

Analisis metode inisialisasi pada algoritma Fuzzy C-Means berbasis singular value decomposition untuk pendeteksian topik = Analysis of initialization methods on a Fuzzy C-Means algorithm based on singular value decomposition for topical detection / Ichsan Mursidah

Ichsan Mursidah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20455981&lokasi=lokal>

Abstrak

ABSTRAK

Pendeteksian topik adalah proses untuk menemukan topik atau pokok pembahasan utama dalam suatu kumpulan dokumen. Untuk data yang besar, pendeteksian topik dengan manual sulit atau bahkan tidak mungkin dilakukan. Sehingga, dibutuhkan metode otomatis yang dikenal dengan istilah Topic Detection and Tracking (TDT). Pada penelitian ini metode TDT yang digunakan untuk masalah pendeteksian topik adalah fuzzy C-means (FCM). FCM bekerja cukup baik pada dimensi data yang rendah, tetapi gagal pada dimensi data yang tinggi. Pada metode fuzzy c-means umumnya dilakukan inisialisasi random yang menyebabkan data konvergen ke satu pusat (centre of gravity) sehingga topik-topik yang dihasilkan antara satu dengan yang lainnya sama. Untuk mengatasi masalah tersebut dibutuhkan inisialisasi yang tidak random, yaitu dengan menggunakan inisialisasi berbasis singular value decomposition (SVD). Hasil akurasi dari metode ini menunjukkan adanya peningkatan lebih baik dibandingkan dengan metode FCM dengan inisialisasi random. Dengan nilai akurasi terbaik untuk FA Cup adalah 0,923, untuk US Elections adalah 0,661 dan untuk Super Tuesday adalah 0,727.

ABSTRACT

Topic detection is the process of finding the main topic or topic in a document. For large data, manual topic detection is difficult or even impossible. Thus, it takes an automatic method known as Topic Detection and Tracking (TDT). In this research the TDT method used for topic detection problem is fuzzy C-means (FCM). FCM works reasonably well on low data dimensions, but fails on high data dimensions. In the method of fuzzy c-means is generally done random initialization that causes data convergent to one center (center of gravity) so that the topics generated from one another are equal. To solve this problem requires non-random initialization, ie by using a singular value decomposition (SVD) based initialization. The accuracy of this method shows a better improvement compared to the FCM method with random initialization. With the best accuracy value for the FA Cup is 0.923, for US Elections is 0.661 and for Super Tuesday is 0.727.