

Analisis kinerja inialisasi fuzzy c-means berbasis singular value decomposition untuk masalah pendeteksian topik = Performance analysis of singular value decomposition based initalization method in fuzzy c-means for topic detection

Deo Lahara, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20459177&lokasi=lokal>

Abstrak

ABSTRAK

Pendeteksian topik topic detection adalah suatu proses yang digunakan untuk menganalisis kata-kata pada suatu koleksi data tekstual untuk menentukan topik-topik yang ada pada koleksi tersebut. Pendeteksian topik pada dokumen yang sangat besar sulit dilakukan secara manual sehingga dibutuhkan metode otomatis. Masalah pendeteksian topik secara otomatis dikenal dengan istilah topic detection and tracking TDT . Suatu metode alternatif TDT untuk masalah pendeteksian topik adalah fuzzy C-means FCM. Pada metode fuzzy C-means, umumnya pusat cluster ditentukan secara acak atau inialisasi random. Namun, terkait dengan masalah dimensi yang tinggi pada inialisasi random akan menyebabkan algoritma konvergen ke satu pusat. Sehingga, topik-topik yang dihasilkan antara satu dengan yang lainnya sama. Untuk itu, diperlukan metode untuk membuat inialisasi yang dapat mengatasi masalah tersebut. Salah satu metode inialisasi yang akan dikembangkan pada penelitian ini adalah metode Singular Value Decomposition SVD . Hasil simulasi menunjukkan bahwa metode inialisasi dapat mengatasi permasalahan fuzzy C-means pada data dimensi yang tinggi sehingga topik-topik yang dihasilkan tidak sama terhadap satu sama lain.

ABSTRAK

Topic detection is a process used to analyze words in a collection of textual data to determine the topics of the collection. Detecting topics on a very large document is hardly done manually so that automatic methods are needed. Automatic method to detect topics in textual documents is known as Topic Detection and Tracking TDT . An alternative method of TDT for topic detection problems is fuzzy C means FCM . In the FCM method, generally the cluster center is random initialization. However, related to the problem of high dimensional random initialization causes the algorithm to converge to one center, it means that all generated topics are similar. For that, a method is needed to create an initialization that resolves the problem. One of the initialization methods that will be developed in this research is Singular Value Decomposition SVD method. The simulation results show that the SVD initialization method can overcome the fuzzy C means problem in the high dimension data so that the resulting topics are not equal to each other.