ALIYA NUGUMANOVA

D. Serikbayev East Kazakhstan State Technical University, Kazakhstan

YERZHAN BAIBURIN

D. Serikbayev East Kazakhstan State Technical University, Kazakhstan

USING SVD FOR TEXT CLASSIFICATION

Abstract:

Singular value decomposition (SVD) is a way to decompose a matrix into some successive approximation. This decomposition can reveal internal structure of the matrix. The method is very useful for text mining. Usually co-occurrence matrix (terms-by-documents matrix) defined over a large corpus of text documents contains a lot of noise. Singular value decomposition allows approximation of the co-occurrence matrix and thereby can reveal internal (latent) structure of text corpus. It decreases information noise, removes the unnecessary (random) links between terms and increases the value of important information. In this paper we apply singular value decomposition to improve text classification. We build co-occurrence matrix and then approximate it by SVD. Obtained matrix is very useful for creating new feature space. We prove our approach by experiments on Reuters Text Classification Collection.