

Kernelisasi metode fuzzy C-means berbasis ruang eigen untuk pendeteksian topik pada Twitter = Kernelized eigenspace based fuzzy C-means for topic detection in Twitter

Yudho Prakoso, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20474317&lokasi=lokal>

Abstrak

Salah satu metode otomatis untuk analisis data tekstual adalah deteksi topik. Eigenspace-based Fuzzy C-Means EFCM adalah metode berbasis soft clustering untuk pendeteksian topik. Pertama, EFCM menggunakan dekomposisi nilai tunggal terpotong untuk mengubah data tekstual dimensi tinggi menjadi data berdimensi rendah. Selanjutnya, proses pengelompokan dilakukan dalam ruang dimensi yang lebih kecil. Namun, proses transformasi itu dapat menghilangkan beberapa fitur penting dari data tekstual. Karena itu, akurasi dapat berkurang.

Dalam penelitian ini digunakan kernel trick untuk mengatasi kelemahan tersebut sehingga proses clustering dapat dilakukan dalam ruang dimensi yang lebih tinggi. Simulasi menunjukkan bahwa pendekatan ini memberikan akurasi yang lebih baik dalam menemukan topik daripada EFCM untuk masalah mendeteksi topik di Twitter.

.....One of automated methods for textual data analysis is topic detection. Eigenspace based fuzzy c means EFCM is a soft clustering based method for topic detection. Firstly, EFCM use truncated singular value decomposition to transform high dimensional textual data to low dimensional data. Next, the clustering process is conducted in the smaller dimensional space. However, that transformation process may eliminate some important features from the textual data. Therefore, the accuracy may be reduced.

In this study used kernel trick to overcome that weakness so that the clustering process is performed in a higher dimensional space. Simulations show that this approach gives better accuracies in term of topic recall than EFCM for the problem of sensing trending topic in Twitter.