Kazi Sabrina Sonnet, Felix Schwock

Milestone B Report

In the project we aim to develop a restaurant review system using HMMs and Part-of-Speech tagging techniques. To do so, we first generate a tagged dataset (corpus) from a list of sample reviews collected in an csv file. After tagging is performed, each word of the sample reviews is assigned to one of the labels GOOD, NEUTRAL, or BAD. Afterwards the corpus is used to build the HMM and compute the transition an emission probabilities. Using the results of the Viterbi and forward algorithm, a new restaurant review can then be evaluated to provide a score for the restaurant. So far, we have started building the tagged dataset and are currently in the process of tagging more reviews to get a large training corpus. Furthermore, Kazi has started building the HMM based on restaurant reviews we have tagged so far. Felix has implemented the Viterbi algorithm and tested its functionality on a very simple sample HMM (sample HMM is taken from this YouTube video https://www.youtube.com/watch?v=kqSzLo9fenk&t=349s).   The next steps are further building the training corpus and creating the final HMM, implementing the forward algorithm, and applying both algorithms on the final HMM.