

Final Project Reviews

26: Team Doggo Lovers

As a cat lover who owns a cat myself, I'm really interested in the harmony of human, pets, and technology that is brought up during the motivation discussion. It is clearly explained that we can expect technology to be developed to the extent that it helps bridge the gap between human and pets, and thus the group attempts to utilize and compare a few algorithms to classify images of pets to give us an insight of what AI and machine learning technology can be applied in the relationships of human and their furry friends. The goal and overall structure of the presentation are also clear to inform viewers what the group has done to reach their results and discussions. Each algorithm is detailed with the feature engineering process, the classification results, and pros/cons along with possible improvements. The group has also researched thoroughly in the field for past literature to learn that neural network is the best to determine relationships of pixels. Before performing classification with SVM, they make use of PCA, which demonstrates their knowledge of materials from lecture.

Overall the presentation is very clear and insightful, while it can be further improved by small modifications. As the presentation is about image classification, more images can be shown to attract viewers' attention, such as the images reconstructed after PCA and the images that represent KNN features. These images not only keep viewers engaged, but also provide a visualization of the classification pre-process. Another suggestion is to elaborate on comparison and problems of the algorithms used during discussion section. For example, the group can discuss what problems exist in KNN/SVM that lead to their lower accuracies for classification of pixels, and what about CNN that makes it the best to be used to determine the relationship between pixels.

36: Team Shallow Learning

The goal of this project is to utilize a few algorithms, including Naïve Bayes Classifier, SVM, and Logistic Regression, to classify the genre of movies. The presentation is very clear and organized. It leads viewers through past literature and methodologies involved in these researches and also the feature engineering process before explaining their models and results. I especially favor the amount of visualization details provided in the presentation, including showing the most frequent words and the graphs for results for each model. These bar graphs and line graphs offer easier ways for viewers to grasp what the team has done. The presentation also covers the mechanism and details behind feature engineering, including formula and definitions of TF-IDF as well as steps to remove common words, which demonstrates their knowledge and understanding of the pre-process taken and also give viewers an insight of how to perform a text analysis. Finally, the quality of results and discussions is high, showing the differences in accuracies of implementing different models and discussing the possible reasons behind the differences as well as possible improvements.

The presentation is informative and clear, while some suggestions can be taken in to account. During text-analysis using TF-IDF, features are often really long. As a result, PCA can be applied in this case. The team can attempt to perform PCA before putting long TF-IDF features

into training and see whether the accuracies have changed much. On the other hand, KNN also seems suitable for classifying genres for movies, as it can form different clusters that represent different genre and movies that have the same genre to have similar word usage. Finally, a personal recommendation is to shorten the section of future improvements (to maybe two of them). Although detailing possible modification is informative for viewers, it can also distract them from the learning and discussion of models and feature engineering process done in this particular project.

19: Team This Cannot Continue

The goal of the paper is to utilize neuro net to perform sentiment analysis and determine if a given text contains offensive insults. Their motivation is also clearly explained, which is to decrease the amount of insulting comments online while not requiring many moderators. I especially like this topic and its motivation as social media platforms are ridden with cyberbully currently, and a successful sentiment analysis will definitely help alleviate the hostility by detecting and controlling insulting comments and posts. An exhaustive research in past literature is done in the paper to determine best algorithm for the classification task. The literature review is comprehensive and informative as it includes several models and methods used in the past for similar sentiment analysis, which demonstrates the team's knowledge gained from these related works.

Since running the implementation is computationally expensive, I would suggest providing some kinds of exploratory data analysis that can offer visualizing insights for readers about the original dataset. The team can also provide a visualization for the neuro net architecture implemented, just as most of their mentioned related works do.

Yu-Chun Chen

Contributions Report

Zhenyu Bi: incremental score: 0

- Model training: Gaussian Naive Bayes Classifier, Bernoulli Naive Bayes Classifier
- Methods writeup for models above

Yu-Chun Chen: incremental score: 0

- Dataset and Exploratory Analysis (including writeup)
- Results writeup
- Discussion writeup

Chengming Dong: incremental score: 0

- Background research
- Introduction and motivation writeup
- Related Work research and writeup

Rui Qin: incremental score: 0

- Models training: KNN, Random Forest, Logistic Regression
- Methods writeup for models above

Shuibenyang Yuan: incremental score: 0

- Models training: Multiperceptron Algorithm
- Methods writeup for models above