Assignment - 2 Report

I used the cross validation using 5 folds to compute accuracy of training model. (So It takes 80% as training data and 20% as dev data).Initially,I calculated wordset , tagset , tagToTag transition count, given tag counts, word count for given tag using training data. I calculated the baseline accuracy for dev data using the word count for given tag.(During Baseline the tag which appears maximum no of times for given word is returned)

I implemented transition , observation matrix using the above calculated counts. I calculated transition matrix using kneser ney smoothing(took d value as 0.75 for absolute discounting). Later I calculated the accuracy using the viterbi algorithm on dev data using transition matrix and observation matrix..

While performing viterbi , if I encountered an unknown word while computing value for the unknown word and the tag cell in the viterbi matrix , I took the value of the word from observation matrix which has least value for the tag for which we are computing above . While computing cross validation using 5 folds , the variance between the 5 accuracies is very less. So the above assumption i made for unknown words should work well as the variance across 5 folds is less.