

Harry Scells
Craig Craci
28 January 2015

Candidate Topics For A Research/Programming Project

Predictive analysis of the weather using twitter	2
Classifying images with colour and natural language	3

Predictive analysis of the weather using twitter

This project aims to use geographic data from twitter to predict the weather based on the content of tweets. Using the geo-data attached to every tweet and lexical analysis of the content of the tweet, it can be mapped to a graph which, with enough training, should be able to determine the weather. With enough user input and data, it should be possible to produce a geographical map with the predicted outside conditions. This project will involve artificial intelligence to analyse text, attempt to determine if the text contains weather conditions, make a generalisation against users tweeting about the same content, and map it to a geographical area.

Analysis of mass amounts of social media is in the realm of big data, which is an emerging area of research and can provide some insight into crowd sourcing information. There have already been projects aiming to make predictive analysis of tweets, including predicting earthquakes and typhoons. The outcome, like similar projects of this kind, should be that of getting results before existing technologies and systems can provide them. It can also benefit people whom are at risk of tornados and snow storms in North America.

The results of the project will be evaluated against current weather conditions. This includes overlaying both datasets above each other to determine how close the predictive analysis is to already established and trusted methods of weather prediction. The text analysis, prediction and mapping will all be implemented in lisp, with the networking in python. The resulting output will be a graph plotting the relationship between the temperature (hot or cold which could infer a heatwave or snow or rain) against the original location of the tweet (east coast to west coast America). The output will also aim to provide a geographical map with overlays, however this can be done well into the end of the project.

Classifying images with colour and natural language

This project attempts to classify images based on the main colours components in them. Images are sorted and classified based on the colours that are contained in them along with english language tags that allow them to be queried. Images can also be analysed and compared to similar images. By using classification techniques to look at the colours and tag images automatically, it is possible to create a database of reference images in which images can be searched and compared by colour.

Learning algorithms will be used to classify new images going into the system. These algorithms will look at similar images and classify them under the colours most present in the image. The techniques being employed here will include image classification, image tagging, colour recognition and computer vision. There are lisp libraries available for looking at images and processing them.

The problem of classification will be the most difficult part of the project, as a colour can have multiple hues and shades. Initially, the program may only sort images into three categories (red, green and blue). However as the learning advances, it should be able to classify with more depth.

The outcome of the project will allow anyone to enter an image as input, and as output, allow them to receive the dominant colours of the image as tags, as well as images that contain similar colours.