School of Information Technology and Engineering

Probabilistic Graphical Models Assignment Four

Submitted By - Abenezer Seifu

Id - GTS/2555/15

Report on Conditional Random Fields and Hidden Markov Model for Part-of-Speech Tagging

Data

I use the Universal Tagset dataset for both conditional random forest and Hidden markov models.The Universal Tagset is a standardized set of POS tags that simplifies cross-lingual and cross-domain POS tagging tasks.

Conditional random Fields

Conditional Random Fields are a type of statistical modeling technique used in various Natural Language Processing tasks , including part-of-speech tagging . in this report i will delve into an overview of the code implementation of CRF-based POS tagging using the Universal tagset dataset with NLTK.

Dataset and Preprocessing

In the code implementation , i used NLK to load the universal tagset dataset and split it into a training set and a test set

Training and Evaluation

I used the NLTK's CRFTagger to train a CRF-based POS tagger on the training data. The tagger learns the relationships between words and their associated POS tags. After training , I evaluated the tagger's accuracy on a separate test set to assess its performance.

Result

The CRF-based Pos tagger achieved an accuracy of 95.24% on the test data.

Hidden Markov Model

Hidden Markov Models (HMMs) are a class of probabilistic graphical models widely used in Natural Language Processing (NLP) for various tasks, including part-of-speech (POS) tagging.

Dataset and Preprocessing

I used the same dataset and procedure to prepare the data for the HMM model as I used for the CRF model. I used NLTK to load the universal Tagset dataset and split it into a training set and a test set.

Training and Evaluation

We utilized MLK's HMM-based tagger to train an HMM model on the training data. During training, the model learns the transition probability between POS tags and the emission probabilities of words given the POS tags. After training, we evaluated the model's performance on a separate test set to measure its accuracy.

Result

The HMM-based POS tagger achieved an accuracy of 93.7% on the test data