

Adapting HMM Approaches for Sentiment Analysis on the Steam Game Review Dataset

<https://github.com/pmukneam/math156-HMM>

Ninja, Ton

March 9, 2023

1 Introduction

Sentiment analysis is the task of identifying and classifying subjective information in text data. In recent years, Hidden Markov Model (HMM) algorithms have become more popular for sentiment analysis due to their ability to model sequential data. In this project, we aim to adapt HMM approaches from the literature for sentiment analysis on the Steam game review dataset and explore the suitability of these approaches for this specific dataset.

2 Objectives

Our objectives are as follows:

- Adapt HMM approaches from the literature for sentiment analysis on the Steam game review dataset.
- Gradually increase the amount of training data, i.e., we need to make sure that it satisfies the necessary assumption to the algorithm, and evaluate the performance of the adapted HMM approaches.
- Draw conclusions regarding the suitability of HMM approaches for sentiment analysis on the Steam game review dataset.

3 Methods

We will carry out the project in the following steps:

1. **Data collection and preprocessing:** We will obtain the Steam game review dataset and preprocess it by cleaning, tokenizing, and labeling the data.

Steam Reviews: <https://www.kaggle.com/datasets/andrewmvd/steam-reviews> by Larxel

2. **Adaptation of HMM approaches:** We will adapt HMM approaches from the literature for sentiment analysis on the Steam game review dataset. Specifically, we will use the following papers as references:
 - A Simple Proposal for Sentiment Analysis on Movies Reviews with Hidden Markov Models by Billy Peralta, Victor Tirapegui, Christian Pieringer, and Luis Caro.
 - Sentiment Analysis of Customer Reviews Based on Hidden Markov Model by Swati Soni and Aakanssha Sharaff.
 - Sentiment Analysis on the Online Reviews Based on Hidden Markov Model by Xiaoyi Zhao and Yukio Ohsawa.
 - Hidden Markov Models for Sentiment Analysis in Social Media by Isidoros Perikos, Spyridon Kardakis, Michael Paraskevas, and Ioannis Hatzilygeroudis.
3. **Training and evaluation:** We will start with a small quantity of training data and gradually increase it as things go smoothly. We will evaluate the performance of the adapted HMM approaches using standard evaluation metrics such as accuracy, precision, recall, and F1-score.
4. **Result analysis and comparison:** We will analyze the results of the adapted HMM approaches and draw conclusions regarding their suitability for sentiment analysis on the Steam game review dataset.

4 Expected Outcomes

We expect the following outcomes from the project:

- Successful adaptation of HMM approaches from the literature for sentiment analysis on the Steam game review dataset.
- Evaluation of the performance of the adapted HMM approaches with gradually increasing amounts of training data.
- Analysis of the results and conclusions regarding the suitability of HMM approaches for sentiment analysis on the Steam game review dataset.
- A deeper understanding of HMM approaches in sentiment analysis.

5 Conclusion

This project aims to advance the understanding of sentiment analysis and HMM approaches by adapting HMM approaches from the literature for sentiment analysis on the Steam