian manual oleh ahli bahasa pada 167 kalimat terjemahan memperoleh nilai akurasi sebesar 54,51%. Berdasarkan pengujian dengan metode penambahan jumlah *epoch* secara konsisten didapatkan hasil uji akurasi *attention* Bahdanau memiliki nilai akurasi uji otomatis BLEU tertinggi dengan pengujian *k-fold cross validation* pada sampel uji6 dengan jumlah *epoch* 50 sebesar 24,26%. Nilai akurasi otomatis BLEU tertinggi pada sampel Uji6 dengan tanpa Attention Bahdanau sebesar 12,41%. Sehingga dapat disimpulkan bahwa dengan mekanisme Attention Bahdanau adalah tingkat hasil akurasi yang terbaik.

Kata kunci: mesin penerjemah jaringan saraf tiruan, mekanisme *attention* Bahdanau, *deep learning tensorflow*, pemrosesan bahasa alami, BLEU, *k-fold cross validation*.

ABSTRACT

Machine translation is an essential branch of natural language processing that aims to translate natural languages using computers. The application of machine translation (MT) in Indonesia has been and is still mostly done on a statistical basis (SMT), especially in regional language translation experiments. In recent years, neural machine translation (NMT) has achieved tremendous success and has become the new method of choice in machine translation practice even now it has been implemented in google. The utilization of machine translators can be used as an effort to preserve regional languages for digital computing. The neural translation machine in this study uses the Bahdanau attention mechanism in Indonesian language to Tiochiu Pontianak language with 5000 sentences of parallel corpus data. The attention mechanism emerged as an improvement over the neural machine translation engine in determining how strongly a word is connected to another word, which then helps the neural machine infer how important the relationship of a word is in the context of the sentence. It is necessary to research the attention mechanism to determine the highest accuracy value based on the BLEU score and linguists. The testing process is carried out by testing the BLEU score of each attention mechanism using the addition method consistently with a total of 10 epochs for seven tests and the method of limited test data with k-fold cross validation. After obtaining the attention mechanism that produces the highest accuracy based on the limited test data method with k-fold cross validation, the test data will be tested again with k-fold cross validation to take samples from the test data that has the highest accuracy then the sample will be tested manual by a linguist. The test results based on the test limited data method with k-fold cross validation obtained an average value of 18.41% with Bahdanau attention, while manual testing by linguists on 167 translated sentences obtained an accuracy value of 54.51%. Based on testing with the addition method consistent with the number of epochs, the results of the attention accuracy test Bahdanau have the highest BLEU automatic test accuracy value with the k-fold cross validation test on the sample test6 with the number of 50 epochs of 24.26%. The highest BLEU automatic accuracy value in the sample test6 with not using Attention Bahdanau is 12.41%. So it can be said that the Attention Bahdanau mechanism is accuracy is the best level of accuracy.

Keywords: neural machine translation, Bahdanau attention mechanism, deep learning tensorflow, natural language processing, BLEU, k-fold cross validation.