Classifying Textual Documents Using Two Dimensional Probabilistic Model

Wai Me Me Than, Nan Saing Moon Kham
University of Computer Studies, Yangon
waiwai.meethan@gmail.com, moonkhamucsy@gmail.com

Abstract

This paper presents the probabilistic model named Two-dimensional Probabilistic Model (2DPM). In this model, terms are seen as disjoin events, and terms and categories are realeated to each other. Since the documents are represented as the union of terms, disjoint event, document and categories are also rreleated. Terms are measured with their presence and expressiveness. The presentce and expressivencess of a term is defined as the peculiarity of that term. A document is defined as set of terms and it also has presence and expressiveness for a category. So, the 2DPM model defines a direct relationship between the probability of a document given a category of interest and a point on atwo-dimensional space. With the points, entire collections of documents are graphed on a Cartesian plane and documents are classifie directly on the two-dimensional representation. To experiment the system, Reuters-21578 newswire dataset is used for text classification.