

Report on smai project

Understanding of the problem:

We present TwitterMonitor, a system that performs trend detection over the Twitter stream. The system identifies emerging topics (i.e. 'trends') on Twitter in real time and provides meaningful analytics that synthesize an accurate description of each topic. Users interact with the system by ordering the identified trends using different criteria and submitting their own description for each trend. We discuss the motivation for trend detection over social media streams and the challenges that lie therein. We then describe our approach to trend detection, as well as the architecture of TwitterMonitor. Finally, we lay out our demonstration scenario.

Workflow:

Data Preprocessing
Feature Extraction
Model Development
Validation
Testing

Data set specifications & description:

1. Name :- Name of the user which is entered in twitter.
2. CountryCode :- Twitter gives code for each and every country.
3. Tweet Data:- Tweet's done by the corresponding user.
4. Parent_id:- parent id of the person who is going to tweet.

Classification/Clustering tools and techniques: -

Various data mining algorithms will be explored and considered. For each algorithm, parameters

will be tuned. The model will be designed using the techniques that give the best Sensitivity,

Specificity, and Median prediction time.

Tools :- Twitter API

Techniques :- Finding bursty keys, PCA, SVD.

Some results in finding brusty keys:-

