

ABSTRAK

Andzar Tsaqif Laksana

NPM 140110190063

Word Embedding adalah suatu metode yang merubah sebuah kata menjadi sebuah vektor. Vektor yang merpresentasikan makna dari sebuah kata tersebut disebut dengan *word vector*. Penelitian mengenai *Word Embedding* sudah sangat jauh berkembang hingga menciptakan beberapa metode yang baik seperti Word2Vec, FastText, GloVe, dan lain-lain. Namun, salah satu dari metode tersebut masih memiliki kekurangan, yaitu terdapat pada model Skip-gram dari Word2Vec untuk fungsi penerjemah. Kekurangan yang dimaksud adalah ketidakkonsistenan antara fungsi objektif untuk mempelajari *word vector*, transformasi linear, serta pengukuran jarak antara *word vector* di kedua bahasa. Salah satu peneliti mengemukakan bahwa langkah awal dari menyelesaikan ketidakkonsistenan tersebut dengan menormalisasikan *word vector*. Penelitian pada skripsi ini terfokuskan pada proses *Word Embedding* dengan model Skip-gram serta menormalisasikan *word vector* yang diperoleh dari model skip-gram dengan *norm*. Jarak antara *Word vector* yang telah dinormalisasi menjadi saling berdekatan. Hasil dari kajian penelitian ini diharapkan dapat memberikan pemahaman terkait proses *Word Embedding* dengan model Skip-gram dan solusi dari ketidakkonsistenan tersebut yang dapat digunakan pada penelitian selanjutnya.

Kata kunci : normalisasi; *word embedding*; skip-gram.

ABSTRACT

Andzar Tsaqif Laksana
NPM 140110190063

Word Embedding is a method that converts a word into a vector. The vector that represents the meaning of the word is called a word vector. Research on Word Embedding has come a long way to create some good methods such as Word2Vec, FastText, GloVe, and others. However, one of the Word Embedding methods called the skip-gram model from word2vec is still lacking for the translator function.. The drawback is the inconsistency between the objective function to learn the word vector, the linear transformation, and the distance measurement between word vectors in both languages. One of the researchers suggested that the first step of resolving this inconsistency is to normalize the word vector. The research in this thesis focuses on the Word Embedding process with the Skip-gram model and normalizing the word vector that obtained from the skip-gram model with norm. The distance between the normalized word vectors becomes closer to each other. The results of this research study are expected to provide an understanding of the Word Embedding process with the Skip-gram model and the solution to the inconsistency that can be used in further research.

Keywords : *normalization; word embedding; skip-gram.*