

Analisis metode inialisasi pada algoritma eigenspace based fuzzy c-means untuk pendeteksian topik berita online Indonesia = Analysis of initialization methods on eigenspace based fuzzy c-means algorithm for Indonesian online news topic detection

Raden Trivan Sutrisman, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20477107&lokasi=lokal>

Abstrak

ABSTRAK

Perkembangan berita online di Indonesia saat ini sudah semakin meningkat sehingga kebutuhan dalam melakukan analisis data berita sangat diperlukan untuk mendapatkan intisari informasi yang akurat dan cepat. Topik merupakan komponen dasar yang sering digunakan untuk menganalisis data dalam bentuk teks seperti berita. Dengan menggunakan pemodelan topik, dapat dilakukan pendeteksian topik secara otomatis pada koleksi dokumen berita yang sangat besar dan sulit dilakukan secara manual oleh manusia. Salah satu pemodelan topik yang dapat digunakan adalah metode clustering menggunakan Eigenspace Based Fuzzy C-Means (EFCM). Metode EFCM pada umumnya menggunakan inialisasi random. Pada penelitian ini akan diimplementasikan metode inialisasi menggunakan Non-Negative Double Singular Value Decomposition (NNDSVD) dan Fuzzy C-Means++ (FCM++) sebagai alternatif metode inialisasi pada algoritma EFCM. Hasil simulasi menggunakan inialisasi NNDSVD dan FCM++ menunjukkan nilai akurasi yang lebih baik dalam hal tingkat interpretabilitas topik daripada metode random.

ABSTRACT

The rapid increasing of online news in Indonesia creates the need for news analysis to obtain information as fast as possible. Topics are basic components that are often used to analyze data in the textual forms, such as the news article. By using topic modeling, topics can be detected automatically on large news documents which are difficult to perform manually. One of the topic modeling that can be used is the clustering-based method, i.e., Eigenspace-based Fuzzy C-Means (EFCM). The common initialization method of EFCM is random. In this research, Non-Negative Double Singular Value Decomposition (NNDSVD) and Fuzzy C-Means++ (FCM++) will be used as initialization methods of EFCM. The simulations show that the NNDSVD and FCM++ methods gives better accuracies in term of interpretability score than the random method.