In The Name Of God

Sentiment Classifier Based On Bayesian Network

Group Members:

MohammadAli Khosorabadi
Parham Heayati

Step 1 : Tokenizing

First of all, tweets must be tokenized. we used nltk library to tokrnize and removing stopwords and lemmatize tweets.
we also used some other methods and library but we got our best result with current method.

Steo 2: Loading Data

we use pandas library to load data and then apply preprocess operation on all tweets of dataset.

Step 3: Training Our Model

We implemented Bayesian Network Model Based on existing formulas.

Calculate_Prior(self, label):

Purpose: Calculates the log prior probability of a class.

- **Input:**
  - `label`: The class label for which the prior probability is calculated.
- **Output:** Returns the log prior probability.

Calculate_Likelihood(self, word, label):

Purpose: Calculates the log likelihood of a word given a class.

- **Inputs:**
  - `word`: The word for which likelihood is calculated.
  - `label`: The class label for which likelihood is calculated.
- **Output:** Returns the log likelihood probability.

Classify(self, features):

Purpose: Predicts the class for a given set of features (words of a tweet).

- **Input:**
  - `features`: List of words in a tweet.
- **Output:** Returns the predicted class.
- **Steps:**
  1. Calculates log prior probabilities for each class.
  2. Calculates log likelihood sums for each class based on the features.
  3. Prints log priors, log likelihood sums, and predictions.
  4. Adjusts predictions based on the calculated `word_lambda_sum`.
  5. Returns the class with the maximum predicted value.

Training Time : around **1 minute**

Step 4: Testing On Eval Dataset

We test our model on eval dataset and obtain overall accuracy and more detailed metrics like precision and recall and accuracy for each class. The results are in a file named `evaluation_result.txt`.

Our overall accuracy on eval dataset is 63.94% which is reasonable according to dataset.

Step 5: Testing On Test Dataset:
finally, we apply our model on test dataset and obtain it's prediction and store it in `result.txt`.