

Pemodelan topik menggunakan eigenspace-based fuzzy C-means dengan metode inisialisasi non-negative double singular value decomposition (NNDSVD) = Topic modeling using eigenspace-based fuzzy C-means with non-negative double singular value decomposition (NNDSVD) initialization method

Hamimah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20475004&lokasi=lokal>

Abstrak

ABSTRAK

Pendeteksian topik adalah proses menemukan topik yang digunakan untuk menganalisis kata dalam suatu kumpulan dokumen. Pendeteksian topik secara manual pada data yang besar sangatlah sulit. Sehingga dibutuhkan metode otomatis yang dikenal dengan pemodelan topik. Salah satu metode pemodelan topik yang sering digunakan adalah metode clustering. Clustering adalah teknik pengelompokan data yang tujuannya adalah untuk mengelompokkan data tersebut sehingga anggota dari grup yang sama lebih homogen atau lebih mirip satu sama lain daripada dengan anggota kelompok yang berbeda. Metode clustering yang sering digunakan adalah Fuzzy C-Means FCM. FCM ini bekerja dengan baik pada data dengan dimensi yang rendah, namun gagal pada data dengan dimensi yang tinggi Winkler, dkk, 2011. Pada data dimensi yang tinggi, algoritma FCM konvergen ke satu pusat centre of gravity, sehingga topik-topik yang dihasilkan antara satu dengan yang lainnya sama. Salah satu pendekatan untuk mengatasi kegagalan metode FCM pada data dimensi tinggi adalah memproyeksikan data pada ruang Eigen dengan dimensi lebih rendah dan metode tersebut dikenal juga dengan Eigenspace-based Fuzzy C-Means EFCM. Pada algoritma EFCM umumnya dilakukan inisialisasi random yang menyebabkan topik yang dihasilkan tidak sama setiap kali algoritma tersebut dijalankan. Untuk mengatasi masalah tersebut dibutuhkan inisialisasi yang tidak random. Untuk itu, pada skripsi ini akan digunakan metode Nonnegative Double Singular Value Decomposition NNDSVD. Algoritma NNDSVD terdiri dari dua proses metode SVD. Hasil dari simulasi ini menunjukkan bahwa nilai akurasi dengan inisialisasi NNDSVD menunjukkan adanya peningkatan lebih baik dibandingkan dengan inisialisasi random dan NNDSVD dapat menyelesaikan masalah EFCM dengan data berdimensi tinggi.

<hr>

ABSTRACT

Detection Topic is a process of finding the topics used to analyze words in a document that a collection of textual data. Detecting topic for a very large document hardly done manually. The topic detection problem is automatically known as topic modeling. One method of topic modeling that are commonly used is clustering method. Clustering is a data grouping technique which purposes is to group the data so members of each group are more homogeneous and more like each other than with different group members. This research will use fuzzy clustering method with Fuzzy C Means algorithm FCM . FCM works well on low data dimensions but it fails on high data dimensions. One approach to overcome the failure of FCM methods in high dimensional spaces is to project data on lower dimensional Eigen spaces and the method is also known as EigenSpace based FCM EFCM. In the EFCM, the algorithm did random initialization that causes the resulting topic was not same every time the algorithm runs. To solve this problem, it requires to implement

non random initialization. In this study, we used the initial Nonnegative Double Singular Value Decomposition NNDSVD. The basis of the NNSVD algorithm is a two processes SVD method. This simulation results show that NNDSVD initialization method can solves the eigenspace based Fuzzy C Means problems in high dimension data and NNDSVD based initialization gives same resulted topic every executed algorithm.