Machine Learning-Proposal

Title: Cross-modal Retrieval

Group Members: Guyu Liu, Zhong Han, Han Jiang

Background:

With the rapid development of mobile devices, social networks, and self-media platforms, multimedia data such as text, images, and videos increasing exponentially in recent years. On social media websites such as Facebook and Twitter, users share a huge amount of text and pictures every day. These different types of data usually describe the same object or event. Therefore, it is crucial important for researchers to find a novel method to combine these different types of data.

Abstract:

The core of Cross-modal Retrieval is to learn a common subspace where the items of different modalities can be directly compared(transferred) to each other.
At present, there are mainly traditional and deep learning methods. CCA(Canonical Correlation Analysis) plays a key role in the traditional methods, is a way of inferring information from cross-covariance matrices.
In the deep learning method, the most basic idea is to extract features from images and texts. The problem is that we cannot combine the text feature and image feature directly, because their features are in different dimensions and representation space.
Inspired by recent advances in adversarial learning, we plan to tackle this problem using deep learning method in this project.

Objectives: Using one of the modalities (such as text) to retrieve other modal results (such as images, video, audio, etc.), that is, cross-modal retrieval.

Method: Adversarial learning

Data Set: Wikipedia, Pascal Sentence