

Fuzzy C-Means Clustering dengan Reduksi Dimensi Convolutional Autoencoder pada Pendeteksian Topik = Fuzzy C-Means Clustering with Convolutional Autoencoder Dimensional Reduction for Topic Detection

Robertus Agung Pradana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20501757&lokasi=lokal>

Abstrak

Pendeteksian topik adalah suatu proses yang digunakan untuk menganalisis kata-kata pada suatu koleksi data tekstual untuk menentukan topik-topik yang ada pada koleksi tersebut, bagaimana hubungan topik-topik tersebut satu sama lainnya, dan bagaimana mereka berubah dari waktu ke waktu. Metod (FCM) merupakan metode yang sering digunakan pada masalah pendeteksian topik. FCM dapat mengelompokkan dataset ke beberapa kelompok dengan baik pada dataset dengan dimensi yang rendah, namun gagal pada dataset yang berdimensi tinggi. Untuk mengatasi permasalahan tersebut, dilakukan reduksi dimensi pada dataset sebelum dilakukan pendeteksian topik. Pada penelitian ini digunakan Convolutional Autoencoder dalam reduksi dimensi pada dataset. Oleh sebab itu, metode yang digunakan pada penelitian ini dalam pendeteksian topik adalah metode Convolutional-based Fuzzy C-Means (CFCM). Data yang digunakan dalam penelitian ini data akun berita nasional pada sosial media Twitter. Tahapan metode CFCM terbagi menjadi dua tahap, yaitu mereduksi dimensi dataset ke dimensi yang lebih rendah dengan menggunakan CAE dan selanjutnya dilakukan clustering dataset dengan menggunakan FCM untuk memperoleh topik-topik. Setelah diperoleh topik-topik, dilakukan evaluasi dengan menghitung nilai coherence pada topik yang diperoleh. Penelitian dilakukan dengan membandingkan nilai coherence pada topik antara metode CFCM dengan satu convolutional layer (CFCM-1CL) dan metode CFCM dengan tiga convolutional layer (CFCM-3CL). Hasil penelitian ini menunjukkan bahwa nilai coherence dari metode CFCM-1CL lebih tinggi dibandingkan metode CFCM-3CL.

<br.

Topic detection is a process used to analyze words in a collection of textual data to determine the topics in the collection, how they relate to each other, and how they change from time to time. The Fuzzy C-Means (FCM) method is a clustering method that is often used in topic detection problems. Fuzzy C-Means can group dataset into multiple clusters on low-dimensional dataset, but fails on high-dimensional dataset. To overcome this problem, dimension reduction is carried out on the dataset before topic detection is carried out. In this study, Convolutional Autoencoder (CAE) is used in the reduction of dimensions in the dataset. Therefore, the method used in this research in topics detection is the Convolutional-based Fuzzy C-Means (CFCM) method. The data used in this study tweets national news account data on social media Twitter. CFCM method are divided into two stages, namely reducing the dataset dimension to a lower dimension using CAE and then clustering the dataset by using FCM to obtain topics. After the topics are obtained, an evaluation is done by calculating the value of coherence on the topics obtained. The study was conducted by comparing the coherence value on the topic between the CFCM method with one convolutional layer (CFCM-1CL) and the CFCM method with three convolutional layers (CFCM-3CL). The results of this study indicate that the coherence value of the CFCM-1CL method is higher than the CFCM-3CL method