The paper we intended to extend from, “Genre Classification in Spanish”, utilized transformers to as part of their machine learning pipeline to automate categorization of books by a Thema-standard categorization. Part of what led to their model being better than others before them was that there were boundaries they established (i.e. their Thema-standard categorization) for real-world applications which they used to guide their models’ classifications. However, for classifying Japanese animated media, assigning labels for what genre each work belongs to is relatively unclear or multifaceted due to the diversity of content each may touch upon or contain. It is such to the extent that even the authors of this work merely categorized manga under a “Comic, manga, and graphic novel” section by just utilizing descriptions – with low success, too. Building on this prior work’s categorization of Spanish literature by genre, I wanted to utilize a variety of means, including transformers, to extend and adapt this study into the domain of anime. Specifically, this work focuses on using a multilingual BERT model and convolutional neural networks to categorize anime by media type as the boundaries for what defines each media type are a little clearer than the genres. By adapting the methods used in book categorization to this task, this study aims to demonstrate the generalizability of transformer-based classification models across domains and provide insights into best practices for applying these models to structured but imbalanced datasets.

The dataset, publicly available on Kaggle, includes information on 12,774 anime from up until the end of 2023. It features data such as user ratings, airing periods, episode counts, anime titles, page URLs, and small image URLs pointing to public cover images hosted on MyAnimeList (MAL), a crowdsourced anime database. I initially focused on shows that were not ongoing or lacked clear airing durations and episode counts. From this subset, I created a new feature, airtime in days, based on recorded months and years. Additionally, I was able to download cover images for 12,641 anime without encountering request errors.

I decided to build and compare several models to predict types of animated show. This involved creating baseline models that predicted the show type based on episode count, user score, and airtime in days. To evaluate potential linear and non-linear trends for our baseline models, we implemented logistic regression models (with a L-BFGS solver for handling larger datasets) and support vector machines, analyzing their performance across each of these 3 variables to gauge the presence of patterns. Additionally, I ran the cover images through multiple CNN models to explore whether there were any specific trends or gaps in how they predicted anime type based on visual content. Finally, I trained a classifier in PyTorch using features extracted from a multilingual BERT (bidirectional encoder representations from transformers) model to predict anime type from the titles, following some degree of text preprocessing. I chose this BERT model due to its ability to

understand/having been trained on Romaji, which is prevalent in a large portion of the titles, making it well-suited for the task.

To some extent, our baseline modeling was successful. The best baseline models for predicting anime type were logistic regression models, which were trained using episode count and airtime in days as features. The highest test accuracies achieved were about 57% for predicting media type based on episode count and about 55% for airtime in days. These results are somewhat consistent with expectations, as it is easier to linearly separate TV shows from other types like Movies and Specials, which typically have single episodes or shorter runtimes. TV shows, in contrast, tend to have full season-length runs. This inherent trend makes it easier to predict a large portion of the data based on these features with this baseline model.

When predicting media type using CNN models, we experimented with data augmentation to expand our training data and improve the model's robustness in classifying different anime types. Despite these efforts, the test accuracy remained relatively low. With data augmentation but no class weights applied, our test accuracy was approximately 42.2%. When class weighting was applied but no data augmentation was used, the test accuracy was slightly lower at 41%. The best result came from combining both data augmentation and class weighting, where the test accuracy reached about 42.5%. Upon examining the classification reports, it became clear that class weighting was necessary, as the unweighted models tended to almost entirely ignore at least one class, while overly favoring TV shows — a trend that did not entirely reflect the actual distribution in the training data. This suggested that, at least with these CNN models, there were no clear visual patterns or cues that could reliably differentiate between certain anime types.

Our final model employed the multilingual BERT transformer model. To prepare the data, we first preprocessed the titles by removing punctuation, converting the words to lowercase, and eliminating English stop-words. After preprocessing, we tokenized the titles, padded them to ensure uniform length, and converted the data into tensors, making them ready for input into the model. The model used is a pretrained version of BERT, which allowed us to leverage its language understanding capabilities for improved performance. Our best model had a test accuracy of about 50.30%. When looking at the classification report, we saw that we had better F1 scores for almost every class besides the ‘TV’ classification. Still, the ‘OVA’ class had the lowest F1 score as before. When evaluating the model on a small sample of data that were in the alternative format (usually as a Romaji or English title) of how they were represented in the dataset, we found that it was capable of correctly predicting almost all the examples presented along with an example outside the dataset.

In conclusion, this study aimed to extend the genre classification framework from Spanish literature to Japanese anime, utilizing transformer models like multilingual BERT and convolutional neural networks. We explored how these models could be applied to the structured yet imbalanced while highlighting some issues with classification. While baseline models using logistic regression and SVMs showed greater test than our other models in part potentially due to existing trends, the CNN models struggled to identify clear visual patterns, highlighting the complexity of the task. Ultimately, the multilingual BERT model provided the most promising results of the two approaches, achieving a test accuracy of 50.30%. Despite the underperformance of the models compared to the baseline models, this work demonstrates the potential of transformer models in domain adaptation. It also demonstrates a necessity for defining clearer boundaries or criteria for identifying animated works in the future. Future work with this dataset would see utilization of web-scraping techniques to extract additional data, such as synopses, to better mirror the work of the original authors of the work we extended from. We might also bolster our data through data augmentation backtranslation of the titles or utilizing alternative names provided by MAL for each show.

Citations:

J. A. Nolazco-Flores, A. V. Guerrero-Galván, C. Del-Valle-Soto and L. P. Garcia-Perera, "Genre Classification of Books on Spanish," in *IEEE Access*, vol. 11, pp. 132878-132892, 2023, doi: 10.1109/ACCESS.2023.3332997.