VOL. 11, NO. 15, AUGUST 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

SIMPLIFIED INTEREST CLASSIFICATION USING SOCIAL MEDIA

Vignesh Rajkumar, Siva Vikram and Kamalesh
Department of Computer Science, Sathyabama University, Chennai, India
E-Mail: vigneshrajkumar@hotmail.com

ABSTRACT

There are many non profit organizations incorporated but only few survive the long run. This is because of the increased financial constraints and lack of volunteers. Considering the fact that the primary source of income for such organizations are donations and the primary source of manpower is by word of mouth and mass communication means, this proposed hypothesis is to connect non profit organizations to potential volunteers and donors who will be interested in the organization's cause and vision, eventually supporting them by donations and manpower during community events hosted by the organization, at an affordable cost.

Keywords: social networks, machine learning applications, non profit organizations.

INTRODUCTION

Social media is always been and is a platform to share a person's interests by sharing status updates and post about things which the user is interested and concerned about. Now that smartphones have become cheap and affordable and on the other hand social networking sites getting more popular and common, we're left us with large sets of data which can be processed to gain insights. One such attempt is to classify user interests by parsing the status updates which the user posts through machine learning algorithms, predict their area of interest and eventually connect them to non profit organizations who work in the same domain as of the user's interest.

LITERATURE REVIEW

Quora, an online forum where you can ask questions and get them answered by the worldwide community uses one such feature where your questions will be automatically classified into their topic categories so that experts and users who are sound in that topic can address the proposed question. Even though Quora uses complex and complicated algorithms, there have to be a simpler way which can be modelled and used by non profit organizations at an affordable cost.

E Commerce sites like EBay and Amazon also uses a similar mechanism to classify user interests which is commonly referred as '*recommender engines*' which processes the history of your online purchase and suggests commodities which you'll be interested in

Facebook, the largest social networking site is no exception to this concept of topic classification. The Facebook platform analyses the statuses and posts the users share and categorize them as *'tending topics'* in our newsfeed.

These initiatives were the motivation for our proposed hypotheses to classify user interests in a simple and intuitive which can be cost efficient.

METHODOLOGY

Existing scenario

By consulting many non profit and non government organizations, it is found that donors are approached irrespective of their interest and eventually leading to the rejection of the proposal. Also this is done through mass communication methods like bulk emailing and word of mouth, which in turn brings down the organization's credibility due spam flags and lack of response.

Proposed system

To test the prosed hypothesis, we chose Facebook as our social media platform and Indico, a machine learning API to parse the status updates of the user.

By virtue, it would be meaningless to parse just one status update to analyze the user's interest. So we pulled the latest 128 status updates and the posts which the user shared, filter out the statuses and messages which are more than 20 characters in length and sorted sorted them to a payload array which will be sent to the 3rd party REST API provided by Indico.io. Refer figure 1 for illustration.

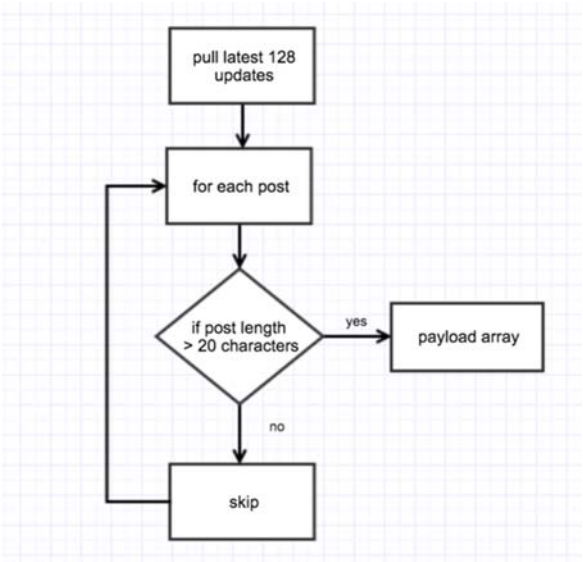

Figure-1. Payload filtering mechanism.

Once the data is processed in the remote API, we'll be getting a response in the form of an associative array with key value pairs of the topic detected and its probability. We'll be aggregating the topics along with their sum weight. The top 3 of the aggregate calculated along with the key value will be assumed as the user interest. Refer Figure-2 for illustration.

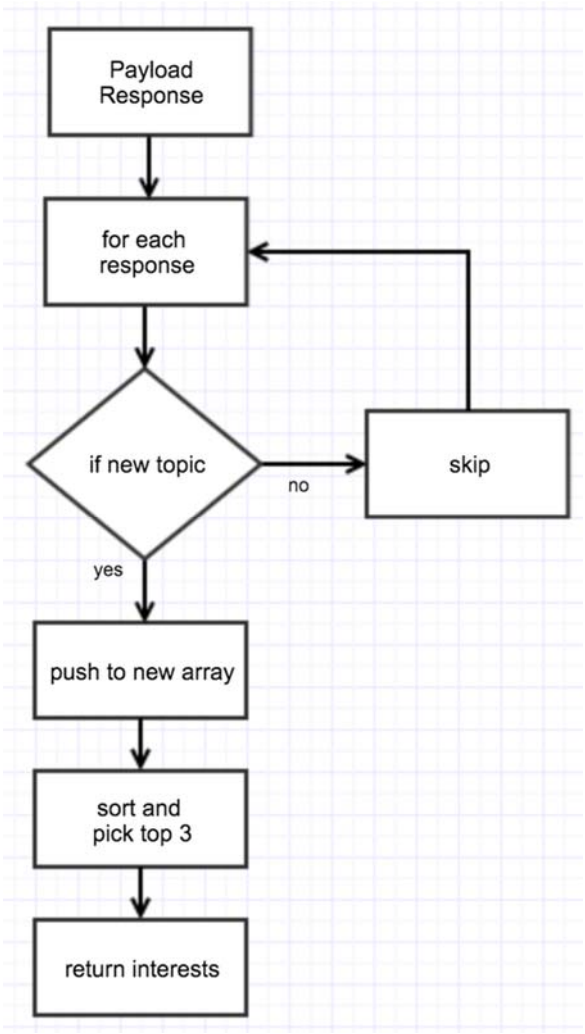

Figure-2. Classification mechanism.

Now that we have the users associated with their interests we can connect them the to non profit organizations who are working in those domains.

IMPLEMENTATION

The readymade availability of Facebook's SDK for PHP is also another reason to consider Facebook as the data source to experiment the hypothesis. The script used to pull data from Facebook's Graph API can be found in the figure 3. The response data from Facebook is of JSON format and hence it has to be decoded for further processing.

```php
// <--- Facebook Graph API Call -->
$fb = new Facebook([
    'app_id' => env('FACEBOOK_ID'),
    'app_secret' => env('FACEBOOK_SECRET'),
    'default_graph_version' => 'v2.5',
]);

try {
    $response = $fb->get('/'.$user->id  .'/posts?limit=128', $user->token);
} catch(FacebookResponseException $e) {
    echo 'Graph returned an error: ' . $e->getMessage();
    exit;
} catch(FacebookSDKException $e) {
    echo 'Facebook SDK returned an error: ' . $e->getMessage();
    exit;
}

$body = $response->getBody();
$jsonBody = json_decode($body);
$userPosts = []; // payload array of status

foreach($jsonBody as $data){
    foreach($data as $d){
        if(property_exists($d, 'message')){
            if(strlen($d->message) >= 15){
                array_push($userPosts, $d->message);
            }
        }
    }
}
```

Figure-3. Script to pull status form Facebook.

Once the statuses are filtered as per the requirements, it has to be parsed to the machine learning API. This process is also made easier by the availability of the Indico's PHP SDK. Refer figure 4 for the script which parses the status.

```php
// <--- Indico API Call -->
IndicoIo::$config['api_key'] = env('INDICO_KEY');
$topics = IndicoIo::text_tags($userPosts, ['top_n' => 1]);

$collection = [];

foreach($topics as $topic){
    $temp = array_keys($topic);
   array_push($collection,$temp[0]);
}

$aggregate = array_count_values($collection);
arsort($aggregate);
$result = array_slice($aggregate,0,3);
$result = serialize($result);
```

Figure-4. Script to parse statuses to Indico.

DISCUSSIONS

Facebook is chosen as the data source due the large online community available. The other reason to consider Facebook for testing this hypothesis is because of the extensive support and service development kit provided. Nevertheless, similar stock and community developed service development kit are available in Github for Facebook and other social networking platforms with developer APIs available.

Similarly, Indico is chosen because of the simple testing process needed. Other APIs and custom algorithm can be developed in frameworks like Flask, Django and Rails. Giving more control and robustness over the deployed code

CONCLUSIONS

The proposed system produces satisfactory results and this can be an affordable solution for non profit organizations to get prospective volunteers and donors. For the purpose of testing the proposed hypothesis,

Facebook and Indico were used. There is no compelling reason to stick on to these services. For instance, Twitter and Instagram can be used as the data source and for analysis and analytics services like Meaningcloud and Aylien can be used or even a custom machine learning algorithm can be written in python and hosted for more flexibility.

The entire concept can be fabricated into a SaaS application and the service can be given to non profit organizations or the concept can be encapsulated into a platform agnostic library and can be hosted in Github for open source access and make it open for contributions from the open source community

REFERENCES

Using Twitter to Understand Public Interest in Climate Change: The case of Qatar - Sofiane Abbar, Tahar Zanouda, Laure Berti-Equille, Javier Borge-Holthoefer Qatar Computing Research Institute Hamad Bin Khalifa University PO

Role of computer applications and tools in the scientific research process-modh jatinkumar,Narmada college of computer application, bharuch-392011. (gujarat) abstract.

http://www.sas.com/it_it/insights/analytics/machine-learning.html

http://www.research.ibm.com/cognitive-computing/machine-learning-applications/decision-support-education.shtml#fbid=Jq0cHD99jZd