

Sorting of Specific Tweets on Twitter

Abstract:

Mining social network data and developing user profiles from unstructured and informal data are a challenging task. The proposed research builds user profiles using Twitter data which is later helpful to provide the user with personalized recommendations. Publicly available tweets are fetched and classified and sentiments expressed in tweets are extracted and normalized. This research uses domain-specific seed lists to classify tweets. Semantic and syntactic analysis on tweets is performed to minimize information loss during the process of tweets classification. After precise classification and sentiment analysis, the system builds user interest-based profiles by analysing user's posts on Twitter to know about user interests. The proposed system was tested on a dataset of almost 1 million tweets and was able to classify up to 96% tweets accurately.

Existing System:

Due to the relatively brief history of the WWW and the social networking services there is still a severe lack of understanding what, e.g., the information provided by microblogs can be used for. For this reason, we provide a review of the literature with a focus on application areas of prediction models that have been developed so far for analyzing data from microblogging services.

By prediction models we mean methods that aim at forecasting new events rather than merely summarizing or describing information contained in data. For instance, among the first studied questions of social media were investigations related to the topological structure of social networks. Specifically, the degree distribution, the community structure and motifs of acquaintance networks representing the "friendships" among members of social networking services, corresponding to nodes in such graphs, have been investigated (Java et al., 2007; Aparicio et al., 2015). Such studies are more descriptive in nature. Instead, in this review we present an overview of the literature that use social media data for classification, regression, or time series prediction problems.

Proposed System:

The proposed classification and sentiment analysis system uses a semantic structure, important keywords, and opinion words from tweets to monitor user interests and then generates personalized healthcare and wellness-related tweet recommendations. These personalized tweets consist of publicly available content which is precisely pre classified by our system. For tweet classification, the proposed system uses a domain-specific seed list which helps to decide which category a particular tweet belongs to. After classification, the proposed system also applies a lexicon-based sentiment analysis approach to extract topic level sentiments in tweets.

To increase the accuracy of tweet analysis, the proposed system also uses synonyms with keywords. The proposed model performs more precise analyses of tweets enriching temporal

patterns and semantics of keywords which optimize filtering results and help to extract more knowledge from tweets. For testing of profile generation, we collected 6000 tweets of users and generated user profiles by extracting health-related keywords, entities, and sentiments. For classification, the system was tested on almost 1,000,000 tweets of different categories. Due to our pre classification strategy and other significant improvements, our current model showed an accuracy of 96% for tweet classification, which is significantly better than our previously published approach, with an accuracy of 89.5% [27]. The proposed system also measured how much information for one category can be extracted from other categories which were ignored by keyword-based search from tweets.

The main contribution of the presented work is complete design and implementation of a personalized recommender system for a user based on his temporal social media history. The proposed system does not just rely on keyword-based interest but it also takes user's temporal sentiments into account. The syntactic and semantic analysis of tweets leads to more complete profile generation and tweet classification.

Software Tools:

1. Jupyter Notebook
2. Anaconda
3. VS Code
4. Colab
5. Scikit-Learn
6. TensorFlow
7. Keras
8. Python3

Hardware Tools:

1. Laptop
2. Operating System: Windows11
3. RAM: 16GB
4. ROM: 4GB
5. Fast Internet Connectivity