Abstract

Language Model smoothing is an imperative technology which deals with unseen test data by re-evaluating some zero-probability n-grams and assign them bare minimum non-zero values. There is an assortment of smoothing techniques employed to trim down tiny amount of probability from the probable grams and share out to zero probable grams within a Language Model. Kneser Ney and Latent Dirichlet Allocation algorithm are two probable techniques used for proficient smoothing. In this paper, a scheme is proposed for effective smoothing by combining Kneser Ney and Latent Dirichlet Allocation approaches. Moreover, another scheme is proposed to measure the reliability of a Language Model and determine the association between entropy and perplexity. These schemes are demonstrated by appropriate examples.

References


Keywords

Smoothing, Pruning, Entropy, Perplexity, Data Sparsity, Statistical Control, Information Retrieval.