Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Predicting Group Success in Meetup

Soumajit Pramanik[1], **Midhun Gundapuneni**[1], **Sayan Pathak**[2] and **Bivas Mitra**[1]
[1]Department of Computer Science & Engineering, IIT Kharagpur, India
[2]Microsoft Corporation, Seattle, USA
soumajit.pramanik@cse.iitkgp.ernet.in, g.midhun95@gmail.com, sayanpa@microsoft.com, bivas@cse.iitkgp.ernet.in

Abstract

Success of groups in Meetup is of utmost importance for members who organize them. However, measures of group success in Meetup is quite vague till now. In this paper, we take a step to quantify the success of Meetup groups. Driven by a comprehensive study of our Meetup dataset, we hand-pick a set of key properties which can potentially regulate a group's success. Finally, we develop a machine learning model leveraging on these features which can predict success of Meetup groups early with high accuracy.

Introduction

Over the recent years, Meetup[1], a popular Event Based Social Networking (EBSN) portal, has provided convenient 'online' platforms for people to create, and organize 'offline' social events. In Meetup, choosing suitable events to attend and selecting proper group to join require a lot of deliberation for the user. In order to mitigate this effort, different recommendation systems have been developed. The prior work stressed on the following three different classes of recommendation systems. (a) Event recommendation - this recommends suitable events to a single or a set of Meetup users based on the past preferences and current context (Luo et al. 2014), (Macedo, Marinho, and Santos 2015) etc. (b) Group recommendation - recommends groups that users are interested to join, considering both implicit and explicit factors, such as users' profile, location and social features etc (Zhang, Wang, and Feng 2013), (Liu et al. 2012). However, most of these systems cater the need of the general Meetup event attendees and group members. Importantly, only a few works like (She et al. 2015), (She, Tong, and Chen 2015) aim to provide proper guidance to the group organizers/coordinators and event hosts (jointly we refer as '*Meetup authorities*') in order to form a successful group or to host a successful event.

Studies show that all the Meetup groups do not survive over a period of time (Lai 2014). Survival of a Meetup group is directly connected to its capability of attracting population. This immediately raises the question- 'Can we develop a tool which can early predict the success of a Meetup group?'. However, there is hardly any consensus on the success measure of a Meetup group. As an EBSN platform, organizing popular events, attracting many attendees, can work as a success measure. On the other side, maintaining a large and consistently growing group could also work as a success yardstick. Hence, defining success measure for the Meetup groups is extremely important for the EBSN community.

In this paper, we develop a framework to predict the success of a Meetup group. The framework has been developed in a step by step manner. First we dissect the Meetup dataset, collected in three US cities, to highlight the major components of Meetup platform. Next we propose a principled approach to measure the success of the Meetup groups. We first identify a set of candidates metrics which may work as group success measure. However, one single metric may not be able to capture the success motive of all these diverse category of groups. So, we classify the different Meetup groups into five categories and identify one success metric for each category. Finally, we develop a machine learning based model for early prediction of the success of a Meetup group. We demonstrate that the proposed model achieves a high prediction accuracy of 0.80 (AUC=0.89).

Dataset

Data Collection

We crawl the Meetup dataset using public API for three cities of the United States namely New York, Chicago and San Francisco during a period of 3 months (from August 2015 to November 2015). We aim to crawl the temporal evolution of the existing groups (say the new joining members) at the higher granularity. However, low sampling rate of the crawler appears to be a significant bottleneck; it takes around $7 - 10$ days to crawl the information of all the events and members of each group in a city. Most of the groups change size at a higher rate than this. On the other hand, events hosted by different groups are less frequent and the details corresponding to them stay for a longer time (in 'Past' event category). In order to address this issue, we develop two different crawlers

(a) **Fast Crawl:** This is a fast crawler (cycle duration of 3 days) which collects only the members of all groups (no detail information) in each city and generates a member-to-group mapping along with timestamps. It does not crawl any

[1]http://www.meetup.com/about/

City	Fast Crawl		Detail Crawl				
	Groups	Members	Groups	Events	Venues	RSVPs	Member Profiles
Chicago	3470	280146	3437	384768	29607	2481365	213830
New York	11884	962006	11758	495581	44206	3473822	504143
San Francisco	8776	649510	8705	461808	43315	3578623	399678

Table 1: Data collected by both crawlers for all 3 cities

event or venue related information. This crawler is designed to collect the member dynamics across the Meetup groups.

(b) **Detail Crawl:** This is a slow but detailed crawler (cycle duration $7-10$ days) which collects the event details of all the groups. Data crawled by this crawler includes group details including member profile, events hosted by them, event RSVPs, event venues etc.

Table. 1 shows the statistics of the number of users, events and groups we crawled. In the following, we introduce the different actors and entities connected to the EBSN dataset.

Dataset: Major Components

Member and group profile: The profile of one member or a group gets specified by the set of Tags, which reflects their respective preferences. Whenever one member joins Meetup, she is asked to select some tags for describing her interest. Similarly, when a Meetup group gets formed by the group organizer, she is asked to select a set of tags which describes the group best.

Event attendance & attendees: In Meetup, for each event, there exists a field called "Headcount" which provides the actual attendance information of an event. However, this count does not provide the details of the individual attendees. On the other hand, details of the individual attendees can be obtained from the RSVP message {"Yes", "No", "Maybe"}. Event attendees for an event e_i are the users who send "Yes" response to RSVPs corresponding to that event.

Group category: During formation, each group is assigned to one of the 33 'official' categories defined in Meetup. For examples, few popular Meetup categories are 'Career/Business', 'Tech', 'Health/Wellbeing', 'Socializing' etc.

Measuring Success of a Meetup group

In this section, we define the success metric of the Meetup groups. As a first step, in the following we introduce a set of potential metrics to realize different signatures of success. Given a wide varieties of Meetup groups, one single metric may not be able to capture the success of all the groups. Next, we judiciously form the metrics to feature success depending on the specific characteristics of the groups.

Candidate Metrics

In this paper, we mostly focus on the popularity centric metrics to feature group success; nevertheless other aspects, such as post event sentiment etc. can also be explored. Popularity of a group can be broadly measured from two perspective - (a) size of the group - if it is able to attract new members to join the group (b) event attendance - if it is able to attract users to attend the events hosted by the group. For a group g organizing events $e_1, e_2, \ldots, e_k$ at

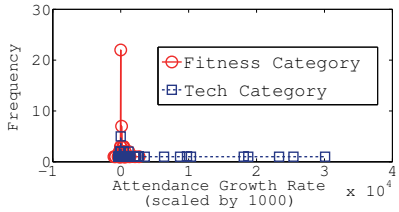

Figure 1: Distribution of Attendance growth rate for groups belonging to 'Tech' and 'Fitness' categories of New York
%vspace-0.1in

times $t_1, t_2, \ldots, t_k$, the candidate metrics can be mathematically defined as,

(a) **Average group size** at t_k, $G_k = \frac{\sum_{i=1}^{k} |g_U^{t_i}|}{k}$ where $g_U^{t_i}$ is the set of group members of group g at time t_i.

(b) **Average event attendance** at t_k, $E_k = \frac{\sum_{i=1}^{k} e_{i,H}}{k}$ where $e_{i,H}$ is the 'Headcount' of event e_i.

However, the aforesaid metrics fail to appreciate the newly created (small sized) groups, having potential to gain popularity in future. Hence, in the following, we introduce the metrics which pay importance to the rate at which the group size and event attendance grow over time

(c) **Event attendance growth rate** at t_k,

$$E_g = \frac{\sum_{i=2}^{k} \frac{e_{i,H} - e_{i-1,H}}{e_{i-1,H}}}{k-1} \quad (1)$$

(d) **Group size growth rate** at t_k,

$$G_g = \frac{\sum_{i=2}^{k} \frac{|g_U^{t_i}| - |g_U^{t_{i-1}}|}{|g_U^{t_{i-1}}|}}{k-1} \quad (2)$$

In summary, we use a candidate suite of 4 metrics $\langle G_k, E_k, E_g$ and $G_g \rangle$ to quantify the success of a group where each of the metric can be measured based on past k events organized by the group.

Category Group	Meetup official categories
Activity	dancing, fitness, sports/recreation, health/wellbeing, games etc.
Hobby	fine arts/culture, fashion/beauty, hobbies/crafts etc.
Social	movements/politics, socializing , singles, parents/family etc.
Entertainment	food/drink, movies/film , music, sci-fi/fantasy etc.
Technical	career/business, tech, education/learning etc.

Table 2: Meetup categories divided into groups

Category Specific Success Metrics

Key Idea: We aim to assign one group success metric for each Meetup category. The basic motivation is that for each category, the distribution of some of the metric values are either very concentrated or skewed. For example, in Fig. 1, we show that the distribution of 'Event attendance growth rate' E_g is highly concentrated for groups belonging to 'Fitness' category whereas for 'Tech' category it is well distributed. In case of the concentrated distribution, most of these groups have very close value of that metric E_g . This in turn says that E_g is unable to make any distinction between the groups of 'Fitness' and hence, should not be used to discriminate "successful" and "unsuccessful" groups. However, for 'Tech' category, E_g may appear as a promising metric.

Methodology: The objective is to identify the most discriminating success metrics for each Meetup category. However, one major challenge is that many Meetup categories contain only a few groups. In order to address this data sparsity issue, we propose to classify the official Meetup categories into the following five classes - a) Activity b) Hobby c) Social d) Entertainment and e) Technical (see Table. 2). Hence, in all the following experiments, we work on this new set of categories.

We use 'Entropy' to characterize the discriminative property of each candidate metric. For example, in case of 'Activity' category, we measure the entropy of 'Average group size' G_k as $\sum_{i=1}^{M} -p_i \log p_i$ where M is the number of different G_k values of groups belonging to 'Activity' category and p_i is the probability of each such value. In Table. 3, we present the entropy of each candidate metric for every category in Chicago. We use a 66.67^{th} percentile (highest one-third) as threshold on this entropy values (7.03 for Chicago) and select the metrics crossing this threshold as the group success metric of that category. In case, none of the metrics are above the threshold, we choose the one with maximum entropy.

Labeling groups: Once the success metrics are chosen for each category, we label the groups belonging to that category as 'successful' or 'unsuccessful'. In this labeling process, for the category with only one success metric say 's', we simply assume that the groups having more than 66.67^{th} percentile value of 's' are 'successful' and less than 33.33^{th} percentile value of 's' are 'unsuccessful'. For the categories with multiple success metrics, we use a 'Veto' strategy. We label a group as successful if it has more than 66.67^{th} percentile value for at least one of the chosen metrics. On the other hand, if no metric labels one group as 'successful' and additionally if it has less than 33.33^{th} percentile value for at least one of the chosen metrics, then that group is labeled as 'unsuccessful'. The number of 'successful' and 'unsuccessful' groups labeled for each category in Chicago is shown in Table. 3.

Category	E_k	E_g	G_k	G_g	**Successful / Unsuccessful**
Activity	4.06	7.46	**7.48**	6.18	133 / 86
Hobby	3.83	**7.03**	6.89	6.32	88 / 55
Social	4.27	**7.50**	7.40	6.43	127 / 77
Entertainment	3.87	**6.07**	5.99	5.63	23 / 24
Technical	4.85	**7.66**	7.40	6.67	136 / 74

Table 3: Entropy values for different sucess metrics for different category of groups in Chicago (considering only groups organizing more than 10 events)

Group Success Prediction

In this section, we introduce (a) Core members and the (b) New members who play an important roles in making a group successful. Next, we define a set of user specific & event specific features which can potentially influence success of a group. Finally, we apply Machine Learning models to leverage on those features and predict the success of a Meetup group.

Key Players

a) Core members: Informally, the core members of a group are the dedicated set of members who have a strong interest overlap with the group. In order to identify the core, we propose a similarity metric between tags of the members and the groups.

Tag similarity ($TagSim$): We represent tags of an individual member/group as a tag vector TV of length N_T where N_T is the number of all possible tags. The coefficient of each component (tag) of this vector is the normalized usage frequency of the corresponding tag. The similarity between two tag vectors $\mathbf{TV_i}$ and $\mathbf{TV_j}$ is calculated as the cosine similarity of these two vectors i.e. $TagSim(\mathbf{TV_i}, \mathbf{TV_j}) = \frac{\mathbf{TV_i}.\mathbf{TV_j}}{\|\mathbf{TV_i}\|\|\mathbf{TV_j}\|}$

We define core members as a subset of group members who have a very high tag similarity with the group tag vector.

b) New members: In Meetup, people search for events and if they intend to participate in one event, they need to join the organizing group first. We define the new members as a set of users who join the organizing group g just before an event e_i (i.e. in between t_i and t_{i-1}) for $i = 1, 2, \ldots, k$. This has been observed that a significant fraction of event attendees join the group just before the event.

Features

The features predicting success of a group g can be broadly divided into the four classes -

- **Semantic or tag related features** - Average tag vector similarity between the organizing group & the group members, Average intra member tag vector similarity etc.
- **Syntactic or count based features** - Average pairwise count of common past events between group members, Fraction of group members sending 'Yes' RSVP etc.
- **Time related features** - Day of week on which the event occurred, duration of the event etc.
- **Location related features** - Average pairwise distance between group members, Average distance between the event venue & the group members etc.

We calculate these features separately for event attendees, core members and new members.

City	Naive Bayes		SVM		Decision Tree		Logistic Regression	
	ACC.	AUC	ACC.	AUC	ACC.	AUC	ACC.	AUC
Chicago	0.66	0.63	**0.71**	0.78	0.66	0.74	0.69	**0.82**
New York	0.78	0.78	0.76	0.87	**0.80**	**0.88**	**0.80**	**0.88**
San Francisco	0.81	0.80	0.80	0.88	**0.83**	0.83	**0.83**	**0.89**
Combined	0.76	0.75	0.77	0.84	**0.79**	0.84	0.77	**0.87**

Table 4: Classification accuracy values (ACC.) and regression AUC values for all 3 cities

Category	Naive Bayes		SVM		Decision Tree		Logistic Regression	
	ACC.	AUC	ACC.	AUC	ACC.	AUC	ACC.	AUC
Activity	0.80	0.77	0.80	0.84	**0.85**	**0.89**	0.83	0.88
Hobby	0.79	0.77	0.77	0.82	**0.81**	**0.84**	0.79	**0.84**
Social	**0.78**	0.74	0.77	0.86	0.76	0.80	**0.78**	**0.87**
Entertainment	0.70	0.67	0.69	0.65	**0.73**	**0.79**	0.72	0.73
Technical	**0.71**	72	0.69	0.80	0.66	0.78	0.69	**0.82**

Table 5: Classification accuracy values (ACC.) and regression AUC values for all 5 categories

Evaluation

We develop three different versions of the prediction model - (a) City specific model where we only consider groups in a specific city (b) Category specific model where we develop models for individual categories and (c) 'Combined' universal model considering all the groups of different cities. We label each group as 'successful' and 'unsuccessful' using the chosen metric, based on the most recent k events organized by that group. In our experiments, we take $k = 5$; however k can be varied from 3 to 5. In order to calculate features, we use past 5 events starting from $k + 1^{th}$ event in the reverse chronological order.

Observations

We demonstrate the prediction results using four standard classifiers - Naive Bayes, Support Vector Machine, Decision Tree & Logistic Regression. The classification accuracy results for city specific models & the 'combined' model are shown in Table. 4 and the results for category specific models are shown in Table. 5. On average, we get around 70% to 80% accuracy for both city specific and category specific models. We follow the standard thresholding procedure and compute the precision-recalls as well as corresponding AUC values (shown in Table. 4 & Table. 5) for different classifiers. On average, we get around 0.75 to 0.85 AUC values for both city specific and category specific models. In Fig. 2, we show PR curves corresponding to all classifiers for the 'combined' model. Here Logistic Regression outperforms other models at the lower recall regions.

Finally, we turn our attention to the relative weight of individual features, which identifies the key features influencing success of a group. We observe that, across different models, tag based and count based features (say intra member tag vector similarity, pairwise count of common past events between group members etc) have shown importance for almost all types of groups.

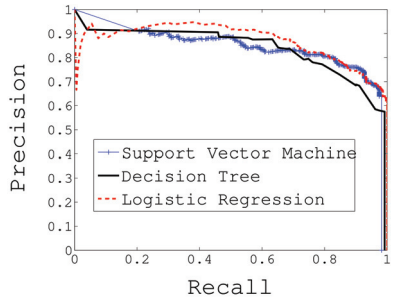

Figure 2: Different ML techniques' performance using 'combined' model.

Conclusion

In this paper, we have developed a framework to predict the success of Meetup groups, considering the diverse objectives of the *Meetup authorities*. We have proposed a principled approach to fix success yardstick of a Meetup group. We have presented a simple machine learning based model to predict group success by leveraging both semantic as well as syntactic features; the model achieves an average accuracy of 80%. Moreover, we have developed individual city specific & category specific models, as well as a 'combined' model. While the 'combined' model performed close to New York and San Francisco, Chicago seems to significantly under perform. Overall, we observe that Logistic Regression with L2 regularization appears as the most suitable model for our experiments.

Acknowledgements

This work has been partially supported by the SAP Labs India Doctoral Fellowship program and DST - CNRS funded Indo - French collaborative project titled "Evolving Communities and Information Spreading".

References

Lai, C.-H. 2014. Can our group survive? an investigation of the evolution of mixed-mode groups. *Journal of Computer-Mediated Communication* 19(4):839–854.

Liu, X.; Tian, Y.; Ye, M.; and Lee, W.-C. 2012. Exploring personal impact for group recommendation. In *CIKM 2012*, 674–683. New York, NY, USA: ACM.

Luo, C.; Pang, W.; Wang, Z.; and Lin, C. 2014. Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations. In *ICDM 2014, Shenzhen, China, December 14-17, 2014*, 917–922.

Macedo, A. Q.; Marinho, L. B.; and Santos, R. L. 2015. Context-aware event recommendation in event-based social networks. In *RECSYS 2015*, 123–130. New York, NY, USA: ACM.

She, J.; Tong, Y.; Chen, L.; and Cao, C. C. 2015. Conflict-aware event-participant arrangement. In *ICDE 2015, Seoul, South Korea, April 13-17, 2015*, 735–746.

She, J.; Tong, Y.; and Chen, L. 2015. Utility-aware social event-participant planning. In *SIGMOD 2015*, 1629–1643. New York, NY, USA: ACM.

Zhang, W.; Wang, J.; and Feng, W. 2013. Combining latent factor model with location features for event-based group recommendation. In *KDD 2013*, 910–918. New York, NY, USA: ACM.