# LOGO: A Long-Short User Interest Integration in Personalized News Recommendation

Lei Li, Li Zheng, Tao Li
School of Computing and Information Sciences
Florida International University
Miami, FL 33199
{lli003, lzhen001, taoli}@cs.fiu.edu

## ABSTRACT

In this paper, we initially provide an experimental study on the evolution of user interests in real-world news recommender systems, and then propose a novel recommendation approach, in which *the long-term and short-term reading preferences of users are seamlessly integrated* when recommending news items. Given a hierarchy of newly-published news articles, news groups that the user might prefer are differentiated using the long-term profile, and then in each selected news group, a list of news items are chosen based on the short-term user profile. Extensive empirical experiments on a collection of news articles obtained from various popular news websites demonstrate the efficacy of our method.

**Categories and Subject Descriptors:** H.3.3[Information Search and Retrieval]: Information filtering

**General Terms:** Algorithm, Design, Experimentation

**Keywords:** Time sensitive weighting, Long-term/Short-term user profiling, User interest evolution

## 1. INTRODUCTION

A key issue of content-based news recommendation is how to construct the user's profile based on his/her reading history, named as "user profiling". To handle this issue, most content-based recommender systems take into account the user's reading history as a whole, and summarize the history to be the user's profile [2, 8]. [7] postulated that a user's preference for particular articles depends not only on the topic and on propositional contents, but also on the user's current context. [11] developed a Bayesian framework for predicting users' current news interests from the activities of that particular user and the news trends demonstrated in the activity of all users. In our previous work [10], we present a scalable news recommender system by modeling news recommendation as a budgeted maximum coverage problem. However, they failed to consider the evolution of users' reading interest. In reality, the general topics that the user is interested in would be *relatively stable or vary slightly in a long-term perspective*, whereas the content accessed by the user might *change along different short-term perspectives*. For example, a user might be interested in "Sports" for a long time; however, he/she might prefer to reading news articles related to "NBA Playoff" in this month, while reading newsfeeds that describe "FIFA U-17 World Cup" next month. It would be more reasonable to explicitly consider interest evolution for providing high-quality recommendation services.

In this paper, we initially provide an experimental study on the evolution of user interests, and show that most users' reading preferences indeed change over time, whereas their long-term interests vary slightly. Then, we propose a news recommender system – LOGO, in which the *LO*n*G*-term and sh*O*rt-term reading preferences of users are seamlessly integrated when recommending news items to individual online users. We construct the long-term profile of a given user based on a **time sensitive weighting** scheme [5], and the short-term profile by analyzing the latest reading history of the user. Both of them can help determine the news recommendation list to individual users.

The contribution of our work is two-fold: (1) We emphasize *the effect of user interest evolution* when recommending news articles, and represent the user's reading preference with *seamless integration of long-term and short-term user profiles*; and (2) We construct a *two-stage news selection strategy*, where the long-term profile is firstly utilized to differentiate news groups with specified preference, and then the short-term profile is applied to filter and recommend specific news articles to individual users. Our work presents *a novel way to better understand the user's reading preference in which the effect of long-term and short-term profiles are being considered simultaneously.*

## 2. METHODOLOGY

We maintain a reading preference model for each online user. This model is essentially a content-based filtering model, involving the long-term and short-term interest profiles. The former is built based on a time sensitive weighting scheme [5], and the latter is represented by the latest reading interest of the user. Our approach assumes that the user's reading preference would evolve over time, and in a long run, the general preference would be relatively stable. Figure 1 depicts an overview of our proposed method.

Given a collection of newly-published news articles, we first construct a news hierarchy purely based on the content of news articles using hierarchical agglomerative clustering algorithm, and then use Dunn's Validity Index [6] to decide the best layer of the dendrogram. In this way, we can avoid


*RecSys'11,* October 23–27, 2011, Chicago, Illinois, USA.

to decide the number of clusters in news dataset. Using the long-term user profile weighted by time decay factor, we can easily filter the news groups that are similar to the long-term user profile. After that, the short-term user profile is utilized to select and rank news items in each news group. Finally, a list of news articles is presented to the end user, where the number of news items from each news group is weighted by the similarity between the news group and the long-term user profile. In the following, the detailed algorithmic information for each module will be introduced.

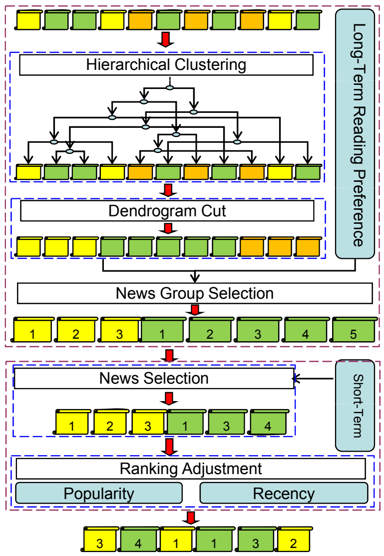


**Figure 1: System overview of *LOGO*.**

## 2.1 User Profiling

In LOGO, we employ Latent Dirichlet Allocation (LDA) [3] as the language model to detect latent topics, and represent the topic distribution of the news collection as a topic vector, each entry of which denotes the weight of the representative words in each topic.

### 2.1.1 Long-Term Profile

The long-term user profile is constructed using a time sensitive weighting scheme. Formally, given the reading history $\mathcal{H}$ of a specific user $u$, $\mathcal{H}$ is initially divided into multiple segments based on a uniform time period $T$, $\mathcal{H} = \{\mathcal{H}_{t_0}, \mathcal{H}_{t_1}, \mathcal{H}_{t_2}, \cdots, \mathcal{H}_{t_n}\}$, where $t_0$ means the current time period. For each time period $t_i$, $i = 0, 1, 2, \cdots, n$, we summarize the corresponding reading history $\mathcal{H}_{t_i}$ using LDA and generate a short-term profile $\mathcal{P}_{t_i}$. Note that we assume the user's reading preference evolves over time. We then define a time function $f(t)$ to find appropriate time weights for each $\mathcal{P}_{t_i}$ in the order that the recent reading histories are able to contribute more to the long-term profile. The user's long-term profile $\mathcal{P}_u$ can be represented as

$$\mathcal{P}_u = f(t_0) \cdot \mathcal{P}_{t_0} + f(t_1) \cdot \mathcal{P}_{t_1} + \cdots + f(t_n) \cdot \mathcal{P}_{t_n}, \tag{1}$$

where $\mathcal{P}_{t_i}$ and $\mathcal{P}_u$ are all topic vectors.

Intuitively, we are concerned with a user's recent reading preference since the recent one could represent the user's current reading interest. More recent reading histories should have higher value in the time weighting scheme. Therefore, $f(t)$ is a monotonic decreasing function, which reduces uniformly with time $t$ and the value of the time weight lies in the range $[0, 1]$. Motivated by [5], we choose an exponential form for the time function, which is able to describe the gradual decay of the importance of past reading histories as time goes [1]. The time function is as follows:

$$f(t) = e^{-\lambda \cdot t}, \tag{2}$$

where $\lambda$ represents the profile decay rate. In the experiment, we choose $\lambda$ as the inverse of the time period $T$ we are using to divide the entire reading history of the user.

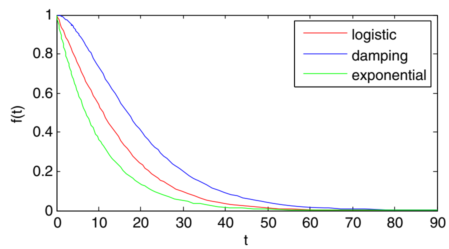


**Figure 2: Comparison of different time functions.**

The exponential function can satisfy our need well. However, there are some other time functions which might also be useful, such as logistic function ($f(t) = \frac{2}{1+e^{\lambda \cdot t}}$), and damping function ($f(t) = (1 + \lambda \cdot t)e^{\lambda \cdot t}$). From Figure 2, we can observe the difference between these three time functions, where $\lambda$ is set to be 1/10. It is straightforward that for all these three functions, the gradient of the curve at the data points close to zero is steeper than that of the data points far away from zero, which perfectly fits the actual scenario of the user's reading history – more recent histories should carry higher weight than the older ones, and also the weight value should decrease slower as time is far from the current time. However, the exponential function shows the greatest decay speed, compared with the other two functions. In the experiment, we provide empirical comparison among the performance using these three functions respectively, and detailed analysis on how to select $\lambda$.

### 2.1.2 Short-Term Profile

Once we obtain the long-term profile for the given user, we can easily deduce the short-term profile we are going to utilize to select specific news items. For simplicity, we choose the latest short-term profile $\mathcal{P}_{t_0}$ to achieve this goal. The reason behind this lies in the fact that $\mathcal{P}_{t_0}$ can represent the user's current reading preference, and therefore the most recent reading history would be more beneficial when filtering news items to the end user.

## 2.2 Group Selection using Long-Term Profile

When dealing with the newly-published news articles, we first divide the article set into distinct news groups using hierarchical clustering algorithm based on the cosine similarity of news content, where the inner cluster similarity is evaluated by average-link metric. To generate groups of news articles, we use Dunn's Validity Index [6] as the metric. This validity measure is based on the idea that high-quality clustering produces well-separated compact clusters. In general, the larger Dunn's Index, the better the clustering. In this way, we do not have to specify the number of clusters when performing clustering on news articles.

Once the collection of news items is well-separated, we summarize each news group using LDA, and present each group as a topic vector to facilitate the group filtering with the long-term user profile. Formally, provided that the summarized news clustering result $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_k\}$, where $\mathcal{C}_i$, $i = 1, 2, \cdots, k$ is a topic vector and $k$ is optimized by Dunn's index, our recommender system automatically calculates the cosine similarity between the long-term user profile $\mathcal{P}_u$ and $\mathcal{C}_i$, and selects the top ranked news groups for further processing. The system sorts the similarity scores in descending order and then selects groups with the similarity higher than the median. Up to this point, we can obtain a list of news groups that the user might prefer in a long-run perspective. It is straightforward that the selected news groups would probably cover the user's general interest based on the topic distribution of the long-term user profile.

### 2.3 News Selection using Short-Term Profile

When filtering news item in each news group, $\mathcal{P}_{t_0}$ (the latest short-term profile) is being considered as the comparison base. Note that the latest short-term profile represents the most recent reading preference of the user, and therefore it is quite reasonable to compare it with news items in each news group. Similar to the previous comparison, each news article is processed using LDA, and then compared with the topic vector of $\mathcal{P}_{t_0}$. In each news group, we adopt a greedy algorithm to sequentially pick up the news article with the largest similarity. To integrate recommended news items from different news groups into the final recommendation list, top ranked items within each group are selected, and the number of items selected in each group is proportional to the interest weight of the user's long-term profile on the corresponding news group.

## 3. EMPIRICAL EVALUATION

### 3.1 Real World Dataset

For experiments, we gather news articles along with users' access history from popular news websites from Jan 15th, 2011 to Apr 16th, 2011. We preprocess the data by removing news articles that are rarely accessed (i.e., the accessed frequency is less than 10 times per day) and by storing users with frequent online reading behaviors (i.e., users who read news articles every day and read more than 10 pieces of news each day). After preprocessing, a total of 103,540 news items are stored, along with 3,430 users, each day in average with 1,125 news articles published.

### 3.2 User Interest Evolution over Time

Recall that our proposed recommendation framework is based on the assumption that the user's reading preference would be relatively stable in a long run, while the content read by the user might change significantly. To validate this assumption, we empirically segment the user's reading history into 3-day time slots and 15-day time slots, respectively, and then examine the profile in 3-day time slot based on word distribution related to general topics (represented by word frequency), and the profile in 15-day time slot based on topic distribution (detected by LDA).

Specifically, we investigate the possibility of interest evolution on different user groups, since users with different news

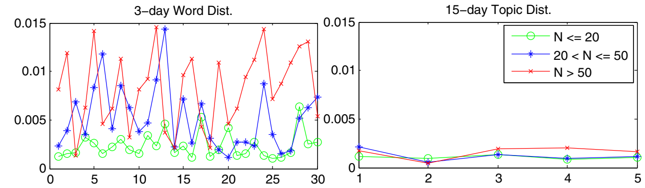


**Figure 3: KL-Divergence of short-term word distributions (topic-related) and long-term topic distributions over different user groups.**

access patterns, such as different reading frequency every day, may have distinct news topic preferences, and therefore the dynamic interest on news articles may vary a lot. To do that, we divide the user pool into three groups based on their reading habits. Suppose a user reads $N$ news articles per day, then the three groups are: (i) $N \leq 20$ (25%); (ii) $20 < N \leq 50$ (38%); (iii) $N > 50$ (37%). To verify whether the user's interest evolves over time, we calculate the KL-Divergence [9] between two profiles of adjacent time slots for each user, and then average the value of different time slots over the three user groups. Figure 3 shows the result, where $X$-axis represents the time slot pair, and $Y$-axis denotes KL-Divergence value. As is evident in Figure 3, all the three user groups exhibit the interest evolution: the general topics (topic distribution in the right figure) are relatively stable in a long run while the specific content (word distribution in the left figure) of those topics that a user might be interested in changes significantly in a short run. Particularly, users with higher access frequency on news articles shift their reading preference more dramatically than the other two groups of users.

### 3.3 Time Function Comparison

In Section 2.1.1, we have analyzed the effects of different time functions. To better understand how the time function influences the final recommendation result, we evaluate the performance of different time weighting schemes (logistic, damping and exponential). For evaluation, we choose 5 time ranges for each user, and use the reading history before the time range to construct the user's profile. Note that we have the ground truth for users in these time ranges. For each approach, we recommend news items (top @10, top @20 and top @30) to 100 randomly selected users, and plot the averaged precision and recall pair of recommendation results over 5 time ranges, where each time range contains 3 days. We construct the user's profile based on the reading history earlier than each time range. Note that the decay factor $\lambda$ is empirically set to 1/3. In next section, $\lambda$ will be tuned based on the recommendation result. As is depicted in Figure 4, besides the higher ratio of precision and recall, the performance distribution of exponential function is more compact than the others, which demonstrates the efficacy and stability of our proposed time weighting scheme.

### 3.4 Parameter Tuning

The value of the decay factor $\lambda$ denotes the decay rate of the user's reading preference over time. The higher the value of $\lambda$, the faster old reading histories decay and the lower the importance of the historical information compared to more recent profiles. To better capture how the recommendation result is influenced by $\lambda$, we use different time periods $T$ to

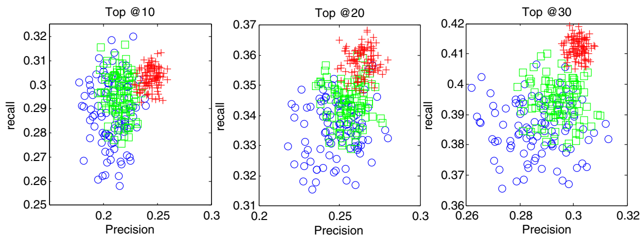


**Figure 4: Precision-recall plot for different time weighting schemes. Remark: "○" represents the performance using damping function; "□" denotes the performance using logistic function; and "+" represents the one using exponential function.**

segment the user's reading history (the corresponding decay factor $\lambda = 1/T$), and adopt the same experimental setup with the above procedure. Instead of using precision and recall, we calculate the F1-score for each selected user, and plot the averaged F1-score for each $\lambda$. The tuning result is shown in Figure 5. As is depicted, for top @10, @20, @30 recommended news articles, the performance achieves the best when $\lambda = 1/3$, meaning that the time slot of each history segment is 3 days. Therefore, we set $\lambda$ as 1/3 for the following experimentation.

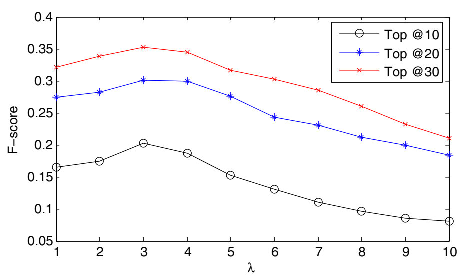


**Figure 5: $\lambda$ tuning curve.**

## 3.5 Comparison with Other Methods

Our proposed recommendation framework takes into account the long-term and short-term interest profiles of the users in an integrated way, where the long-term profile is designed to filter preferable news groups, and the short-term one aims to retrieve the attractive news items in each selected news group. In order to verify the effectiveness of the two-stage news selection strategy in LOGO, we use the long-term (LT) and short-term (ST) profiles, respectively, to perform the recommendation task, as two baseline methods. In each baseline, the selection of news groups and news articles are all based on a single profile, either the long-term one or the short-term one. In addition, we implement two existing approaches, [4] (Goo) and [11] (ClickB) for comparison. The former is a collaborative filtering based method, whereas the latter is a content-based method. The experimental setting is the same as the previous procedure. We then recommend news items (top @10, top @20 and top @30) to 100 randomly selected users. For comparison, we compute the averaged precision, recall and F1-score of recommendation results for these users over the 5 time ranges. Table 1 shows the comparison results.

| Method | Top @10 | | | Top @20 | | | Top @30 | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F |
| LT | 0.1552 | 0.2134 | 0.1741 | 0.2364 | 0.3137 | 0.2606 | 0.2601 | 0.3511 | 0.2917 |
| ST | 0.1628 | 0.2206 | 0.1850 | 0.2401 | 0.3054 | 0.2612 | 0.2659 | 0.3306 | 0.2845 |
| Goo | 0.1894 | 0.2362 | 0.2061 | 0.2559 | 0.3288 | 0.2889 | 0.2783 | 0.3673 | 0.3125 |
| ClickB | 0.1808 | 0.2387 | 0.2013 | 0.2528 | 0.3196 | 0.2835 | 0.2806 | 0.3520 | 0.3046 |
| LOGO | 0.2103 | 0.2480 | **0.2148** | 0.2736 | 0.3618 | **0.3051** | 0.3135 | 0.4021 | **0.3411** |

**Table 1: Comparison among different recommenders.**

From the result, we observe that: (1) Simply using one single profile (long-term or short-term) cannot guarantee the double-effect of the general topic interest and the recent reading preference; (2) Treating the user's reading interest over different topic categories independently, like ClickB, cannot effectively capture the exact reading interest of users; and (3) Our proposed recommender LOGO outperforms other methods by providing a seamless integration of long-term and short-term user profiles.

# 4. CONCLUSION

In this paper, we initially provide an experimental study on the evolution of user interests in real-world news recommender systems. To better capture the interest evolution issue, our proposed recommender LOGO, seamlessly integrates the long-term and short-term reading preferences of users when recommending news items. The time sensitive weighting scheme on the long-term profile, along with the two-stage news recommendation framework, shows promising performance compared with other existing methods.

# ACKNOWLEDGEMENTS

The work is supported in part by NSF grants IIS-0546280, CCF-0939179 and HRD-0833093.

# 5. REFERENCES

[1] C.C. Aggarwal, J. Han, J. Wang, and P.S. Yu. A framework for projected clustering of high dimensional data streams. In *Proc. of VLDB*, pages 852–863, 2004.

[2] D. Billsus and M.J. Pazzani. A hybrid user model for news story classification. *Internatinal Center for Mechanical Sciences*, pages 99–108, 1999.

[3] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] A.S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *Proc. of WWW*, pages 271–280. ACM, 2007.

[5] Y. Ding and X. Li. Time weight collaborative filtering. In *Proc. of CIKM*, pages 485–492. ACM, 2005.

[6] J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Cybernetics and Systems*, 3(3):32–57, 1973.

[7] J. Jancsary, F. Neubarth, and H. Trost. Towards context-aware personalization and a broad perspective on the semantics of news articles. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 289–292. ACM, 2010.

[8] H.R. Kim and P.K. Chan. Learning implicit user interest hierarchy for context in personalization. In *Proc. of IUI*, pages 101–108. ACM, 2003.

[9] S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[10] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan. Scene: a scalable two-stage personalized news recommendation system. In *Proc. of SIGIR*, pages 125–134. ACM, 2011.

[11] J. Liu, P. Dolan, and E.R. Pedersen. Personalized news recommendation based on click behavior. In *Proc. of IUI*, pages 31–40. ACM, 2010.