TweetUp: A Twitter and Meetup Fusion

Shaima AbdulMajeed
Department of Computer Science
University of Illinois at Urbana-Champaign
shaimaa2@illinois.edu

Dipali Ranjan
Department of Computer Science
University of Illinois at Urbana-Champaign
dipalir2@illinois.edu

***Abstract*—This paper proposes a new method of recommending events and twitter accounts to follow, by fusing the best features of two social medias- Twitter and Meetup. We aim to connect people possible having similar social interest, who can potentially meet and increase the impact in their communities. This would help in increasing the number of social movements and social reform campaigns in community. We answer three main questions in this paper, - Do people share same ideas on both social media, Can we recommend meetups based on users profile, Is it better to recommend users from twitter or from meetup**

The recommednations - meetup events and twitter accounts - are based on the user's twitter history and data mined from Meetup. We use a multi-class classification approach to classify the tweets and LDA topic modeling to see correlation between the same account's meetup and twitter profiles, and suggest accordingly.

***Index Terms*—Twitter, Meetup, Recommendation System, Classification, Topic Modeling**

I. Introduction

Some scholars have referred to the United States as a social movement society because the collective actions associated with social movements play such an important role in bringing about social change in political, religious, educational, health, corporate, government, and other institutional arenas [3]. Social Media can help tremendously in getting people together offering solutions to world's unpromising problems, a good example for that can be the Egyptian revolution in 2011, It all started from a Facebook page that was used at first for venting outrage and grievance, to an event coordinating 1.8 million by that time, followed by the infamous hashtag on Twitter #Jan25th leading to a crowd of 80,000 to the street of Cairo on that day [4].

According to the study of the Egyptian revolution, American Scholar Linz put forward that there from the ways that can affect collective action: Make the disgruntled citizens more coordinated take some public action and through the information cascade to improve the predictive chance of success [5]. Social media revolutions can also have a negative influence on the social movement. Malcolm Gladwell defined the SNS(sympathetic nervous system) activity as weak ties and low level organization structure, and put forward that the social relations constructed through the Internet is very difficult to translate into collective action [6]. A recent study reveals the some of the motivations that participants cited for sharing information on social media are - to support a cause or issues they feel strongly about; to define themselves; to grow and nourish relationships [2]. In this paper we fuse two social medias that can serve the purpose of strengthening the online based relations and giving the ability to its users to turn their beliefs on the Internet into actions in their community.

Twitter and Meetup are different social media sites, serving two different purposes. Twitter is a huge microblogging site having 330 million users per month, and 500 million tweets per day resulting in broadcasting enormous amount of data, enabling users to share their thoughts and beliefs.

On the other hand, Meetup functions as an IRL(In Real Life) social network, enabling communities to organize events and bring people physically together.As of Aug 2017 Meetup has 32.30 M users, 288,716 groups, 614,764 events/month and 3.90 M RSVP/month [1]. The service allows users to create groups acting as organizer and setting up events. To attract users to the right event, Meetup asks the users to select few topics from the predefined set of categories like Outdoors & adventure, Tech, Family etc and recommends events and groups based on the selection. To engage users, it allows an 'Explore' option allowing users to browse events with a radius of their current location. To recommend better events to users, Meetup has started a campaign to 'tag' topics to events/group. This is done by the organizers to aptly chose the topics to describe their events, allowing users to match their interest based on keyword like topics. Meetup doesn't have a dynamic way to update user interest information besides showing Related events, for example Pinterest, Twitter have a trending page which gives users a new way to follow most fresh updates on the topic they are interested in.

As with other social systems, efficient information discovery is one of the major hurdles for continuous growth. In this paper, we propose a recommendation system based on Meetup and Twitter to help users to discover interesting groups and topics on Meetup and users to follow on twitter that they can potentially meet later on.

II. Related Works

There is no work done fusing Meetup with Twitter, so we will be listing here the work done on Twitter to enhance the recommendation system. Also we will be presenting one study fusing Twitter with Pinterest, that inspired us in our work and evaluation methods.

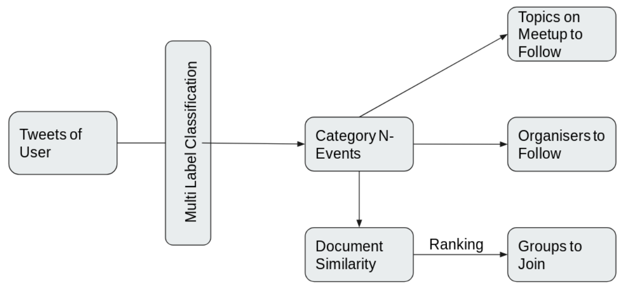

Fig. 1. Proposed System Flow

A. Twitter Topics Mining

A work from uppsala university proposed a topic categorization method but limited its dataset to scientific conversations only on twitter [8], Also another work by Kevin Dela Rosa et al. proposed a method to categorize topics for a predefined set which was News, Sports, Entertainment, Science, Technology, Money, and Just for Fun [9] However, one of the problems in that area is the dataset, there is no topic annotated dataset for twitter.

B. Twitter and Pinterest

The paper uses a Multi-Label classification, similar to our model, to map Twitter user followees to pinboard topics and visual diversication to recommend pinboards given user interested topics. [10] User portfolio is the foundation of any recommendation system, search engine and advertisements targeting. Previous work has been done by Geng et al. building Fashion ontology and a multi- task CNN to map Pinterest for images recommendation on Pintereset. [7]

III. Design

Our proposed Algorithm is shown in Fig. 1, where we take the user's tweets, clean it then we input the clean data in the multi-classification model, the model classifies the set of tweets into category(i.e. social interest) the user is interested in, based on that result we get the events in this category at the user's specific location, we get the organizers' of these events twitter data, and apply topic modeling the ones with the most frequent words resembling the category are suggested to the user.
We also rank the events in descending order starting from the events the most similar to user's twitter profile.
Last thing we recommend, is topics that they can follow on meetup, which can increase their meetup activity.

In the following subsection we explain the proposed method in more details.

A. Gathering Data

To address this problem we needed to define what do we mean by *social interests*, and break it into further categories. Firstly, we tried mining hashtags from Twitter that resembles social interests such as - women rights(#MeToo and #neveragain #EverydaySexism), environment(#recycling,#pollution), poverty, health care etc. We were able to categorize 10,000 tweets per category. After preprocessing the data (the hashtag from each tweets), we found this way to be not be robust to capture our model. The Twitter data is highly dynamic and Twitter API doesn't allow for collection of tweets which are more than 7 days old. This posed a problem for us as it was difficult to get highly relevant tweets in such a short interval. Additionally, it didn't prove to be scalable to other social interests easily, and needing a lot of manual handling which was adding bias to dataset.

As a result, we decided to get our data from Meetup. Data from Meetup is less noisy, tagged (multiple subtopics, location), and better access to data with no constraints. Following we will explain how we gathered this data:

1) Meetup has a list of predefined categories (Fig.2) from which we chose the categories that can potentially include social interests, for example Movements, Environment, Health and Wellness.
2) For each category we used Meetup python API [11] to get all the topics in this category [12]. The topics here refer to Meetup's internal categorization which is called "Topics" to user. These topics are much more specific than the bigger categories shown, eg- Black Women, Transgender women, Women's equality, Feminisin etc. This saves us the trouble of tagging specifically on the basis of specific interests.
3) We filter out topics from this list manually. We faced two major problems in this. Firstly, Meetup returned us result based on a keyword search. If we input environment, it would return groups like Montessori environment for kids, Relaxing environment at work. Secondly, the topics were often out of context, "Social" refrred to Social Justice as well as Social lesiure.
4) We manually filtered through the topics selected and related topics generated by Meetup API. The resulting topics are regrouped into our own defined social categories (class labels) (depicted in Fig.3), for the course of this project we focused on three categories namely political and human rights, environment, support groups.
5) Using meetup python API we gathered all the groups tagged with these subtopics for SAN Francisco and New York cities (most popular on Meetup) [13]. To strengthen our categories we also mined topics from multiple zip codes
6) We Remove the duplicate groups. Same group can be retrieved through different queries for example group attached to multiple ZIP codes, group that is tagged with multiple topics.
7) For each group we get the description, organizer ID, topics it is tagged with (as in Fig.4
8) For each organizer ID, we check if twitter is connected with their profile, If so we use Tweepy to get the user's profile and retrieve the last 300 tweet(twitter limitation) [14].

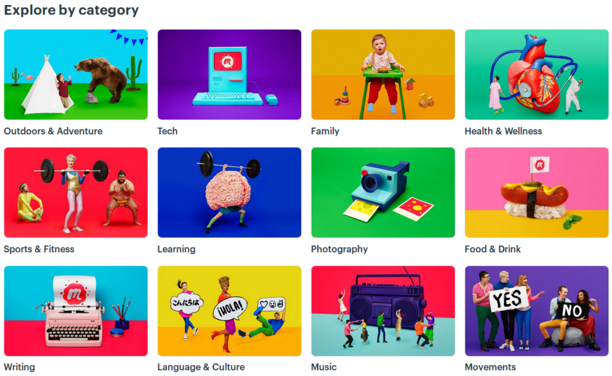

Fig. 2. Example of Meetup categories

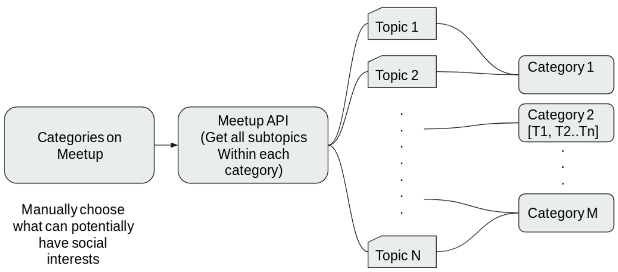

Fig. 3. Getting Categories

B. Classification

To accurately categorize our user to a social interest category, we build a Multi - Classification. It's trained on text based data of Group description collected from Meetup in the previous step. We test this model on data of user's Twitter history. The Tweets are converted to a text file and assigned a category and it's associated subtopics.

Multiple feature extraction method are used for classification. Count Vector is a matrix representation of the dataset in which every row represents a document, every column represents a term from the corpus, and every entry represents the frequency count of a term in corresponding document. TF-IDF score is calculated using the normalized Term Frequency (TF), and the Inverse Document Frequency (IDF). It is the logarithm of the number of the documents divided by the

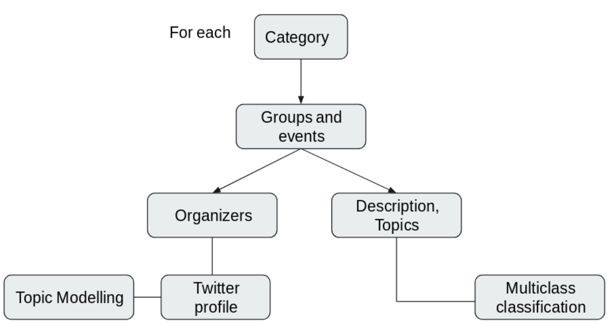

Fig. 4. Getting Data from the groups

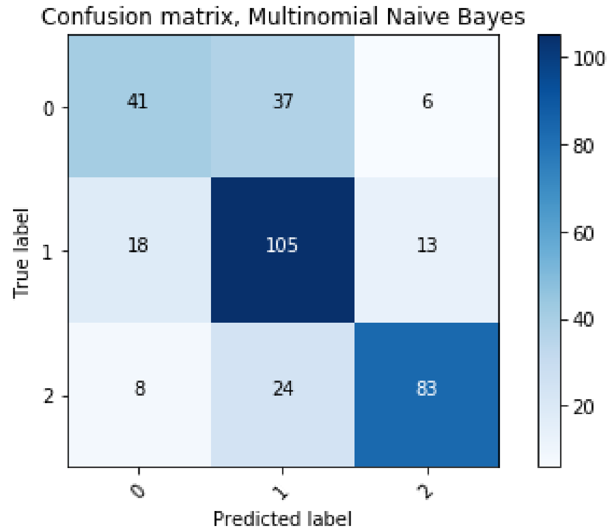

Fig. 5. Naive Bayes Confusion Matrix

number of documents where a term appears. Word Level TF-IDF represent tf-idf scores of every term in different documents. N-gram Level TF-IDF combine N terms together. Character Level TF-IDF represent tf-idf scores of character level n-grams.

For this problem we tried the most popular machine learning techniques for text classification problems namely Multinomial Naive Bayes, Logistic Regression and Support Vector Machine. At this time we have limited our model to classical machine learning models rather than the improved deep learning models for text classification, as deep learning methods usually gives better results with larger corpus.

For feature generation, we use Count Vectors, WordLevel TF-IDF and N-Gram Vectors. Lastly we compare all the models with all the different features, and choose the best accuracy model among all. We choose the model with best accuracy which is Multinomial Naive Bayes and Count Vector feature. We can see the Confusion matrix (Fig.5) doesn't hint any anomaly. We use this classification model to classify the user to a predefined category.

C. Ranking of Recommendations

Once we know the category a user belongs to, we rank the events using document similarity as shown in Fig.1, We compare every event with the User Tweets and get a similarity score. We use TF-IDF vectorizer and Cosine Similarity to generate the score.

Since meetup is not very straight forward when it comes to following topics, it only provides one page with a very long list of topics, which makes it a hassle to find all topics related to one category, tp prove this point, we considered a couple of keywords and used meetup api to get all topics having those keywords that was specifically related to politics and human rights, we provided 13 keyword, and got 447 topic from meetup, so this shows how difficult it is to find all topics to follow on meetup, that's why our proposed methed suggests topics to follow on meetup, we generate the Top 20 frequent subtopics in the recommendation list.

D. Meetup and Twitter User Profiles Correlation

To suggest accounts to follow on twitter, we tried two methods, First we tried scraping twitter, getting popular users in specific location having most tweets about specific social interest, and checking their followers, however this method proved to be very time consuming, and also it needed a lot of manual tagging for ground truth data.

The second method that we finally settled upon, is using meetup database, which has many advantages that we listed before, but the most important would be that we know the ground truth, if we get the organizer of a meetup about environment then that user is definitely interested in environment issues. Moreover many meetup users connect their twitter account to their meetup profile, giving us the ability to scrape this specific account's tweets, and we already know their interest from meetup, so instead of being a multi label classification problem, it is a binary problem, after analyzing the account's twitter profile we either say this person share about environment or not and accordingly we decide to recommend this account or disregard it.

We do this as following:

1) After getting group organizer's twitter profile from their meetup account, we use Tweepy API to get the last 300 tweets.
2) We clean those tweets using NLTK python package essentially removing URLs, hashtags, stopwords, punctuation, HTML special entities (e.g. &), tickers, small words (e.g. RT).
3) We perform stemming and lemmatization using spaCy.
4) Then Latent Dirichlet Allocation (LDA) model is applied, which is an unsupervised machine learning technique to identify latent topic information from large document collection. It uses a bag of words assumption, we used scikit learn Library [15] for LDA and applied tf-idf transformer to the bag of words matrix. We set the number of topics to be produced to six.
5) One method that has been used in the past to address the poor performance of LDA on shorter documents is the aggregation of documents into longer pseudo-documents. This is done in an effort to increase the amount of relevant word co-occurrences within a document if a Twitter user often talks about the same subjects, then combining their tweets will result in an increase in the co-occurrence of the relevant terms. When working with Twitter data, this has often been achieved by combining tweets that were created by the same author.
6) We then check the most frequent words and compare it with the words usually found in the corresponding predefined social interest category.

IV. Evaluation

In this section we will show our classification results, same user's twitter and meetup profiles similarities results and final results (the recommendations).

A. Classification Model Results

In this section we present all classification models with different features accuracies, as we mentioned before we chose Naive Bayes with count vector since it gave the highest accuracy.

Model	Feature	Accuracy
Naive Bayes	Count Vector	0.701
Naive Bayes	WordLevel TF-IDF	0.596
Naive Bayes	N-Gram Vectors	0.559
Naive Bayes	CharLevel Vectors	0.548
Logistic Regression	Count Vector	0.615
Logistic Regression	WordLevel TF-IDF	0.660
Logistic Regression	N-Gram Vectors	0.555
Logistic Regression	CharLevel Vectors	0.624
Support Vector Machine	Count Vector	0.447
Support Vector Machine	WordLevel TF-IDF	0.44
Support Vector Machine	N-Gram Vectors	0.434
Support Vector Machine	CharLevel Vectors	0.447

TABLE I
MULTI- CLASSIFICATION RESULTS

B. Does the Group Organizer Talk about the Same thing on Twitter and Meetup?

In order to suggest twitter accounts we retrieve all group organizer's twitter account Ids from meetup, so a good question here would be does the same person talk about the same beliefs on twitter that they made a meetup for.

To answer that question we retrieved 130 organizer twitter profiles for 3 social interests categories namely Environment, Political and Human Rights and Support Groups, we applied LDA on the full set of cleaned tweets of each organizer, then we aggregated all topics produced from all the organizers and visualized them through word cloud to better understand how similar a user's though is.

In Fig. 6 these are the tweets resembling the organizers' tweets in the political and human rights category, the most frequent words are Trump, Change, Women, rights, NRA, civil ...etc

In Fig. 7 these are the tweets resembling support groups, the most frequent words are support, love, live, cancer, way, good, great.

In Fig. 8 these are the tweets resembling environment groups, the most frequent words are climate, green, world, help.

The above analysis shows that people talk about the same topic to a great extent in both social medias, making it quite safe to suggest organizer's of events as potential accounts to follow on twitter.

C. Final Results

In this section we will present 2 use cases, we chose two random user's on twitter who are socially motivated and pretty active. In Fig.9 we show both users last 300 tweets topic

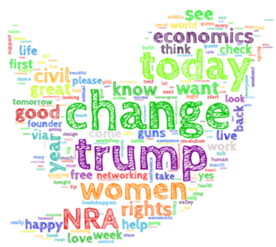

Fig. 6. Political and Human Rights Groups Organizers' Tweets LDA Result

Fig. 7. Support Groups Organizers' Tweets LDA Result

Fig. 8. Environment Groups Organizers' Tweets LDA Result

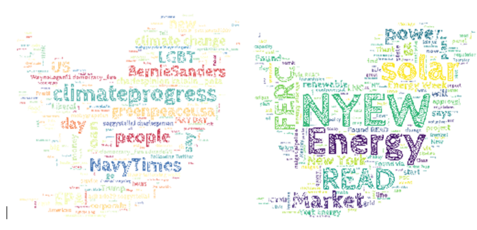

Fig. 9. User Twitter Representation

Fig. 10. Twitter Accounts Suggested to User 1

Content Proportionality	Topics
0.64	Civil Rights, Human Rights, Social Justice, Activism, Local Politics, Political Activism,Social Movements
0.27	Green Living, Eco-Conscious, Sustainability, Environment, Healthy Living, Environmental Awareness
0.09	Mental Illness Awareness, Trauma Survivors, Adult Survivors of Child Abuse, Cancer Support.

TABLE II
USER 1 TWEETS ANALYSIS

modeling results as a *word cloud*.
Next we show the recommendation our system makes to each of these users. For user 1 our model classifies the user as one interested into politics and human rights, giving the following analysis, depicted in table II

Accordingly our system suggests events such as (Actual events from meetup shown in Fig.11): "Democratic club of sunnyvale" which we consider pretty strong result since the tweets had Bernie Sanders pretty frequently, and the system didn't recommend any republican meetup

The twitter accounts suggested to this user are shown in Fig 10, the most important observation here that all the accounts suggested are not verified accounts, and it is not a must to have something related to politics in their biography section of twitter, aas our proposed system doesn't look at twitter account's biography, since the social interests change by time and shown in user tweets, however rarely a user changes their biography to continuously show their current ever changing social interest.

Similarly, for User 2 (Fig. 9) the word representation suggests that the user is involved in Energy, environment. Our system recommends environment specific events such as solar benefits and New York energy week (actual events are shown in Fig. 12) and the twitter account suggested for user 2 are

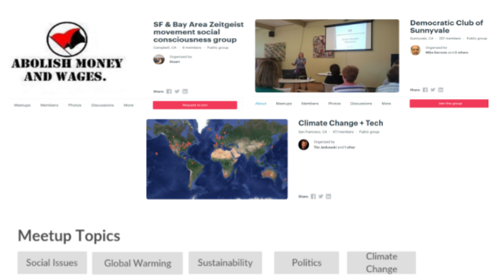

Fig. 11. User 1 results

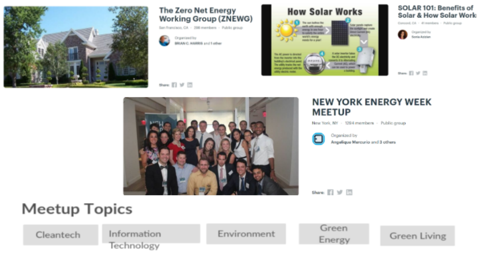

Fig. 12. User 2 results

Fig. 13. User 2 results

shown in Fig. 13

Thus our proposed recommendation system shows strong results that can potentially connect set of people who cannot connect relying on twitter or meetup only, Also giving meetup the opportunity to use twitter's large database to increase activity on its own.

V. Challenges

The project shows promising results, however from its main challenges is the sparsity of twitter data, which leads to confusing results, and needs a lot of cleaning, as data gathering step took a lot of time.

Our model shows a promising accuracy however we feel it be enhanced with Deep Learning models to improve classification, however this needs large data corpus. Also the model confuses with the overlapping subtopics of various categories. The Meetup generated topics are often overlapping for example climate change is considered political and environment event at the same time.

Also one problem is the same word can be used with different meanings, for example environment, can be considered as environmental issue when used like sustainable environment, and it can be classified as support group if was like "less stressful environment". These problems can be better enhanced with larger more clean data, and better labeled.

VI. Conclusion

In this paper we propose an enhanced recommendation system that suggests potential twitter users to follow in your area, and events that serves the same social interest the user have. This is done by fusing twitter and Meetup platforms.

We found that the people who organize events on meetup, and have their twitter account connected, mostly discusses the same problems on twitter, we are attributing this to meetup being a paid service, to organize a meetup you need to subscribe to a monthly plan, that's why whoever creates meetups believes so much in the idea and pretty active on social medias advocating for their beliefs.

When our system proposed accounts to follow on twitter, surprisingly these accounts wasn't popular or verified, and many of them didn't even have in their biography section any information about their social interests, however their tweets are all about social interests. based on the previous information, twitter's usual search would never suggest these accounts, since twitter's search engine is based on keyword matching with the account's biography, and biased towards the popular/verified accounts.

The results and the above observations suggest that our recommendation system is much stronger than just searching on twitter, as it connects regular people together, specially those who potentially wants to meet, also it depends on tweets content rather than keyword matching of twitter account's biography section.

We hope this will lead to more campaigns and action plans in real life of our system user's based on their virtual lives online.

References

[1] http://www.meetup.com/about/
[2] New York Times Consumer Insight Group, Why Do People Share Content?, 2016.
[3] http://www.soc.ucsb.edu/research/social-movements-revolutions-social-change
[4] Murphy, Dan (January 25, 2011). "Inspired by Tunisia, Egypt's protests appear unprecedented". The Christian Science Monitor.
[5] Lynch, Marc. After Egypt: The limits and promise of online challenges to the authoritarian Arab state. Perspectives on Politics.
[6] Gladwell, Malcolm (Oct 4, 2010). "Small Change: Why the revolution will not be tweeted".
[7] X. Geng, H. Zhang, Z. Song, Y. Yang, H. Luan, and T.-S Chua. One of a kind: User proling by social curation.MM'14.
[8] David Jderberg "Sentiment and topic classification of messages on Twitter "
[9] Kevin D., Rushin S., Bo L., Anatole G., Robert F. Topical Clustering of Tweets
[10] Yang, Li., Yuo., Pinterest Board Recommendation for Twitter Users (2015)
[11] https://meetup-api.readthedocs.io/en/latest/
[12] https://www.meetup.com/topics/
[13] https://www.aggdata.com/free/united-states-zip-codes
[14] http://docs.tweepy.org/en/v3.5.0/api.html
[15] http://scikit-learn.org/stable/modules/generated/sklearn.decomposition