# Bag-of-Discriminative-Words (BoDW) Representation via Topic Modeling

**Abstract:**

Many of words in a given document either deliver facts (objective) or express opinions (subjective) respectively depending on the topics they are involved in. For example, given a bunch of documents, the word “bug” assigned to the topic “order Hemiptera” apparently remarks one object (i.e., one kind of insects), while the same word assigned to the topic “software” probably conveys a negative opinion. Motivated by the intuitive assumption that different words have varying degrees of discriminative power in delivering the objective sense or the subjective sense with respect to their assigned topics, a model named as discriminatively objective-subjective LDA (dosLDA) is proposed in this paper. The essential idea underlying the proposed dosLDA is that a pair of objective and subjective selection variables are explicitly employed to encode the interplay between topics and discriminative power for the words in documents in a supervised manner. As a result, each document is appropriately represented as “bag-of-discriminative-words” (BoDW). The experiments reported on documents and images demonstrate that dosLDA not only performs competitively over traditional approaches in terms of topic modeling and document classification, but also has the ability to discern the discriminative power of each word in terms of its objective or subjective sense with respect to its assigned topic.

**Index Terms**—Topic Modeling, Latent Dirichlet Allocation, Objective and Subjective Classification, Bag-of-Discriminative-Words Representation..