Content

1. Abstract
2. Motivation
3. Introduction
4. Formulation of the problem
5. Literature & Related Work
6. Design

5.0 Design principles

* 1. Crawling Data
  2. Influence Models
     1. Influence Score and Strength
     2. Influence Area Detection
     3. Activity Based Model
     4. User Based Model
  3. Viewing Results

1. Study case
   1. Choice of social network
   2. Crawling techniques
2. Technical Solution

6.0 Design Overview

6.1 Crawling Data

6.2 Influence Models

* + 1. Influence Strength
    2. Influence Classification
    3. Activity Model
    4. User Model
    5. User Interface

1. Analysis
   1. Crawling Analysis
   2. Influence Strength Analysis
   3. Influence Classification Analysis
2. Further Improvements
3. Conclusion
4. References
5. Appendix
6. Abstract
7. Motivation

Social Influence has been one of our main human characteristics since the dawn of history, becoming more sophisticated and interesting as an object of research, fuelled by the increase of density and human population across the globe. The fact that every individual can both influence and be influenced by other humans in many aspects during life is driven by an internal need to adapt to different environments and situations that might face us which in turn increase our chances of survival and self-realization.

We are currently standing at the very beginning of a technological and digital revolution which will change most of our life aspects and social influence is no exception. Which is simply defined as how humans as social individuals interact and influence each other. Many social media networks have been introduced at the beginning of this revolution and it is still gaining more popularity among both current and potential users, rising a challenge to be able to accommodate the amount of digital information generated to use for enriching the overall experience and understanding of societies.

Many digital platforms such as Facebook, Twitter or Reddit has already started to realize the importance of studying the behavior of each user on the network and how we can interpret the actions between users to establish a ground understanding of how strong the influence between users is, and in which category a certain influence is to be classified. Which opens a road of challenges and opportunities for computer science and engineering to build reliable influence models that can help detect social influence and reveal the hidden power of data generated by these social networks.

1. Introduction

The aim of this research is to build a reliable digital system that can request data from a certain social network platform using its API access endpoints. The data will then be preprocessed to suit our algorithms and then used to build a desired influence model that visualizes network users as graph nodes (Influencers) and the interaction between them as a set of weighted directed edges (influences).

At the start of this research project, we will review some of the most relevant related work in the field of social media analysis associated with computer science. This will help us to widen our knowledge horizon and keep us updated of the latest technology achievement in this field, and perhaps providing a source of inspiration to build a set of some reliable and useful influence models.

After establishing a ground study of related work, we will proceed in defining the problem we are trying to solve in satisfying technical details to help clarifying the driving idea behind this research project.

We will then start the solution process by analyzing how we are going to automate the data flow from social networks feeding into our application, and then dive into the data analysis to be able to find a common pattern in the data properties that can be obtained from most popular social platforms, which will help us build a flexible influence model that can adapt to different data schemes from different social media platforms, making it possible to compare or merge data from different social networks into one model containing useful information on who are the influencers, what areas they do influence in and how strong are these influences are.

In the analysis process, we will evaluate the performance of our solution by viewing statistics measured under different processes in the system periodic life cycle. Based on such information, we will be able to test and reveal the strengths and weakness or this solution.

1. Formulation of the problem

The main purpose behind this research project is to be able to identify and classify influence between users of a social media network, finding out who are these users, how much influence they have on each other and what are their area of influence; weather it is politic, entertainment or any other defined category. Based on this main idea, we can detail the desired functionality of our final system as follows below:

1. Information Gathering

The first operation to take place is gathering information about a certain network or group within this network, this row data can be obtained using different methods, such as manual uploading, fetching a database or signing up for an API service account that enables our system to crawl data from our studied network. Under this process, we will try to define a common data structure that applies to most social networks to maximize the flexibility of the system, making it more portable to be used for as many social media networks as possible. Another advantage of having a unified data structure is its ability to combine or merge information from different social networks for the application of comparison of influence activity across multiple social networks at once.

1. Influence modelling

This process represents the core of our system where data is being processed through an specialized algorithm that helps detects influences between users based on their activities and organizing the users and the influences between them as a graph with nodes representing participating users and edges that show the direction of influence, along with a series of diverse scores stored in each edge which tells us how strong the influence is most likely to be from the source user node to the target user node of one edge.

1. Data Storage

A key feature of this system is its ability to crawl and store data and built influence models on fixed time intervals (daily for instance). The system will then have to be designed to keep track of when the data has been gathered from remote source and store this data in a suitable and easy to access database using a time driven scheme.

Another important feature of the process of data storage is the desired persistent of data in the database, where data is kept in a unified data structure even after using it for building an influence model. This measure will enable us to be able to edit our influence model algorithm on previously gathered data without having the need to decode the new structure resulted from the previously used influence model algorithm. Having more than one influence model is also another advantage of keeping the initially gathered data stored in a unified data structure.

1. Data Storage

Visualizing the influence models is highly important, for this reason it is desired to have a web user interface that enables the public to view, examine and monitor different influence models without having to worry about internal technical details carried out in the previous steps.

This web user interface should implement some tools and functionalities such as graph visualization, influence shortest path finding between users, influence score filtering and influence area filtering along with monitoring statistics and details about the 2 most crucial processes; information gathering and influence models.

1. Related Work

Among the community of computer science, a wide variety of studies has focused on extracting information from the available social media platforms, and a big amount of effort has been dedicated to reveal influences between users to better understand the behaviour of individuals for many both commercial and non-commercial purposes. Research of social influence takes different forms and vary in size and scope, while some researchers take on the very fundamentals of detecting social influence others dive through to reveal specific influences and hidden behaviour patterns on different levels.

A good fundamental approach is described by a social network analysis carried out by Y. Guo, J. Cao & W. Lin. The fellow researchers are dividing the influence evaluation models into 2 main categories; the first category is based on network topology which measure social influence of a certain user by considering his degree, shortest path and random walk characteristics, while the second category bases the influence between users on their interactions through different activities organized in tree-like structures like submissions and multilevel comments. However, and despite the reasonably good classification and overview these researchers offers, their paper lacks some proven results of an experimental approach [1]

A short but interesting experimental approach is introduced by a paper on Data-driven Influence Learning in Social Networks published by F. Wang, W. Jiang, G. Wang & D. Xie. In this paper, the process of influence diffusion is divided into two parts: the launcher (influence strength) and the receiver (influence threshold) which can generate an accurate and finer grained influence diffusion model according to this research. Furthermore, the researchers highlight the importance of having a solid criterium when scoring the strength and threshold properties of social influences. Another important acknowledgment is the difficulty and complexity associated with detecting influence relationships between users as a by-product of big datasets that usually include a considerable amount noisy or less important datapoints, making it essential for any algorithm used in learning and testing the influence models to perform a minimal scan over the data in the most efficient way possible. [2]

1. Design

5.0 Design principles

* 1. Crawling Data
  2. Influence Models
     1. Influence Score and Strength
     2. Influence Area Detection
     3. Activity Based Model
     4. User Based Model
  3. Viewing Results