

Fuzzy C-Means Clustering dengan Reduksi Dimensi Deep Autoencoders untuk Pendeteksian Topik pada Data Tekstual Twitter = Fuzzy C-Means Clustering with Deep Autoencoders Dimensional Reduction for Topic Detection on Textual Data from Twitter

Natasha Rosaline, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20501795&lokasi=lokal>

Abstrak

Pendeteksian topik merupakan suatu teknik untuk memperoleh informasi dengan cara mengekstrak topik-topik dari kumpulan data yang sangat besar. Salah satu metode yang digunakan untuk pendeteksian topik adalah metode clustering, yaitu Fuzzy C-Means (FCM). Namun, kinerja dari FCM menjadi buruk saat harus melakukan clustering pada data yang berdimensi tinggi. Kelemahan dari FCM tersebut dapat ditanggulangi dengan cara melakukan reduksi dimensi. Pada penelitian ini, digunakan suatu metode deep learning, yaitu Deep Autoencoders (DAE), untuk mereduksi dimensi dari kumpulan data. Metode FCM clustering dengan reduksi dimensi DAE ini disebut Deep Autoencoders-Based Fuzzy C-Means (DFCM). Metode DFCM dibagi menjadi dua tahapan, yakni mereduksi dimensi kumpulan data yang berdimensi tinggi menggunakan Deep Autoencoders, dan melakukan FCM clustering pada data yang telah direduksi. Hasil dari metode DFCM adalah topik-topik. Topik-topik tersebut dievaluasi menggunakan nilai coherence. Pada penelitian ini, dibangun dua metode DFCM, yaitu FCM berbasis DAE dengan satu lapisan tersembunyi (DFCM-single hidden layer) dan FCM berbasis DAE dengan multi lapisan tersembunyi (DFCM-multi hidden layers). Hasil dari kedua metode ini menunjukkan bahwa topik-topik pada DFCM-single hidden layer memiliki nilai coherence lebih tinggi dari topik-topik pada DFCM-multi hidden layers.

.....Topic detection is a technique to find out information by extracting topics from big data. One method used for topic detection is the clustering method, namely Fuzzy C-Means (FCM). However, the performance of FCM becomes worse when clustering on highdimensional data. That weakness is resolved by dimensional reduction. In this research, deep learning method is used to reduce the dimensions of the data set, namely Deep Autoencoders (DAE). FCM clustering method with DAE dimensional reduction is called Deep Autoencoders-Based Fuzzy C-Means (DFCM). DFCM is divided into two parts. First, reducing the dimensions of high-dimensional data collection using Deep Autoencoders. Second, performing FCM clustering on the reduced data. Results of DFCM are topics. These topics are evaluated using the value of coherence. In this research, two DFCM methods were built, namely DAE with one hidden layer based FCM (DFCM-single hidden layer) and DAE with multi-hidden layers based FCM (DFCMmulti hidden layers). The results of these two methods show that the topics in DFCMsingle hidden layer have a higher coherence value than the topics in DFCM-multi hidden layers.