

**Project Title:**

Comparison of Machine Learning Algorithms for Multi-label Classification on Youtube Videos

**Authors:**

Zexi Han ( [han.ze@husky.neu.edu](mailto:han.ze@husky.neu.edu) )

Chengbo Gu ( [gu.chengbo@husky.neu.edu](mailto:gu.chengbo@husky.neu.edu) )

Hengfang Deng ( [deng.he@husky.neu.edu](mailto:deng.he@husky.neu.edu) )

**Description of the problem:**

The objective of this project is to develop a multi-label video classifier that assigns labels based on given features of the video using the Google's recently released YouTube8M V2 (YT8M V2) dataset. This project is also in parallel with the Google Cloud & YouTube8M Video Understanding Challenge on Kaggle.

**Summary of the data:**

YT8M V2 is a large scale labeled video dataset that consists of over 7 million YouTube videos and includes video labels from a vocabulary of 4716 classes (3 to 4 labels per video on average). It comes with both pre-extracted audio & visual features. Each video level entry is represented by quantized 1024 dimensional visual features and 128 dimensional audio features.

**Methods:**

In order to compare the performances of standard machine learning algorithms for this problem, **Logistic Regression**, **Support Vector Machines** and **Neural Networks** will be applied and then assessed. Under the circumstance of large scale dataset, we use online learning algorithms to implement model updates. There is a separate validation set to be used for validation purpose and the criterion used is Global Average Precision (GAP). One key python library package used is **Tensorflow** which is commonly used for numerical computation of data flow graphs. Due to the complexity and maturity of the data, the features extracted from the videos are assumed to be optimal independent variables of the video and hence not to be cleaned or pre-processed. We would focus on the modeling and analysis process and hoping to shed some lights on each learning algorithm in terms of their performances for classifying multi-label videos.

**Preliminary Results:**

Based on the literature review and with such procedures implemented, we expect the GAP to be in a range of 60% to 80%. We foresee that SVM would outperform Logistic Regression due to its kernel-based nature and we also want to further discover the capabilities of Neural Networks by optimizing the parameters (e.g. number of neurons and layers) and hence to propose the optimal classifier.

**References:**

[1] Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., & Vijayanarasimhan, S. (2016). Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*.