Towards a Social Recommendation System

Joel DesArmo

Introduction

In the physical world humans often solicit advice from others when making decisions. Sometimes this can be asking a worker at a restaurant what they recommend on the menu, or other times it can be calling up a friend and asking them what kind of vacuum cleaner they use. Seeking recommendations could be as basic as reading a review for a movie before deciding what to see on a Friday night date, or as involved as talking with multiple neighbors about where to take a car to get repaired. A form of getting recommendations could come in the form of becoming aware of similar options. For example, when listening to a radio station, the music generally follows a certain genre of music. The concept of item similarity can also be applied to automatic recommendation systems.

Wouldn't it be great if people could use the power of the advice of trusted friends in combination with the ease of an automated system to deliver enhanced recommendations? This paper will review the main approaches to computerized automatic recommendation systems and survey some of the key literature on this topic, in a build-up to a new proposed social recommendation system for delivering better recommendations of everything from Amazon products, to Netflix movies, to Yelp restaurants. This paper proposes a new system, dubbed the FAIR System (Facebook Automatic Integrated Recommendation), which has the potential to revolutionize recommendations as known today.

Two Approaches

Modern computerized recommendation systems come in two main flavors: a collaborative filtering approach or a content-based filtering approach.

Collaborative Filtering

Among the first articles published discussing recommendation systems was "Using Collaborative Filtering to Weave an Information Tapestry" published in 1992 (Goldberg, Nichols, Oki, & Terry, 1992). This paper discussed research by scientists at Xerox for a type of recommendation system for email. This was not really a recommendation system as thought of in modern times, but was really simply a filtering system. The scientists were trying to address the growing problem of office workers being deluged by a new communication technology: email. So they set out to develop an automatic method of filtering emails that would be of most interest to workers. They accomplished this by creating the Tapestry email system in which users could rate the email they received. The tapestry system then used these ratings to filter out undesirable email and thereby limit the total amount of email received by users of the system. Thus the concept of collaborative filtering was born. The new FAIR System, to be proposed later in this paper, relies upon existing collaborative filtering techniques.

Collaborative filtering is a technique for automatically making predictions (the filtering aspect) of user interests based on the interests of other users (the collaborative aspect). The basic premise of collaborative filtering is that given X has the same opinion as Y, then if X has opinion *i* for item *j*, then Y will also have opinion *i* for item *j.*

One could loosely say mathematically:

$$\text{Given } X = Y,\ i_j \in X \rightarrow i_j \in Y$$

This principle arises from humans often share multiple similar interests. It is not a hard and fast rule, but can be considered to be a guiding principle of likelihood, relative to human preferences.

In broad terms, collaborative filtering can be accomplished as follows:

1. Multiple users $X_{1,}$ $X_{2,}$..., X_n give a rating to items $i_{1,}$ $i_{2,...,}i_n$. These could be physical products such as a laptop computers, or services such as restaurants, or intellectual property such as movies.
2. User Y gives ratings to multiple items.
3. The system will cluster users with similar preferences on the items.
4. The system will recommend items to User Y that Y has not yet rated, based on the ratings of other uses now in Y's cluster.

Collaborative filtering recommendation techniques generally require explicit input from users in the form of entering some kind of rating, whether it is thumbs up or down, or a Likert scale, and so on.

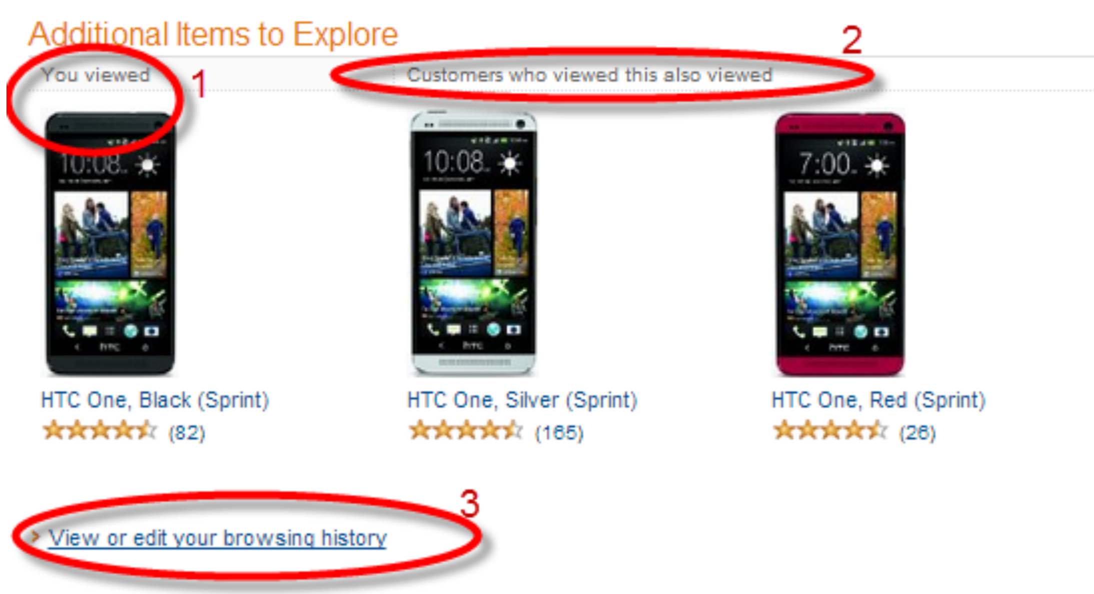

Figure 1. Collaborative filtering in Amazon clearly informs user 1) why they are being shown recommendations; 2) further explanation why they are being shown these recommendations; and 3) ability to alter the data that Amazon uses to provide this recommendation.

Disadvantage: One of the disadvantages of this method of recommendation is that Y will not receive any recommendations until Y has also made some ratings. Until such time as Y rates a threshold level of items, the system will be unable to match User Y with other users, and thereby with other user preferences.

A second disadvantage of collaborative filtering is that it is dependent on users to be actively engaged in the recommendation process. If users do not rate items, then there is no data from which the system to base recommendations.

Advantage: Collaborative filtering can give recommendations about a variety of items. For example, this recommendation method can easily cut across movie genres or types of ethnic restaurants.

Content-based Filtering

The content-based filtering approach uses the intrinsic properties of the item to make recommendations to users. This method generally relies on the person in control of the items being recommended to make sure that the system has some metadata about the items in the collection (Adomavicius & Tuzhilin, 2005). This can often be seen in text-based systems such as Google News. In this case the system can automatically perform a feature extraction in the form of keywords.

The system generally collects data in a more implicit fashion in a content-based approach. It may rely on logging items in its collection that a user views, or use the users purchase history in order to recommend similar items.

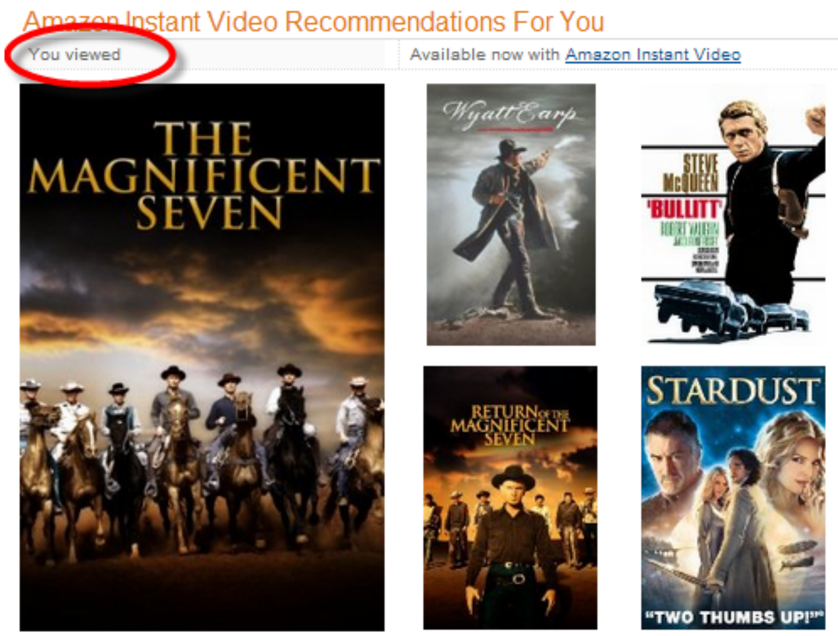

Figure 2. Content-based filtering in Amazon. #1) informs user why they are being shown recommendations.

Disadvantage: A disadvantage of this recommendation method is that the technique can require significant upfront work on behalf on the collection owner. This technique lacks the ability to recommend items that are dissimilar.

Advantage: Many users may be hesitant to take the time required to share a rating. With content-based filtering, the users can be totally passive and still receive useful recommendations from the system.

Key Research

To get an idea of the amount of research related to recommendation systems, the Association for Computing Machinery (ACM) Digital Library was consulted in December 2012. The query "recommendation systems" yielded 1,565 results. Sorting the results by most recent, reveals that 254 scholarly works have been published in 2013 that are in the ACM collection.

Rank	Article Title	Year	Citations
1	"Evaluating collaborative filtering recommender systems" (Herlocker, Konstan, Terveen, & Riedl, 2004)	2004	723
2	"Propagation of trust and distrust" (Guha, Kumar, Raghavan, & Tomkins, 2004)	2004	251
3	"Latent semantic models for collaborative filtering" (Hofmann, 2004)	2004	249
4	"Recommender systems in e-commerce" (Schafer, Konstan, & Riedl, 1999)	1999	245

Table 1. Top ACM Digital Library results for query term "recommendation systems" sorted by number of citations.

It can be seen by the results in this list of most cited works on recommendation systems that collaborative filtering is an important approach. Also noteworthy is the fourth most cited work in the ACM collection related to "recommendation systems" is also related to e-commerce and was published in 1999 at the height of the dot-com frenzy occurring in the United States.

CASE STUDY: The Netflix Prize and the technical challenge of recommendation systems

The Netflix Prize was a contest created by the Netflix movie rental company to try to improve their movie recommendations to their users. The contest began Oct 2, 2006 and offered $1,000,000 to any team that could create a recommendation system that was 10% more accurate than Netflix's own. Only two days later a team of two researchers had already produced better results than Netflix's own recommendation algorithm, which was at that time called Cinematch. However, the contest was seeking a solution that made a 10% improvement over Netflix's current recommendation algorithm, so the contest continued.

The performance of new algorithms was determined by calculating the root mean squared error (RMSE). The training data consisted was enormous; it consisted of over 100 million movie ratings of approximately 18,000 movie titles from over 480,000 users (Bennett & Lanning, 2007). Netflix had an additional approximately 3,000,000 ratings that it set aside to use as a test data. The task was to use this large training data set provided by Netflix to build a model to predict the ratings of the approximately 3,000,000 ratings that were held in reserve.

Although it was only days before a team had improved upon the performance of the Netflix system, the improvement was only slight. It would be nearly three years before researchers had finally achieved a 10% improvement. At the end of 2007, a small prize was awarded by Netflix for the team that currently in the lead of the contest. The BellKor team's solution was a mixture of 107 models (Bell, Koren, & Volinsky, 2007). The team had a RMSE=0.8712, but the challenge was to beat the accuracy performance of Netflix's own Cinematch recommendation algorithm by 10%, which required had a RMSE= 0.9525, so the contest continued, despite the improvement and the vast amount of models that went into the solution. According to the Netflix website, there were 41,305 teams composed of 51,051 contestants from 186 different countries. Of those, 5,169 teams submitted 44,014 potential solutions (Netflix, 2009). The amount of people working on this problem was immense. The contest was finally won in September 2009 by team BellKor's Pragmatic Chaos, a joint combination effort of three teams. In the end, another team also had produced exactly the same recommendation accuracy, by submitted their solution just 20 minutes later (Lohr, 2009). Koren and Bell continue to be at the forefront of recommendation systems research. Their chapter on advances in collaborative filtering is a great overview at the multitude of techniques that go into a modern recommendation system, taking into account both explicit user input and implicit user input, as mentioned earlier in this paper, as well as topics used to solve the Netflix Prize such as used to solve this problem included K-nearest neighbor, singular value decomposition, matrix factorization, and neural networks (Koren & Bell, 2011).

FAIR Recommendation System

I propose an enhanced recommendation system based on existing collaborative filtering systems, but incorporating a user's social network, called the Facebook Automatic Integrated Recommendation (FAIR) System. Word of mouth is the oldest form of recommendation system. Throughout history, people have relied on their friends to recommend consumer products to one another. There is an old saying "birds of a feather flock together." Current recommendation systems, such as those employed by Amazon or Netflix, have enormous training data sets that they can use to predict the preferences if users. However, better recommendation results may be able to be achieved by giving greater weight to the preference data associated with known friends of a given user.

This could be achieved through Facebook integration with online consumer websites such as Amazon, movie websites such as Netflix, or restaurant review websites such as Yelp. This is admittedly as much of a business challenge as it is a technical challenge, and perhaps the business challenge is the more difficult. However, in May 2011, there were 721 million active Facebook users, defined as having at least 1 Facebook friend and having logged into the website at least once in the previous 28 days. Furthermore, also in May 2011, there were 149 million Facebook users in the United States, while at the same time there were approximately 260 people in the United States over the age of 13, which is the minimum age required to create a Facebook account (Ugander, Karrer, Backstrom, & Marlow, 2011). Those figures represent 57% of people residing in the United States that are eligible for a Facebook account actually being active users of Facebook. According to the Facebook Data Science Team, the average Facebook user has 190 friends on Facebook (Backstrom, 2011). Research has shown a correlation between the similarities of users and trust in online communities (Ziegler & Lausen, 2004).

Conclusion

Leveraging the ubiquity and the built-in trust of the Facebook network with existing collaborative filtering recommendation systems may provide a meaningful enhancement to the current state-of-the-art. Additionally, gamification techniques along with optional social notifications could be made available to users as an enticement to actively engage in rating items and thereby further strengthening the accuracy of automatic recommendations in the proposed FAIR recommendation system.

Works Cited

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions On*, *17*(6), 734–749.

Backstrom, L. (2011, November 21). Anatomy of Facebook. Retrieved from https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859

Bell, R. M., Koren, Y., & Volinsky, C. (2007). The BellKor solution to the Netflix prize. *KorBell Team's Report to Netflix*. Retrieved from http://www2.research.att.com/~volinsky/netflix/ProgressPrize2007BellKorSolution.pdf

Bennett, J., & Lanning, S. (2007). The netflix prize. In *Proceedings of KDD cup and workshop* (Vol. 2007, p. 35). Retrieved from http://su-2010-projekt.googlecode.com/svn-history/r157/trunk/literatura/bennett2007netflix.pdf

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Commun. ACM*, *35*(12), 61–70. doi:10.1145/138859.138867

Guha, R., Kumar, R., Raghavan, P., & Tomkins, A. (2004). Propagation of Trust and Distrust. In *Proceedings of the 13th International Conference on World Wide Web* (pp. 403–412). New York, NY, USA: ACM. doi:10.1145/988672.988727

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5–53.

Hofmann, T. (2004). Latent Semantic Models for Collaborative Filtering. *ACM Trans. Inf. Syst.*, *22*(1), 89–115. doi:10.1145/963770.963774

Koren, Y., & Bell, R. (2011). Advances in Collaborative Filtering. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook* (pp. 145–186). Boston, MA: Springer US. Retrieved from http://www.springerlink.com/index/10.1007/978-0-387-85820-3_5

Lohr, S. (2009, September 22). A $1 Million Research Bargain for Netflix, and Maybe a Model for Others. *The New York Times*. Retrieved from http://www.nytimes.com/2009/09/22/technology/internet/22netflix.html

Netflix. (2009). Netflix Proze Leaderboard.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender Systems in e-Commerce. In *Proceedings of the 1st ACM Conference on Electronic Commerce* (pp. 158–166). New York, NY, USA: ACM. doi:10.1145/336992.337035

Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the facebook social graph. *arXiv Preprint arXiv:1111.4503*. Retrieved from http://arxiv.org/abs/1111.4503

Ziegler, C.-N., & Lausen, G. (2004). Analyzing correlation between trust and user similarity in online communities. In *Trust management* (pp. 251–265). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-24747-0_19