

Multi-modal Entity Clustering
By Kushal Arora
karora@cise.ufl.edu

Problem Statement

Given multi-modal data (images and text) such that each image is associated with a small portion of the text, cluster entities (objects, people etc) such that data mentions (text mentions and image segments corresponding to entity) are clustered into single cluster center.

Example

can M4A2 carrying a Japanese Type 94 light-**tank** on its back, Namur, 1944.

A #Marine M1A1 Abrams **tank** with the @24thMEU sends rounds down range during exercise Eager Lion 2015 in Jordan.

The Landkreuzer P. 1000 Ratte, Germany's super-heavy **tank**. The project was canceled by Albert Speer in 1943. #WW2

Today in "First World Problems", **Barrack Obama** creates a twitter and follows each Chicago team except the #Cubs.

President **Barrack Obama** Finally Joins Twitter | Gained 1M Followers Within 5Hours -

In a world of infinite possibilities, somewhere you are **Barrack Obama**

General Approach

- Detect entities for each modalities separately example object segmentation (or face detection) for images and NER for text
- If a blurb has a textual mention and object segment, link all them with each other with weight 1.
- Link textual mentions and image segments on the basis of similarity with weight (0,1)
- This will give a very dense graph, do thresholding to make graph sparse. (For example remove all weight less than 0.4)
- Now we have a graph of multi-modal entities, apply heirarichal clustering to detect multi-modal cluster centers

Brazil Protest Dataset Clustering Problem

- The dataset contains tweets send out in Brazil on June 12, 2014 during protest against Brazil World Cup
- The objective here is to cluster all the tweet corresponding to people mentioned or in images to classify them as participants or non-participants in the protest
- This problem is similar to multi-modal clustering problem we just mentioned. Here modalities are text and images and entities to be clustered are people

Approach

- Name Entity Recognition on Tweet text to detect names (with Person Tag)
- Face detection for images
- If tweet contains names and images, add edges with weight 1 among them
- Add similarity edges between names using Jaccard Similarity or Jaro-Winkler distance
- Similarity between the faces using LBP(Local Binary Pattern) descriptor

Approach(2)

- Thresholding to make graph relatively sparse.
- Hierarchical clustering to detect clusters in data

Face Detection and Pre-processing

- Opencv provides the HaarCascade classifier for frontal face detection.
- We still need to do some more pre-processing steps
 - Crop out anything except for face. For a image with multiple people, this will give us an array of faces
 - Histogram equalization for handling brightness and contrast differences

Previous Work

- Face clustering problem has been considered in literature previously especially for the case of videos.
- A tutorial on the same can be found at [here](#).
- Some papers that tackle similar problem are
 - Automatic Detection and Clustering of Actor Faces based on SpectralClustering Techniques
 - A mutual information based face clustering algorithm for movie content analysis

Thank You!