Analysis of Trending YouTube Videos Across Different Countries

Olivia Long, Catherine Nguyen, Annie Zou

Introduction

- YouTube is one of the most popular social media platforms, and YouTubers can heavily influence their audiences.
- We are interested in discovering insights and analyzing differences between trending videos in different countries.

Related Work

- Siersdorfer discovered strong dependencies between YouTube comment sentiment and video content. [1]
- Krishna found a correlation between trends in YouTube comment sentiment and real world events with Naive Bayes models. [2]

Regression

We used the US bag-of-words representation to predict video views. We used a Linear Regression model on an 80-20 train-test split and got a coefficient of determination R^2 on the test prediction of 0.6561.

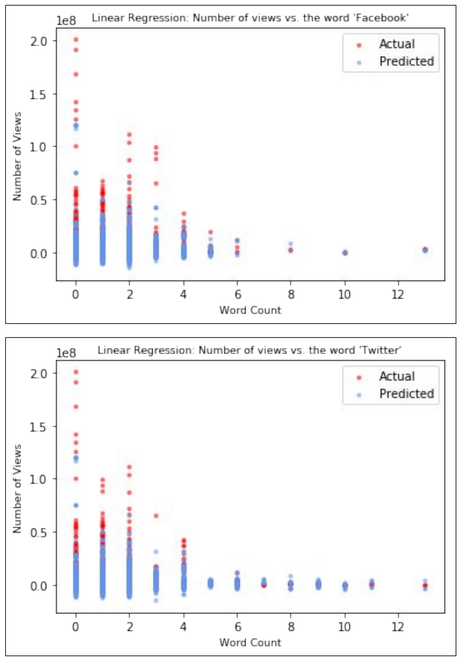

Figure 1: **Visualizations of fitting multiple linear regression to the US dataset, plotted as number of views versus word counts for the top weighted words from US LDA.**

Latent Dirichlet Allocation (LDA)

We ran LDA on bag-of-words representations of video tags, titles, and descriptions to discover latent structure in the datasets.

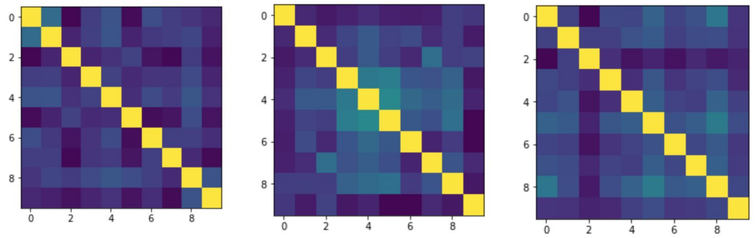

Figure 2: **Correlation between 10 latent topics for the US, Canada, and the UK, respectively.**

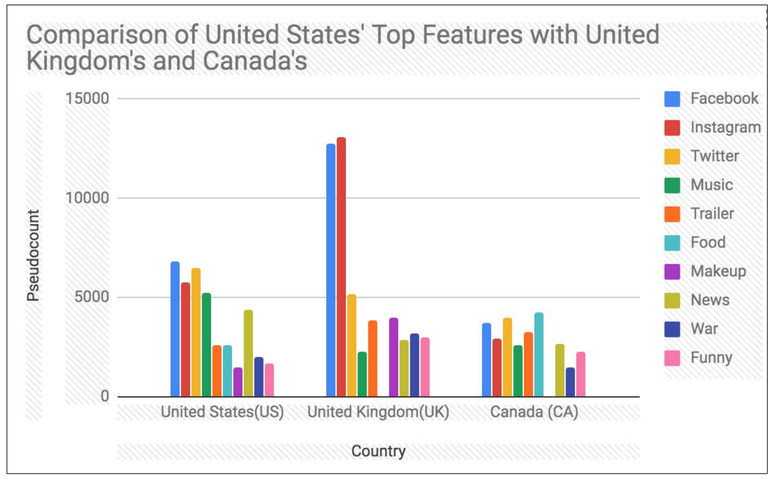

Figure 3: **Distribution of 10 component LDA pseudocounts across the US, Canada, and the UK, using the top weighted words we got across all the topics from US LDA.**

Classification

To explore multi-class classification, we used the US bag-of-words representation to predict video category labels. We trained multinomial Naive Bayes, Linear SVM, and Logistic Regression models, shown below.

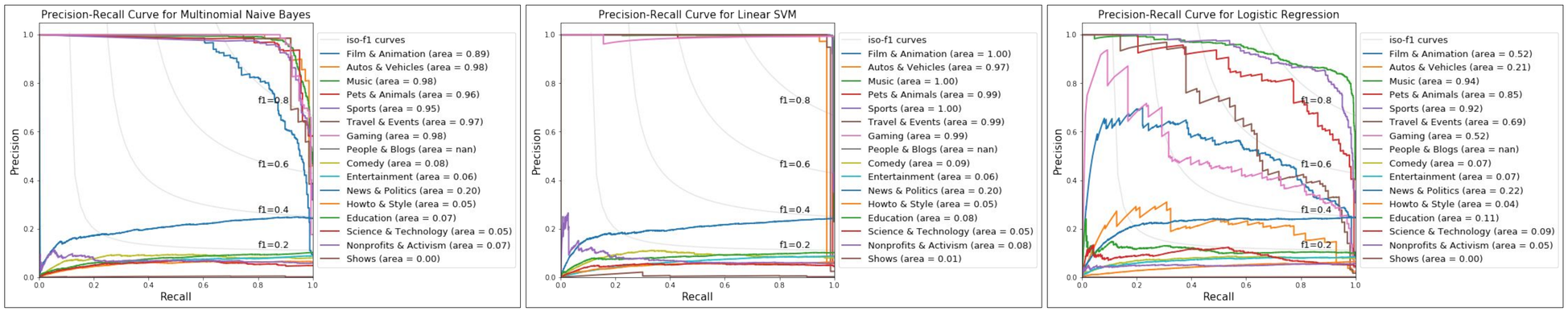

Figure 4: **Precision-Recall curves for multi-class classifiers on the US dataset.**

Data Exploration

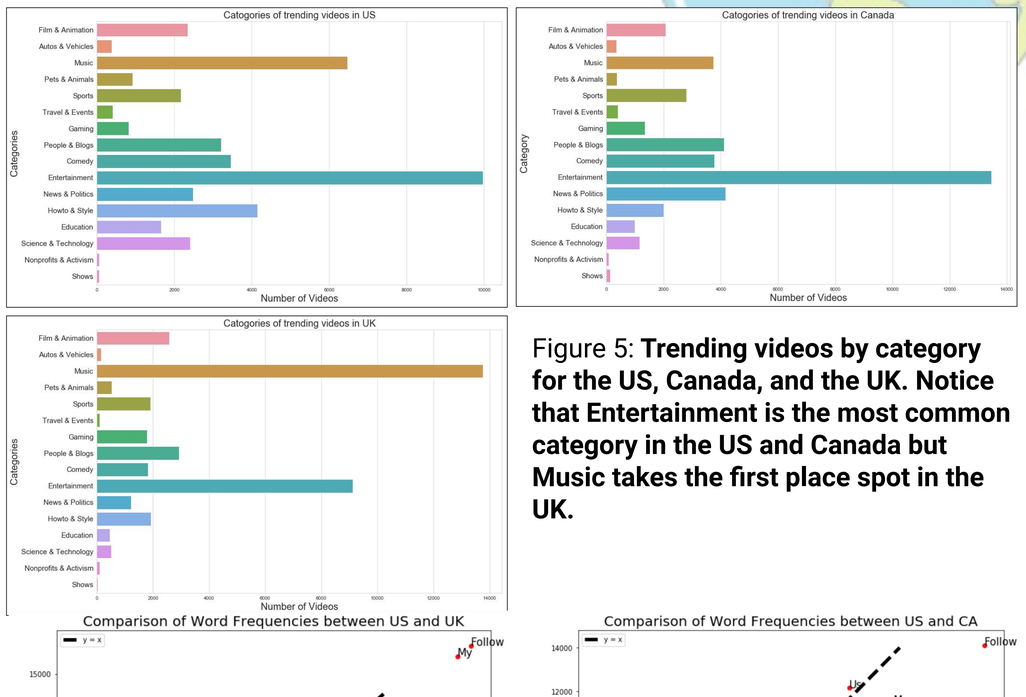

Figure 5: **Trending videos by category for the US, Canada, and the UK. Notice that Entertainment is the most common category in the US and Canada but Music takes the first place spot in the UK.**

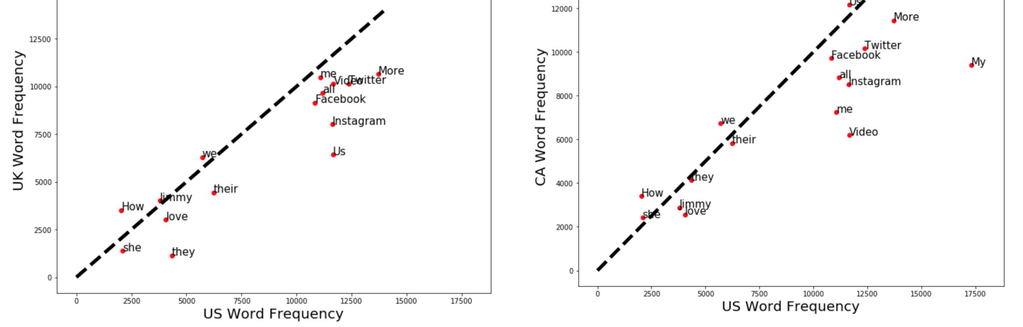

Figure 6: **Comparative scatterplots of popular word frequencies in different countries.**

Citations

[1] Siersdorfer, Stefan, et al. "How useful are your comments?: analyzing and predicting youtube comments and comment ratings." *Proceedings of the 19th International Conference on World Wide Web.* ACM, 2010.
[2] Krishna, Amar, Joseph Zambreno, and Sandeep Krishnan. "Polarity trend analysis of public sentiment on YouTube." *Proceedings of the 19th International Conference on Management of Data.* Computer Society of India, 2013.