

Analisis metode penentuan kata anchor menggunakan singular value decomposition berbasis variansi kata pada topik untuk separable nonnegatif matrix factorization = Analysis of anchor words selection method using singular value decomposition based on word variance on topic for separable nonnegative matrix factorization / Naufal Khairil Imami

Naufal Khairil Imami, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493847&lokasi=lokal>

Abstrak

ABSTRAK

Deteksi topik adalah proses yang digunakan untuk menganalisis kata-kata dalam kumpulan data tekstual untuk ditentukan topik dalam koleksi, bagaimana mereka saling berhubungan, dan bagaimana topik ini berubah dari waktu ke waktu. Salah satu metode yang digunakan untuk mendeteksi topik adalah Nonnegative Matrix

Metode Factorization (NMF) berdasarkan metode langsung atau disebut Separable Nonnegative Matriks Faktorisasi (SNMF). Dalam penelitian ini, tweet data diambil dari akun berita nasional pada aplikasi Twitter, dan topik terdeteksi menggunakan metode SNMF. Ada tiga tahap dalam metode SNMF, yaitu, menghasilkan kata co-kejadian

matriks, menentukan kata jangkar, dan memulihkan untuk mendapatkan matriks topik-kata. Penentuan dari kata anchor dilakukan dengan menggunakan tiga metode berbeda, yaitu Convex Metode berbasis lambung, metode SVD berdasarkan kata maksimum pada topik, dan SVD

metode berdasarkan perbedaan kata pada topik. Dalam deteksi topik, jumlah topik diproduksi bervariasi tergantung pada jumlah kata jangkar yang dihasilkan. Setelah mendapatkan topik, akurasi dihitung menggunakan unit topik word2vec. Hasil yang diperoleh menggunakan

tweet data akun berita nasional menunjukkan bahwa metode SVD berdasarkan perbedaan kata pada topik memiliki evaluasi topik yang buruk dibandingkan dengan dua topik lainnya.

ABSTRACT

Topic detection is the process used to analyze words in a textual data set to determine the topics in a collection, how they are related, and how these topics change over time. One method used to change the topic is the Nonnegative Matrix The Factorization (NMF) method is based on the direct method or called Separable Nonnegative Factor Matrix (SNMF). In this study, tweet data is taken from national news accounts on the Twitter application, and topics taken using the SNMF method. There are three methods in SNMF, namely, generating co-occurrence words matrix, determine anchor words, and recover to get a topic-word matrix. Determination of the word anchor is done using three different methods, namely Convex the stomach-based method, the SVD method based on the maximum words on the topic, and SVD method based on differences in words on the topic. In topic detection, the number of topics created varies depending on the number of anchor words produced. After getting the topic, it is calculated using the word2vec topic

unit. The results obtained using National news account data tweets show the SVD method based on word differences on the topic has a worse evaluation topic compared to the other two topics.