17637 Project Proposal

Crowdsourcing Tagging System

Team Members: Kuixi Song (kuixis), Yubin Gao (yubingao), Di Lu (dilu)

Project Description

Our project is designing and deploying a website implementing crowdsourcing tagging functions. To be specific, **crowdsourcing** is a distributed problem solving or production model and the act of applying tags is **tagging**.

We chose the idea because crowdsourcing has been a major trend harvesting social wisdom and the project can make the most of technologies we learnt from class. By crowdsourcing, one task can be divided into a large number of simple sub-tasks with small workload, which are then distributed to netizens through the network platform to complete. In our project, the task will be **image entity tagging** and natural language **part of speech (POS) tagging**.

For image tagging, our website will provide a list of images and users can browse through the image collections, choose tagging entities and drag on the page to draw a box to apply tags (like a traffic light, a cat, etc.) to these images.

For natural language part-of-speech tagging, we will provide a number of sentences and a collection of part-of-speech tags (like noun, verb, etc.) and users can drag and drop these tags on every individual token in the sentence.

For both tasks, simple statistics will be available to users once tagging is complete. Some statistics include **number of tags**, **detailed tag information (like image coordinate list and POS tags)**, **tag statistics and distribution** and **number of completed tasks by the user**.

There could be extensions on the type of tasks using the website, as long as the task has the nature of crowdsourcing and hand-tagging. For example: labeling human behavior of static images, sentence sentiment analysis and handwritten recognition.

Some of the pages we will build include:

- Login and Register page
- All Tasks browsing page
- Tasks introduction and contents detail page
- Image Tagging page
- POS Tagging page
- Profile & Statistic page

Technologies

We will use Django, Ajax, JavaScript, HTML and Bootstrap to build the website. For image tagging, we will use HTML Canvas to draw a circle or rectangle on the image and use JavaScript to store the x, y offsets. For POS tagging, we will use NLTK to preprocess the sentences and tokenize them into lists of words.