An Online News Recommender System for Social Networks

Manish Agrawal  
Department of Computer Science  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801  
magraw@microsoft.com

Maryam Karimzadehgan  
Department of Computer Science  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801  
mkarimz2@uiuc.edu

ChengXiang Zhai  
Department of Computer Science  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801  
czhai@cs.uiuc.edu

ABSTRACT
In this paper, an online news recommender system for the popular social network, Facebook, is described. This system provides daily newsletters for communities on Facebook. The system fetches the news articles and filters them based on the community description to prepare the daily news digest. Explicit survey feedback from the users show that most users found the application useful and easy to use. They also indicated that they could get some community specific articles that they would not have got otherwise.

1. INTRODUCTION
Web 2.0 has provided a new way for people to communicate with each other. Nowadays, there are hundreds of sources delivering content in various forms and more content is generated from users through blogs, reviews and comments than through professional writers. Users are often faced with an information overload. Even in the days before Internet, people found it difficult to decide what book to read, or what movie to watch. They were often guided with the opinions and recommendations of their friends. But with today’s ever increasing information overload, opinions and recommendations have become more important than ever before.

There are some situations, when users do not know exactly what they want or they may not even be actively looking for information. Generally, users have some particular interests and they will like to have WebPages, news articles, blogs or events related to their interests delivered to them. However, they are currently forced to visit multiple sources and scan through irrelevant content before finding useful information. For such a long-term information need, the best way to help users is to recommend information based on user profile and their history.

With a number of social networking Web sites such as Facebook, MySpace, Orkut and LinkedIn available, it is most desirable to have a system application that could integrate information from multiple sources and filter out irrelevant content to provide customized information for a community, i.e. a group of users sharing some common interests. These social network communities allow us to get feedback from users and develop algorithms that most users would benefit from. However, little research has been done on recommending information to social networks.

In this paper, we present an online news recommender system for Facebook and study users’ feedback. Our preliminary results indicate that most users find such an application useful and easy to use.

2. A FACEBOOK NEWS RECOMMENDER SYSTEM

2.1 Facebook
Facebook is one of the fastest growing social networks. It consists of many networks, each based around a school, workplace, or a region. It has users ranging from college students to working Professionals. More than 100 million users log on to Facebook at least once each day [4].

Each user has a personal profile and a user may belong to multiple networks. Also, a user may join various interest-based groups on Facebook or may even start a new group.

In general, two users may be connected in three different ways: 1) User a is a friend of User b. 2) User a and User b belong to the same network. 3) User a and User b belong to the same group.

Connection 1 and 2 are strong connections, in the sense, that a user can see the profiles of all his friends and also the profiles of all the people in his network(s). If the user finds someone on Facebook, who is neither his friend nor in his network, he will only be able to see a restricted profile.

We refer to both network and group as community, and the main task of our recommender system is to recommend information to the users of a community through delivering a customized newsletter (i.e. a list of recommendations) for each community. The recommendation was initially based on manually created keyword description of a community. After obtaining user feedback, the system would improve recommendation decisions based on feedback information.

2.2 Content-based recommendation
Each community has a set of sources associated with it. These sources are RSS feeds from community related websites or news agencies. Data of the specified sources for the communities are crawled. This kind of focused crawling has low coverage of the Web because only a small subset of the available news sources is crawled. As a result, we also crawl news articles from Yahoo! News\(^1\). For each community, we search Yahoo! News by using keywords from the community.

Once the articles are crawled, we construct an inverted keyword index over the entire text for all the articles. For each community, we construct a separate index for each day using the Indri indexing [9]. For each community, a set of queries with weighted terms are constructed. Indri has a rich query language that provides flexibility in building such queries. As an example, the following shows the set of Indri queries used in our system for the “University of Illinois” community:

\[
\begin{align*}
&\text{#uw(#syn(#2(University Illinois) UUUC #1(U of I)) news)} \\
&\text{#uw(#syn(#2(University Illinois) UUUC #1(U of I))...} \\
&\text{#syn(students academic events))} \\
&\text{#uw(#syn(#2(University Illinois) UUUC #1(U of I))...} \\
&\text{#syn(#uw(fighting illini athletics football basketball)))}
\end{align*}
\]

As shown, the terms in the #syn block are used as synonyms. \#2(University Illinois) matches "University * Illinois" (where * is any word or null). \#1(U of I) exactly matches "U of I". \#uw stands for unlimited unordered window and indicates that all constituents must occur within the current context in any order [9]. So, it is possible to construct complex queries using the Indri query language.

When the user registers a new community, it may not be always possible to manually construct such high quality queries. So, in this case, a query is constructed with all keywords having equal weights. Our experiments show that even such simple queries are pretty effective for content based filters.

Different queries may return the same document as a relevant result. Also, different sources may report the same news article. These factors produce redundant articles in the result set that should be clubbed together. This is achieved by clustering the documents based on their content using Indri’s clustering routine. We used the default centroid-based agglomerative clustering algorithm that uses cosine similarity as the distance measure. The clusters are sorted on the similarity threshold high (0.70) to allow only highly similar documents to merge together and avoid generating large and general clusters.

Presently, the Daily Newsletter is prepared mainly based on the retrieval results using such queries (i.e., a content-based filter on the news feeds). However, the Popular Articles list is prepared by aggregating the user feedback (i.e., collaborative filtering). Thus our system integrates content-based filtering with collaborative filtering. Below we describe how feedback is done.

3. COLLABORATIVE USER FEEDBACK

The system is designed to capture both implicit and explicit user feedback. This feedback later, can be leveraged to provide useful features in the application and improve the filtering accuracy. \(F = \{(u_i, a_j, f_{ij})\}\) is a feedback where \(u_i\) is the user, \(a_j\) is the article and \(f_{ij}\) is a parameter that aggregates all kinds of feedback for article \(a_j\) from user \(u_i\). The user feedback is captured in three different ways:

1. Clickthroughs of articles (\(f_1\)).
2. Ratings of articles (\(f_2\)).
3. Recommendation of an article by a user to another user (\(f_3\)).

For each article \(a_j\), we have potentially three different kinds of feedback from each user \(u_i\). We then need an algorithm to aggregate these feedbacks together. Let \(w\) be the base rating of the articles recommended by our system. Only if a user rating is more than \(w\), it will be considered positive feedback from the user. We use a very simple algorithm for feedback.

\(f_{1ij}\) denotes the feedback through clickthrough for article \(a_j\) from user \(u_i\). \(f_{2ij}\) denotes the feedback through rating of article \(a_j\) from user \(u_i\). \(f_{3ij}\) denotes the feedback through recommendation of article \(a_j\) from user \(u_i\).

The parameters are defined using the ternary operator. If the predicate in between the brackets is true, the variable takes the value after ”?”, otherwise the variable takes the value after ”:”. \(m\) denotes the scale on which the feedback is measured. So, \(m\) is the maximum feedback value that an article can get and \(r\) is the rating given by user \(u_i\) for article \(a_j\).

\[
\begin{align*}
&f_{1ij} = (u_i \text{ clicked article } a_j) \ ? m: 0 \\
&f_{2ij} = (u_i \text{ rated article } a_j) \ ? (r - base) * (m / (m - base)) : 0 \\
&f_{3ij} = (u_i \text{ recommended article } a_j) \ ? m: 0
\end{align*}
\]

And \(F_{ij}\) aggregates all these feedbacks as follows:

\[
F_{ij} = \sum_{k=1}^{3} \lambda_k * f_{kij}
\]

where \(\lambda_k\) denotes the weight we want to put on each kind of feedback.

\(F(a_j)\) is the feedback for article \(a_j\) aggregated for all users which is defined as follows:

\[
F(a_j) = \sum_{\text{Contact}} W_i * F_{ij}
\]

where \(W_i\) is the weight we want to put on each user.

Let \(support\) be the number of users used for calculating the feedback. We set \(\lambda_k = 1/3 \ \forall \ k\); this means that we are equally weighing the three kinds of feedback and \(W_i = 1/support\); which implies the feedback from all users are given equal weight. \(m\) and \(base\) are set to 5 and 2, respectively.

Let us see some scenarios to better understand the feedback model. We have dropped the subscript \(i\) and \(j\) to make the scenarios more readable. Note that \(F\) means \(F_{ij}\).

Case 1 (No feedback available): \(f_1 = 0; f_2 = 0; f_3 = 0\) \(\Rightarrow F = 0 \text{ (No Feedback)}\)

Case 2 (Bad article): User does not even click the main article but based on the synopsis, the user decided not to read it. The user rates it as 1 (minimum rating). \(f_1 = 0; f_2 = (1 - 2) * (5 / (5 - 2)) = -5/3; f_3 = 0\)

\(F = (f_1 + f_2 + f_3)/3 = -5/9 \text{ (Negative Feedback)}\)

\(^1\)http://news.search.yahoo.com/
Case 3 (Best article):
User reads the article, likes it, recommends it and rates it as 5.
\[ f_1 = 5; f_2 = (5 - 2) \times (5/(5 - 2)) = 5; f_3 = 5; \]
\[ F = (f_1 + f_2 + f_3)/3 = 5; \]
(Maximum Positive Feedback)

These scenarios show how the parameter \( F(a_j) \) can be interpreted to understand the collaborative user feedback. In the Pilot Study, we use this feedback for Popular Articles feature, where Articles are ordered by \( F(a_j) \) as long as \( \text{support}(a_j) > \text{support}_{\text{min}} \).

3.1 Newsletters System
The Newsletters application can be accessed by visiting the URL:
http://apps.facebook.com/news_letters/

The application uses Facebook API to find out the networks of the user. The Newsletters for the networks are available as tabs on the top of the page. The newsletter is presented as a list of articles, each with its title, news synopsis, links to original article and the locally cached page. Users can rate an article on a scale of 1 to 5. In the newsletter, the clusters are sorted on the score of the most relevant document in each cluster. Only one result per cluster is presented to the user, but the user can look at the other results by navigating through the “Similar pages” link. Only the top 5 results are presented to the user. The users can also recommend particular news articles to their friends through a recommendation button provided beside each news result. When a user makes a recommendation, his friend is sent a notification. The snapshot of the system is shown is figure 1.

Figure 1: University of Illinois NewsLetter

4. PILOT STUDY
We conducted a pilot to understand how users respond to such a recommendation system. The Newsletters application was launched on Facebook for three months. The application was advertised amongst university students by "word of mouth" publicity. During this period, sixty people have added the application on Facebook from a number of universities. There were initially 3 seed communities on the application with available newsletters but during three months, the number of communities has increased to 25. All the new communities are user initiated.

A set of University of Illinois (U of I) students, were asked to use the application on a regular basis. They used the application for about one month and then a user survey was conducted to garner explicit feedback from them. The basic aim was to evaluate the utility, accuracy and scope of the application. 22 users participated in the survey. Before presenting the user survey results, let us first look at the user data automatically captured by our system. Table 1 shows the basic statistics of user clickthroughs and ratings. We can see that the average rating of a news article is 3.41 (out of 5), and it is higher for clicked and recommended articles.

<table>
<thead>
<tr>
<th>Community</th>
<th>No. of people registered</th>
<th>No. of clicks</th>
<th>No. of recommendations</th>
<th>Average rating (out of 220 ratings)</th>
<th>Average rating for clicked article</th>
<th>Average rating for recommended articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>U of I</td>
<td>60</td>
<td>350</td>
<td>15</td>
<td>3.41</td>
<td>3.71</td>
<td>4.07</td>
</tr>
</tbody>
</table>

Table 1: Usage Statistics

After we got the first user feedback and the usefulness of the application was advertised amongst university students by "word of mouth" publicity. During this period, sixty people have added the application on Facebook from a number of universities. There were initially 3 seed communities on the application with available newsletters but during three months, the number of communities has increased to 25. All the new communities are user initiated.

In the Pilot Study, we used this feedback for Popular Articles feature, where Articles are ordered by \( F(a_j) \) as long as \( \text{support}(a_j) > \text{support}_{\text{min}} \).
this application has the potential to help them socialize. In the future, more features can be added to the application that may extend the users’ social networking experience.

We asked the users that if they had a perfect news recommender system, what delivery methods are most preferred by them. 14 users (63.64%) will like to receive newsletters on Facebook. So, the users see Facebook as the best platform for receiving such content. 10 users also expressed interest in receiving newsletters through email. This facility can be integrated in our system where the users can register specific newsletters for delivery via email.

5. RELATED WORKS

Recommender systems have been studied for many years. Adomavicius et al. [1] have categorized recommendations systems in three main categories: Content-based filtering, collaborative filtering, and Hybrid approach which combines the previous two methods. Collaborative filtering systems, recommend items based on similarity of users and their preferences. Examples of such a method are [8, 2, 6]. In Content-based recommendation, items similar to the ones that a user has seen in the past are recommended. Examples of such systems are [7] for news recommendation and [10] for book recommendation.

Although there has been some previous work on recommendations in social networks [11, 5, 3], and many applications have been developed for Facebook, there has been no application for recommending news information on Facebook, which is our goal.

6. CONCLUSION AND FUTURE WORKS

In this paper, we presented a novel system for recommending news articles to communities of a social network. Our main contribution is building a novel news recommender system and integrating it with Facebook and gathering user feedback. Our recommendation approach is a combination of content-based filtering and collaborative filtering. To the extent of our knowledge, this is the only application of its kind on Facebook. The system is automatic, scalable and sustainable. User studies indicate that most users of this application find it useful and efficient, demonstrating the feasibility of recommending information through social networks.

For future, we plan to do the followings:

Presently the newsletter is community specific. It can be personalized for individual Facebook user by exploiting the user profile and their friends’ profiles. Adding some features to help users socialize is another interesting direction. Tagging entities in the news articles would also be interesting.

In addition, we plan to develop and evaluate more advanced feedback algorithms based on machine learning and incorporate them with the system.

7. REFERENCES