

Analisis akurasi fuzzy C-means berbasis kernel untuk pendeteksian topik pada data Twitter = Accuracy analysis kernelized fuzzy C-means for topic detection on Twitter

Praditya Nugraha, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494614&lokasi=lokal>

Abstrak

Salah satu metode otomatis untuk analisis data tekstual adalah deteksi topik. Fuzzy C- Means di Ruang Eigen (EFCM) adalah metode berbasis soft clustering untuk pendetek- sian topik. Pada Algoritme EFCM adanya reduksi dimensi data awal menjadi lebih kecil. Namun, proses reduksi itu dapat menghilangkan beberapa fitur penting dari data tekstual. Sehingga, akurasi dapat berkurang. Dalam mengatasi hilangnya fitur penting digunakan bantuan Kernelisasi Fuzzy C-Means di Ruang Eigen sehingga proses clustering dapat di- lakukan dalam ruang dimensi yang lebih tinggi. Dalam penelitian ini akan dicek akurasi dari metode EFCM dan KEFCM dan perbandingannya dengan metode standar seperti Latent Dirichlet Allocation (LDA) dan Nonnegative Matrix Factorization (NMF) dalam masalah pendeteksian topik. Simulasi menunjukkan bahwa KEFCM memberikan akurasi yang lebih baik dalam menemukan topik daripada metode standar LDA dan EFCM namun tidak lebih baik dari NMF untuk masalah mendeteksi topik berita online di Twitter.

<hr>

One of automated methods for textual data analysis is topic detection. Fuzzy C-Means in Eigenspace (EFCM) is a soft clustering-based method for topic detection. In, EFCM Algorithm there is a step to transform high dimensional textual data into lower dimensional data. However, that transformation process may eliminate some important features from the textual data. Therefore, the accuracy may be reduced. To overcome in losing important features Kernelized Fuzzy C-Means in Eigenspace (KEFCM) is needed, so that clustering process can be done in higher dimensional space. In this study the accuracy of EFCM and KEFCM will be evaluated and these methods will be compared by any standard method such as Latent Dirichlet Allocation (LDA) and Nonnegative Matrix Factorization (NMF) for topic detection problem. Simulations show that KEFCM gives better accuracy to find topics than LDA and EFCM method. However, these methods fail to give better results than NMF for the problem of sensing trending topic in online news in Twitter.