CS 6320 Natural Language Processing
Homework 2 N-gram Language Model

I. Language model application in

a. Machine Translation:

Machine translation is a process where text or speech of one natural language is translated to another. At the basic level it substitutes simple words in one natural language to another language. Problem arises when the translator has to translate a sentence or phrase. This is solved by using corpus and statistical techniques. The language model (LM) is responsible for finding probability that the produced sentence is grammatically and semantically correct without looking at the source sentence and plays important role in translating into morphologically rich languages.

b. Information retrieval:

The goal of an information retrieval (IR) system is to rank documents optimally given a query so that relevant documents would be ranked above nonrelevant ones. Thus, not surprisingly, seeking an optimal scoring function (retrieval function) has always been a major research challenge in information retrieval. A retrieval function is based on a retrieval model, which formalizes the notion of relevance and enables us to derive a retrieval function that can be computed to score and rank documents. Probabilistic models treat the process of document retrieval as a probabilistic inference. Similarities are computed as probabilities that a document is relevant for a given query. Probabilistic theorems like the Bayes' theorem are often used in these models.

II. Report:

Perplexity: 57.2505598244

Number of words in test file: 3225

N1: 141875

N2: 35133