Probabilistic News Recommender Systems with Feedback

Shankar Prawesh and Balaji Padmanabhan
Information Systems and Decision Sciences
College of Business, University of South Florida
4202 E. Fowler Avenue, Tampa, FL 33620
{shankar1, bp}@usf.edu

ABSTRACT
In prior work we addressed a major problem faced by media sites with popularity based recommender systems such as the top-10 list of most liked or most clicked posts. We showed that the hard cutoff used in these systems to generate the "Top N" lists is prone to unduly penalizing good articles that may have just missed the cutoff. A solution to this was to generate recommendations probabilistically, which is an approach that has been shown to be robust against some manipulation techniques as well. The aim of this research is to introduce a class of probabilistic news recommender systems that incorporates widely practiced recommendation techniques as a special case. We establish our results in a special case of two articles using the urn models with feedback mechanism from probability theory.

Categories and Subject Descriptors
H.2.8 [**Database Applications**]: Data mining

General Terms
Algorithms, Top-N

Keywords
News recommender systems, probabilistic sampling, feedback

1. INTRODUCTION
There has been growing evidence of the influence of news recommender systems (NRS) on users. It is considered an important source of news articles for readers, articles which otherwise may get lost due to dynamic environment of news cycles driven by continuous arrival of news articles [12]. It has been noted that once a story appears in most popular list (widely used by media sites), there is an abrupt increase in its popularity and advertising than other stories [1, 7]. As it becomes visible to more readers the number of votes grows at a faster rate.

In prior work we showed that popularity based NRS such as $most\ emailed$ or $most\ popular$[1] is susceptible to amplifying negligible differences in the initial counts of N^{th} and $(N+1)^{th}$ article. Hence, $(N+1)^{th}$ article which may have "just" missed making the cutoff is often unduly penalized in the most-popular NRS.

These systems are also easily susceptible to manipulation. With some initial effort if a manipulator can get an article into a "most popular" list then the self-reinforcing nature of such lists makes the article remain there with little additional effort. Recently there have been a lot of popular press articles that have highlighted exactly this problem, with manipulation in particular attracting a lot of attention.

Probabilistic NRS has been proposed as a solution to address the aforementioned issues [10]. Briefly, such systems generate recommendations by sampling probabilistically from the pool of articles that could be recommended. An article's probability of being chosen in the recommended list is proportional to its current count/popularity.

This method still generates good recommendations, but permits all articles to have some chance of being recommended. Such a mechanism does not penalize the marginal next articles that might have just missed a hard cutoff in a traditional Top-N list. This mechanism is also more robust against manipulation since it does not suffer as much from the self-reinforcing nature of the hard cutoff lists.

However, there are some limitations of the probabilistic NRS presented in [10]. For one, this approach may select some articles that are not as popular, thereby potentially sacrificing short-term clicks or readership. In an era where page views translate proportionally to advertising revenue this can be a concern in implementing this. Second, giving articles probabilities proportional to their counts is just one method for probabilistically sampling articles and does not provide the media client with any flexibility in implementing such a sampling scheme.

In the present research we propose a novel solution to these problems of the prior probabilistic NRS through a class of probabilistic NRS with *feedback* and discuss various notable properties of it. Feedback models [4] are used in applications where the behavior of the system creates either positive or negative feedback that affects future behavior of the system. For example, the typical "Top N" recommender has a positive feedback mechanism for the articles in the list.

The probabilistic selection mechanism with feedback introduced in this research can be considered as a unified model of selection techniques used for different news recommendation mechanisms based on the count of articles. For example, random selection, probabilistic selection in [10] and Top-N selection can be considered as special cases of a generalized probabilistic NRS introduced in the present research. In general the recommendation probability of an article with count n is proportional to $f(n) = n^{\gamma}, \gamma \in \mathbb{R}$. For a special case with two articles we provide theoretical insights of the proposed recommendation process using results from classical urn models in probability theory [4].

[1] It is also called Top-N NRS.

RecSys'12, September 9–13, 2012, Dublin, Ireland.

2. RELATED WORK

There has been growing attention towards the study of articles in the most-popular list or articles which are promoted to the $front\ page$. For example, Berger and Katherine [1] have addressed the issue of virality of most-emailed list at the New York Times; while Lerman and Ghosh [7] have studied the distribution of popularity for articles promoted to the front page of Digg- a popular social new aggregator. In somewhat different approach Prawesh and Padmanabhan [10], have discussed some specific characteristics of Top-N NRS using a thought experiment.

The count of evolution of articles in a probabilistic NRS with feedback exhibits similar behavior as discussed by Khanin and Khanin [6]. They have used a probabilistic model of $positive\ feedback$ to study the pattern of neuron growth. In their context several "neurites" are known to exhibit a pattern of growth and contraction until one of them rapidly grows to become an "axon". The probability that a neurite grows in a time period is modeled to be proportional to its length in the previous period and also depends on the level of competition from other neurites. Our proofs have been adapted from the work of Khanin et al. [6].

Other growth processes where feedback mechanisms have been observed are: the technology dominance of QWERTY and Microsoft's operating systems monopoly [4]. Metcalfe's law- used to value a telecommunication network is also considered a special case of feedback mechanism [11]. In business, positive feedback mechanisms can help a company reach monopoly status by starting with some initial advantage over competitors. Some types of positive externalities particularly contribute to this effect where the values of some systems increase super-linearly with the number of users [4].

In a slightly different context, Pandey et al. [9] have studied the issue of *exploration* and *exploitation* of web-pages using controlled randomness into search result ranking methods. They show that modest amount of randomness leads to improved search results.

3. MODEL

Let us assume that a media site maintains a comprehensive list (CL) of articles. From CL, N articles are selected as "recommendations". The selection of articles is based on probabilistic sampling without replacement. At any given time t the probability that an article-a will be selected in display list (DL) is given by

$$p_a(t) = \frac{{c_a}^{\gamma}(t)}{\sum_j {c_j}^{\gamma}(t)} \qquad (1)$$

Where $C_a^{\gamma}(t)$ represents the count of an article 'a' raised to the exponent γ. While, $\sum_j C_j^{\gamma}(t)$ represents the sum of counts of articles (those are not yet selected for DL) at time t to the exponent γ. This sampling process is repeated N times to generate the N recommendations in DL.

In the rest of this section we derive analytical results that can illustrate the workings of this selection mechanism in a formal manner. But before we do so, it is worth observing how this selection mechanism can work in practice.

With $\gamma = 1$ we have the probabilistic selection mechanism implemented in [10]. With $\gamma > 1$ we will have a system with positive feedback for the articles with higher counts. These are now going to have an even higher (i.e. more than proportional) of being in a recommended list, which can result in rapidly increasing counts.

With very high γ it is easy to see that the (sampling without replacement) mechanism is similar to the current "Top N" selection. At each stage the article with the highest count is most likely to be selected and the process will end up with a list identical to the Top N.

For values in-between we have varying degrees of positive feedback. It is also interesting here to consider what might happen when $\gamma = 0$. In such a case it is easy to see that all articles have the same probability of being recommended, essentially simulating a random recommender.

From a practical perspective systems like this can permit the site to dynamically manage a recommender list, alternating between exploration and exploitation as needed. Below we present formal theoretical results.

3.1 Analytical Results

To study the count evolution process of articles for the proposed probabilistic NRS, we make following assumptions.

1. Two articles are available for recommendation (article-1 and article-2).
2. Reader upon arrival reads the recommended article with probability p or reads the other with probability $1 - p$.
3. An initial count of articles before the recommender system was implemented is given by $c_1(0)$ and $c_2(0)$; $(c_1(0) > c_2(0)\ without\ loss\ of\ generality)$ respectively.
4. NRS has fairly strong influence on reader's reading behavior (i.e. $p \sim 1$).

The count of two articles at time t has been denoted by $c_1(t)$ and $c_2(t)$ respectively. Let us denote the discrete time points by integer values. At each time, upon arrival of a reader, an article is read and its count is increased by 1. The total count of articles in the system at time t is deterministic and it is given by $c_1(0) + c_2(0) + t$. We focus on the article '1' for subsequent derivation; we also note that theoretical results for article '2' can be obtained in similar way. Let us denote the probability $p[c_1(t+1) = c_1(t) + 1]$ as $p_{1t}(read)$.

In probabilistic NRS, article-1 can be read in two ways. The article is in the recommended list (with probability $p_1(t)$) and the reader chooses to read the recommended article (with probability p). Or, article- 1 can be in the other list $= (CL \setminus DL)$ (with probability $1 - p_1(t)$) and the reader chooses to read the un-recommended article (with probability $1 - p$).

Specifically, the probability that an article '1' is being read at time t is given by

$$p_{1t}(read) = p * p_1(t) + (1 - p) * (1 - p_1(t)) \qquad (2)$$

Substituting the expression for $p_1(t)$ in the above expression from equation (1) we have,

$$p_{1t}(read) = p * \frac{1}{1 + \left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}} + (1 - p) * \frac{1}{1 + \left(\frac{c_2(t)}{c_1(t)}\right)^{(-\gamma)}} \qquad (3)$$

So, the process described in generalized probabilistic NRS for the article-1 can be understood as a processes generated through mixture of two processes defined by $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}}$ and $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{(-\gamma)}}$ respectively. γ can take any real value between

$(-\infty, \infty)$. However, due to symmetric nature of $p_{1t}(read)$ with respect to γ (equation 3), we will discuss the case when $0 < \gamma < \infty$. Similar, analysis can be extended for $-\infty < \gamma < 0$.

3.1.1. $\gamma = 0$

In this case equation (3) is given by $p_{1t}(read) = \frac{1}{2}$. This is equivalent to an article being read randomly at each time step, irrespective of reader's preference (i.e. p). Hence, we do not use any mechanism to incorporate the reader's preference for article recommendation in such a system.

In the following sections for completeness we will discuss the processes generated through the both expressions $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}}$ and $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{(-\gamma)}}$. However, it should be noted that with $assumption\ 4$, equation 3 takes the form of

$$p_{1t}(read) \sim \frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}}$$

In this case, count evolution process based on the probabilistic NRS, $p_{1t}(read) = \frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}}$ will suffice for the discussion.

3.1.2. $\gamma = 1$

The reading probability of the article '1' is given by

$$prob_{1t}(read) = p * \frac{c_1(t)}{c_1(t)+c_2(t)} + (1-p) * \frac{c_2(t)}{c_1(t)+c_2(t)}$$

$$= p * \{share\ of\ article\ '1'at\ time\ t\} + (1-p) * \{1 - (share\ of\ article\ '1'at\ time\ t)\}$$

The processes generated in this case can be understood as a combination of Pólya urn mechanism and Friedman urn mechanism respectively [5]. Where Pólya urn mechanism corresponds to the probability function $\frac{c_1(t)}{c_1(t)+c_2(t)}$ (i.e. $p = 1$) and a Friedman urn corresponds to the probability function $\frac{c_2(t)}{c_1(t)+c_2(t)}$ (i.e. $p = 0$).

The formulation of Pólya urn process is defined as follows. Let an urn initially contains $c_1(0)$ black balls and $c_2(0)$ white balls. Each time, a ball is drawn randomly from the urn and it is replaced back in the urn with another ball of same color. Using the martingale property of share of the black balls in the urn after time t, (X_t) [5]. It can be shown that

$$\mathbb{E}(X_t) = \frac{c_1(0)}{c_1(0)+c_2(0)}$$

Further, X_t converges with probability 1 to a limiting random variable X_∞ as $t \to \infty$. The distribution of X_∞ depends on the initial share of the black ball. In Friedman urn model each time a ball is drawn randomly from the urn and it is replaced back in the urn with another ball of *different* color. Let us denote the share of black ball in Friedman urn model at time t as Y_t. Then Y_t converges with probability 1 to $\frac{1}{2}$ [5].

One notable property of this probabilistic mechanism is that, it is robust towards maintaining the share of all articles in the system for an influential NRS ($p \sim 1$).

3.1.3. $1 < \gamma < \infty$

From equation 3 we have,

$$p_{1t}(read) = p * \frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}} + (1-p) * \frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{-\gamma}}$$

In this case the processes generated by probability function $p_{1t}(read)$ can be considered as a mixture distribution of two different processes given by $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{\gamma}}$ and $\frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{-\gamma}}$. When the probability of article '1' being read is given by, $p_{1t}(read) = \frac{c_1(t)^{\gamma}}{c_1(t)^{\gamma}+c_2(t)^{\gamma}}$, (i.e. $p = 1$). The count evolution processes of articles correspond to a generalized Pólya scheme [6]. This phenomenon is also understood as systems with positive feedback in economics, biology and chemistry [6, 8].

In the context of NRS it leads to a situation where two articles (article-1 and article-2) compete until one article obtains non-negligible advantage in the count share; eventually leading to dominance of the NRS by a single article. The strength of feedback is modeled through the parameter γ. In this case after a random moment of time an article will be always recommended. More precisely, with probability 1 there exists a time t^* such that an article is recommended for all $t > t^*$.

The distribution function $p_{1t}(read) = \frac{c_2(t)^{\gamma}}{c_1(t)^{\gamma}+c_2(t)^{\gamma}}$ (i.e. $p = 0$) corresponds to the case with $negative\ feedback$. In this case an article with high initial share will become less popular and an article with low initial share will become more popular over time. So, in the present setup the total reading probability (equation 3) is a mixture distribution of positive and negative feedback mechanism.

3.1.4. $0 < \gamma < 1$

The path followed by the probability function $p_{1t}(read) = \frac{c_1(t)^{\gamma}}{c_1(t)^{\gamma}+c_2(t)^{\gamma}}$ generates the subcritical regime in which both articles will have counts of the same order [6]. More precisely, the ratio of counts for articles tends to 1 as $t \to \infty$.

$$\frac{c_1(t)}{c_2(t)} \to 1\ as\ \ t \to \infty \qquad (4)$$

It follows from (4) that, $c_i(t) = \frac{t}{2} + o(t); i \in \{1,2\}$ (5)

The behavior of $o(t)$ depends on γ. If $\frac{1}{2} < \gamma < 1$ then there exists nonzero random constant k, such that $\frac{o(t)}{t^{\gamma}} \to k$ as $t \to \infty$. In this case an article with high count (i.e. article-'1') will maintain higher share for all large enough t.

When $0 < \gamma \leq \frac{1}{2}$, for both articles there exists a sequence $t_n \to \infty$ such that the article will have higher count at time t_n. $o(t)$ in equation (5) are of the order $\sqrt{t}$ if $0 < \gamma < \frac{1}{2}$ and of the order $\sqrt{t \ln t}$ if $\gamma = \frac{1}{2}$.

For the probability function, $p_{1t}(read) = \frac{c_2(t)^{\gamma}}{c_1(t)^{\gamma}+c_2(t)^{\gamma}} = \frac{1}{1+\left(\frac{c_2(t)}{c_1(t)}\right)^{-\gamma}}$, almost surely, $\lim_{t\to\infty} c_i(t) = \frac{t}{2}$. As, when $\gamma < 1$

then counts of both articles grow at the same rate asymptotically [3, 4].

Again, it should be noted that with $assumption\ 4$, steps discussed in previous sections for the expression $\frac{c_1(t)^\gamma}{c_1(t)^\gamma + c_2(t)^\gamma}$ (i.e. $positive\ feedback$) will suffice for the derivation of results.

3.1.5. $\gamma \to +\infty$

Finally, we discuss the situation in which proposed NRS will behave like the most-popular NRS. In this case, the total reading probability of the article '1' can be approximated as $p_{1t}(read) \sim p$. This is equal to the reading probability of the recommended article, when recommendation is based on high counts.

In a more general sense this is equivalent to the NRS that uses articles corresponding to the highest count for recommendation (Top-N NRS). The selection process of N articles generated through a processes defined by equation (1) is given by

$$p_a(t) = \frac{{c_a}^\gamma(t)}{\sum_j {c_j}^\gamma(t)} = \frac{1}{1 + \sum_{j \neq a} \left(\frac{c_j(t)}{c_a(t)}\right)^\gamma}$$

Without loss of generality we assume that all articles have different count. Further, it can be easily observed that $\forall j$ such that $\frac{c_j(t)}{c_a(t)} < 1, \lim_{\gamma \to \infty} \frac{c_j(t)}{c_a(t)} \to 0$. So, for the article with highest count (among those which are not yet selected for DL) the selection probability in DL will be 1, i.e. $\max\{p_a(t)\} \to 1$. Hence, N probabilistic selections correspond to selection of N article with decreasing order of their counts.

4. CONCLUSION

There has been growing awareness towards the various limitations of the "most popular" lists in recommendation systems. For example, in a recent article [2] in New York Times, it has been noted that, being 11th on a top 10 list on the recommendation system is a lot different than being 10th on that list. Nick Bilton in [2] writes "*Being at the top of these lists can generate substantial windfalls. The iTunes App Store, where apps like Angry Birds, Words With Friends and Pages have spent months at the top of the charts, help the app makers collect hundreds of thousands of dollars in revenue, while those who cannot get that visibility founder in obscurity*".

In a broader context, the widespread use of Top-N based NRS is leading us to less choice of news articles [2]. In several cases it has been also observed or suspected [2] that manipulators artificially inflate the popularity of the items of their interest. A further reason for count amplification in these top lists is the propagation of recommendations over social networks. Once an article (or app) makes such a list they are more likely to be picked up and propagated through social networks.

While some form of increased attention to "good" articles or content is a plus, when this is done so in a disproportional manner, at the expense of other (possibly equally good) articles or apps the mechanism starts creating phenomena that are clearly undesirable. The system influences readership or what succeeds by virtue of an artificial cutoff. It is such weaknesses that attract manipulators to potentially game the system. At the extreme such systems therefore are prone to a high degree of noise.

The use of the proposed probabilistic NRS is one possible solution to deal with these limitations of Top-N NRS. The absence of a hard cutoff eliminates a key component of the self-reinforcing nature of such lists. However, it should be noted that in this new paradigm the term "most popular" will not be appropriate in all cases. Instead it has to be replaced by more appropriate term such as, popular articles.

Possibly more important, probabilistic mechanisms can potentially sacrifice short term revenue for the media site if it selects unpopular articles that are not likely to be read. Our study of feedback functions in this paper makes a novel contribution in the field of NRS by proposing a recommendation technique that elegantly addresses this drawback by allowing the users to control the extent of the feedback in the system.

If a site wants to promote diversity and minimize artificial amplification the model can be parameterized appropriately to select articles. At some point if the goal is to maximize short-term revenue the parameter can be adjusted to mimic the Top-N systems in behavior. The specific feedback function proposed in this paper for NRS therefore permits elegant combination of exploration and exploitation. Dynamically fine-tuning this and experimental results are aspects we are working on in current work.

REFERENCES

[1] Berger, J. and Milkman, K. L. 2011. What Makes Content Viral? *Journal of Marketing Research* (To Appear).

[2] Bilton, N. 2012. Disruptions: Top 10 Lists Lead to Less Choice on the Web. *The New York Times*, April 1, 2012.

[3] Chung, F., Handjani, S. and Jungreis, D. 2003. Generalizations of Pólya's Urn Problem. *Annals of Combinatorics* 7 (2), 141-153.

[4] Drinea, E., Frieze, A. and Mitzenmacher, M. 2002. Balls and Bins Models with Feedback. *In Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete Algorithms (SODA 2002)*.

[5] Freedman, D. A. 1965. Bernard Friedman's Urn. *The Annals of Mathematical Statistics*, vol. 36(3), 956-970.

[6] Khanin, K. and Khanin, R. 2001. A Probabilistic Model for the Establishment of Neuron Polarity. *Journal of Mathematical Biology*, 42, 26-40 (2001).

[7] Lerman, K. and Ghosh, R. 2010. Information Contagion: an Empirical Study of Spread of News on Digg and Twitter Social Networks. *In Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM'10)*.

[8] Mitzenmacher, M., Oliveira, R. and Spencer, J. 2004. A Scaling Result for Explosive Processes. 11 (2004), *The Electronic Journal of Combinatorics*.

[9] Pandey, S., Roy, S., Olston, C., Cho, J. and Chakrabarti, S. *Shuffling a stacked deck: The case for partially randomized ranking of search engine results*. VLDB Endowment, Norway, 2005.

[10] Prawesh, S. and Padmanabhan, B. 2011. The "top N" News Recommender: Count Distortion and Manipulation Resistance. *In Proceedings of the fifth ACM Conference on Recommender Systems* (*RecSys'11)*. ACM, New York, NY, USA, 237-244.

[11] Shapiro, C. and Varian, H. 1999. Information Rules, *Harvard Business School Press*. 1999.

[12] Weber, T. E. 2010. Cracking the New York Times Popularity Code. *The Daily Beast*, December 19, 2010.