Intentional analysis of medical conversations for community engagement

Saurav Sahay
School of Interactive Computing
Georgia Tech
E-mail: ssahay@cc.gatech.edu

Hua Ai
School of Interactive Computing
Georgia Tech
E-mail: hua.ai@cc.gatech.edu

Ashwin Ram
School of Interactive Computing
Georgia Tech
E-mail: ashwin@cc.gatech.edu

Abstract

With an explosion in the proliferation of user-generated content in communities, information overload is increasing and quality of readily available online content is deteriorating. There is an increasing need for intelligent systems that make use of implicit user generated knowledge in communities for community engagement. We describe our approach based on modeling user utterances in communities to proactively target the community for exchange of questions and answers. We envision a system that automatically encourages user engagement and participation by routing relevant conversations to users based on individual and community activity levels. In this paper, we analyze health forum conversations from WebMD, a popular health portal consumer site, and classify them in different acts of speech using Verbal Response Modes (VRM) theory. We describe our approach for modeling an intelligent community recommender to engage participants based on observations from our analysis.

Introduction

Virtual communities have emerged due to the recent advances in computer-mediated communication infrastructures and web technologies. Nowadays, people are always connected, whether in their daily lives or in their activities online. At the same time, the highly networked environment has trigged researchers to study how individuals behave differently in on-line social environments versus in more traditional face-to-face contexts. Online social communities are found to exhibit more uneven participation distributions. Within small group sessions, it is common for the top few active participants to account for 50-75% of the communication activity, while the less active participants contribute very little relatively. Some early research shows that participation differentials may be due to status differences(Saunders, Robey, and Vaverek 1994)(Weisband and Connolly 1995) and differences in individuals expectations regarding participation(Rojo 1995). Research studies have looked into different ways to motivate community contributions. (Kollock and Smith 1996) use a theory of effectively managing group resources as design principles to analyze the successes and failures of Usenet. A significant amount of research has been devoted to enhance online communities through expertise finders systems(Krulwich and Burkey 1996)(Kautz, Selman, and Shah 1997). Such systems(Lenz, Hbner, and Kunze 1998) identify people who have expertise to answer certain types of questions.

In this paper, we describe techniques of our socio-semantic recommendation system(Cobot) for health related conversation routing. Intelligent information agents in the community route conversations to relevant users with engagement models for balanced recommendations. One of the advantages of this balanced recommendation approach is sustained engagement with less community dropout rates. We have crawled and studied existing conversation patterns of a medical community and analyzed them to guide our approach for modeling recommendations with different engagement models. These engagement models depend on user's activity and Verbal Response Modes.

System Description

Verbal Response Modes is a principled taxonomy of speech acts that can be used to classify literal and pragmatic meaning within utterances(Lampert, Dale, and Paris).[1] In this learning task, utterances are classified into disjoint sets comprising Question(Q), Disclosure(D), Edification(D), Advisement(A), Acknowledgement(K), Reflection(R), Interpretation(I) and Confirmation(C). We crawled 12000 conversations from WebMD forums consisting of 3260 users to train and test our VRM classifier.

Choice of Features The choice of features to predict the type of utterances is extremely important. We have used a mix of contextual, syntactic and semantic features for our data. We have extracted the following features for our task: *Number of words*, *First word*, *Last word*, *word bigrams*, *word dependencies (1st/2nd/3rd person subject, inverted subject-verb order and imperative verbs.)*, *Morphology*, *Hand constructed word lists*, *Wh words*, *Top n words*.

In order to develop the VRM classifier that could categorize the conversations at sentential level, we (two annotators) manually tagged 175 conversational sentences for a total of 1941 instance training set including the VRM training data. This includes conversation data annotations from the

[1] We thank the authors for providing us with the data they used in their research

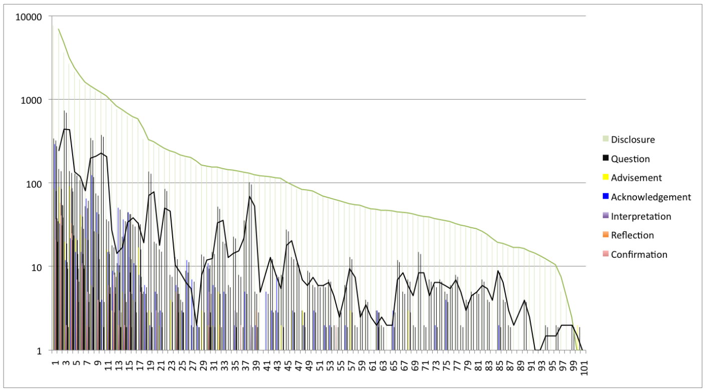

Figure 1: Community Intentional Analysis

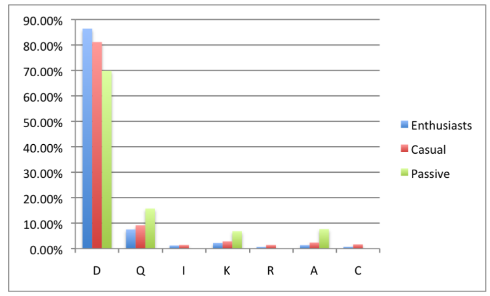

Figure 2: Community Intentional Parameters

original VRM codes augmented with our domain specific annotations. We report our 10-fold cross validation accuracies as follows. Our SVM based classifier achieves 10 fold cross-validation precision of 72.3%, recall of 75.3% and a F-Measure of 73%. In our classification task, we combined Disclosure and Edification categories into one as our classifier was confusing with these categories. The classifiers are good at capturing the literal meanings while the Edification category seemed more like a pragmatic concept to us which could be expressed in terms of other categories. Our top features in this classification task were domain independent features such as '?', length of words, 'you', 'i','okay', 'well', etc.

Community Modeling

After training, we ran the classifier on the crawled data and analyzed the top 100 most active users in our data. Since we had crawled only a few forums, we didn't know the overall activity of users hence we sampled from the top 100 users having at least 5 different utterances. Our goal is to create three different user engagement models, an enthusiasts model, a casual model and a passive model of community participants. We can route conversations to users based on their actual and expected activity levels in these groups thereby reducing information load on topic experts and preventing community dropouts.

Our observations are summarized in the Figure 1 (with Trend lines). The first quartile (42.5 users) consists of very active users (note the Y-axis logarithmic scale), the next quartile consists of casual users and the third and last quartiles consist of passive user community. Not so surprisingly, many of the top users were WebMD forum personnel who actively sought to increase participation in the community. We also noticed that people in casual and passive user model groups asked more questions compared to the enthusiasts in this context as shown in Figure 2.

We propose the following high-level approach for engagement based recommendation modeling:

1. Classify users' responses into questions (Q) (Question VRM category), answers (A) (Disclosure, Edification and Advisement VRM category) and miscellaneous (O) (Acknowledgement, Interpretation and Reflection VRM categories)
2. Learn a question-answers-miscelleneous proportion(Q:A:O) of enthusiasts, casual and long tail user models (with A more than Q more than O constraint)
3. Categorize each user into the enthusiast, casual or passive model.
4. Given a question, get top n topically relevant users (based on topic relevance models)
5. Re-rank users based on their actual engagement levels (if a user is under engaged, prefer him before others, don't choose this user if he has already met his desired engagement level)
6. Recommend users
7. Update activity after users have responded.

We are deploying our system and testing our recommendation approach in a large user community for a systematic community engagement based social recommendation system.

References

Kautz, H.; Selman, B.; and Shah, M. 1997. Referral Web: combining social networks and collaborative filtering. *Commun. ACM* 40(3):63–65.

Kollock, P., and Smith, M. 1996. Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. Amsterdam: John Benjamins. 109–128.

Krulwich, B., and Burkey, C. 1996. Contactfinder agent: answering bulletin board questions with referrals. In *The 1996 13th National Conference on Artificial Intelligence, AAAI 96. Part 1 (of 2)*, 10–15.

Lampert, A.; Dale, R.; and Paris, C. Classifying speech acts using verbal response modes.

Lenz, M.; Hbner, A.; and Kunze, M. 1998. Textual cbr. In Lenz, M.; Burkhard, H.-D.; Bartsch-Sprl, B.; and Wess, S., eds., *Case-Based Reasoning Technology*, volume 1400 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 115–137. 10.1007/3-540-69351-3$_5$.

Rojo, A. 1995. *Participation in scholarly electronic forums*. Ph.D. Dissertation, University of Toronto.

Saunders, C. S.; Robey, D.; and Vaverek, K. A. 1994. The persistence of status differentials in computer conferencing. *Human Communication Research* 20(4):443–472.

Weisband, Suzanne P., S. S. K., and Connolly, T. 1995. Computer mediated communication and social information: Status salience and status differences. *Academy of Management Journal* 38(4):1124–1151.