

Fuzzy c-means pada ruang eigen untuk pendeteksian topik = Fuzzy c means in eigen space for topic detection / Triyana Muliawati

Triyana Muliawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20433793&lokasi=lokal>

Abstrak

ABSTRAK

Seiring perkembangan teknologi informasi dan komunikasi, pemenuhan kebutuhan informasi dapat diperoleh melalui media sosial, seperti Twitter. Banyaknya pengguna internet telah memicu aliran data yang sangat besar dan cepat, sehingga membuat analisis secara manual sulit atau bahkan tidak mungkin dilakukan. Metode otomatis diperlukan untuk menganalisis data tersebut yang salah satunya yaitu dengan topic detection and tracking (TDT). Suatu metode alternatif lain dari TDT untuk masalah pendeteksian topik selain latent dirichlet allocation (LDA) adalah fuzzy clustering dengan menggunakan algoritma fuzzy Cmeans (FCM). FCM pada pendeteksian topik dapat memenuhi asumsi bahwa suatu dokumen pada Twitter dapat terdiri dari beberapa topik. FCM bekerja cukup baik di dimensi data yang rendah, akan tetapi gagal dalam dimensi data yang tinggi. Oleh karena itu, dibutuhkan suatu metode untuk mereduksi dimensi ruang eigen yang tinggi ke dimensi yang lebih rendah. Salah satu metodenya adalah singular value decomposition (SVD) dengan menggunakan truncated SVD. Pada penelitian ini, dilakukan prosestruncated SVD kemudian FCM yang dinamakan fuzzy C-means pada ruang eigen (Eigen FCM). Hasil akurasi dari metode ini menunjukkan peningkatan lebih baik dibandingkan FCM dan LDA pada pendeteksian topik.

ABSTRACT

As the information and communication technology developed, the fulfillment of information can be obtained through social media, like Twitter. The enormous number of internet users has triggered fast and large data flow, thus making the analysis manually is difficult, or even impossible. The automated methods for data analysis is needed now, one of which is the topic detection and tracking (TDT). An alternative method other than TDT for topic detection problem other than latent dirichlet allocation (LDA) is a fuzzy clustering algorithms using fuzzy C-means (FCM). FCM in topic detection meet the assumption that a document on Twitter can consists of several topics. FCM works pretty well in low-dimensional data, but fail in high-dimensional data. Therefore, we need a method to reduce the dimension of the high-dimensional eigenspace into lower dimension. One method to do that is the singular value decomposition (SVD) using truncated SVD. This paper carried out the truncated SVD process then FCM called fuzzy C-means on

the eigenspace (Eigen FCM). The results of the accuracy of this method shows an increase is better than FCM and LDA on topic detection.