Trust Metrics for Online Social Media

Abstract

The vast amount of information available online makes it hard to trust the information that is presented to us online. This is particularly true with crisis situations, or other pandemics where many people post updates and information about the truth of the event and some may or may not be entirely truthful. This project looks at Twitter specifically and analyses tweets posted by people and asses the trustworthiness of the tweets. The aim of this project is to device an algorithm that would be able to differentiate a trustworthy tweet from an untrustworthy one.

Motivation

The proliferation of Internet usage has resulted in huge amounts of information being made available online. A big contributor to this are social media networks such as Facebook, Twitter, Google +, Question and Answer platforms like Yahoo Answers, Quora and Stackoverflow which act as public boards of discussion. However, whilst the scale and variety of data has considerable value for commercial and social purposes, the quality, provenance, trust and validity is often questionable [1]. Opinions are frequently posted to social media sites and are often a mixture of fact, speculation or rumour whereas user-driven sites such as Wikipedia are often questioned for their trustworthiness. Blogs are used as sound boards for average users to disseminate information to their social circles and favourable opinion becomes fact which is in turn regarded as trustworthy information. Trust is also a primary factor in influence. Verified Twitter and Facebook profiles go some way to ensure that the identity of an entity is verified independently therefore ensuring the trustworthiness of that entity. This project looks at ways of predicting trustworthiness for information posted on twitter.

Aim

The aim of this work is to predict the trustworthiness of tweets on twitter to a reasonably (>80%) level of accuracy. This project is based on the assumption that well known news websites publish trustworthy material.

Goals

The goals of this project is to gather data from news websites for good data. Gathering data from twitter accounts which consistently post false tweets for bad data. Apply a machine learning technique to the data gathered and chose the technique that most accurately predicts the real result.

Approach

We looked at tweets from news agency twitter accounts, tried to learn what a trustworthy tweet would look like from these tweets and apply the learning to any other tweets that we get. We also chose a couple of websites that publish non-trust worthy data and collected data from here for bad data. Data collection was done over a week, on 10 twitter accounts. Once this is done, we then use the data collected and pass train it with a Machine Learning algorithm. I have chosen two algorithms to test on - Support Vector Machines and Logistic Regression. This will then be tested with new data and the best predicting algorithm will be chosen. We will have to play around with different features etc to understand or best predict something. However this is the basic approach to solving this problem.

Bibliography

Bibliography :

[1] Jason R. C. Nurse, Syed S. Rahman, Sadie Creese, Micheal Goldsmith, Koen Lamberts, "Information Quality and TrustWorthiness : A Topical State-of-the-Art Review," in 2011 International Conference on Computer Applications and Network Security (ICCANS 2011)