Movie Review Translation and Sentiment Analysis

Syed Ali, Zach Friedman, Rebecca Moore, Arefa Patwary

Github   Dataset

Description

Movies are something that people love across cultures and languages. By translating movie reviews between English and Spanish, we will allow more people to connect over their mutual love (or hatred) of a movie. We will perform neural machine translation to convert Spanish reviews to English, and vice versa. We will then use sentiment analysis to characterize the reviews as either positive or negative.

Goals & Objectives

- **Motivation**

  Movie reviews are a great way for viewers to explore new genres and participate in a global creative community. Film is a form of storytelling, and its message may take on different meanings when viewed through different cultural lenses. "While we don't all share the same experiences, we do all share a similar array of emotions, fears and hopes," (Hennebury, 2020). But language barriers are a limiting factor. Critically acclaimed foreign films may be missed entirely because of them. Machine translation offers an efficient way to translate digital media and the discussion around it, while sentiment analysis can help gauge reception abroad. Our goal is to facilitate and encourage inclusion in online discussions, in this case about movies. While niche, we hope to demonstrate that the techniques we use can be applied to achieve that goal in a wide range of digital environments.

- **Significance**

  Movies are a very effective means of shaping culture and a major backbone of cultural exchange. With global box office revenue "[weighing] in at $26 billion in 2022," (Frater, 2023), the medium offers more than just entertainment. The lessons garnered from these audiovisual works of art are often thought provoking, memorable, and transcend beyond cultural boundaries. In an increasingly connected world, a global audience now has the ability to share their thoughts about a piece, and that feedback can be used as a springboard for creating more refined media. By translating reviews from Spanish to English and vice versa, we are broadening the connection between consumers and cultures and allowing information to be used and shared by people that it was previously inaccessible to.
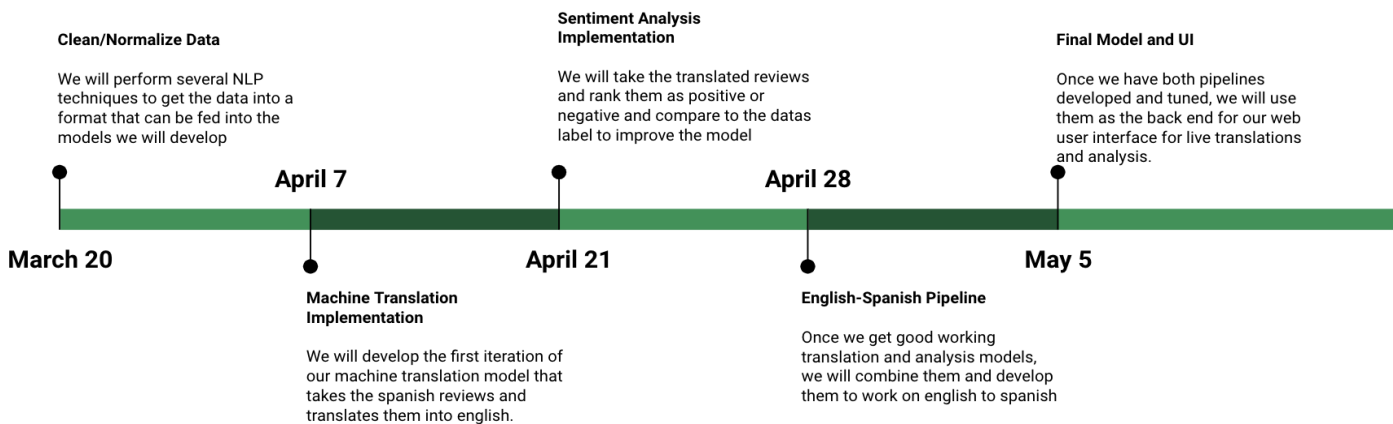
- **Objectives**
  - Implement a machine translation system to…
    - Translate Spanish movie reviews to English with high accuracy.
    - Translate English movie reviews to Spanish with high accuracy.
  - Implement a sentiment analysis system that indicates whether a review is positive or negative.
  - Evaluate the machine translation system by scoring how much of the original meaning of the review is kept. This will be done using the C-DAC GIST MT tool which is designed to evaluate machine translations. Adjust the model accordingly.
  - Evaluate the sentiment analysis system by comparing the predicted score (positive or negative) to the labeled score and observe a basic confusion matrix of true positives, true negatives, false positives and false negatives to evaluate what errors we may have in the model.

- **Features**
  - Take Spanish movie reviews and translate to English
  - Take English movie reviews and translate to Spanish
  - Mark reviews as positive or negative in Spanish and English
  - Allow users to input their own review to be translated, and have it analyzed as positive or negative
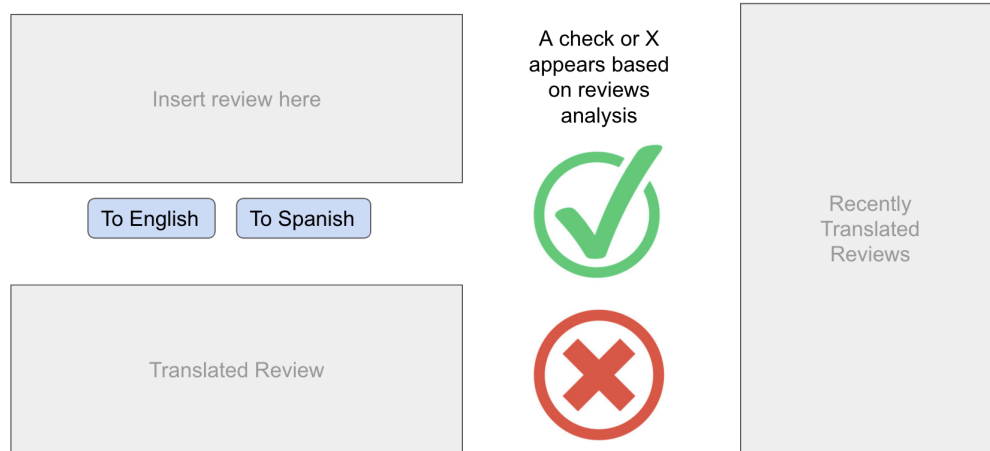
- **Workflow and Deadline**

Clean/Normalize Data

We will perform several NLP techniques to get the data into a format that can be fed into the models we will develop

Sentiment Analysis Implementation

We will take the translated reviews and rank them as positive or negative and compare to the datas label to improve the model

Final Model and UI

Once we have both pipelines developed and tuned, we will use them as the back end for our web user interface for live translations and analysis.

April 7 **April 28**

March 20 **April 21** **May 5**

Machine Translation Implementation

We will develop the first iteration of our machine translation model that takes the spanish reviews and translates them into english.

English-Spanish Pipeline

Once we get good working translation and analysis models, we will combine them and develop them to work on english to spanish

- **UI Design Mockup**
  - Text box for user to paste text
  - Button for submission of text
  - Responds with translation in appropriate language and indicates weather review is positive or negative

Movie Review Translation and Analysis

References

1. Frater, P. (2023, January 6). *Global box office notched 27% gain in 2022 to hit $26 billion total, research shows*. Variety. Retrieved February 18, 2023, from https://variety.com/2023/data/news/global-box-office-in-2022-1235480594/

2. Li, S. (2018, June 25). *Neural machine translation with python*. Towards Data Science. Retrieved February 15, 2023, from https://towardsdatascience.com/neural-machine-translation-with-python-c2f0a34f7dd

3. Q-Success. (n.d.). *Usage statistics of content languages for websites*. W3Techs. Retrieved February 17, 2023, from https://w3techs.com/technologies/overview/content_language

4. Hennebury, C. (2020, March 29). *Storytelling is not just entertainment. it's a fundamental part of being human | CBC news*. CBCnews. Retrieved February 20, 2023, from https://www.cbc.ca/news/canada/newfoundland-labrador/storytelling-is-human-1.5511027

5. *Gist MT (Machine Translation) Evaluation Tool*. Centre for Development of Advanced Computing. (n.d.). Retrieved February 20, 2023, from https://cdac.in/index.aspx?id=mlc_gist_nlp_mt#:~:text=GIST%20MT%20(Machine%20Translation)%20Evaluation%20Tool&text=C%2DDAC%20GIST%20MT%20