THE UNIVERSITY OF TEXAS
AT AUSTIN

CS388 NATURAL LANGUAGE PROCESSING

Programming Assignment 01

Edited by LaTeX

Department of Computer Science

STUDENT

Jimmy Lin

xl5224

COURSE COORDINATOR

Prof. Raymond J. Mooney

UNIQUE NUMBER

51360

RELEASE DATE

Feb. 01 2016

DUE DATE

Feb. 18 2016

TIME SPENT

10 hours

February 14, 2016

1. Introduction

In this experiment, we will examine the predictive powers of Bigram models of various flavors. They are respectively Forward Bigram Model, Backward Bigram Model, and Bidirectional Bigram Model.

Forward Bigram Model (FBM) is a language generation model based on a natural linguistic production methodology (left-to-right context). Differing from FBM, a Backward Bigram Model (BBM) is opposite to the natural way of generates language, which in terms of right-to-left context. Both FBM and BBM will take advantage of smoothing techniques through combining the probability estimates of Unigram and Bigram models. The smoothing ratios for both FBM and BBM are fixed at 0.1 and 0.9. The Bidirectional Bigram Model (BiBM) can be seen as a sort of combination of FBM and BBM. Its decision of certain token at a particular position comes from the probability interporlated from that of FBM and BBM.

The experimented datasets are *atis, wsj, brown*, all of which are sourced from Linguistic Data Consortium (LDC)'s Penn Treebank collection. By default, every dataset will be split into training set and testing set with a ratio of $9:1$. That is, nine of ten sentences will be used for training Bigram Models and the leftover one will be employed for performance evaluation. For the fairness of the comparisons, the training set and testing set are the same for all models.

2. Comparative Results

Comparisons of word perplexity between various Bigram Models can be found at Table 1.

Table 1: Word Perplexity Comparisons between Various Bigram Models

Datasets	atis		wsj		brown	
Bigram Models	training	testing	training	testing	training	testing
Forward	10.59	24.05	88.89	275.12	113.36	310.67
Backward	9.11	21.71	71.04	205.75	89.30	218.79
Bidirectional	9.37	17.54	48.60	130.48	63.74	172.75

Note that the Bidirectional Model employs even interporlation from Forward Model and Backward Model in the normal probability space (not log probability space).

From the Table 1, it can be observed that Backward Model and Bidirectional Model reach lower word perplexity (which is better) than Forward Model in any training/testing set derived from three experimental datasets. Furthermore, Bidirectional Model outperforms Forward Model in any scenario except the case of the training set from atis. This could be out of the overfitting phenomenon.

3. Discussions

In this section, we are intended to respond to questions of minimal requirements.

(a) How does the "Word Perplexity" of the backward bigram model (for both training and test data) compare to the normal model? Discuss the reasons for any diferences or similarities found.

Backward v.s. Forward In all scenarios, Backward Bigram Model outperforms Forward Bigram Models. This evidence could serve as a powerful support for the idea that the backward (right-to-left) linguistic generation modelling can better explain the natural language generation process. If *Bidirectional v.s. Backward* is not considered, no indicator reveals the existence of overfitting for Backward Bigram Model according to the presented word perplexity value for training and testing sets.

(b) How does the "Word Perplexity" of the bidirectional model (for both training and test data) compare to both the backward model and the normal model? Discuss the reasons for any diferences or similarities found.

Bidirectional v.s. Forward Bidirectional Bigram Model gets lower word perplexity values than Forward Bigram Models in all ways. This shows that Bidirectional Bigram Model can also serve as a strong modelling for the natural language generation process.

Bidirectional v.s. Backward This pair of comparison is more interesting. As we can see from Table 1, the bidirectional modelling works better than Backward modelling except the training set of atis. This phenomenon can be explained by the overfitting theory: Backward method overfits the training set of atis, such that it works relatively worse in atis' testing set. The good thing is Bidirectional Bigram Model successfully escape from just-so-so decision of BBM and proposes a better explanation for testing sentences.

4. Critical/Practical Issues

In this section, we discuss a few practical issues that we might suffer from in the experiments.

Shuffling We are supposed to shuffle each experimental dataset before spliting it into training set and testing set. This is because a training/testing split without shuffling might cause the sentences of different styles unevenly distributed to training and testing set and thus makes comparison between techniques undistinguishable. But in this asignment, the reported results above are not preprocessed with shuffling for the purpose of comparing ours to others' implementation in a more fair way. That is, we assume that the provided corpus have been shuffled in advance.

Parameter Tuning Although both backward method and bidirectional method derives better results under experimented settings, it is still hard to conclude that backward and bidirectional methods are more powerful than the forward method. It is possible that the forward model could reach the lowest word perplexity over all other models in some particular contexts (e.g. training/testing split, unigram-bigram interpolation). Thus, without a more comprehensive experiments, it is unreasonable to conclude that the FBM is less powerful thant BBM and BiBM. But at least, we can say that *under the experimented/common settings*, Bidirectional Bigram Models with even interporlation outperforms Forward Bigram Model and Backward Bigram Model.