

Abstrak

A Deep Learning Approach with Neural Networks to Recognise Javanese Dialects

oleh:

Alvin Januar Ramadan
18/425514/PA/18406

Sudah terdapat banyak topik riset berbahasa Jawa di berbagai jurnal dan dunia akademis. Walaupun begitu, bentuk riset yang dilakukan sebagian besar merupakan studi perbandingan antara dua dialek yang berbeda atau di suatu daerah. Terlebih juga, layanan penerjemah juga sudah mulai menawarkan Bahasa Jawa sebagai salah satu bahasa yang didukung dalam layanan tersebut. Dengan lebih dari 98 juta penutur, sudah sepantasnya dibutuhkan sebuah model yang mampu membedakan berbagai macam dialek Bahasa Jawa. Berbagai macam pihak seperti perseorangan dan juga penyedia jasa seperti translator dapat memahami konteks yang berbeda ketika pihak-pihak tersebut memahami dialek yang sedang digunakan.

Riset ini bertujuan untuk membuat sebuah model yang mampu memahami satu-satu dari berbagai macam dialek Bahasa Jawa. Hal seperti ini sangat dimungkinkan dengan menggunakan berbagai macam metode Deep Learning untuk mencari tahu macam-macam pola dari masing-masing dialek Bahasa Jawa. Dengan tersedianya subset dari dataset korpora EVA dari Max Planck Institute yang berfokus kepada pemetaan dialek Bahasa Jawa, riset ini berfokus dalam pembuatan sebuah model jaringan syaraf tiruan yang berbasis penelitian dari bahasa-bahasa lainnya.

Berbasis kepada arsitektur jaringan syaraf tiruan dari berbagai macam artikel jurnal terpublikasi, kami membandingkan berbagai macam tipe jaringan syaraf tiruan. Dengan menggunakan beberapa model seperti fully-connected deep neural network, convolutional neural network hingga convolutional recurrent neural network, model yang kami gunakan mampu mengidentifikasi berbagai macam dialek Bahasa Jawa dengan ketepatan validasi mulai dari 30% hingga 97.6%. Dibandingkan dengan berbagai macam artikel jurnal yang menjadi basis dari penelitian kami, performa yang diraih oleh riset ini disebabkan oleh beberapa keputusan dalam tahap preprocessing dan penggunaan lapisan jaringan syaraf tiruan, seperti normalisasi suara, regularisasi Ridge, batch normalisation, dan lapisan Dropout.

Kata kunci — *bahasa jawa, identifikasi dialek, dialek, jaringan syaraf tiruan, deep learning*

Abstract

A Deep Learning Approach with Neural Networks to Recognise Javanese Dialects

by:

Alvin Januar Ramadan
18/425514/PA/18406

The research topic of the Javanese language is abundant in journals and the academic world. However, such research projects turn out to be mostly about comparative studies of either two dialects or in a particular area. Moreover, translation services are also offering Javanese as a supported language. With over 98 million speakers, the need for a Javanese dialect identification model has been more significant than ever. People and service providers such as translators can better understand the context with the dialects in mind.

This research aims to recognise dialects from one of the many available dialects in the Javanese language. It is possible by using deep learning methods to figure out patterns in each of the Javanese dialects. With the readily available subset of the EVA corpora by the Max Planck Institute on Javanese dialect mapping, this research focuses on constructing a neural network model based on breakthroughs in other languages.

By referring to neural network architectures from established journal articles, we compare different types of neural networks. Using several models ranging from fully-connected deep neural networks, convolutional neural networks up to convolutional recurrent neural network, our inferred models can identify Javanese dialects very well with validation accuracy ranging from 30% up to 97.6%. Compared to articles cited in this article, this is due to several preprocessing steps and layer usage decisions, such as sound normalisation before input, ridge regularisation, batch normalization and dropout layers.

Keywords: *javanese, dialect identification, dialects, neural network, deep learning*