The Oblivion Problem: Exploiting forgotten items to improve recommendation diversity

Fernando Mourão, Claudiane Fonseca, Camila Araújo, Wagner Meira Jr.
Department of Computer Science - Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos 6627 - ICEx - 31270-010, Belo Horizonte, Brazil
(fhmourao,fonseca,camilaaraujo,meira)@dcc.ufmg.br

ABSTRACT

Recommender Systems (RSs) have become a crucial tool to assist users in their choices on various commercial applications. Despite recent advances, there is still room for more effective techniques that are applicable to a larger range of domains. A major challenge recurrently researched is the lack of diversity in the recommendation lists provided by current RSs. That is, besides being effective to suggest interesting items to users, a good RS should provide useful and diversified items. In order to address this problem, we evaluate the use of forgotten items in recommendation. By forgotten items, we mean items that have been very relevant to users in the past but are not anymore. Therefore, we formally define the **Oblivion Problem**, which is the problem of recommending forgotten items, propose a methodology for verifying it in real scenarios, and perform a deep characterization of this problem in a relevant music domain, the *Last.fm* system. Applying our methodology to Last.fm has demonstrated the existence of the oblivion problem in practice, as well as showed the utility of this methodology. Further, the behavior exhibited by forgotten items in *Last.fm* suggests that defining techniques that incorporate such items into RSs consists in a promising research direction.

Categories and Subject Descriptors

H.4.m [**Information System Applications**]: Miscellaneous; H.m [**Information System**]: Miscellaneous

General Terms

Recommendation, Formalization, Characterization

1. INTRODUCTION

Recommender Systems (RSs) are becoming increasingly important tools for many commercial applications due to their ability to filter a huge and growing volume of options, showing only what may be interesting to users [8]. We define RSs as any system that is designed to produce individualized recommendations as output, or to guide users through a huge variety of options [3, 7]. Intuitively, the growing demand for such tools may be explained by the so-called **Paradox of Choice** [15], which states that, as the number of options grows, the effort required to make a wise decision also increases, making the possibility of choosing a burden, instead of an advantage.

Despite the numerous strategies proposed for RSs, current systems still lack effectiveness in terms of identifying not only accurate but also diversified lists of recommended items [19]. The problem of such lack of diversity is that recommending over and over again the same items, even being relevant ones, to the same users is likely to annoy them, decreasing their interest in interacting with the RS over time. It was observed that, although the domains where RSs operate present a wide diversity of items, the recommendations are, in general, poor in terms of diversity [19], as a consequence of a huge concentration of users around few popular products followed by a much smaller demand around the other products, a phenomenon known as *Long Tail* [2]. As a result, only a small portion of items obtain enough ratings and, therefore, is considered suitable for recommendation [13]. In addition, an important factor that affects the items popularity, and consequently the lack of diversity on recommendations, is their aging, since the probability of an item to be recommended is inversely proportional to its age [2]. In summary, diversity means accuracy loss in most application scenarios, and achieving both in a wide range of real-world scenarios such as travel and financial services, among others, is a constant challenge.

Traditionally, diversity in RSs increases with the arrival of new items in the system. Nevertheless, new items have few evaluations and do not contribute immediately to improve recommendations [1]. Another source of diversity that is assessed in this paper is to use forgotten items. We define as forgotten items any item that used to be relevant and of frequent interest to a particular user in the past, and now it is not. The main hypothesis of this work is that forgotten items, that appear in the tail of the popularity distribution of items, may increase the recommendation diversity while keeping its accuracy. The main premise here is that user recommendation profiles are defined as a function of their most consumed items in a given period of time, and despite the relevance loss associated with aging and competition associated with new items, their importance in the past may guarantee a good recommendation. Further, given the amount of information associated with forgotten items, they represent a richer source of information, compared to new ones, becoming more suitable for RSs. In addition, rescuing these items may be surprising and bring back good memories in scenarios where old items may remain interesting for the users over time.

There are two key issues related to use forgotten items that are addressed in this work. The first one consists of validating the main hypothesis, that is, the utility of forgotten items, since both users and the domain as a whole evolve over time. For instance, considering the music domain, someone who used to like *Cindy Lauper* in the past not necessarily would like her nowadays. Further, forgotten items may have a very long past, but no recent information, since they are no longer consumed. The second key issue is to assess the utility of a forgotten item, which is difficult because items are forgotten at different ages by different users.

One interesting observation is that the problem of recommending forgotten items, which we call the **Oblivion Problem**, is similar to the *Cold Start* problem in RSs [14], once the difficulty of recommending new items also lies in the absence information, but now due to lack of past information. Such similarity may allow the use of current solutions for the *Cold Start* in the task of recommending new and forgotten items, in order to appropriately improve the diversity in traditional RSs.

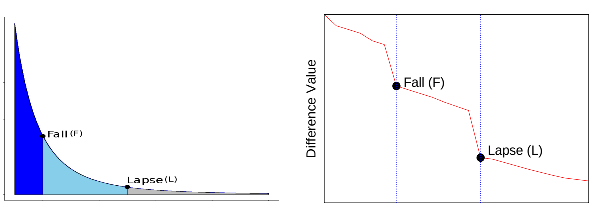
We restrict the evaluation of such problem to the music recommendation domain, due to an intrinsic feature of this domain: the high repetition rate in music consumption over time. That is, users tend to listen to a given song repeatedly more often than watch a given movie, reinforcing this work hypothesis. Further, this behavior shows that the requirement that old items may remain interesting for the users over time could be expected in such scenario.

In summary, the aims of this work are restricted to define and to characterize the Oblivion problem, rather than proposing solutions for it. In this sense, we highlight the following contributions: **(1)** the formalization of a new problem in RSs, namely the **Oblivion Problem**; **(2)** the proposal of a methodology to verify whether this problem occurs in real domains; and **(3)** a deep characterization of the Oblivion Problem in a real and relevant scenario, *Last.fm*. Our analyses use *Last.fm* as workload, since it represents one of the largest musical community in the world, comprising more than 12 million distinct artists and 30 million active users at the gathering moment. Besides this huge size, the availability of most data on the WEB make this system a promising data source for studies on music recommendation. Results from our methodology on a *Last.fm* sample demonstrate the existence of the Oblivion Problem in this scenario, as well as show the practical utility of this methodology. Further, the behavior exhibited by forgotten items in this domain suggests that defining techniques that incorporate such items into RSs represents a promising research direction.

2. BACKGROUND CONCEPTS

In this section we present some key concepts for formalizing the Oblivion Problem. We start by discussing the role of long tail distribution on the recommendation domain, since it is one of the main motivations for the Oblivion Problem. After, we introduce some definitions derived from the analysis of this distribution.

Recently, the so-called *Long Tail* distribution [13] is regarded as one of the most recurrent data models for various commercial applications. It is defined as a distribution of items ordered decreasingly by the number of distinct users who have consumed each item in a given period of time. This distribution indicates a sharp and strong interest for a restricted set of popular products, followed by fast demand decrease that extends for a long and low tail, associated with increasingly unpopular products [2]. Figure 1 *(a)* presents an example of this distribution.



(a) Example Distribution (b) Example of Distribution Partition

Figure 1: Long Tail

The relevance of this distribution to RSs stems from a property discussed by Anderson [2], which states that the long tail defines a market of a large number of non-popular items that rivals with the popular ones. In fact, the emergence of very specific users' niches, made explicit by the long tail, increases the need for RSs that are more capable of providing diverse and accurate recommendations, since recommending just popular items is not enough to reach a large portion of those niches. However, long-tailed distributions pose some challenges for RSs and the main one refers to the scarcity of information about many of the items, since most of them are consumed infrequently.

As illustrated in Figure 1 *(a)*, we divide long-tail distributions into three parts. The first part represents the head of distribution and is composed of the most popular items of the system, and, controversially, the smallest number of unique items. The second part is the body of the distribution, which includes a larger number of distinct items but with a lower popularity compared to the items in the first part. Finally, we have the tail, which represents the vast majority of existing items in commercial domains. Each of these items, however, is consumed only by a negligible portion of distinct users of the domain. We define the *Fall Point* (F) as the point in the distribution that separates the head from the body, and the *Lapse Point* (L) as the point that splits the body from the tail. In general, the identification of such points is performed through some transformation functions on the distribution. For example, we can plot the absolute values of the differences between adjacent points in the distribution. In this plot, those points are identified by possible existing elbows, as illustrated in Figure 1 *(b)*.

For recommendation purposes, we determine the distribution for a delimited period of time. That is, for a given period of time we evaluate which items are more or less popular, obtaining different distributions for distinct periods. Thus, the dynamics of real scenarios is captured and items that are no longer relevant for the users over time are not considered in the distribution. The peculiarity of recommendation scenarios is that, regardless the period of analysis, the generated distributions are long-tail as a consequence of the following facts. First, most of the novel items that appear during each observation period remain at the tail of the distribution. Second, there are always "migration" movements of items along the distribution. That is, some items become more popular, migrating toward the head, while other items that are no longer relevant become less popular, moving toward the tail of the distribution. By considering both the three parts and possibility of popularity changes over time, we denote the items from the head as *successful* items, items from the body as *transition* ones, and items belonging to the tail as *unpopular* ones. In fact, item migration along the long tail distribution over time is the starting point for the Oblivion Problem, which is described in the next section.

3. THE OBLIVION PROBLEM

In the last section we discussed as the popularity distribution of an item stems from the consumption habits of all its users. Conversely, we may adopt a similar concept when modeling the popularity of products for a given user, that is, we may define individual relevance distributions for each user. That is, let S be the set of all items consumed by a particular user U_i in a given period of time M_c. Given a non-increasing frequency distribution of the consumption of those items by U_i during M_c, the most consumed items, thus the most relevant ones, are on the distribution head while the remaining are distributed along the tail, following a long-tail shape. This distribution is a simple model for determining "successful" items for user U_i during M_c.

Considering the age of each item I_j as a relevant factor that affects how often I_j is consumed by U_i, the older I_j is for U_i the less relevant it becomes, moving gradually toward the tail of the distribution defined for U_i. Thus, the temporal dynamics, the variety of items that arise, and the competition among them for the user preference results in successful items being forgotten, which makes them less and less visible over time. Therefore, we define as **oblivion** this natural shift of successful items from the head of individualized distributions toward the tail over time.

More formally, we define the Oblivion Problem based on the items movement on the long tail distribution of items relevant to each user. As discussed in Section 2, we divide this type of distribution into three distinct regions, delimited by the *Fall point* (F) and the *Lapse point* (L). From these definitions, we may better model the movement of items along this distribution over time. In this sense, we can define a probability density function $f_{i,j}(t)$ of each item I_j be consumed by a user U_i at a given time t, as illustrated in Figure 2. Such function is hypothetically impossible to be exactly defined, since it may include many unmeasured or even unknown variables, as well as subjective aspects inherent to the users.

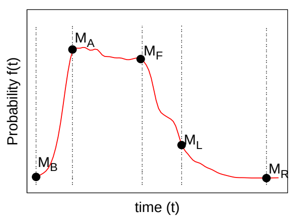


Figure 2: Hypothetical Probability Density Function for a user U_i listening to an item I_j over time.

It is interesting to notice that, despite the possible approximations to function $f_{i,j}(t)$, we can define five distinct critical time moments for this function. The first one refers to the *Birth Moment* M_B of I_j for U_i, and represents the first moment at which I_j was consumed by U_i. The second moment, called *Ascension Moment* M_A, comprises the moment at which the frequency that U_i consumes I_j exceeds the value found at the *Fall point* on the items relevance distribution of U_i, as illustrated in Figure 1 *(a)*. The third moment M_F refers to the *Fall Moment* and it is related to when I_j cross back the *Fall point* towards the tail, losing its relevance. The fourth moment represents the *Lapse Moment* M_L and it is associated with the moment that I_j has crossed over the *Lapse Point* L on the relevance distribution towards the tail. Finally, we have the *Rescue Moment* M_R, which refers to the moment at which the probability $f_{i,j}(t)$ becomes again greater than the probability found at the moment M_L. It is important to mention that $f_{i,j}(t)$ may present distinct behaviors along these five moments. Between M_B and M_A, we have the displacement period of the item toward the head, showing that I_j is becoming relevant to U_i. In the period between M_A and M_F, I_j is among the most frequently consumed items by U_i. Subsequently, we have between M_F and M_L, the period in which I_j is consumed much less frequently. After, the period between M_L and M_R is defined as the period during which I_j becomes unpopular to U_i, presenting a very low probability of being consumed. Thus, this represents the period during which the items become forgotten. In fact, even though, theoretically, the probability of I_j be consumed during this period is greater than zero, in practice I_j is not consumed, becoming effectively forgotten. Finally, we have the period after M_R when I_j again becomes potentially relevant to U_i.

It should be highlighted that not all items I_j will necessarily present the five defined moments. In fact, the vast majority will not reach the *Ascension Moment* and some of them will not have a *Rescue Moment*. Probably a significant part of the items that exceed the *Lapse Moment* may never be rescued simply by representing a mismatch w.r.t. the user taste, which changes dynamically. Further, we are not assuming that the function $f_{i,j}(t)$ exhibits a monotonic behavior. In fact, some items may have more than one *Ascension Moment*, for instance, presenting a periodic behavior. At this way, we can define the Oblivion Problem as the problem of determining the *Rescue Moment* M_R for each item I_j that has achieved, at least once, the *Ascension Moment* M_A in the past.

Clearly this problem represents a challenging task, since it consists of predicting **which** items and **when** they must be recommended to users again. Since not all items in the tail are likely to be rescued, it is also important to know the exact moment to recommend them. Premature recommendations may be ineffective, since users may still be "tired of" the recommended items. On the other hand, late recommendations may no longer be of users interest given the evolution of their tastes. Moreover, the task of identifying a subset of relevant items from the tail is a challenge by itself. In addition to choose among a huge range of items, the recommendations domain is inherently dynamic, since both environment and users evolve over time. Thus, in a music scenario, for example, people who used to like *Cindy Lauper* long ago no longer like listening to her today. New songs come out and remixed versions of old songs may become more attractive to a given user. Finally, it is important to consider the trade-of between recommending forgotten items and new ones, since the overuse of forgotten items may make the recommendation even less diversified.

A relevant aspect to point out is that the Oblivion Problem may be defined considering two perspectives: individual and global. From an individual perspective, we are interested in finding the *Rescue Moment* of items for each user, considering the individualized relevance distributions. Thus, the same item might have different rescue moments to distinct users, or even not be relevant to others. The rescue of forgotten items, in this case, improves the ability of personalizing recommendation services. In turn, the global perspective aims to identify, at each moment t, what the best items, forgotten by the system as whole, to be rescued are, given the traditional popularity distribution of items in the system at the moment t. That is, we are interested in identifying the items that were collectively successful in the past and exhibit a high probability of being successful again. In the music scenario, for instance, this perspective has a great utility for record labels, assisting them in the following question: Which music to be remixed at the time t is likely to become a hit?

In this work, we evaluate specifically the Oblivion Problem in music recommendation domains, given an important characteristic inherent to this sort of domain. Considering the scenario in which the focus is on the individualized preferences, it is believed that, in general, the frequency that a user might listen to a particular song is significantly higher than his or her willingness to watch a given movie or to read a given book again. Therefore, rescuing a particular song seems to have a different impact, probably more promising, than retrieving an old movie or book. Probably, the reason for such difference is related to the time required to listen to a song compared to watch a movie or to read a book.

4. METHODOLOGY OF CHARACTERIZATION

In this section, we present a characterization methodology, for recommendation domains, that assesses empirically the two main issues related to the Oblivion Problem: its existence and utility in real scenarios. Therefore, we divide our methodology into 2 main steps, namely: **Problem Verification** and **Utility Analysis**. It is noteworthy that, since the Oblivion problem is stated regarding a global and an individualized perspective, such methodology can be applied for evaluating both of them. All metric descriptions were designed considering the global perspective, but for individualized analyses all we need to do is to ignore the summarization process executed by mean or median calculations and distributions.

4.1 Problem Verification

Our first task consists in verifying the existence of the Oblivion Problem in real scenarios. Thus, measuring and understanding how this problem manifests itself is a primary goal. To this end, we define five metrics that, together, provide evidences of the natural process of forgetfulness that occurs in recommendation scenarios. Table 4.1 defines each of these metrics and we discuss them next.

The *Mean Individualized Relevance* allows us to verify whether, in fact, items become less frequently consumed by users over time. The *Mean Inter-consumption Interval*, in turn, aims to identify how the interval between consecutive consumption of an item behaves over time. If this interval increases, it means that users tend to consume the item less and less frequently, until they effectively stop consuming it. Considering now, the *Mean Items popularity*, we aim to verify whether items become less popular or not over time in the system. That is, in addition to a given item becoming less relevant to each user individually, we also investigate whether it becomes globally forgotten by the system, since fewer users continue to consume it over time. In a complementary way, the *User Mean Age of Items* verify whether there is an increasing, stable or decreasing consumption behavior of the "age" of the consumed items over time, considering the moment at which each user first consumes each item. If such behavior is stable, decreasing or increasing at a rate lower than the actual aging rate of items and users in the system, we have a scenario in which items known for a long time by the users become forgotten. Similarly, the *System Mean Age of Items* analyzes this aging rate of the consumed items, but now considering the first time that each item appeared in the system. In this case, a decreasing behavior demonstrates that users also tend to focus their consumption on recently released items. Finally, the *Age Correlation* aims to verify whether items consumed by the first time by a given user are new or old in the system, showing up their interest in finding old items.

For sake of analysis, if a domain presents a decreasing relevance of items over time; an increasing inter-consumption time interval between consecutive consumptions of an item for a same user; such items still exhibit a descending popularity over time; and users present a stable consumption pattern focused on recent items in the system; we have a clear picture of oblivion. In this case, the rescue of forgotten items might be an important strategy to diversify the recommendation.

4.2 Utility Analysis

Once we verify that the Oblivion Problem happens in a given domain, the next step consists in checking the relevance of the forgotten items to RSs. In fact, may forgotten items be useful to recommendation in this sort of domain? This methodology step is concerned with answering this question. We define in Table 4.2 the main metrics related to the Utility Analysis and discuss them next.

Through the *Percentage of Successful Items*, we aim to analyze the diversity of items regarded as relevant for each user at each moment, based on the premise that a moderate diversity is better for popular items. It should be verified since, for a large diversity of successful items, users may behave dynamically, and, as a consequence, most of such items may not stay relevant for many users. In this case, recommending these items might worsen the accuracy of the recommendations. On the other hand, a very restricted range of successful items helps very little in diversifying recommendations, given the small number of distinct items. Considering the *Probability of Continuous Return*, we aim to determine whether,

Metric	Description
Mean Individualized Relevance (**PV-1**)	First, we define for each item I_j its "birth" moment in the system as the first moment at which I_j has been consumed by any user. Later, for each age A_t of I_j we count how many times, on average, I_j was consumed by the subset of users who have consumed it on age A_t. Then, we normalize the frequency found for each item I_j, on each age, by the highest value found for I_j throughout the period of analysis. Finally, we define the mean individualized relevance on each item age A_t as the mean of the normalized frequencies of all items analyzed at A_t.
Mean Inter-consumption Interval (**PV-2**)	Let I be the set of all items consumed by a given user U_i. For each item $I_j \in I$, we define its "birth" to the user U_i as the first moment at which I_j has been consumed by U_i. Then, we define the Inter-consumption Interval for each age A_t of I_j (i.e., A_t is seen as the user age of I_j) as the number of time units between the age A_t and the last moment at which I_j was consumed by U_i. We repeat this process for all users of the domain and, finally, we calculate a mean time interval between consecutive consumptions at each age A_t considering all items I_j analyzed for all users.
Mean Items popularity (**PV-3**)	For each item I_j, first, we define its system age at each moment M_l of analysis, considering as its "birth" moment in the system the first moment at which I_j was consumed by any user. Thereafter, we identify the distinct number of users who have consumed I_j at each distinct age A_t. Then, we normalize the value found on each age for each item I_j by the highest value obtained for I_j in a single age. Finally, we define the mean for all items at each age A_t as the Mean Items popularity at A_t.
User Mean Age of Items (**PV-4**)	For each user U_i, we define his age in the system since the first time he or she has consumed an item. Considering the items, we set its age since the first time it was consumed by U_i in the system. Therefore, for each user age A_t we calculate the mean age of all items consumed by U_i at A_t. Finally, we define the mean age among all users at each age A_t as the User Mean Age of Items at A_t.
System Mean Age of Items (**PV-5**)	Again, we assign to each user his or her age in the system since the first time he or she has consumed an item. However, we define the age of each item considering the first time it has been consumed in the system by any user. Therefore, for each user age A_t we calculate the mean age of all items consumed by U_i at A_t. Finally, we define the mean age among all users on each age A_t as the System Mean Age of Items at A_t.
Age Correlation (**PV-6**)	For each item I_j, we define its system birth as described before. Then, for each user U_i, we define the user birth of I_j as the first moment at which I_j has been consumed by U_i. Later, we calculate the differences between both moments of birth of each item I_j for each user U_i. Finally, we generate a probability distribution of these values, which we call Age Correlation.

Table 1: Metrics for Problem Verification (PV)

compared to the behavior presented by new items, old items still consumed by the users are likely to be consumed for a longer time. Although these old items do not consist of forgotten ones, this analysis demonstrates the relevance of old items compared to new ones for each user individually. If an RS rescues forgotten items with similar relevance, it would be expected a similar behavior for the probability of these items be listened continuously as well. We may say the same about the analysis of the *Probability of Continuous Frequency*. We aim to verify whether old items are consumed more often than new ones, assuming that the same might occur with forgotten items properly selected.

The Analysis of *User Age Distribution of Items* aims to verify whether a set of old items, consistently, belongs to the consumption set of the users in different periods. If so, in fact, rescuing old items not only diversifies the items consumed by users, but also this particular subset of old items. In a complementary way, the analysis of *System Age Distribution of Items* aims to verify whether items considered new for each individual user are also new to the system, or if they are actually old items just discovered by the users. This analysis makes possible to contrast individual behaviors to global trends in terms of consumption of new and old items. Therefore, from this set of metrics we determine a relevance measure of forgotten items for each scenario, allowing us to check the potential of techniques focused on rescuing forgotten items.

It should be noted that our methodology is based on comparative analyses between distinct periods. Thus, for verifying if a given scenario suffers from the Oblivion Problem it is necessary some assessments over time, in order to identify behavioral trends. Absolute values of these metrics, by themselves, are not enough for such purpose. In order to demonstrate the applicability of the proposed methodology, we evaluate it on a relevant music scenario in the next section.

5. LAST.FM: A CASE STUDY

5.1 Dataset

Metric	Description
Percentage of Successful Items (**UA-1**)	For each user U_i, at each age A_t in the system, we define the ratio between the total number of distinct items recognized as successful ones for U_i at A_t and the total number of distinct items consumed by U_i at A_t. An item I_j is considered successful for a user U_i, at a given moment A_t, if their frequency of consumption is X standard deviations greater than the mean frequency of consumption of all items consumed by U_i at A_t.
Probability of Continuous Return (**UA-2**)	First, we define two distinct sets of items for each user U_i at each moment M_a of analysis. The first set comprises the new items for U_I, since they were consumed by U_i for the first time on less than X time units before the moment M_a. In the second set, we have the old items, consumed by U_i for the first time on more than Y time units before M_a, with $Y > X$. In each set, we define for each item I_j the largest continuous time interval, considering all distinct moments M_b (such that $a >= b$), that I_j has been consumed by U_i. Then, we generate a probability distribution of these interval values found in each set for all users, obtaining two distinct curves.
Probability of Continuous Frequency (**UA-3**)	As defined for the metric **UA-2**, we assign to each user a set of new items and other of old ones. In each set, we define, for each item I_j, the mean frequency of consumption at the time interval identified as the largest continuous one for I_j, starting from the moment of analysis M_a. Then, we generate a probability distribution for the frequency values found in each set for all users, obtaining two distinct curves.
User Age Distribution of Items (**UA-4**)	Initially, we select, for each user U_i and time moment M_a, a set S of the K items most consumed at M_a. After, for each item $I_j \in S$, we define the age of I_j at M_a, considering as "birth" moment the first moment at which I_j has been consumed by U_i. Finally, we plot the percentage of occurrence of each age in S at each analyzed moment M_a.
System Age Distribution of Items (**UA-5**)	This analysis is exactly the same as described for the metric **UA-4**, except that we consider, to calculate the age of each item $I_j \in S$, its "birth" as the moment that I_j has been first consumed in the system.

Table 2: Metrics for Utility Analysis (UA)

In this section we present the dataset used in our analysis. We use a dataset from the *Last.Fm* system[1], which is a UK-based Internet radio and music community website, founded in 2002. It has claimed over 30 million active users when we collected the dataset. It was also estimated that *Last.fm* has more than 27 million different tracks and 12 million distinct artists in its database[2]. As *Last.Fm* represents one of the largest musical community in the world, and since all data are readily available in the WEB, it is a good data source for music recommender systems.

Our analyses were performed on a data sample from *Last.Fm*. These data were collected through an API provided by *Last.fm*[3]. This API allows us to collect information related to several data entities such as artists, albums, tracks, and users, among others. We consider as relevant to our analysis only information related to users, artists, and tracks. Such information was collected for a set of 104,770 distinct users, randomly selected, 217,774 different artists and about 2 millions distinct tracks listened to by the collected users, spanning the period from 11/12/2008 to 04/26/2009.

5.2 Characterization of the Oblivion Problem

In this section, we present the results from the application of the methodology presented in Section 4 to our *Last.fm* data sample.

5.2.1 Problem Verification

Starting our analysis by the Verification Problem step, we obtained the results shown in Figure 3. Analyzing, first, the measure **PV-1**, Figure 3 *(a)*, we observe that *Last.fm* appears as an environment with decreasing frequency of song consumption over time. That is, the longer the songs are in the system, the lower the frequency users listen to them. Furthermore, by analyzing the measure **PV-2**, Figure 3 *(b)*, we note that the interval between consecutive consumptions becomes larger as the song ages for each user. Therefore, the longer a user knows a song, in general, the higher is the interval between consecutive consumptions, until the user stops listening to these songs. Clearly, these results show that, over time, old songs loose relevance for each user at an individual basis.

Considering the *Mean Items Popularity* (**PV-3**), as shown in Figure 3 *(c)*, we note that the songs also become less popular globally in our *Last.fm* sample. That is, besides becoming less relevant over time for each user individually, in fact, the songs become less relevant globally for the system. Thus, over time, most of these songs may become forgotten, both by users and by the system as a whole, and, in general, even by the RSs. It is also noteworthy in this analysis the growth in popularity observed for the oldest songs (i.e., songs older than 19 weeks). Such behavior occurs as a result of the existence of a very restricted set of songs that are continuously listened by the users. Some of these songs are from artists such as *Beatles*, *U2* or *Michael Jackson*, which, interestingly, present a more stable popularity over time.

Our next analysis refers to the *Mean Age of Items*, considering both user (**PV-4**) and system (**PV-5**) perspectives. According to the users perspective, Figure 3 *(d)*, we found that consumed items age much more slowly than their actual aging rate in the system. That is, in general, the users consumption is focused on new items, rather than on items already known by him. Thus, the average age of consumed items in each week has becoming increasingly distant from the actual aging rate of users and items in the system (i.e., solid line in the plot). It is also important to emphasize that, although the consumed items age slowly, they do get old since the mean age curve is not a straight line parallel to the X axis, which shows that, even focusing their consumption on new items, users in *Last.fm* consume some old items. Thus, we may conclude that old items are part of the consumption behavior of users. Similar conclusions are obtained considering the system perspective, as shown in Figure 3 *(e)*. Over time the mean age of consumed songs is lower than the actual users and items age, defining a curve below the continuous one plotted in the graphic. This shows that, besides focusing their consumption on items that users themselves know for a short time, some of those items are also recent in the system as a whole. In this case, however, the difference between the two aging rates is not significant. Thus, in general, old songs are slowly being abandoned, or "forgotten", and as new songs are consumed as soon as they are released.

Finally, we assess the correlation between the items' age for each user and the actual items age in the system. The plot in Figure 3 *(f)* represents this analysis. We observe that 35% of the differences between the actual items age and the items age recognized by each user, as defined in section 4, are smaller than 4 weeks. That is, more than one third of the items consumed by the users in our sample were in the system for less than 1 month. On the other hand, only 14% of the consumed items are older than 18 weeks (i.e., 4 months and a half) when a user has first consumed it. Therefore, based on the Verification Problem analysis, we describe *Last.fm* as a scenario in which there are clear evidences of the existence of the Oblivion Problem. Its users consume more often new items in the system and, over time, consume a given item less and less frequently. In addition, the items, in general, become quickly less popular as they age. We also found that the users consumption, in its majority, consists of new items in the system that were first consumed recently.

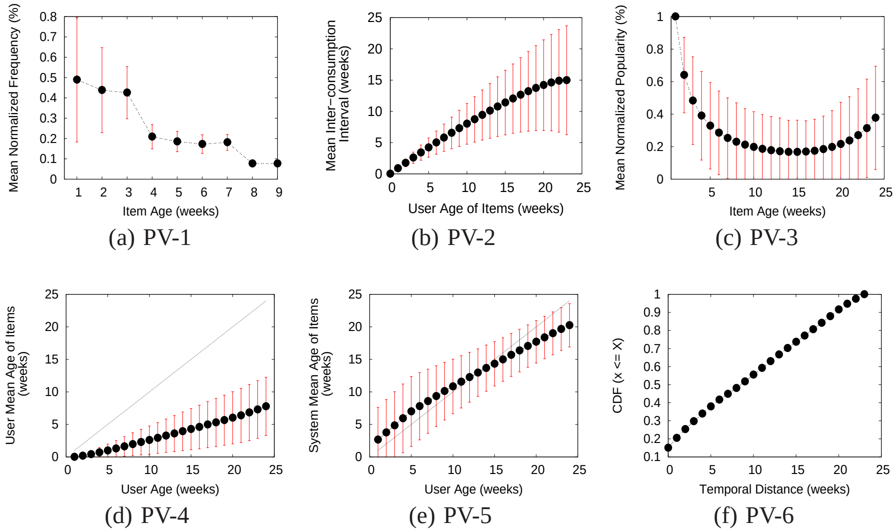
5.2.2 Utility Analysis

We start our Utility Analysis by the *Percentage of Successful Items* (**UA-1**), such as shown in Figure 4 *(a)*. For this analysis, we consider as time granularity one month and the number X of standard deviations equal to 2. That is, for a song being successful in a given month M for a user U_i, it should exhibit a frequency of occurrence, at least, two standard deviations higher than mean frequency of consumption of U_i at M. As we can observe, the mean percentage of distinct successful songs for each user of our sample is about 8% of the song set listened by him or her each month. However, we also found higher standard deviation values. Although we can identify only very few successful items for some

[1]Available at http://www.last.fm/

[2]This information were retrieved from the *Last.Fm Radio Announcement*, on 03/25/2009, available at http://blog.last.fm/2009/03/24/lastfm-radio-announcement

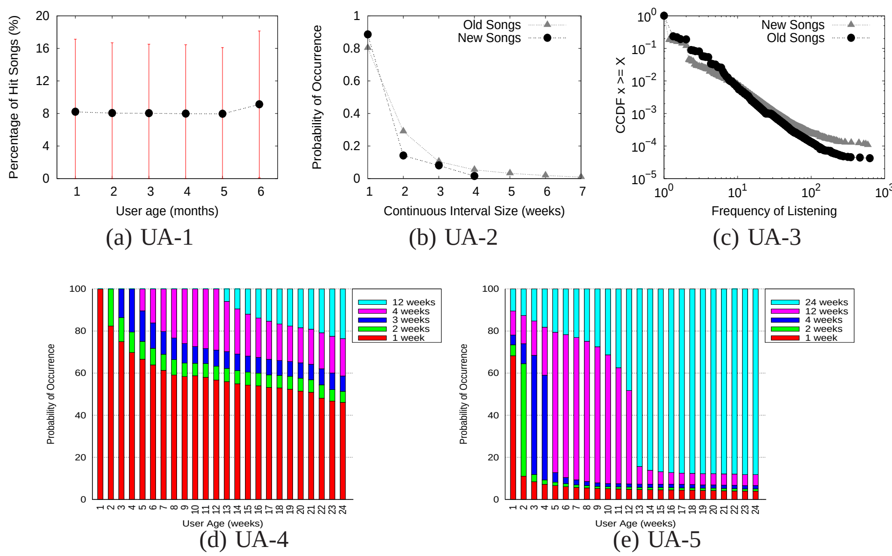
[3]http://www.last.fm/api



(a) PV-1 (b) PV-2 (c) PV-3

(d) PV-4 (e) PV-5 (f) PV-6

Figure 3: Results for Problem Verification



(a) UA-1 (b) UA-2 (c) UA-3

(d) UA-4 (e) UA-5

Figure 4: Results for Utility Analysis

users, in general, it is possible to identify a meaningful number of successful items, defining a rich set of relevant items over time for diversity strategies.

Considering the *Probability of Continuous Return* (**UA-2**), we found a very distinct behavior, regarding old and new songs, as shown in Figure 4 *(b)*. While new items tend to be listened for a short period of time, songs known for a long time are consumed continuously for larger periods. This fact demonstrates that the subset of older songs, even not being the majority, represents more consistently the users taste over time. In a complementary way, the plot in Figure 4 *(c)* shows the complementary cumulative distribution function (CCDF) for the *Continuous Frequency* Analysis (**UA-3**). As expected, new songs present a higher probability of being listened more frequently, since the users are still “learning" them. Contrasting this findings to those presented in the plot of Figure 3 *(a)*, we can conclude that, although forgotten songs are likely to be listened, they will not be listened as intensively as the moment when the users knew these songs.

Considering now the Analysis of *Age Distribution of Items* for both user (**UA-4**) and system (**UA-5**) perspectives, we also have interesting results, as shown in Figures 4 *(d)* and *(e)*, respectively. From the user perspective, we found that, despite the consumption behavior of *Last.fm* users is focused on recently discovered items, as the users age they tend to listen to old songs more often. That is, over time, users tend to keep listening to songs known for a long time. This finding is particularly relevant for us, since it demonstrates that a portion of songs being consumed by a user is composed of old songs. Consequently, forgotten items would be relevant even to diversify this specific subset of songs.

Finally, we focus on the system perspective, as shown in Figure 4 *(e)*, which requires a careful analysis. First, it is important to remember that we do not know the actual age of the items in the system. As we consider as the “birth moment” of each item the first time it has been consumed by any user of our sample, it is expected that at age 1 most of the users consumes 1 week old songs. However, we can see that over time users fail to bring new items to our sample and start to listen to items that already belong to it. That is, although users look for new items in the system, they increasingly consider novel items already present in the system. Further, as the number of existing items grows over time, the probability of a user to find a recent addition to the system that are also relevant to her or him decreases.

In summary, we can conclude that, despite the continuous and fast changes in terms of user taste, expressed through the listened songs, the few remaining old songs that he or she still listens regularly represent exactly the subset of songs listened to by a longer period of time. In addition, over time, *Last.fm* users tend to anchor their consumption more and more on the older songs in the system. Consequently, these songs represent a relevant set of information about users, demonstrating the significance of old items in the music recommendation domains. Thus, we believe that rescuing relevant forgotten songs would be particularly promising for diversifying not only the overall list of songs listened by each user but also this subset of old songs.

6. DISCUSSION

Obviously, once we have shown the existence and relevance of the Oblivion Problem, our next step is to identify forgotten items, and then select a subset of most relevant ones for RSs. A straightforward way of identifying the items forgotten by a user, or by the system as a whole, can be derived from the problem definition. After identifying the Ascension and Lapse moments of each item for each user, we need only to select the items that have had at some moment, at least, one *Ascension Moment* and are no longer being consumed by users (i.e., are beyond the *Lapse Moment*). The selection process of a subset of relevant items, since not all forgotten items are likely to be recommended, is more challenging. As the probability function $f_{i,j}(t)$, of a user U_i consumes an item I_j, itself is, in practice, impossible to be modeled, we need to define and evaluate heuristic strategies, defining such probabilities over time and then select the items with highest probabilities at each moment t. Further, considering the trade-off of recommending old and new items, in order to achieve high accuracy and diversity rates simultaneously, a final step is related to incorporate the selected forgotten items into traditional RSs.

It is also interesting to discuss novelty issues related to the Oblivion Problem. As the primary goal is to recommend items already known by users, our first aim is to consider that there are no novelty gains in such recommendations. But, actually do forgotten items represent any sort of novelty to users? In the literature, novelty of a piece of information refers to how different it is regarding "what has been previously seen", by a specific user, or by a community as a whole [4]. However, we argue that it is an incomplete statement, missing a relevant issue: for how long an information is known. We can redefine novelty of a piece of information as how different it is with respect to "what has been previously seen **in a recent moment properly identified**". That is, relevant items that used to be consumed long ago could represent, in some sense, a degree of novelty, since we are assuming that hardly the users would remember by themselves most of those "lost" items. Therefore, a promising direction consists in investigating in deep the novelty issues related to the Oblivion Problem.

6.1 The Cold Start Duality

Going one step further in the discussion about solving the Oblivion Problem, we emphasize an interesting view that could help us in this task: its duality with the Cold Start problem [14]. Considering the Oblivion Problem, the role of recommendations extrapolates the task of indicating to users items that he or she may eventually like. Now it includes rescuing relevant items that, due to the RSs inadequacy for it, the natural growth of the tail and the competition process among the items by the user attention, get lost amid a wide range of options. Such rescuing aims to bring up old relevant items and to provide, besides the diversity improvement, surprise and good memories to the users. However, the challenge of recommending forgotten items relies on the fact that both domain and users evolve over time. Furthermore, such items have usually a very long past, but no recent information. As current RSs take into account, in general, only most recent information for providing the recommendations, forgotten items, controversially, suffer from lack of information. Therefore, the Oblivion Problem may be seen as the dual of the *Cold Start* problem, in which the difficulty of recommending new items also occur due to lack of information, but now due to lack of past.

Cold Start is a classic problem in traditional collaborative filtering recommendation [14]. It is hard to generate recommendations for new items because there is not enough experience data about the new items to make reliable correlations with other items. Pure user-oriented collaborative filtering cannot help in a cold-start setting, since no user preference information is available to form any basis for recommendations. A usual solution for this problem consists in using a content based recommendation approach, or even a hybrid approach, combining collaborative filtering with content based techniques [1].

Given this duality between the Cold Start and the Oblivion Problem, a first attempt to address the later would be to apply the same solutions currently used for Cold Start. For instance, using content-based techniques we can derive weights for the forgotten items according to their similarity with items currently consumed by the users. Also, hybrid methods may also be useful to retrieve forgotten items. However, the application of such methods to this scenario is even more challenging, since the attributes that would be selected to describe the content of the items must have a time-invariant character. This is an important requirement since the users taste evolves. Therefore, using information such as artist name and music genre among others may not be a good choice given that users might change their interests, in particular w.r.t. artists or genres over time. Thus, temporal correlations between items is an aspect to be incorporated into hybrid methods, in order to address the Oblivion Problem. Further, we believe that most of the evaluation metrics and strategies proposed for Cold Start problem could, with some changes, be applied to evaluate RSs designed to deal with the Oblivion Problem.

7. RELATED WORK

The growing relevance of Recommender Systems (RSs) in various domains have boosted research on this topic recently. Despite the advances achieved, there are still several challenges associated with this task, such as a proper user taste modeling, huge volume of items, and sparsity or lack of information about users [1]. Among these challenges, we highlight two that have received prominent attention: diversity on recommendation lists and temporal dynamics inherent to this kind of domains.

Recent studies found that diversity, combined with a high accuracy rate, is a relevant requirement associated with the usual taste similarity issue [16]. In [20], for instance, the diversification of recommendation lists is extensively addressed. The authors propose a similarity metric using a taxonomy-based classification and use it to compute an intra-list similarity metric that determines the overall diversity of the recommended list. Another study has examined the conditions in which diversity can be increased without loss of similarity, and presented an approach to determine such similarity, preserving increases in diversity when possible [12]. In [19], the diversification goal is seen as a binary optimization problem and a solution strategy to this problem consists in relaxing it to a trust-region problem. Further, the role of diversity in traditional recommender systems is clarified in [11], highlighting the pitfalls of naively incorporating current diversity enhancing techniques into existing recommender systems.

Considering efforts on temporal dynamics, it is almost consensual that accurately capturing user preferences over time is a major challenge for RSs. Therefore, numerous studies have tried to characterize, to model and to propose new strategies to deal with this problem without penalizing accuracy [10]. Ding [6] presented a novel algorithm to compute the temporal weights for items so that older items get smaller values. This approach has a disadvantage due to latest data are not always important while old data are not trivial all the time. Recently, Koren [9] predicted movie ratings for Netflix by modeling the temporal dynamics via a factorization model. Analogously, many time-evolving models [17] introduced time as a universal dimension shared by all users. It is remarkable that simple correlations over time are typically not meaningful, since users change their preferences due to different external events. Some studies argue that the time dimension is a local effect and should not be used for comparison among users [18].

Unlike other work, we address both aforementioned challenges, diversification and temporal dynamics, simultaneously in the context of the Oblivion Problem. That is, we address temporal dynamics using a new strategy that enhances the diversity in RSs by exploiting the subset of items consumed by users in the past (i.e., forgotten items), thus providing a differentiated and promising source of recommendations compared to traditional techniques. In fact, most techniques that focus on temporal dynamics aim to define a set of items that best match the most recent behavior or desires of the users [5]. We argue, however, that some of these users' wishes might be met by old items. Further, given the large amount of existing information about those forgotten items, we believe that such items may enhance the usual trade-off between diversity and accuracy in RSs. A study that deserves a deeper analysis, given its similarity to our work w.r.t. goals, is presented in [10]. However, its authors propose techniques that, by looking at recent recommendations provided to users, avoid that the same items are recommended to users over and over again through time.

8. CONCLUSIONS AND ONGOING WORK

In this paper, we present a differentiated and promising strategy to increase diversity in recommendation lists, based on the temporal dynamics inherent to recommendation domains. Such strategy defines a new problem for recommender systems, the Oblivion Problem, that aims to identify which items have been successful in the past for a given user (i.e., forgotten items) and exhibit a high probability of being consumed by this user again in the present. Besides formalizing this problem, we propose a methodology for its identification in real scenarios. The application of our methodology on a sample of *Last.fm* demonstrates the existence of the Oblivion Problem in this scenario, as well as the potential usefulness of recommending forgotten items in this case.

An immediate step of our work consists in developing techniques that are able to automatically identify the forgotten items relevant to each user, individually, or to the system as a whole. Later, we aim to build a recommender system that incorporates such items to recommendation lists, without penalizing the accuracy, in order to increase the diversity in the RSs. Finally, we highlight as another relevant direction the problem verification in distinct scenarios, such as recommendation of contacts in social networks, which present characteristics different from those observed in music recommendation.

9. ACKNOWLEDGMENTS

This work is partially supported by CNPq, Finep, Fapemig and InWeb – the Brazilian National Institute of Science and Technology for the Web.

10. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE TKDE*, 17(6):734–749, 2005.

[2] C. Anderson. *The long tail*. Gramedia Pustaka Utama, 2006.

[3] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–470, 2002.

[4] P. Castells, S. Vargas, and J. Wang. Novelty and diversity metrics for recommender systems: Choice, discovery and relevance. In *Proc. of International Workshop DDR*, pages 29–37, 2011.

[5] T. Cebrián, M. Planagumà, P. Villegas, and X. Amatriain. Music recommendations with temporal context awareness. In *Proc. of the 4th RecSys*, RecSys '10, pages 349–352, New York, NY, USA, 2010. ACM.

[6] Y. Ding and X. Li. Time weight collaborative filtering. In *Proc. of the 14th C*, pages 485–492. ACM, 2005.

[7] K. Hoashi, K. Matsumoto, and N. Inoue. Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proc. of the 11th ICM*, pages 110–119. ACM New York, NY, USA, 2003.

[8] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *Proc. of ICDM-08, 8th IEEE ICDM*, pages 263–272, 2008.

[9] Y. Koren. Collaborative filtering with temporal dynamics. In *Proc. of the 15th ACM SIGKDD*, pages 447–456. ACM New York, NY, USA, 2009.

[10] N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal diversity in recommender systems. *Proc. of SIGIR'10*, pages 210–217, 2010.

[11] L. McGinty and B. Smyth. On the role of diversity in conversational recommender systems. *Case-Based Reasoning Research and Development*, 2003.

[12] D. McSherry. Diversity-conscious retrieval. *Advances in Case-Based Reasoning*, pages 27–53, 2002.

[13] Y. Park and A. Tuzhilin. The long tail of recommender systems and how to leverage it. In *Proc. of the ACM RecSys*, pages 11–18. ACM, 2008.

[14] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In *Proc. of the 25th SIGIR*, pages 253–260. ACM Press, 2002.

[15] B. Schwartz. *The paradox of choice: Why more is less*. Harper Perennial, 2005.

[16] B. Smyth and P. McClave. Similarity vs. diversity. *Case-Based Reasoning Research and Development*, pages 347–361, 2001.

[17] J. Sun, C. Faloutsos, S. Papadimitriou, and P. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In *Proc. of the 13th SIGKDD*, pages 687–696. ACM, 2007.

[18] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. Sun. Temporal recommendation on graphs via long-and short-term preference fusion. In *Proc. of the 16th ACM SIGKDD*, pages 723–732. ACM, 2010.

[19] M. Zhang and N. Hurley. Avoiding monotony: improving the diversity of recommendation lists. In *Proc. of the ACM RecSys*, pages 123–130. ACM, 2008.

[20] C. Ziegler, S. McNee, J. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proc. of the 14th WWW*, pages 22–32. ACM, 2005.