Twitter Analysis for Journalism

Vishnu Hundi Kamath

MSc-Big Data

INTRODUCTION

Social media is mainly about networking and communication through text, video , blogs, pictures on sites such as Facebook, Twitter etc. What makes social media of particular interest to journalism is its influence as a tool for communicating breaking news.

PROJECT AIM

The aim of this project is to implement a probabilistic model that determines something unusual in live tweets. Journalists, instead of reading all tweets, receives only those tweets which could possibly make news.

IMPLEMENTATION

A single tweet contains data like - username, time and location of tweet, number of retweets etc. A model is built which calculates the probability of occurrence of each word in the tweets. This probability is compared with the words occurring in the incoming tweet. If the overall probability of the tweet turns out to be very low, that tweet seems to be unusual. All tweets with overall probability below a threshold are ranked, the most unusual tweets appear at the top of the list. These tweets are then presented to journalists along with the percentage of how unusual the tweet is.

WHAT MAKES THIS PROJECT DIFFERENT?

There are analytical tool out there in the market, which gives information about latest trends, what celebrities are tweeting , visualization of tweets etc. Breaking news could be someone famous tweeting or something unusual. This application mainly concentrates on unusual tweets which could potentially be breaking news.