

Documents Classification by Using Bi-Dimensional Probability Model

Myat Su Oo

Computer University, Pyay
myatsuoo.mso@gmail.com

Abstract

This paper proposes to implement Bi-Dimensional Probabilistic Model (BDPM). This model is to find a bi-dimensional representation of textual documents for the problem of text categorization. The main idea is to consider the importance of a word. In this model, terms are seen as disjoint events, and terms and categories are related to each other. Terms are measured with their presence and expressiveness. The presence and expressiveness of a term is defined as the peculiarity of that term. BDPM is document classification model which can represent a document in two dimensional schemes. It can also be used for both visualization and classification of textual documents. To experiment the system, Reuters-21578 newswire dataset is used. This bi-dimensional model has the advantage: no explicit need for feature selection, to reduce dimensionality, since documents are represented by only two dimensions.