Activity and Topical Influence in Multi-interest Communities

Zubin John
Department of Computer Science & Engineering
The Ohio State University
Columbus, Ohio 43210

Abstract

Online communities are growing constantly at a rate where major communities are now at a size comparable to small nations. The information exchanged in these communities adds to the human store-house of knowledge. The nature of this information is a subject of a great deal of research. The problem we want to focus on is more pertinent to users in these communities. The knowledge exchange is possible only due to the users who interact and exist as nodes in this gigantic network. In online communities now-a-days a recurring problem is retaining interest of users and ensuring that users remain active. We propose a novel framework which integrates a number of models to predict the activity of new or sophomoric user in a multi-interest community namely Yelp. We introduce Topical Influence Models in Activity prediction which go on to out-perform traditional methods and our baseline by a wide-margin.

Introduction

Activity, Churn and Relevant Jargon

Online user communities are a thriving sector of Internet and social media. This digital incarnation of special-interest groups has evolved to a stage where the average population of users has grown to the order of millions, millions seeking peer generated opinion or a channel to reach out to a pool of like-interested users and express this opinion. There are online communities dedicated to followers of a single interest like Goodreads, IMDB and last.fm these are termed as special-interest online communities. Others like Yelp and Foursquare provide a number of different interests for users to rate and interact about. Lastly, another group of user-communities generating interest on the web currently is termed as CQA or question answering.

For any class of online communities the users play an important role in their sustenance and evolution. In order for the community to thrive it must generate interest, draw users and retain their attention. Marketing and growth plans deal with the first two aspects of these communities. The latter requires a lot of statistical analysis and planning as well as a reward-based system for retention of community members. Community administrator must follow a practice known as activity prediction or churn prediction to try and predict the activity of certain users in the community. This allows an understanding of how vibrant the community currently stands, and identifies potential exiting members for the community. The work described in our paper deals with predicting the activity of newly joined members in order to make their interaction more enjoyable and prevent the loss of these members from the community. One of the motivation behind this is that in a community it is quite difficult to breed contributing participants, and hence an easier task to identify such members and prolong their activity as a contributing members.

As a parallel contribution of this paper we are able to identify many of the parameters which influence and effect the quality of interaction the members of the community may experience. An online community in essence symbolizes an investment-payoff model where the continued interest of a user is guaranteed by the correct reward for her efforts. This can be imagined as very similar to a game scenario, where the user is motivated in her endeavors by significant payoffs along the road providing an en-richening experience. By way of learning to predict the activity of a user we are simultaneously also exploring her motivations for continued interest. We hope in culmination the work suggests not only how long a user would contribute to the community but also the factors that assure her patronage.

Related Work

There is recent work that has been conducted in this broad area of learning to predict churn and/or activity life cycle for users of such communities. Dror et al. For instance were able to predict churn in Yahoo Answers a community based QA portal using features extracted from profile information such as age, gender and time required to answer along with social relationship elements. With multiple classifiers they achieved the best result using a Random Forest approach with a F-1 score of 0.755.

Similarly Yang et al. (2012) focuses on identifying factors for sustainable motivation of community members, following a series of statistical approaches they were able to report patterns in experiences and participation that affect a user's behavior on sites such as Yahoo Answers!, Baidu and Naver-Knowledge-iN South Korea's larger QA community. Their proposal of first post experience influence is an intuitive indicator of counter-active behavior modeling. Within the scope of our experiments as well, we have verified that interaction-based features from first few experiences are perhaps the most informative in the task of forecasting activity levels.

Prior to exploration in the online community sphere churn prediction has been the source of investigation in other peer-based and user-based bases. In their work Birtola et al. (2013) explore churn analysis in a e-commerce aimed at supporting merchants in highlighting at-risk churn customer. Borbora and Srivastava (2012), model user behavior using game logs for participants in Mass-Multiplayer Online Roleplaying Gaming, another trend set to break the Internet. MMORPG's as they are colloquially known record extensive user activity information in logs which were exploited to model player churn behavior and life-cycles. Activity traits of regular players were successfully used to discriminate players with declining interests from the population, outperforming traditional systems.

More recent work has introduced a lot of new outlook and possibilities in achieving accurate forecasting of potential drop-outs. Emotions and sentiment are potential indicators for user satisfaction and should forge a direct consequence to her longevity. Along these lines Garcia et al. (2013) ran puts forward an unconventional setup to understanding the social dynamic to emotions in an open-source software (OSS) collaboration project community (Gentoo). The work terms bug trackers and forums as special social media outlets in their case-study and through Bayesian classification are able to predict real-time contributor turnover in OSS projects.

Most influential to our work is perhaps Danescu-Nicelescu-Mizil et al. (2013) which discusses implications of linguistic evolution in online communities and their and how the subsequently affect a mature user leading to eventual abandonment. The work introduces a somewhat revolutionary paradigm to churn and activity life-cycle understanding in a single interest online community. The work here also lead to study in understanding behavioral differences between neophytes and experienced users in a community as illustrated in McAuley & Leskovec (2013).

Aspects of user growth and personal development are interesting directions, which however have not been pursued in our work. Our work strives to study users in a more sophomoric phase, to understand the characteristics motivating their interest and the communal rewards that aid in their membership. Again, since our work focuses on a multi-interest user community, which introduces new challenges of its own it is intractable to model linguistic preferences. Inspired by Ferrera et al. (2014) which studies topical interests in another single-interest community, namely Instagram we observe that topical interests are a great tell on what's trending; in addition it mirrors the individual and community-as-a-whole interest preferences. While looking at Yelp, Business categories for each review are provided, however these features compare to traditional methodology,. We propose a novel topical study based framework to identify topical meta-attributes or very specific sub-topics trending in the popular reviews in the community. Exploiting these meta-topics we gain an insight into the reviewing norms that are propagating and receiving appreciation consequently we can categorize user activity-levels in just how in-sync they are with these undercurrents.

Data and Models

Yelp Academic Dataset

The Yelp Academic Dataset (2014) has been published for academic research purposes and collects information regarding businesses, Yelpers (Yelp users), reviews and check-ins. The description page claims that the data has been collected in and and around Arizona for about 1000000+ reviews. Of particular interest to us for the task of activity prediction, we require access to users and their subsequent reviews. We have defined what we like to call a Two Year Forecasting Model which will be described in more detail a little later. As part of this model we filtered out all results for years apart from 2013 and 2014 (for motivation behind this, refer the next section). Our resultant dataset hence contains information form about 6000+ Yelpers and their subsequent reviews.

There are other preprocessing steps taken to better fit the data to our framework. Since our objective is to study the abandonment in relatively new users we look at the users that have joined in our two years spanning from January 2013 to August 2014. To select data native to early experiences that users face in their first few posts we restricted our data-points to considering only each user's first five reviews irrespective of the number of their posts available. A limitation that we found with the data was that a lot of the user data-points had keys in their list of reviews however the values of these keys when used to extract the corresponding reviews were non-existent in the data set. We believe there may be two reasons responsible for this inconsistency; firstly, it is possible that the Yelper profile information contains a log of all posts that were made at a point however the reviews could not e pulled out as they have been deleted. Secondly, since the reviews and user data-points are only location specific to Arizona, it is possible that certain reviews that have been posted for businesses outside of AZ by a AZ centered user have not been included in th dataset. In any case, the volume of 6000+ data-points and 50,000 reviews performs to our satisfaction according to our Two Year Forecasting Model.

The Two Year Forecasting Model

Online Communities are ever in a state of flux, with new content being brought in by new users and focus of trends or temporal interests evolving and changing. With this in mind, it would not seem wise to perform forecasting for a wide span of time. It is more ideal to divide a huge time-span into a time series of multiple models each covering a limited time-frame of two years. We call this our Two Year Forecasting Model. We can build a user topic model network for two years which holds sensitive information for its corresponding interval. With prediction for every two years it is quite easy to calculate a gradient of decline in activity frequency i.e posting over time. Consequently, we at present looking to predict activity for only sophomoric users and hence, two years is an apt time-span.

First Experience Modeling Paradigm

A unifying principle for Behavior Models that we are considering in online communities is that the first experience plays an important role in both developing a user's interaction with the community, her growth and her life-cycle in the community. Our task adheres uniformly with this paradigm in feature selection and lays heavy emphasis on modeling interaction. A lot of parameters are associated with the user's first experience criticism/feedback, emotions involved, linguistic style, outlook and connectivity in social network. We have weighed in our choice of models with these in mind. Modeling sentiment involved takes care of the emotional range of the Yelper, however with the nature of identifying sentiment being a challenging problem in itself, this proves to be difficult feature to integrate. Inconsistencies and errors with sentiment analysis may trickle forward in the training of our prediction models and result in poor performance. Sentiment must hence be handled with care.

Linguistic style and network connectivity have been weighed down in our model for reasons. The scope of our work did not involve integration of the former which requires expert analysis of its own. It is quite possible for future work to extend on our framework by taking the said feature into consideration and delivering results. Network Connectivity was overlooked because of the static nature of such models. Like content the network graph is ever in flux, the task of analyzing the network demographic of the user at the time of each of her five initial posts would require capturing these dynamics with a framework which would have been unsuitable to our task. For the sake of computational and space efficiency we again ignored the network-relevant information. Consequently, the Yelper preferred post and feedback behavior has been captured by the topic and interaction models respectively.

Statistical Features

For computation of activity predictions our first setup is built on a purely statistical model. One composed of statistical features descriptive of the review/post shape of each user. These descriptive features have been prepared and presented in the dataset itself. The dataset offers user descriptors such as number of friends, number of fans, number of reviews, average review rating and so on. We simply extracted these features from the dataset to build a model composed of two statistical features a average review rating computed for first 5 reviews and secondly the average size of first three reviews.

Interaction Models

To capture the nuances of first interaction experience and feedback we take into consideration interaction models built around the response that users receive on their first few reviews. These integrate feedback on how cool, funny or useful people thought the review posted was. Users offer this feedback in terms of a vote on Yelp. Following our rewards analogy, we believe this to be a motivation which fuels the user's interest in the community. These votes denote a positive experience for the Yelper and follow a direct relationship. In our analysis, the useful vote category seemed to be highly correlative to a successful first experience, which again ties in with (McAuley & Leskovec 2013),(Danescu-Nicelescu-Mizil et al. 2013); suggesting funny and cool votes are relative to user maturity and hence while sophomoric users do receive them, they might not amount to a great value in this early phase.

Sentiment Models

Relatively there is no dearth of research indicating strong ties between user emotion and satisfactory synchronization with on line communities; (Garcia et al. 2013) for instance. However,we choose to tread very careful with our sentiment model as it is unreliable to say so in the very least. Off the-shelf tools for sentiment analysis do not promise high levels of accuracy for one; the reason being that polarity and valence scores in this domain are highly dependent on the knowledge base. There is no all-cure solution which can promise accurate results irrespective of domain. Perhaps, what we needed for this purpose was a model trained on review data in the absence of which we were skeptical of our expectations.

Topic Influence Models

The intent with topic models was to expose the inherent meta-topics in the document cluster. Building individual topic models for each of the three voting categories of funny, cool and useful we expected to capture that most discussed, and propagating trend genomes. By a trend genome we envision a unit that follows an almost viral pattern in carrying forth popular descriptive patterns and critiquing/ appreciatory styles in new Yelpers.

For instance, new Yelpers or novice reviewers tend to include an *inclusive* property a lot in their reviews. This basically describes a review that talks about every aspect of the business being reviewed, however in our topical influence study: reviews with the *inclusive* trait are considered neither funny, cool nor useful. On the other hand properties such as describing *food_service*, *amenities* are definitely considered useful.

Our models constructed topical clusters to exhibit the influence of each of the three individuals model. Since three is better than one and offers more diversity we decided to go with three unique models supporting individual contributing micro-properties under each review type. The 10 most significant clusters are listed in Table 1 in descending order of influence quotient the largest denoting a statistically dominant property.

Funny	Cool	Useful
food-service	food-service	food-service
amenities	diversity	frndsnfam
punctuality	revisit	disappoint
meat	order	time
price	frndsnfam	inclusive
happy-hour	amenities	revisit
revisit	location	amenities
parking	room-service	diversity
location	time	parking
inclusive	portions	management

Table 1: Top 10 Micro-topic clusters for each review vote attribute

Latent Dirichlet Allocation was used to build the topic clusters; over 100 iterations with the per-decided cluster size of 20 clusters. In the future we intend to explore the possibilities with finer level of granularity. Currently, however there was some manual labeling involved on behalf of one of the authors. With clusters beyond the size of 20 there was observed overlap among st the topics. We decided to preemptively adhere to using 20 clusters in order to ensure that hand-labeling them with the decided size would be less noisy.

Experimental Setup

Ridge Regression

We decided to go forth with a ridge regression model for training and classification. A thresholding function was then used to map regression scores into class labels for evaluation. The data being strictly linear in nature fits well to such a training model. It behaves similar to a multi-class categorizer however is more efficient in terms of computational cost for training. For the purpose of evaluation we divided our data into a training and testing set by ratio of 1:3 with points being divided randomly. As a result it is possible that evaluation scores may vary from run to run; the scores and performance reported in our work is the best performance that was observed in 10 different runs of the code. The pattern of improvement with ablation of features however is consistent for our framework and validates our results. The translational mapping for converting regression scores into class labels has been recorded in the table below.

Model	Features
Statistical	Average Rating, Average Length
Interaction	Votes for Funny, Useful, Cool
Sentiment	Average Sentiment
Topical Influence	Jaccardian Similarity

Table 2: Models - Features

Statistical and Interaction Features

As mentioned the statistical and interaction features are first and second order features calculated directly from values reported in the data set, following minimal preprocessing they have been brought down to suitable ranges for the purpose of a clean and uniform feature space. It is quite possible that unseen test features might exceed theoretical thresholds established while training, however we believe our model deals well with such anomalies and tunes them out by virtue of training.

Sentiment Features

To translate sentiment scores for sentences into usable features, the setup iteratively calculates sentiment for each sentence in the review space of the first 5 reviews, then averages this value down to assign it as the average sentiment associated with the user review-space. The values are normalized to a range of -1 to 1, the boundaries indicating a higher degree of negativity or positivity evident in language respectively.

Translating Topic Models to Topic Influence Ratings

To indicate heavy influence of a micro-topic in a Yelper's reviewing style we rely on the Jaccardian distance measure. The Jaccardian distance is calculated individually as a feature for the broad vote inspired category it falls into. For instance each user has a $Funny_{jacc}$, $Cool_{jacc}$ and $Useful_{jacc}$ denoting correlation between their personal topical preference over the topical influence/ preference of the community as a whole.

$$A_{jacc} = \frac{|A_{topics} \cap Community_{topics}|}{|A_{topics} \cup Community_{topics}|}$$

Results and Discussion

We used a baseline of predicting every Yelper as a rarely active member of the community. This staring baseline gave us a F-Score of 0.167. Working incrementally we evaluated the data with each of our feature models to achieve successful predictions with a highly significant increase in F-score. The best performance was observed with the interaction and topical influence models.

Features	Precision	Recall	F-Score
Baseline	0.219	0.133	0.167
+ Statistical	0.697	0.424	0.527
+ Interaction	0.707	0.428	0.533
+ Sentiment	0.707	0.428	0.533
+ Topical Influence Ratings	0.745	0.451	0.562

Table 3: Performance with different feature models

Why Topics and Interaction are indicative

The results from the above said models are quite in line with the expectations. Interactions as mentioned earlier speak a lot about the early inclination of a young Yelper towards the community in staying active. The rewards/payoff functionality is perhaps embodied best by this model.

Topical preference on the other hand is highly insightful in exposing true reviewing intent and patterns of the user. It speaks of a stable cast that the user relies upon to get his opinions about a business though. The social influence dynamic is not as ubiquitous in multi-interest communities as their single-interest counter-parts. For instance we repeatedly observed for a user there is a high correlation amongst the reviews voted either funny, useful or cool. The writing cast enumerates almost the same properties for a user irrespective of the votes her reviews received.

Why did Sentiment fail?

Quite frankly, this can can be attributed to the significant difference between linguistic sentiment in review and emotion involved with the community. The first is indicative only of a repeating trend in the author's writing style and their inclination towards a stand-point. Reviews in general are almost never neutral in nature , heavy criticism or extolling is not indicative of the writer's mood. In contrast while social media is a great indicator of the author's mood, reviewers are normally detached. Rants and praises mean nothing at all in the environment in question. This nullifies any possible correlation that can be drawn between users in a multi-interest community their activity and attached emotion; in light of the unavailability of such information. In our perspective this quite possibly puts the debate of influence of emotional states in churn prediction at-least with respect to linguistic sentiment analysis.

Conclusion

We proposed a framework that suggests the use of four models based on first-order statistical features, interaction features, sentiment and topical influence features. The use of topic models in churn prediction is a novel previously unexplored system which intuitively performs much to our expectations. Consequently, as a result of this work we would like to believe there are some sound observations that contribute to understanding parameters associated with abandonment. We believe that psychological analysis to this end can help its application in preventing the loss of young community members. Lastly, we believe that positive results with topic models are highly encouraging in their continued used as as analytical models in on-line communities.

Future Work

It is possible to envision quite a few enhancements to our work. The first as stated before is analysis of micro-topics in higher-granularity; with further dedicated effort it is quite possible to identify a larger number of potential meta-attributes to serve as topical properties and add greater variance to our topical influence models. Second, it would be interesting to cascade multiple Two Year Interaction models together to try and observe the activity life-cycle of users in a multi-interest online communities similar to the work in (Danescu-Niculescu-Mizil 2013). Finally, another potential down-stream application of our task would be to try and predict elite users; Yelp rewards certain users with a Elite title for their continued activity on an yearly basis. The Elite would be a subset of our frequently active class users and our models and data could be easily expanded to rank top k users as elite.

References

Borbora, Z. H., & Srivastava, J. 2012. User behavior modelling approach for churn prediction in online games. *In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)* (pp. 51-60). IEEE.

Birtolo, C., Diessa, V., De Chiara, D., & Ritrovato, P. 2013. Customer Churn Detection System: Identifying customers who wish to leave a merchant. In *Trends in Applied Artificial Intelligence* (pp. 411-420). Springer Berlin Heidelberg.

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 307-318). International World Wide Web Conferences Steering Committee.

Dror, G., Pelleg, D., Rokhlenko, O., & Szpektor, I. 2012. Churn prediction in new users of Yahoo! answers. In *Proceedings of the 21st international conference companion on World Wide Web* (pp. 829-834). ACM.

Ferrara, E., Interdonato, R., & Tagarelli, A. 2014. Online popularity and topical interests through the lens of instagram. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 24-34). ACM.

Garcia, D., Zanetti, M. S., & Schweitzer, F. 2013. The Role of Emotions in Contributors Activity: A Case Study on the GENTOO Community. In *Cloud and Green Computing (CGC), 2013 Third International Conference on* (pp. 410-417). IEEE.

McAuley, J. J., & Leskovec, J. 2013. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 897-908). International World Wide Web Conferences Steering Committee.

Rehurek, R. & Sojka, P. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45-50).

Yang, J., Wei, X., Ackerman, M. S., & Adamic, L. A. 2010. Activity Lifespan: An Analysis of User Survival Patterns in Online Knowledge Sharing Communities. In*ICWSM*.